



HAL
open science

Etude des longs ARNs non codants dans la leucémie aiguë myéloblastique à caryotype normal

Etienne de Clara

► **To cite this version:**

Etienne de Clara. Etude des longs ARNs non codants dans la leucémie aiguë myéloblastique à caryotype normal. Cancer. Université Paul Sabatier - Toulouse III, 2015. Français. NNT : 2015TOU30280 . tel-01877708

HAL Id: tel-01877708

<https://theses.hal.science/tel-01877708>

Submitted on 20 Sep 2018

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.

Université Fédérale



Toulouse Métropole Pyrénées

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)

Présentée et soutenue par :

Etienne DE CLARA

le 26 Novembre 2015

Titre :

Etude des longs ARNs non codants dans la leucémie aiguë myéloblastique à
caryotype normal

École doctorale et discipline ou spécialité :

ED BSB : Biotechnologies, Cancérologie

Unité de recherche :

UMR1037 CRCT

Directeur/trice(s) de Thèse :

Pr. Pierre Brousset

Dr. Marina Bousquet

Jury :

Pr. Estelle Espinos, Président du Jury

Dr. Claire Francastel, Rapporteur

Dr. Olivier Bernard, Rapporteur

Dr. Jean-Marc Egly, Rapporteur

« La vie, ce concept mystérieux, est ramenée à la présence d'ADN. Il n'y a plus de frontière entre la matière animée et inanimée. Tout n'est qu'une question de degré de complexité. »

Albert Jacquard

REMERCIEMENTS

Yo !

Je tiens tout d'abord à remercier le Professeur Pierre Brousset, pour m'avoir donné la chance de réaliser cette thèse au sein de son équipe et de m'avoir permis de finaliser mes travaux, au-delà de mon financement initial.

Je remercie particulièrement tous les membres de mon jury de thèse, le Dr. Claire Francastel, le Dr. Olivier Bernard et le Dr. Jean-Marc Egly, d'avoir accepté de juger mon travail. Je vous suis profondément reconnaissant d'avoir pris ce temps, d'avoir été présent pour la soutenance et pour toutes vos remarques qui m'ont permis d'améliorer le manuscrit et de prendre un certain recul sur les différentes thématiques abordées. Je tiens également à remercier le Professeur Estelle Espinos pour avoir accepté de présider ce jury ainsi que pour tous ses conseils au cours de ces trois années.

Bien entendu, je tiens profondément à remercier le Dr. Marina Bousquet, pour m'avoir donné ma chance (peu de gens l'aurait fait avec un skype à l'envers), pour m'avoir soutenu, pour ta disponibilité, pour m'avoir supporté malgré mon obstination (passagère) et peut-être même ma folie (également passagère). Ça n'a pas été toujours simple, ça a été génial également, et je ne peux que te remercier pour tout ce que tu m'as appris, scientifiquement et pour tout le reste. J'aurais du mal à le dire mais je le couche sur cette page, tu m'as apporté énormément et je pense que tu as été et tu es un (sacré) bon modèle ; la preuve en est par ta réussite (amplement méritée) au concours cette année. Je vais rester soft et éviter le pathos pour finir sur merci, merci merci et re-merci !!

Merci à toute l'équipe, passée, présente et future, pour toute votre aide au cours de ces trois ans et pour m'avoir permis de monopoliser la radio. Will et Charlotte, les deux anciens voyous qui m'ont donné une envie d'ailleurs. Merci Cyril pour ces questions très pointilleuses et pour les conseils pré-stress. Merci à Camille et Laure, les deux parisiennes qui sont passées maîtres dans l'art de la signature. Merci à Cathy l'indétrônable rockeuse aux doigts de fée toujours de bons conseils sur quasiment tout...sauf le rap. Julie la vosgienne, non pardon la lorraine, merci pour tous ces rires de craquage qui m'ont beaucoup aidés, surtout les tiens face à mes multiples medley un peu prosélytes d'ailleurs. Courage pour la dernière ligne droite, je reviendrai pour ta soutenance et je chanterai même pour fêter ça « smiley clin d'œil » (non non pas d'émoticônes dans une thèse voyons). Merci également Avédis pour ta parfaite imitation de Zeus sur l'Olympe, j'ai re-lavé le cadeau Kinder pour être sûr. Merci Coralie pour tes conseils de bio mol, je t'offre bien volontiers mon bac à glace maintenant et (dès que je le trouve) un petit dictionnaire d'expressions de d'jeuns de sorte à ce qu'on continue de communiquer. Je te souhaite bonne chance pour cette année à venir et ces concours !! Merci Ouafa pour tes conseils de sérénité que j'ai essayé d'appliquer et toutes tes histoires qui m'ont bien détendu les midis. Merci Fabienne, Sylvie, Laurence pour votre disponibilité et votre gentillesse, surtout Fabienne pour m'avoir fait remarquer mes cernes de ces derniers mois, j'ai acheté de la crème Q10 du coup. Merci également aux p'tits nouveaux Céline, Annabelle, Morgane, Romain. Ah oui merci Christian aussi. J'hésitais mais bon, je suis apparemment obligé. Naan, merci à toi, ton aide, tes conseils, ton soutien, tes farces de filou (je reste soft) cette dernière année qui aurait été beaucoup plus compliquée sans toi. C'est la tête haute que je m'en vais avec une victoire 6 à 5 dans notre bataille qui s'est avérée épique, pleine de rebondissements et avec peu d'honneurs. Merci à tous encore une fois,

pour toute votre aide, votre soutien, et votre humour (parce qu'il en faut), vous allez beaucoup beaucoup me manquer, YO !, pardon, PEACE.

Merci également à toutes les autres personnes de l'unité qui m'ont apporté à un moment donné, un rire, un conseil, une clope, voir un verre.

Merci à mes amis de m'avoir soutenu tout au long de cette thèse, avec une pensée particulière pour les Strasbourgeois qui se sont motivés pour faire 1000km et entendre parler une autre langue pendant 2h. Au top les mecs ! Ce week-end post-thèse a été parfait grâce à vous !! Narimène, Victoria, Ben, NiCotelette, Cédric, PPdernierMot, merci. Je remercie également tous ceux qui n'ont pas pu venir et qui m'ont accueilli à Strasbourg pour mes décompressions et le lâchage des fauves, pour les visites impromptues, les rires, les concerts, les balades à cheval au clair de lune. Merci à mon tatoueur Toulousain pour avoir semé le doute dans l'esprit de mon entourage en particulier de celui de mon directeur de thèse (Non Pierre, ce n'est pas cabalistique).

Je finirais par un énorme, gigantesque, gargantuesque merci pour ma famille ! Ma mère, mon père qui, de fait, ont rendu tout cela possible. Merci à la meilleure des grands-mères qui enchante ma vie Rock'n Roll Erica, à mes tantes et oncles, Rachel, David, Stéphane, Corinne, Jackie, et ma regrettée tante et marraine Linda. Bien sûr, un grand merci à toutes les cousines !! Maya, Sarah, Emma, Fanny, Léa et Thibault et mon petit Sammy, le petit BG de la famille. Mes frères Théo et Nicolas, et ma sœur Marylou, si tout ceci a été possible c'est grâce à vous, votre soutien, notre complicité, nos rires, nos engueulades, notre bonheur d'être ensemble, je vous aime profondément.

Peace. 

Table des matières :

<i>Introduction Générale</i>	15
PARTIE I : LA LEUCEMIE AIGUË MYELOBLASTIQUE OU LAM	19
D) Généralités et Classification des LAMs.....	19
a. Epidémiologie	19
b. Classification FAB et OMS.....	21
II) Anomalies Génétiques	24
a. Les mutations de Type 1	26
b. Les mutations de Type 2	28
c. "Dynamique" de l'apparition des mutations	31
i. Au cours de la phase pré-leucémique.....	31
ii. Au cours de la rechute.....	33
III) Symptômes, Caractérisation et Diagnostic	34
IV) Traitements des LAMs.....	35
a. Phase d'induction.....	35
b. Phase de consolidation	36
PARTIE II : LA NUCLEOPHOSMINE NPM1	39
D) Généralités	39
a. Structure	39
b. Modifications post-traductionnelles	40
c. Localisation cellulaire	40
d. Fonctions	41
i. Biogenèse des Ribosomes	41
ii. Chaperone d'histones	42
iii. Duplication des centrosomes.....	44
iv. Réparation de l'ADN.....	45
v. Régulation de la transcription	45
vi. Régulation de l'apoptose	46
II) NPM1 dans les cancers	47
a. NPM1 en tant qu'oncogène	47
b. NPM1 en tant que tumeur suppresseur.....	48
III) La mutation NPM1 dans les LAMs	50
PARTIE III : LES LONGS ARNS NON CODANTS OU LNCRNAs	55
D) Historique de la "matière noire" du génome.....	55
II) Structure et conservation des lncRNAs	56
a. Structure des lncRNAs	56
i. Généralités	56
ii. Structure primaire	57
iii. Structure secondaire.....	59
iv. Structure tertiaire et quaternaire.....	60
b. Conservation au cours de l'évolution.....	60
III) Organisation génomique et classification des lncRNAs	61
a. Classification relative à un gène codant	61
b. Classification relative à une région génomique.....	63
c. Autres classes	64
IV) Expression et localisation	64
a. Généralités.....	64
b. Expression spécifique au cours du développement	65

c.	Expression tissu spécifique	67
d.	Localisation cellulaire	68
V)	Fonctions des lncRNAs.....	69
a.	Dans la régulation transcriptionnelle.....	70
i.	Régulation de la machinerie de transcription	71
ii.	Séquestration/Recrutement des facteurs de transcription	72
iii.	Régulation de l'activité des facteurs de transcription	73
iv.	Autres régulations transcriptionnelles	74
b.	Dans la régulation post-transcriptionnelle	75
i.	Dans l'épissage	75
ii.	Dans l'édition	77
iii.	Dans la dégradation/stabilité des ARNm	78
c.	Dans la régulation traductionnelle.....	79
d.	Dans la régulation épigénétique	81
i.	Dans l'empreinte parentale	81
ii.	Modifications post-traductionnelles des histones.....	83
iii.	Modifications de l'ADN	84
iv.	Mécanisme de ciblage?	85
e.	En tant que réservoirs à petits ARNs non codants.....	88
i.	Réservoirs à microARNs	88
ii.	Réservoirs à snoARNs	90
iii.	Autres petits ARNs non codants	90
f.	Dans l'assemblage de structures nucléaires	91
VI)	Implications des lncRNAs dans les pathologies humaines	93
VII)	Implications des lncRNAs dans les cancers.....	94
a.	Prolifération.....	94
b.	Maintien de la croissance	95
c.	Apoptose et sénescence	95
d.	Angiogenèse.....	96
e.	Invasion et Métastases.....	97
f.	Immortalité répliquative	97
VIII)	Mécanismes de dérégulations des lncRNAs dans les cancers	99
IX)	Biomarqueurs et potentiel clinique	99
X)	lncRNAs et hématopoïèse normale	101
XI)	lncRNAs et hémopathies malignes.....	104
XII)	lncRNAs dans la leucémie aiguë myéloblastique.....	105

Objectifs de la thèse..... 109

Article en cours de rédaction..... 113

Résultats Supplémentaires..... 147

I)	Quel est l'impact de la mutation NPM1 sur l'expression des lncRNAs	149
II)	Quelle est la localisation cellulaire des lncRNAs associés à la mutation NPM1?.....	153
III)	Quelles peuvent être les fonctions de ces lncRNAs?.....	154
IV)	Etude du lncRNA XLOC_087120	156
V)	Quel est l'impact de la dérégulation du XLOC_087120 sur l'expression des histones?	158
VI)	Est-ce que XLOC_087120 est capable de recruter PRC2?.....	160
VII)	Quel pourrait être l'impact de NPM1 mutée sur XLOC_087120?.....	161

VIII) Quel est l'impact de la relocalisation de XLOC_087120 induite par NPM1 sur l'expression des histones?	163
<i>Discussion et Perspectives</i>	165
<i>Illustrations supplémentaires</i>	183
<i>Annexe</i>	193
<i>Bibliographie</i>	197

Abréviations:

1/2-sbsRNAs: Half-STAU1-binding site RNAs

2-HG: 2-Hydroxyglutarate

ADAR: Adenosine Deaminase Acting on RNA

ADN: Acide désoxyribonucléique

AF9: ALL1-fused gene chromosome 9

AIR: Acute insulin response

ALK: Anaplastic Lymphoma Kinase

ANCR: Anti-differentiation ncRNA

ANRIL: Antisense non-coding RNA in the INK4 locus

APEX1: Apurinic endonuclease

ARF1: ADP Ribosylation/Actor 1

ARN: Acide ribonucléique

ARNm: ARN messenger

ARNr: ARN ribosomique

ARNt: ARN de transfert

Ascl2: Achaete-scute family bHLH transcription factor 2

ATM: Ataxia Telangiectasia Mutated

ATP: Adénosine triphosphate

ATR: Ataxia Telangiectasia RAD3 related

ATRA: Acide tout-trans-rétinoïque

BACE1: β -secretase -1

BARL-2: B-ALL-associated long RNAs-2

BAX: BCL2-associated X protein

BBC3: BCL2 binding component 3

Bcl-2: B-cell CLL/lymphoma 2

BCL6: B-cell CLL/Lymphoma 6

BGL3: Beta Globin Locus 3

BNA: 2',4'-Bridged Nucleic Acid

BRCA2: Breast cancer 2

CAD: Caspase-activated DNase

CAGE: Cap analysis gene expression

CBF: Core Binding Factor

CBX7: Chromobox homolog 7

CCAT1: Colon Cancer Associated Transcript 1

CCT4: Chaperonin containing TCP1, subunit 4

CD: Cluster of Differentiation

Cdkn: Cyclin-dependent kinase inhibitor

CEBPa: CCAAT/Enhancer-Binding Protein Alpha

CENP: Centromere Protein

ceRNA: Competing Endogenous RNA

ChiP-seq: Chromatin immunoprecipitation-Sequencing

ChIRP: Chromatin Isolation by RNA Purification

CHRF: Cardiac Hypertrophy Related Factor

ciRNA: Circular intronic RNA

circRNA: Circular RNA

CKII: Casein Kinase II

CSF: Codon Substitution Frequency

CSF1: Colony stimulating factor 1

CTCF: CCCTC-binding factor

DAVID: Database for Annotation, Visualization and Integrated Discovery

Db: Double brin

DC: Dendritic cell

DFS: Disease Free Survival

DHFR: Dihydrofolate reductase

DISC1: Disrupted in schizophrenia 1

DLX: Distal-less homeobox

DNMT3A: DNA (Cytosine-5-)-Methyltransferase 3 Alpha

DSB: Double-Strand DNA Break

DSIF: DRB sensitivity-inducing factor

E/CDK2: Cyclin E-dependent kinase 2

EGO: Granule Ontogeny lncRNA

ENCODE: Encyclopedia of DNA Elements

EPO: Erythropoïétine

EPS: Erythroid ProSurvival lncRNA

ER : Ecotropic Receptor

ERa: Récepteur alpha aux œstrogènes

ERBB4: Erb-b2 receptor tyrosine kinase 4

eRNA: Enhancer RNA

ESTs: Expressed sequence tag

EZH2: Enhancer of Zeste Homolog 2

FAB: French-American-British

FLT3: Fms-like tyrosine kinase 3

FPKM: Fragments per Kb per Million

GAGE6: G antigen 6

GAPDH: Glyceraldehyde-3-phosphate dehydrogenase

GAS5: Growth arrest-specific 5

GM-CSF: Granulocyte-Macrophage Colony Stimulating Factor

GO: Gene Ontology

GRE: Glucocorticoid Response Element

GWAS: Genome-Wide Association Studies

HAR1: Highly accelerated region 1

HDAC: Histone désacétylase

HEXIM1: Hexamethylene bis-acetamide inducible 1

HIFa: Hypoxia Induced Factor Alpha

HLA: Human Leukocyte Antigens

hnRNP: Heterogeneous nuclear ribonucleoprotein

HOTAIR: HOX Transcript Antisense RNA

HOTAIRM1: HOXA transcript antisens RNA, Myeloid specific 1

HOTTIP: HOXA distal transcript antisense RNA

HOX: Homeobox

HULC: Hepatocellular carcinoma up-regulated long non-coding RNA

ID2: Inhibitor of DNA binding/Differentiation 2

IDH: Isocitrate Deshydrogenase

IGF: Insulin-like growth factor

IGFR: Insulin-like growth factor receptor

IGS RNAs: Intergenic Spacer RNAs

IL: Interleukine

inv: Inversion

IQGAP1: IQ motif containing GTPase activating protein 1

IRES: Internal Ribosome Entry Site

IRF1: Interferon regulatory factor-1

ITD: Internal Tandem Duplication

ITPR1: Inositol 1,4,5-TrisPhosphate Receptor type 1

K-RAS: Kirsten Rat Sarcoma Viral Oncogene

Kb: Kilobase

Kcnq1: Potassium channel, voltage gated KQT-like subfamily Q, member 1

Kcnq1ot1: KCNQ1 overlapping transcript 1

KLF1: Kruppel-like factor 1

KPNB1: Karyopherin (importin) beta 1

LAGC: Lymphomes anaplasiques à grandes cellules

LAL-B: Leucémie aigue lymphoblastique B

LAL-T: Leucémie aigue lymphoblastique T

LAM: Leucémies aiguës myéloblastiques

LAP: Leucémies aiguës promyélocyitaires

LMC: Leucémie myéloïde chronique

LNA: Lock Nucleic Acid

lncRNA: Long noncoding RNA

LOI: Loss Of Imprinting

LSD1: Lysine-Specific Histone Demethylase

LT-HSC: Long Term Hematopoietic Stem Cell

LUNARI: Leukemia-induced Non-coding Activator RNA

MALAT1: Metastasis associated lung adenocarcinoma transcript 1

MAPK: Mitogen-Activated Protein Kinases

mascRNA: MALAT1-associated Small Cytoplasmic RNA

MBD1: Methyl-CpG-binding domain protein

MBP: Major Basic Protein

MECP2: Methyl-CpG binding protein 2

MEG3: Maternally expressed 3

MGG: May-Grünwald-Giemsa

MLF1: Myeloid Leukemia Factor 1

MLL: Mixed Lineage Leukemia

MLN: Motilin

MOPP: Méchlorométhine, Oncovin, Procarbazine, Prédnisolone

MPO: Myéloperoxidase

MYH11: Myosin, Heavy Chain 11

N-RAS: Neuroblastoma Rat Sarcoma Viral Oncogene

NAT: Natural Anti-sens Transcript

NEAT1: Nuclear Paraspeckle Assembly Transcript 1

NEK2A: NIMA-related kinase 2

NELF: Negatif Elongation Factor

NER: Nucleotide Excision Repair

NF90: Nuclear Factor 90

NF-κB: Nuclear factor of kappa light polypeptide gene enhancer in B-cells

NF-YA: Nuclear transcription factor Y alpha

NFAT: Nuclear Factor of Activated T cells

NGF: Nerve Growth Factor

NMD: Nonsense mediated decay

NPM1: Nucléophosmine

NRON: Non coding RNA, repressor of NFAT

NSE: Non Specific Esterase

NSG: NOD/SCID/Gamma

NSMCE2: Non-SMC element 2

nt : Nucléotide

OMS: Organisation Mondiale de la Santé

ORF: Open Reading Frame

Osbp15: Oxysterol binding protein-like 5

PANDA: p21 associated ncRNA DNA damage activated

PARP: Poly(ADP-ribose) polymerase

PARS: Parallel Analysis of RNA Structure

PAX5: Paired box 5

Pb: Paire de base

PBX3: Pre-B-cell leukemia homeobox 3

PCA3: Prostate cancer associated 3

PCAT: Prostate Cancer-Associated ncRNA Transcript

PCBP1-AS1: Poly(rC) binding protein 1-Anti Sens 1

PCGEM1: Prostate-specific transcript 1

PCR: Polymerase Chain Reaction

Phlda2: Pleckstrin homology-like domain, family A, member 2

PI3K: Phosphatidylinositol-3-Kinases

PIC : Complexe de pré-initiation de la transcription

PIG: Puro IRES GFP

PINT: p53 induced transcript

PIP: Phosphatidylinositol (3,4,5)-triphosphate

Plk2: Polo-Like Kinase 2

PMAIP1: Phorbol-12-myristate-13-acetate-induced protein 1

PML: Promyelocytic leukemia

PNA: Peptid Nucleic Acid

Pol : Polymérase

POU5F1: POU class 5 homeobox 1

PRC: Polycomb Repressive Complex

pRNAs: Promoter-Associated Long Non Coding RNAs

PROMPTS: PROMoter uPstream Transcripts

PSF: Protein associated splicing factor

PTBP3: Polypyrimidine tract binding protein 3

PTEN: Phosphatase and tensin homolog

R3C: RNA-guided Chromatin Conformation Capture

RACE : Rapid Amplification of cDNA Ends

RARa: Récepteur à l'acide rétinolique alpha

RAT: Reverse Transcription-associated trap

REPA: Repeat A

RIP : RNA ImmunoPrecipitation

RP: Ribosomal Protein

RPKM: Reads per Kb per Million

RT-qPCR: Rétrotranscription-PCR quantitative

RUNX1: Runt related transcription factor 1

RUNXOR: RUNX1 overlapping RNA

sb : Simple brin

SBB: Noir Soudan

SCF: Stem Cell Factor

SHAPE: Selective 2'-Hydroxyl Acylation analyzed by Primer Extension

SHP-1: Src Homology region 2 domain-containing phosphatase-1

Slc: Solute carrier family

SMD: STAU1-mediated Messenger RNA Decay

snARNs: Small Nuclear RNAs

SnHG: Small Nucleolar RNA Host Gene

snoARNs: Small Nucleolar RNAs

Sox2: SRY (sex determining region Y)-box 2

SPHK1: Sphingosine kinase 1

SPRY4: Sprouty RTK signaling antagonist 4

SR: Sérine/Arginine riches

SRA ARN: Steroid receptor RNA activator

SRG1: SER3 regulatory gene 1

ST-HSC: Short Term Hematopoietic Stem Cell

STAT3: Signal transducer and activator of transcription 3

STAU1: Staufen double-stranded RNA binding protein 1

SUZ12: Suppressor of zeste 12

t: Translocation

TAF(1)48: TATA box binding protein (TBP)-associated factor

TAL1: T-cell acute lymphocytic leukemia 1

TBP: TATA Binding Protein

TERC: Telomerase RNA component

TERRA: Telomeric repeat-containing RNA

TERT: Telomerase Reverse Transcriptase

TET2: Tet oncogene Family member 2

TKD: Tyrosine Kinase Domain

TLS: Translesion DNA Synthesis

TNF: Tumor Necrosis Factor

TPO: Thrombopoïétine

Trx: Trithorax

TSS: Transcription Start Site

Tssc4: Tumor suppressing subtransferable candidate 4

UCA1: Urothelial carcinoma-associated 1

UCHL1: Ubiquitin Carboxy-Terminal Hydrolase L1

UCR: Ultra Conserved Region

UHG: U22 Host Gene

USP36: Ubiquitin specific peptidase 36

UTR: Untranslated Region

UV: Ultra Violet

WDR5: WD repeat domain 5

WT1: Wilms Tumor

*XACT: X Active Coating
Transcript*

Xi: X inactive

*XIST: X-inactive specific
transcript*

YY1: Yin Yang 1

Introduction
Générale

On estime actuellement que seul 1% du génome code des protéines, or l'avancée des technologies de séquençage à haut débit a permis de démontrer que près de 80% du génome était transcrit en ARNs non codants. Parmi ces derniers, on peut citer certaines classes d'ARNs bien connues telles que les ARNs de transfert (ARNt), les ARNs ribosomiques (ARNr), les petits ARNs nucléolaires (snoARNs) ou les petits ARNs nucléaires (snRNAs). Mis en évidence plus récemment, il existe également des ARNs tels que les microARNs ou les longs ARNs non codants (lncRNAs). Le nombre exact de longs ARNs non codants chez l'homme est pour l'instant inconnu. Cependant, de récentes analyses transcriptionnelles et bioinformatiques suggèrent l'existence de plusieurs dizaines de milliers de ces lncRNAs. La recherche dans ce domaine a permis de définir des fonctions spécifiques pour un certain nombre de lncRNAs et de démontrer leur intervention à tous les niveaux de régulation, transcriptionnelle, post-transcriptionnelle et même épigénétique. De plus, des lncRNAs ont été décrits comme dérégulés et impliqués dans de nombreux processus physiologiques, du développement aux pathologies, et plus particulièrement dans les cancers. Toutefois, nos connaissances quant à l'implication des lncRNAs dans la leucémie restent limitées.

Les leucémies aiguës myéloblastiques ou LAMs sont des pathologies clonales liées à un arrêt de la différenciation myéloïde associée à une augmentation de la prolifération cellulaire. Les LAMs résultent de l'acquisition de mutations dans les cellules souches ou dans les progéniteurs hématopoïétiques. Elles peuvent être divisées selon des critères cytogénétiques. On y retrouve des patients avec des anomalies chromosomiques récurrentes, des caryotypes complexes, des translocations chromosomiques rares. Toutefois, la moitié des cas de LAMs est représentée par des caryotypes normaux pour lesquels il existe peu de facteurs pronostiques car les événements oncogéniques sont inconnus. La mise en évidence de lncRNAs dérégulés dans ces pathologies pourrait nous permettre de mieux comprendre leurs fonctions d'un point de vue fondamental, mais également leurs rôles dans les processus oncogéniques.

Au cours de cette thèse, je m'attacherai à décrire l'état de la littérature concernant la leucémie aiguë myéloblastique et les longs ARNs non codants. Pour plus de clarté quant aux résultats de ce travail de thèse, j'insisterai également dans l'introduction sur l'implication de la nucléophosmine NPM1 dans les LAMs, ainsi que sur les régulations épigénétiques modulées par les lncRNAs.

PARTIE I : La Leucémie Aiguë Myéloblastique ou LAM

Les leucémies sont des hémopathies malignes caractérisées par l'accumulation de cellules hématopoïétiques anormales dans la moelle osseuse et le sang. On distingue les leucémies chroniques, caractérisées par une progression lente et l'accumulation de cellules de la moelle osseuse à un stade avancé de différenciation, des leucémies aiguës, caractérisées par une progression rapide et par l'accumulation dans la moelle osseuse et dans le sang de cellules immatures. Dans le cas où cette accumulation de cellules anormales concerne la lignée lymphoïde on parlera de leucémies lymphoblastiques, s'il s'agit de cellules de la lignée myéloïde on parlera de leucémies myéloblastiques.

I) Généralités et Classification des LAMs

Les leucémies aiguës myéloblastiques (LAMs) sont des hémopathies malignes résultant d'une prolifération aberrante et maligne de précurseurs myéloïdes bloqués à différents stades de la différenciation du lignage myéloïde. Le stade de blocage de la différenciation, plus ou moins précoce, est considéré notamment comme un paramètre permettant la classification des LAMs (Vardiman et al., 2009). Le site initial de prolifération leucémique se situe au sein de la moelle osseuse avec potentiellement des infiltrations extramédullaires. L'origine cellulaire des LAMs serait une petite population de cellules souches leucémiques $CD34^+$ et $CD38^-$ et d'une "side population" $CD123^+$ qui servirait de réservoir à la production de cellules leucémiques, ou blastes, et à la reconstitution post-traitement de la population leucémique totale (Saito et al., 2010)(Moshaver et al., 2008). Les cellules leucémiques présentent fréquemment des anomalies chromosomiques et/ou mutationnelles. Ces anomalies observées sont également un outil de classification des sous-types de LAM. Ainsi, l'OMS (Organisation Mondiale de la Santé) a réparti les LAMs en 7 groupes selon un consortium de paramètres cliniques et biologiques (Vardiman et al., 2009).

a. Epidémiologie

En occident, les LAMs sont la forme majoritaire de leucémie, représentant 25% des leucémies. En effet, l'Institut de Veille Sanitaire estime à environ 3000 le nombre de nouveaux cas de LAM par an en France (Estimation nationale de l'incidence des cancers en France entre 1980 et 2012, InVS). Le taux d'incidence sur la population mondiale est de 2,6 pour 100 000 chez l'homme et de 2,3 pour 100 000 chez la femme. L'incidence est relativement basse chez les sujets de moins de 40 ans avant d'augmenter plus rapidement à partir de 60 ans. Il s'agit en effet d'une pathologie touchant essentiellement le sujet âgé avec un taux d'incidence supérieur à 30 pour 100 000 chez l'homme et supérieur à 19 pour 100 000 chez la femme de 85 ans et plus (Figure 1). L'âge médian de survenue de ces pathologies est de 71 ans et celles-ci restent de très mauvais pronostic avec une survie relative à 5

ans qui ne dépasse pas 20 %. Ce sont des cancers mortels pour un grand nombre de cas et plus particulièrement chez les patients âgés, plus difficilement traitables par chimiothérapies conventionnelles, les conditions de comorbidités comme les maladies cardiaques ou les insuffisances rénales empêchant l'administration de fortes doses d'agents thérapeutiques. De plus, les patients âgés sont plus susceptibles de développer une infection en cours de traitement.

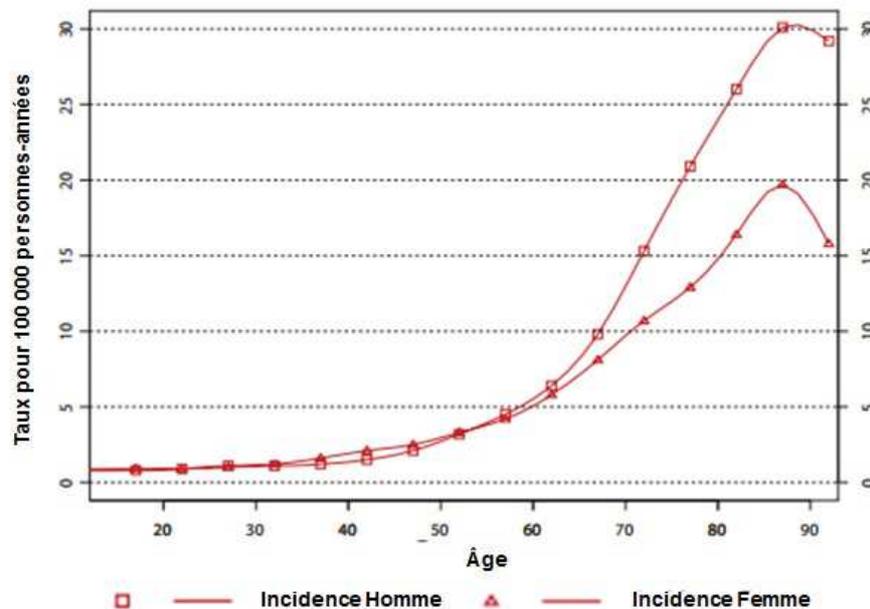


Figure 1 : Incidence par âge des Leucémies aiguës myéloïdes en France (Institut de Veille Sanitaire).

L'origine des LAMs est diverse : elle peut survenir consécutivement à un traitement cytotoxique dirigé contre une tumeur solide (LAMs secondaires), elle peut être *de novo* ou encore faire suite à un autre désordre hématopoïétique tel que des syndromes myélodysplasiques ou, plus rarement, suite à des syndromes myéloprolifératifs. Certaines maladies génétiques sont associées à un plus grand risque de développer une LAM. C'est le cas de la trisomie 21, de l'anémie de Fanconi, du syndrome de Bloom, et de la neurofibromatose (Belson et al., 2007). Bien que l'origine exacte de la maladie reste encore inconnue, plusieurs facteurs sociaux/environnementaux semblent influencer le processus de leucémogénèse. On peut citer quelques exemples identifiés comme les radiations, le benzène via la formation d'adduits et l'induction du stress oxydatif, les solvants industriels et les agents chimiothérapeutiques (Pierce et al., 2012)(McHale et al., 2012)(Godley and Larson, 2008). En effet, nous savons depuis plusieurs années que la chimiothérapie peut induire des leucémies aiguës secondaires de deux types. Les LAMs secondaires du premier type sont les plus fréquentes et surviennent après l'utilisation d'agents alkylants comme le traitement MOPP (Méchlorométhine, Oncovin™, Procarbazine, Prédnisolone) dans la maladie de Hodgkin. Le second type de LAM secondaire est le résultat de l'utilisation d'agents inhibiteurs de la topoisomérase II comme les

étoposides et anthracyclines pouvant être utilisés dans le traitement des tumeurs solides. Celles-ci sont associées à des anomalies cytogénétiques du bras long du chromosome 11, des translocations entre les chromosomes 8 et 21 (t(8;21)) et t(15;17) et des inversions du chromosome 16 (inv(16)). Les LAMs secondaires présentant ce type d'anomalies présenteraient le même pronostic que les LAMs *de novo*, mais plus sévère en cas de caryotypes complexes.

Les LAMs sont des hémopathies malignes issues de progéniteurs de la moelle osseuse associées à une accumulation clonale de blastes myéloïdes fonctionnellement déficients et bloqués à différents stades de leur maturation. Dans la moelle osseuse, la cellule souche myéloïde multipotente peut se différencier en progéniteurs spécifiques de chaque lignage sous l'impulsion de signaux, notamment cytokiniques (EPO, GM-CSF...). Ces progéniteurs continuent leur différenciation terminale en cellules myéloïdes différenciées et matures : granulocytaire, monocyttaire, érythrocytaire et mégacaryocytaire (Figure 2). Les blocages de différenciation dans ces différents lignages conduit à des pathologies clonales immunophénotypiquement et morphologiquement identifiables, et permettant ainsi une première classification et caractérisation des LAMs.

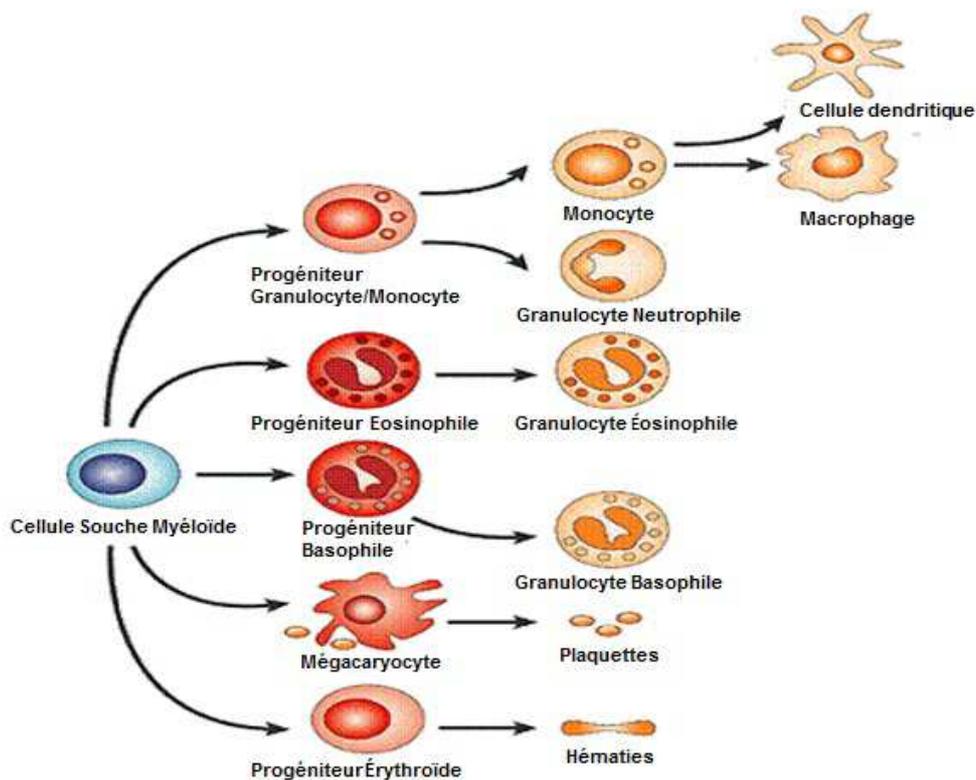


Figure 2 : Représentation schématique de l'hématopoïèse myéloïde.

b. Classification FAB et OMS

L'intérêt premier de ces classifications est bien évidemment d'identifier des sous types de LAMs, de les classer et de stratifier le risque afin d'adapter les traitements.

La classification FAB des LAMs a été proposée en 1976 par le groupe French-American-British (FAB) en se basant sur différents paramètres comme la morphologie ou l'état de différenciation des cellules. Ce système de classement a été complété en 1985 avec l'immunophénotype définissant ainsi 8 types de LAMs, de M0 à M7 (Bennett et al., 1976)(Neame et al., 1986)(Vardiman et al., 2009) (Tableau 1). La classe M0 se caractérise par des blastes indifférenciés ; la classe M1 sans maturation se caractérise par une prédominance de myéloblastes avec moins de 10% de différenciation granulocytaire ; la classe M2 avec maturation présente plus de 30% de myéloblastes, plus de 10% de différenciation granulocytaire et une différenciation monocyttaire inférieure à 20% ; la classe M3 se caractérise par des promyélocytes hypergranuleux (classe particulière présentant la translocation t(15;17)(q22;q12) et la protéine de fusion *PML-RAR α*) ; la classe M4 myélomonocytaire présente des cellules monocytaires avec au moins 30% de monocytes différenciés ; la classe M5 monocyttaire est caractérisée avec au moins 80% de monocytes faiblement (monoblastes) ou fortement différenciés ; la classe M6 érythroleucémique présente plus de 50% d'érythroblastes et plus de 30% de myéloblastes non érythroïdes ; et la classe M7 mégacaryoblastique est plus polymorphique et est caractérisée par des blastes vacuolés pouvant présenter des expansions cytoplasmiques.

FAB	Caractéristiques	Marqueurs immunophénotypiques
M0	blastés indifférenciés	CD34 ⁺ CD33 ⁺ CD13 ⁺ CD15 ⁺ CD16 ⁻
M1	sans maturation ; prédominance de myéloblastes ; moins de 10% de différenciation granulocytaire	CD34 ⁺ HLADR ⁺ CD33 ⁺ CD13 ⁺ CD117 ⁺
M2	avec maturation ; plus de 30% de myéloblastes ; plus de 10% de différenciation granulocytaire ; différenciation monocyttaire inférieure à 20%	CD34 ⁺ HLADR ⁺ CD16 ⁺ CD33 ⁺ CD13 ⁺
M3	avec promyélocytes hypergranuleux ; classe présentant la translocation t(15;17)(q22;q12) et la protéine de fusion <i>PML-RARα</i>	CD13 ⁺ CD34 ⁻ HLADR ⁻ CD117 ⁺
M4	Myélomonocytaire ; au moins 30% de monocytes différenciés	CD34 ⁺ HLADR ⁺ CD14 ^{+/-} CD36 ^{+/-} CD45 ⁺
M5	monocyttaire ; au moins 80% de monocytes faiblement (monoblastes) ou fortement différenciés	CD34 ⁺ HLADR ⁺ CD14 ⁺ CD36 ⁺ CD15 ⁺
M6	érythroleucémique ; plus de 50% d'érythroblastes ; plus de 30% de myéloblastes non érythroïdes	CD34 ⁺ HLADR ⁺ GPA ⁺ CD235 ⁺
M7	Mégacaryoblastique ; avec blastés vacuolés pouvant présenter des expansions cytoplasmiques ; blastés parfois agrégés aux plaquettes	CD34 ^{+/-} HLADR ⁺ CD41 ⁺ CD42 ⁺

Tableau 1 : Classification FAB des LAMs.

Au fur et à mesure des années, il a été observé des récurrences d'anomalies génétiques associées aux différentes classes présentant par ailleurs des valeurs pronostiques. Il était ainsi pertinent de les intégrer dans une nouvelle classification par l'OMS (Tableau 2). De plus, les observations de LAMs post-thérapies et de LAMs découlant de syndromes myéloidyplasiques, ont mené à l'addition de deux nouveaux types : LAMs secondaires et LAMs avec myéloidyplasie "multilignée". La classification par l'OMS définit donc 4 groupes : les LAMs avec anomalies chromosomiques récurrentes, les LAMs secondaires, les LAMs avec myéloidyplasie "multilignée", et les LAMs qui n'entrent dans aucune de ces catégories (Leymarie et al.). Pour ces dernières, la classification FAB est toujours utilisée. La classification de l'OMS comprend aussi 3 catégories définissant des néoplasies myéloïdes, le sarcome myéloïde, la prolifération myéloïde liée au syndrome de Down et les néoplasies de cellules dendritiques blastiques plasmacytoïdes.

À partir de 2008, l'OMS a ajouté dans ce classement des entités particulières avec de fortes valeurs pronostiques comme les LAMs présentant des mutations récurrentes dans les gènes *NPM1* (Nucléophosmine), *CEBP α* (CCAAT/Enhancer-Binding Protein Alpha) et *FLT3* (Fms-like tyrosine kinase 3). Les LAMs avec mutations *NPM1* et *CEBP α* sont considérées comme des entités provisoires dans la classe de LAMs avec anomalies génétiques récurrentes, au contraire des LAMs avec la mutation *FLT3*, laquelle est souvent associée à d'autres anomalies mais présentant une forte valeur pronostique (Vardiman et al., 2009). L'introduction de ces entités provisoires, caractérisées par ces mutations récurrentes dans la LAM, a essentiellement permis de mieux stratifier les LAMs à caryotype normal (LAM-CN) représentant environ 50% des LAMs.

Classement OMS des LAMs
LAMs avec anomalies cytogénétiques récurrentes
LAMs avec t(8;21)(q22;q22); fusion <i>RUNX1-RUNX1T1</i>
LAMs avec inv(16)(p13.1q22) ou t(16;16)(p13.1;q22); fusion <i>CBFB-MYH11</i>
LAPs avec t(15;17)(q22;q12); fusion <i>PML-RARA</i>
LAMs avec t(9;11)(p22;q23); fusion <i>MLLT3-MLL</i>
LAMs avec t(6;9)(p23;q34); fusion <i>DEK-NUP214</i>
LAMs avec inv(3)(q21q26.2) ou t(3;3)(q21;q26.2); fusion <i>RPN1-EVII</i>
LAMs avec t(1;22)(p13;q13); fusion <i>RBM15-MKL1</i>
Entité provisoire : LAMs avec mutation <i>NPM1</i>
Entité provisoire : LAMs avec mutation <i>CEBPα</i>
LAMs avec myélodysplasie "multilignée"
LAMs secondaires
LAMs sans spécificité
LAMs faiblement différenciées
LAMs sans maturation
LAMs avec maturation
Leucémies aiguës myélomonocytaires
Leucémies aiguës monocytaires/monoblastiques
Leucémies aiguës érythroïdes
Leucémies aiguës mégacaryoblastiques
Leucémies aiguës à basophiles
Panmyéloses aiguës avec myélofibrose
Sarcomes myéloïdes
Proliférations myéloïdes liées au syndrome de Down
Néoplasies dendritiques blastiques plasmacytoïdes

Tableau 2 : Classification OMS 2008 des LAMs (Adapté de Vardiman et al., 2009).

II) Anomalies Génétiques

On distingue plusieurs types d'anomalies génétiques telles que les insertions, les duplications et les translocations chromosomiques, très fréquentes dans les hémopathies malignes. Les translocations chromosomiques font intervenir des recombinaisons et des juxtapositions de gènes codants ou non, normalement séparés génétiquement. On distingue les anomalies qualitatives et les anomalies quantitatives. Les anomalies qualitatives font intervenir des points de cassure situés à l'intérieur de deux gènes, le plus souvent introniques, et la fusion des deux régions peut conduire à la formation d'une protéine chimérique potentiellement oncogénique. Il existe par exemple dans les

LAMs une forte récurrence des translocations $t(15;17)(q22;q12)$ et $t(9;11)(p22;q23)$ générant les protéines de fusion *PML-RAR α* et *MLL-AF9*, respectivement (Figure 4). Les anomalies quantitatives sont quant à elles caractérisées par des points de cassure situés à proximité de proto-oncogènes et ces translocations provoqueront un changement de profil d'expression des gènes, devenant oncogènes. La translocation $t(2;11)(p21;q23)$ par exemple induit une forte surexpression de l'oncomiR miR-125b (Bousquet et al., 2008).

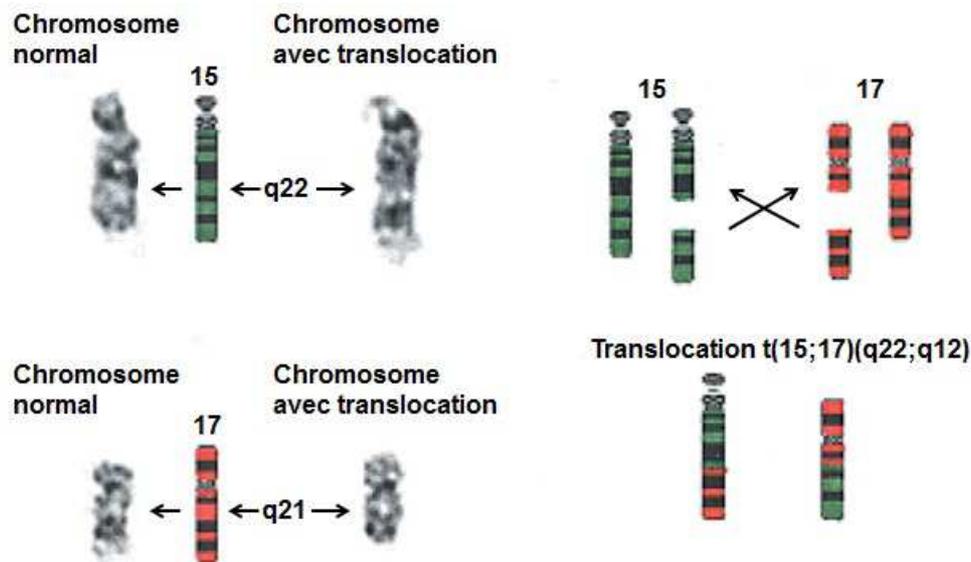


Figure 4 : Mécanisme de translocation chromosomique entre le chromosome 15 et 17 générant la translocation $t(15;17)(q22;q12)$ et la protéine chimérique *PML-RAR α* .

Une grande proportion des translocations chromosomiques sont considérées comme des entités à part entière et font intervenir des cofacteurs et facteurs transcriptionnels impliqués dans la différenciation hématopoïétique.

Lors de ma thèse, je me suis intéressé aux LAMs à caryotype normal (LAM-CN) représentant environ 50% des LAMs. L'absence de caractéristiques cytogénétiques pour permettre une meilleure classification et compréhension de la pathogénèse de ces cas a conduit à la recherche d'anomalies de séquences plus fines (Patel et al., 2012). Initialement, les mutations se sont classiquement distinguées en deux classes de mutations récurrentes dans la LAM: les mutations de type 1 qui provoquent l'activation de voies de signalisation conférant un avantage prolifératif aux progéniteurs myéloblastiques, et les mutations de type 2 qui regroupent des gènes induisant *in fine* des blocages de la différenciation myéloïde.

a. Les mutations de Type 1

Les mutations de type 1 sont des mutations dans des gènes apportant un avantage prolifératif ou une amélioration de la survie des progéniteurs myéloïdes en dérégulant des voies de signalisation comme les gènes *FLT3*, *NRAS*, *KRAS* (Neuroblastoma et Kirsten Rat Sarcoma Viral Oncogene) et *KIT*.

Mutations somatiques	Fréquence dans la LAM-CN
<i>KIT</i>	25-30%
<i>NRAS</i>	5-10%
<i>KRAS</i>	5-10%
<i>FLT3-ITD</i>	25-30%
<i>FLT3-TKD</i>	7%
<i>CEBPα</i>	10-18%
<i>TET2</i>	12-27%
<i>WT1</i>	10-13%
<i>DNMT3</i>	25%
<i>IDH</i>	30%
<i>NPM1</i>	55%

Tableau 3 : Fréquences des anomalies génétiques récurrentes dans la LAM.

Les mutations du gène *KIT* sont présentes dans 25-30% des LAM-CN. Le gène *KIT* code la protéine c-KIT, un récepteur tyrosine kinase liant le ligand SCF (Stem Cell Factor). Après interaction avec son ligand, c-KIT se dimérise et active des voies de signalisation intracellulaires comme la voie des MAPs Kinases (Mitogen-Activated Protein Kinases) ou PI3K (Phosphatidylinositol-3-Kinases), impliquées dans la survie cellulaire, la prolifération et la différenciation (Yarden et al., 1987)(Edling and Hallberg, 2007). Ces mutations sont également présentes chez des patients présentant une inv(16) ou une translocation t(8;21), des anomalies cytogénétiques généralement associées à un pronostic plus favorable. Cependant, les mutations de *KIT* dans la LAM-CN ne semblent pas être corrélées à des valeurs pronostiques.

Les mutations dans les gènes de la famille RAS, *KRAS* et *NRAS* sont retrouvées dans 5 à 10% des LAM-CN. Ces gènes codent des GTPases, activées par certaines cytokines et intervenant dans la transduction de signaux modulant la prolifération et la survie des progéniteurs hématopoïétiques. Les mutations affectant ces gènes conduisent à une activation constitutive des voies de signalisation pro-prolifération (Tyner et al., 2009). Cependant, ces mutations ne sont associées à aucune anomalie

cytogénétique particulière, ni aux caractéristiques cliniques ou pronostiques des LAMs (Kumar, 2011)(Kiyoi et al., 1999).

Le gène *FLT3* est un des gènes les plus fréquemment mutés dans les LAMs avec une occurrence dans 30 à 35% des LAM-CN. *FLT3* code une tyrosine kinase membranaire exprimée dès les premiers stades de l'hématopoïèse. Son activation entraîne une stimulation de la croissance et la survie des progéniteurs notamment via la voie des MAPs kinases. De manière générale, *FLT3* est très exprimé dans les LAMs mais il existe aussi deux mutations activatrices récurrentes :

-FLT3-TKD (Tyrosine Kinase Domain) est une mutation ponctuelle dans le domaine tyrosine kinase avec une valeur pronostique discutable (Figure 5).

-FLT3-ITD (Internal Tandem duplication) est une mutation de type duplication en tandem de taille variable dans le domaine juxtamembranaire du gène (Figure 5A). Cette mutation est plus étudiée, plus fréquente (25-30% des LAM-CN) et est associée à une valeur pronostique forte (Gilliland and Griffin, 2002) (Figure 5B).

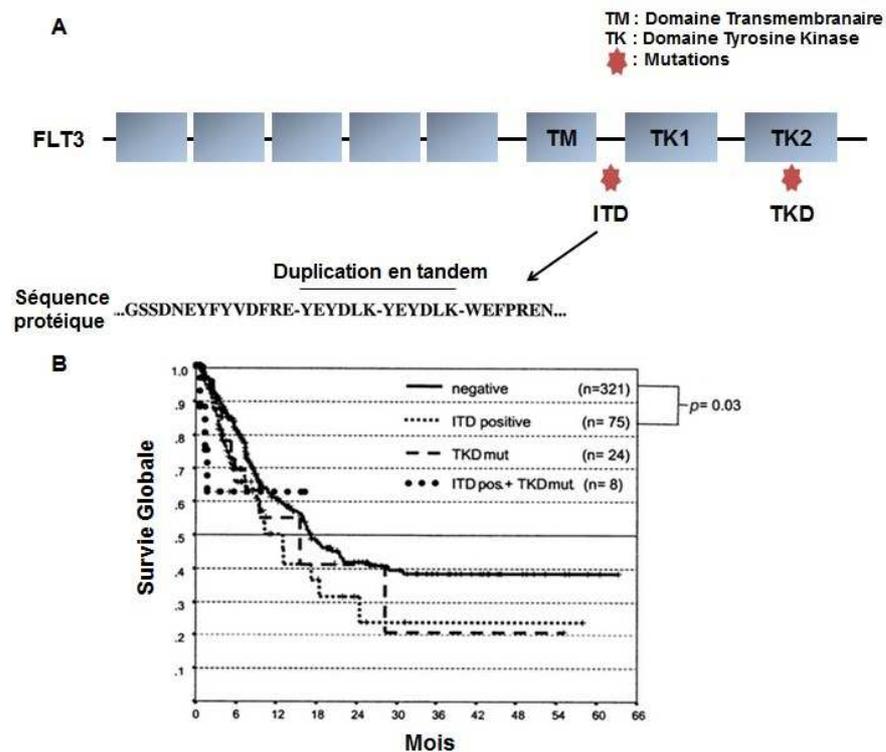


Figure 5 : Mutations dans le gène *FLT3*. A. Localisation de la mutation FLT3-ITD dans le domaine juxtamembranaire (et la conséquence de la duplication en tandem au niveau protéique) et de la mutation TKD dans le deuxième domaine tyrosine kinase (Adapté de Gilliland and Griffin, 2002). B. Analyse de la survie globale des patients LAMs en fonction du statut mutationnel de *FLT3*. Les patients présentant la mutation FLT3-ITD sont associés à un mauvais pronostic (Adapté de Thiede et al., 2002).

Les deux mutations FLT3-ITD et TKD conduisent à une activation constitutive de ce récepteur tyrosine kinase et des voies en aval, se traduisant par une prolifération accrue et un effet anti-apoptotique (Spiekermann et al., 2003). La transduction de FLT3-ITD dans des cellules hématopoïétiques primaires murines conduit à d'importants syndromes myéloprolifératifs (Kelly et al., 2002).

Les mutations de type 1 peuvent mener seules à des pathologies myéloprolifératives mais généralement avec des cellules myéloïdes toujours différenciées et fonctionnelles. En revanche, les mutations de type 2 permettent de perturber le processus de différenciation hématopoïétique.

b. Les mutations de Type 2

Ces mutations interviennent dans des gènes impliqués dans le processus de différenciation hématopoïétique. Il existe par exemple des mutations récurrentes dans les gènes *WT1* (Wilms Tumor), *TET2* (Tet oncogene Family member 2), *CEBP α* ou *NPM1* (Dombret, 2011). Plus récemment, des mutations ont été mises en évidence dans les gènes *DNMT3A* (DNA (Cytosine-5-)-Methyltransferase 3 Alpha), *IDH1* et *IDH2* (Isocitrate Deshydrogenase) (Ribeiro et al., 2012)(Mardis et al., 2009)(Gross et al., 2010).

Les mutations du gène *WT1* sont présentes majoritairement dans 10% à 13% des cas dans les LAM-CN. Ce gène code un facteur de transcription exprimé dès les stades les plus indifférenciés de l'hématopoïèse, mais absent dans les cellules matures. WT1 a été impliqué dans la prolifération, la survie et la différenciation et peut se comporter comme un suppresseur de tumeur ou un oncogène suivant le tissu et le transcrit étudié (Lindstedt et al., 2014). L'impact pronostique est incertain mais quelques études montrent un effet délétère des mutations de *WT1* sur la survie des patients (Owen et al., 2010).

Les mutations du gène *TET2* sont retrouvées dans 12 à 27% des LAMs avec une valeur pronostique sujette à débat, même si ces mutations semblent être de mauvais pronostic (Gaidzik et al., 2012). Les protéines TET ont un rôle dans la déméthylation de l'ADN et pourraient participer à l'hyperméthylation régulièrement observée dans la leucémie et les cancers en général. Les mécanismes permettant le ciblage de TET2 sur des gènes impliqués dans la leucémogénèse ne sont pas connus, mais Wang *et al.* ont suggéré que l'impact de TET2 dans la LAM était dépendant de son interaction avec WT1 (Wang et al., 2015a).

Les mutations dans le gène *CEBP α* sont retrouvées majoritairement dans les LAM-CN (10 à 18%) et peuvent être divisées en deux groupes: un groupe avec une seule mutation dans le gène (1/3 des cas) et un deuxième groupe avec deux mutations (2/3 des cas). Le deuxième groupe est généralement associé à un meilleur pronostic (Marcucci et al., 2011) (Figure 6A). CEBP α est un

facteur de transcription impliqué dans l'hématopoïèse en réprimant l'expression de *MYC*, et en régulant positivement l'expression de gènes spécifiques de la lignée granulocytaire. L'étude de modèles murins de knock-out de *CEBPα* a montré que *CEBPα* était essentielle au développement des granulocytes matures et que sa perte de fonction facilitait la leucémogénèse par un blocage de cette différenciation (Zhang et al., 1997).

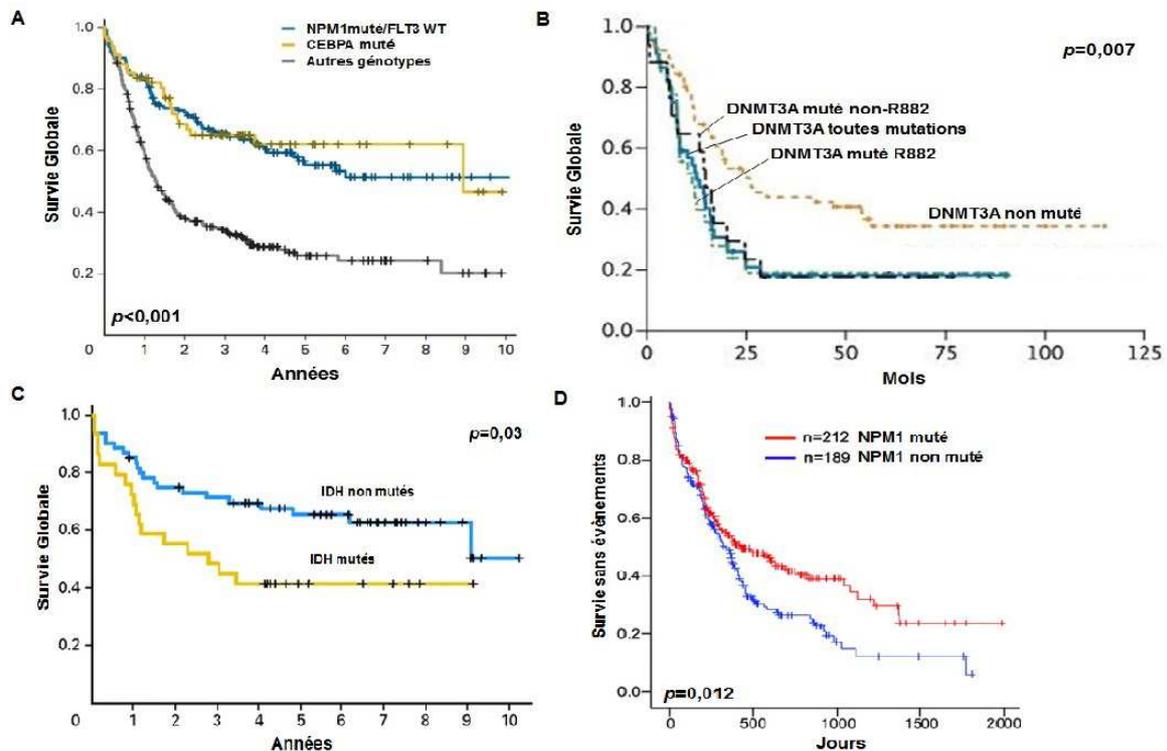


Figure 6 : Impact des différentes mutations récurrentes dans la LAM sur la survie globale. A. La mutation dans le gène *CEBPα* est associée à un bon pronostic (Adapté de Marcucci et al., 2011). B. Les mutations dans le gène *DNMT3A* sont associées à un mauvais pronostic (Adapté de Ley et al., 2010). C. Les mutations dans les gènes *IDH* sont associées à un mauvais pronostic (Adapté de Paschka et al., 2010). D. La mutation dans le gène *NPM1* est associée à un bon pronostic (Adapté de Schnittger et al., 2005).

Les mutations de *DNMT3* sont retrouvées dans approximativement 25% des cas et sont associées significativement à un mauvais pronostic (Figure 6B) et aux mutations dans les gènes *NPM1*, *FLT3* (*FLT3-ITD*) et *IDH1* (Ley et al., 2010). Les DNMTs catalysent la réaction de méthylation des îlots CpG, régulant ainsi l'expression de nombreux gènes. Il a été suggéré que l'impact de cette mutation dans *DNMT3* se traduisait par une dérégulation des gènes *HOX* dans l'hématopoïèse. Cependant, les LAMs présentant des mutations dans *DNMT3* ne présentent pas de signature épigénétique forte et spécifique par rapport aux LAMs *DNMT3* non muté (Ribeiro et al., 2012).

Des mutations des gènes *IDH1* et *IDH2* ont récemment été identifiées par séquençage à haut débit (Mardis et al., 2009). Elles sont souvent associées aux LAM-CN, aux mutations du gène *NPM1* et sont de mauvais pronostic (Paschka et al., 2010)(Boissel et al., 2010) (Figure 6C). Les protéines *IDH1* et 2 sont impliquées dans le métabolisme de la synthèse lipidique, la protection du stress oxydatif et la transduction de signaux. De plus, l'expression d'*IDH1* et 2 mutées provoque une accumulation du métabolite oncogénique putatif 2-HG (2-Hydroxyglutarate), et induit un blocage de la différenciation myéloïde *in vitro* (Figuroa et al., 2010).

Les mutations dans le gène de la nucléophosmine *NPM1* sont les plus récurrentes dans les LAM-CN (55% des LAM-CN) et associées à un bon pronostic (Schnittger et al., 2005) (Figure 6D). La nucléophosmine a de nombreux rôles dans la cellule : elle régule l'assemblage des ribosomes, la duplication des centrosomes, l'assemblage des histones, la réparation de l'ADN et l'apoptose. Je détaillerai la fonction et l'impact de *NPM1* dans la partie II "La nucléophosmine *NPM1*".

Cette précédente classification des mutations dans les LAMs peut apparaître comme binaire et peut être trop simpliste. Le groupe The Cancer Genome Atlas Research (TCGA) a rapporté en 2013 des résultats de séquençage (génomique complet et exome) de 200 cas de LAM (Cancer Genome Atlas Research Network, 2013). Les auteurs de ce travail ont identifié des mutations dans un grand nombre de gènes, déjà recensés ou nouveaux, dont certains pourraient être considérés comme "driver" dans la LAM. Ces gènes ont été classés dans 9 catégories basées sur leurs fonctions connues ou prédites telles que les facteurs de transcription, méthylation de l'ADN ou de la chromatine, ou encore des gènes impliqués dans la formation des complexes cohesin et du spliceosome. La forte prévalence de mutations dans des gènes tels que *DNMT3A*, *TET2*, *IDH1/IDH2*, *ASXL1* ou même *EZH2* (Ernst et al., 2010) suggère qu'un troisième groupe de mutations impliquées dans l'épigénome pourrait compléter le modèle pré-existant du "two-hit" décrit précédemment. De manière intéressante, dans le travail du groupe TCGA, les mutations au sein d'une même classe de gènes présentent une tendance à être mutuellement exclusives et chaque cas de LAM présente en moyenne 5 mutations dans des gènes fréquemment mutés. Ces observations soulèvent plusieurs questions pertinentes :

Comment ces mutations s'accumulent-elles dans un seul lignage cellulaire?

Quelles sont les mutations fondatrices et initiatrices de la pathologie?

Il a alors commencé à émerger l'hypothèse que ces mutations s'accumulent de manière sériée dans les cellules souches hématopoïétiques, qui ont une durée de vie plus longue que la plupart des cellules de la différenciation myéloïde. Selon ce modèle, on trouverait chez les patients LAM de rares compartiments résiduels de cellules souches hématopoïétiques comprenant des cellules normales sans mutation, mais également des cellules pré-leucémiques contenant une ou plusieurs des mutations

retrouvées dans le blaste leucémique correspondant. Cela reflèterait un historique de l'acquisition des mutations ainsi que la présence d'une phase pré-leucémique résultant d'une expansion hématopoïétique clonale, dont les études récentes semblent confirmer l'existence (Jan et al., 2012).

c. "Dynamique" de l'apparition des mutations

i. Au cours de la phase pré-leucémique

Différents travaux sur le sujet ces dernières années ont montré que les LAMs surviennent à partir d'une cellule hématopoïétique immature (cellules souches hématopoïétiques et les progéniteurs) transformée à la suite d'une accumulation de plusieurs étapes de changement génétique et épigénétique. Cette série de changements est supposée générer une cellule souche pré-leucémique précédant la cellule souche leucémique transformée. Si la cellule souche leucémique est fonctionnellement définie par sa capacité à donner naissance rapidement à une leucémie avec une forte pénétrance dans des modèles murins de transplantation, la cellule pré-leucémique peut quant à elle conduire à une leucémie *in vivo* avec une latence variable et sous la dépendance de l'accumulation graduelle d'événements génétiques additionnels.

Le travail de Génovese *et al.* publié en 2015, a permis d'analyser les mutations et leurs fréquences à partir d'échantillons sanguins sur une cohorte de plus de 12 000 personnes ayant développé des cancers hématologiques ou non (Genovese et al., 2014). Les auteurs ont pu mettre en évidence une expansion clonale de l'hématopoïèse plusieurs mois avant le diagnostic de LAM. De manière particulièrement intéressante, cette expansion est associée à la présence de mutations somatiques dans les gènes *DNMT3A*, *ASXL1* et *TET2*, alors que les mutations dans *FLT3* ou *NPM1* ne sont pas retrouvées. Cette observation suggère que les mutations de *DNMT3A*, *ASXL1* et *TET2* seraient des mutations initiatrices, maintenues sur de longues périodes sans manifestation clinique majeure, et que les mutations *FLT3* et *NPM1* apparaîtraient plus tard, peut être en tant qu'élément déclencheur du phénotype leucémique proprement dit. Il semblerait ainsi que le séquençage d'échantillons sanguins pourrait permettre une détection précoce des cancers hématologiques. Cependant, Genovese *et al.* restent prudents quant à cette possibilité. En effet, les auteurs ont observé ces mutations somatiques dans les cellules hématopoïétiques normales dans 10% des cas, alors que la fréquence du passage de l'expansion clonale en cancer était modeste (1% par an).

Ces observations semblent être en accord avec de précédents travaux, notamment ceux menés par Challen *et al.* (Challen et al., 2011). Afin d'analyser l'impact de la mutation *DNMT3A* dans l'hématopoïèse *in vivo*, les auteurs ont généré un modèle murin inductible et conditionnel de knock-out *DNMT3A* et transplanté les cellules souches hématopoïétiques de ces animaux chez des animaux sauvages et irradiés. Cette étude a permis d'observer une augmentation progressive de la proportion des cellules hématopoïétiques *DNMT3A* *-/-* dans le sang de ces souris au fur et à mesure des

transplantations successives. La moelle de ces souris présente également un défaut d'hématopoïèse avec une expansion du compartiment des cellules souches hématopoïétiques. Cependant, aucune de ces souris n'a développé de maladies myéloprolifératives, indiquant certainement la nécessité de mutations additionnelles. Chez l'homme, Shlush *et al.* ont pu mettre en évidence chez des patients LAM *DNMT3A* muté/*NPM1* muté, une forte prévalence de mutations de *DNMT3A* dans les cellules souches hématopoïétiques de patients LAM et sans mutations de *NPM1* (Shlush *et al.*, 2014). De plus, les xénogreffes de ces cellules souches hématopoïétiques de patients ont montré que ces cellules *DNMT3A* muté et *NPM1* non muté présentaient un avantage sélectif puisqu'elles sont associées à une expansion clonale. Cela suggère vivement que la mutation *DNMT3A* apparaît précocement dans les cellules souches hématopoïétiques et leur confère un avantage compétitif pouvant expliquer l'expansion clonale de cellules souches pré-leucémiques.

Il en est de même pour les mutations de *TET2* pour lesquelles plusieurs études semblent supporter le modèle que l'inactivation de *TET2* est un événement précoce dans la leucémogénèse et conduit à l'expansion clonale de cellules pré-leucémiques chez l'homme. Par exemple, le génotypage de cellules $CD34^+/CD38^-$ de patients avec syndrome myéloprolifératif a montré que les mutations récurrentes de *JAK2* n'étaient pas observées en l'absence de mutation de *TET2*. De plus, l'analyse des cellules souches hématopoïétiques suggère que la mutation *TET2* apparaît en premier lieu (Delhommeau *et al.*, 2009). Jan *et al.* ont également mis en évidence la présence de mutation de *TET2* dans des cellules souches leucémiques résiduelles de patients LAM représentant potentiellement des cellules pré-leucémiques selon le modèle décrit précédemment (Jan *et al.*, 2012). Afin de mettre en évidence l'acquisition séquentielle des mutations dans les cellules souches leucémiques des patients LAMs, les auteurs de ce travail ont décrit l'historique des événements génétiques en génotypant les colonies formées à partir d'une seule cellule souche hématopoïétique non leucémique de patients LAMs. Chez deux patients avec des mutations bi-alléliques de *TET2*, ils ont mis en évidence que cet événement est un changement précoce, car la majorité des cellules souches présentaient une ou les deux mutations *TET2*. À l'inverse, les mutations *FLT3* ont uniquement été détectées dans les cellules leucémiques. De plus, ces cellules pré-leucémiques présentant des mutations bi-alléliques *TET2* sont plus représentées que la population hétérozygote pour *TET2*, elle-même plus représentée que les cellules souches ne présentant pas la mutation *TET2*. Ces résultats indiquent que la mutation *TET2* pourrait conférer un avantage compétitif par rapport aux cellules ne portant pas la mutation *TET2*, et conduire à une expansion clonale au cours du temps.

Ces différents travaux suggèrent que le développement de la leucémie débute dans des cellules souches hématopoïétiques qui vont acquérir des mutations dans des gènes impliqués dans divers mécanismes de régulation (méthylation de l'ADN, modifications de la chromatine). Il en résulte un avantage compétitif par rapport aux cellules ne présentant pas de mutations et une expansion clonale qualifiée de phase pré-leucémique. Cependant, malgré cette expansion clonale, ces cellules souches

pré-leucémiques génèrent des clones hématopoïétiques incluant des progéniteurs et des cellules myéloïdes matures et fonctionnelles. Il est envisageable que ce soit dans ces progéniteurs que les autres mutations rapportées, telles que celles dans des gènes impliqués dans la prolifération et/ou dans la différenciation myéloïde, puissent apparaître afin d'établir la cellule leucémique proprement dite.

ii. Au cours de la rechute

Ce modèle d'évolution clonale de la leucémie pourrait avoir de fortes implications cliniques, par rapport au diagnostic mais également par rapport à la rechute. La leucémie consiste probablement en la prédominance d'un clone leucémique avec plusieurs sous-clones minoritaires. Un grand nombre de patients atteignent une rémission complète après chimiothérapie d'induction, cependant cette rémission est souvent de courte durée et conduit à une rechute. On peut alors se poser des questions par rapport au mécanisme d'évolution clonale dans ce contexte de rechute et l'impact des différentes mutations recensées :

Est-ce la résurgence du clone leucémique dominant non éradiqué par la chimiothérapie?

Est-ce la résurgence d'un clone minoritaire présentant une résistance chimiothérapeutique?

Est-ce une évolution clonale avec de nouvelles mutations dans un clone leucémique pré-existant?

Est-ce une évolution clonale avec de nouvelles mutations dans un clone pré-leucémique pré-existant générant une rechute au phénotype distinct?

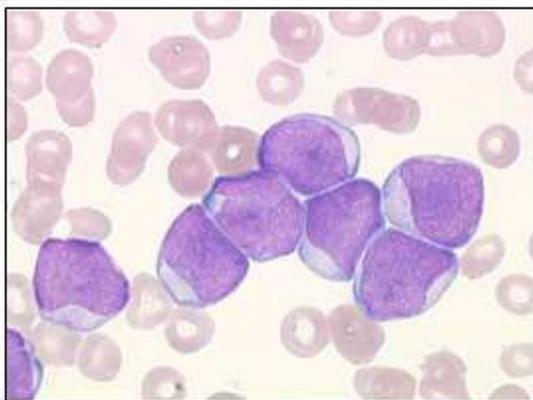
Afin de répondre à ces questions, plusieurs études ont analysé le génotype des cellules leucémiques entre la maladie au diagnostic et à la rechute. Dans le travail de Shlush *et al.* précédemment décrit, les auteurs mettent en évidence la persistance du clone pré-leucémique *DNMT3A* muté après traitement. Ce clone pourrait donc être un réservoir pour la reconstitution de la population leucémique après rechute (Shlush *et al.*, 2014). De plus, les auteurs ont décrit des rechutes comme étant la conséquence de la ré-émergence du clone leucémique originel ou l'émergence d'un nouveau clone présentant les mêmes mutations. Il n'y a cependant pas de caractérisation de nouvelles mutations dans ce travail. De la même manière, Ding *et al.* ont par exemple montré sur une étude de 8 cas, que la rechute correspondait à deux voies principales, l'acquisition de nouvelles mutations dans le clone à l'origine de la maladie primaire ou dans un sous-clone pré-existant ayant survécu à la chimiothérapie (Ding *et al.*, 2012). En effet, les auteurs retrouvent les mutations récurrentes connues telles que *DNMT3A*, *FLT3*, *NPM1*, mais également des mutations nouvellement identifiées telles que *WAC*, *SMC3*, *DIS3*. Bien que cette évolution et cette acquisition sériée de mutations puissent être inhérentes à la cellule leucémique, ces résultats peuvent également mettre en lumière l'induction de ces nouvelles mutations par les agents chimiothérapeutiques.

III) Symptômes, Caractérisation et Diagnostic

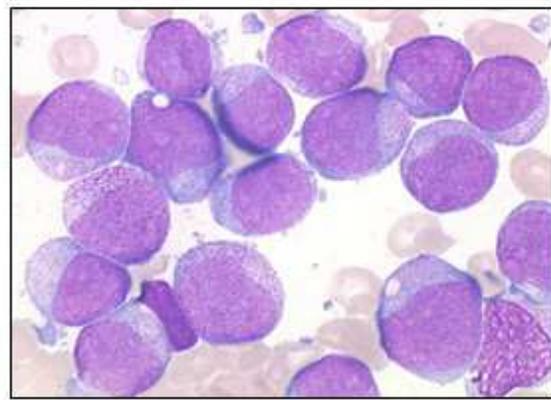
L'accumulation d'anomalies génétiques au niveau des différents stades de la différenciation myéloïde provoque d'une part un blocage de la différenciation, et d'autre part une prolifération accrue. Ceci se produit dans la moelle osseuse au détriment des cellules hématopoïétiques normales. Cette insuffisance médullaire entraîne anémie, neutropénie, thrombopénie avec comme conséquences une plus grande fatigue, une plus grande susceptibilité aux infections et des hémorragies cutanées et des muqueuses (Estey and Döhner, 2006). D'autres symptômes résultent de l'infiltration extramédullaire des blastes par le sang vers le foie, la rate, les ganglions ou la peau avec des signes tels que la splénomégalie, des lésions cutanées de type nodule ou une hypertrophie gingivale.

Le diagnostic de LAM nécessite un examen des frottis médullaires et sanguins couplé aux études morphologiques, immunophénotypiques, cytogénétiques et moléculaires des blastes myéloïdes.

L'analyse morphologique est réalisée par des méthodes histologiques. Les cellules blastiques, prélevées à partir de la moelle, sont colorées au MGG (May-Grünwald-Giemsa), coloration utilisée principalement en hématologie pour distinguer les cellules sanguines et les blastes : ces derniers sont caractérisés par une chromatine plus fine et plus granuleuse et un cytoplasme réduit. Pour identifier la lignée myéloïde impliquée, des tests cytochimiques sont utilisés, tels que des marquages MPO (Myéloperoxydase)(Figure 3), SBB (Noir Soudan), ou aux NSE (Non Specific Esterase). Ces caractéristiques permettent une première classification par rapport au FAB. Pour un diagnostic de LAM, la proportion de blaste doit être supérieure à 20%.



LAM sans maturation : blastes sanguins



LAM sans maturation : moelle osseuse

Figure 3 : Observation de blastes sanguins et de la moelle osseuse au microscope à fond clair (Cytochimie de la myéloperoxydase MPO).

L'immunophénotypage par cytométrie en flux est essentiellement utilisé pour les cas difficiles à classer d'un point de vue cytologique, en analysant des marqueurs spécifiques (Cf Tableau 1). Ceci permet d'établir un diagnostic pour les LAMs avec différenciation minimale, les LAMs mégacaryocytaires et les leucémies aiguës pour lesquelles il y a une ambiguïté de lignées.

L'analyse cytogénétique est aujourd'hui indispensable au diagnostic des LAMs. 50% des patients présentent des anomalies caryotypiques et plusieurs translocations et inversions chromosomiques récurrentes qui sont classées dans la catégorie LAM avec anomalies génétiques récurrentes.

IV) Traitements des LAMs

Depuis plusieurs décennies, le traitement des LAMs consiste en une combinaison d'agents chimiothérapeutiques comme la daunorubicine ou idarubicine (anthracyclines) et de cytarabine (ARA-C). La thérapie se déroule en deux phases : une phase d'induction et une phase de consolidation. Le but de la phase d'induction est d'atteindre une rémission complète avec moins de 5% de blastes dans la moelle, un taux de neutrophiles supérieur à 1000/mm³ et un taux de plaquettes supérieur à 100 000/mm³ (Tallman, 2005). Pour les patients jeunes atteints de LAM, le taux de rémission complète est d'environ 75%, mais pour les patients de plus de 60 ans, le taux de rémission complète est de 40 à 65%, avec 85% de rechutes dans les 3 ans (Scholl et al., 2009)(Erba, 2010). Dans de nombreux cas, la phase d'induction ne suffit pas et conduit à une rechute. C'est pourquoi une phase de consolidation est appliquée afin de permettre de diminuer le nombre de ces rechutes. La meilleure réponse des patients jeunes est en partie due à une meilleure tolérance à la thérapie et à la présence chez ces patients de moins de facteurs de risque. En 2009, Rao *et al.* ont montré une activation plus élevée des voies oncogéniques RAS, SRC et TNF (tumor Necrosis Factor) chez les patients âgés, pouvant contribuer à une survie moins élevée et à une plus grande résistance aux traitements (Rao et al., 2009).

a. Phase d'induction

Le traitement d'induction consiste en 3 jours d'anthracycline (anthracenedione mitoxantrone, 10-12 mg/m² ou daunorubicine au moins 60mg/m² ou idarubicine 10-12 mg/m²) et 7 jours de cytarabine en intraveineuse, 100-200 mg/m². Des doses plus importantes de cytarabine peuvent être utilisées pour augmenter le temps de survie sans rechute, mais elles n'augmentent pas le taux de rémission complète et de survie globale. Un troisième agent peut être ajouté comme l'étoposide ou la thioguanine, avec des résultats mitigés selon les études (Bradstock et al., 2005)(Hann et al., 1997).

b. Phase de consolidation

La phase de consolidation précoce de la rémission complète obtenue après chimiothérapie d'induction consiste habituellement en une utilisation de cytarabine à forte dose pouvant être associée à une anthracycline. Plusieurs types de stratégies peuvent être appliqués pour prévenir la rechute : plusieurs cycles de chimiothérapie intensive ou une chimiothérapie suivie d'allogreffe ou d'autogreffe de cellules souches hématopoïétiques.

La chimiothérapie intensive repose essentiellement sur de hautes doses de cytarabine. Le nombre de cycles de traitement, la dose appropriée, le calendrier d'administration ainsi que l'intérêt de combinaison entre plusieurs agents restent à déterminer.

Dans le cas de la greffe de cellules souches hématopoïétiques, plusieurs études ont évalué l'intérêt de l'autogreffe, mais bien qu'elle soit associée à un effet anti-leucémique, elle n'améliore pas la survie globale. La greffe à partir d'un parent compatible est utilisée depuis presque 30 ans, cependant l'agressivité du traitement et les risques de réaction du greffon contre l'hôte conduisent à un débat quant à savoir quel bénéfice le patient en tire par rapport à la survie globale, (Koreth et al., 2009). La décision se fait au cas par cas et porte sur le bénéfice-risque pour le patient au regard de l'agressivité de la chimiothérapie nécessaire à la greffe (aplasie importante) et l'efficacité de cette greffe. Dans le cas des LAMs avec un pronostic favorable (LAMs avec des anomalies génétiques impliquant le gène *CBF* (Core Binding Factor), LAMs avec mutation *NPM1* sans la mutation *FLT3-ITD* et LAMs avec mutations de *CEBP α*), 3 cycles de cytarabine à haute dose ou plus sont bénéfiques pour les patients et la greffe n'est généralement pas envisagée pour atteindre la première rémission. Le choix de stratégie de traitement prend donc en compte les facteurs pronostiques associés aux différentes anomalies récurrentes de la LAM. Ainsi, les anomalies génétiques récurrentes dans la LAM ont été classées selon leurs valeurs pronostiques (Tableau 4).

Groupes pronostiques	Caractéristiques
Favorable	t(8;21)(q22;q22); fusion <i>RUNX1-RUNX1T1</i> inv(16)(p13.1q22) ou t(16;16)(p13.1;q22); fusion <i>CBFB-MYH11</i> Mutation <i>NPM1</i> sans <i>FLT3-ITD</i> (caryotype normal) Mutation <i>CEBPA</i> (caryotype normal)
Intermédiaire I	Mutation <i>NPM1</i> et <i>FLT3-ITD</i> (caryotype normal) <i>NPM1</i> sauvage et <i>FLT3-ITD</i> (caryotype normal) <i>NPM1</i> sauvage sans <i>FLT3-ITD</i> (caryotype normal)
Intermédiaire II	t(9;11)(p22;q23); fusion <i>MLLT3-MLL</i> Anomalies cytogénétiques non classées comme favorables ou défavorables
Défavorable	inv(3)(q21q26.2) ou t(3;3)(q21;q26.2); fusion <i>RPNI-EVII</i> t(6;9)(p23;q34); fusion <i>DEK-NUP214</i> t(v;11)(v;q23); réarrangement <i>MLL</i> -5 ou del(5q); -7; abnl(17p); caryotype complexe (≥ 3 anomalies à l'exclusion des anomalies récurrentes de la classification OMS)

Tableau 4 : Classification pronostique des LAMs.

La mise en place de système de classement des LAMs et de leurs pronostics, se fondant sur des caractéristiques biologiques et cytogénétiques, mène à une meilleure prise en charge des patients. Les LAM-CN représentent la moitié des cas de LAM, et la découverte de mutations récurrentes avec des valeurs pronostiques a permis d'affiner plus encore la stratification du risque. La mutation dans le gène *NPM1* est la mutation la plus fréquente dans les LAMs et plus particulièrement dans les LAM-CN. Cette mutation est clairement associée à un pronostic favorable en l'absence de la mutation *FLT3-ITD*, ce qui lui a valu d'être classée comme une entité provisoire au classement OMS en 2008. Cependant, de nombreux patients rechutent après la première rémission et le pronostic des patients *NPM1* muté reste encore relativement hétérogène. Depuis lors, de nombreuses études tentent de mettre à jour les mécanismes moléculaires et cellulaires impliquant *NPM1* mutée dans le processus de leucémogénèse, afin d'améliorer potentiellement la prise en charge et la réponse aux traitements de ce type de LAM.

PARTIE II : La nucléophosmine NPM1

I) Généralités

NPM1 (Nucléophosmin ou Numatrin, ou B23) est le membre de la famille NPM le plus étudié et est directement impliqué dans le développement des cancers chez l'homme. La forte prévalence de mutations de ce gène dans les LAMs et plus particulièrement dans les LAM-CN en fait un candidat d'étude majeur.

a. Structure

NPM1 est une protéine ubiquitaire très bien conservée entre les espèces. Chez l'homme, le gène fait 24kb contient 12 exons et est situé sur le chromosome 5q35. Cette protéine de 294 acides aminés et d'un poids moléculaire de 35 kDa, est caractérisée par une structure et des domaines distincts nécessaires à ses fonctions et pour sa localisation cellulaire (Hingorani et al., 2000) (Figure 7).

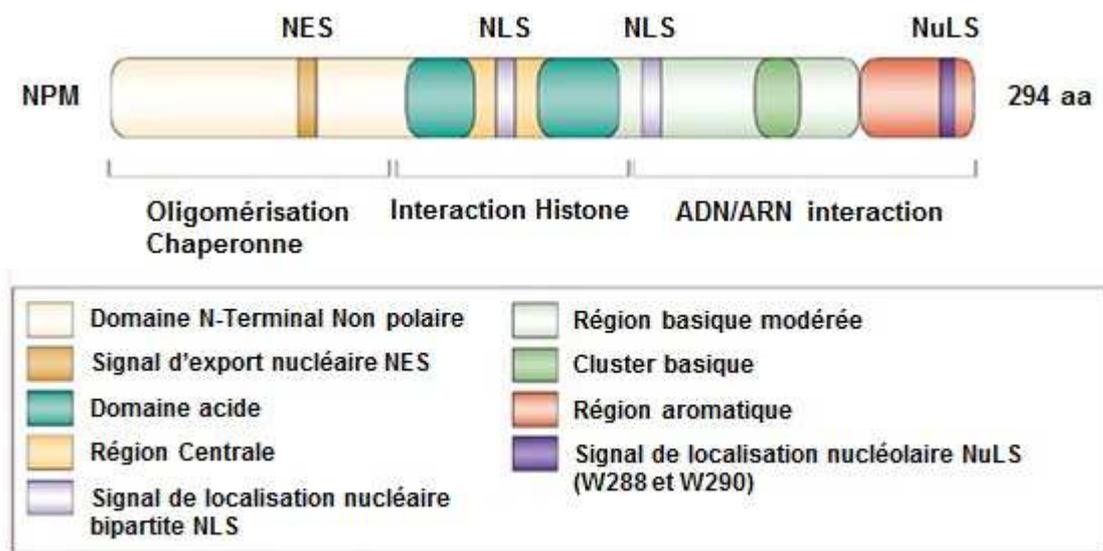


Figure 7 : Représentation schématique de la structure de NPM1 (Adapté de Grisendi et al., 2006).

La région N-Terminale intervient dans l'oligomérisation de NPM1 en pentamère et est nécessaire à sa fonction de chaperone de protéine. La région centrale est quant à elle nécessaire à la fonction de chaperone d'histone de NPM1. En effet, la région centrale permet l'interaction avec les histones H3, H4, H2A et H2B pour réguler la formation des nucléosomes (Okuwaki et al., 2001a). La région C-Terminale contient des acides aminés basiques générant une charge positive suivie d'une

région riche en acides aminés aromatiques. Ces séquences interviennent dans l'interaction entre NPM1 et les acides nucléiques (Wang et al., 1994), l'ATP (Chang et al., 1998), l'activité ribonucléase (Hingorani et al., 2000) et la localisation nucléolaire (Zirwes et al., 1997). De plus, NPM1 possède plusieurs séquences nécessaires à son activité de protéine navette : un signal d'export nucléaire (NES) localisé en N-Terminale, un signal de localisation nucléaire (NLS) bipartite localisé dans la région centrale et un signal de localisation nucléolaire (NuLS), constitué notamment de deux résidus tryptophane essentiels (W288 et W290), localisé dans la partie C-Terminale.

b. Modifications post-traductionnelles

La protéine NPM1 peut être soumise à des modifications post-traductionnelles comme la phosphorylation, l'acétylation et l'ubiquitylation. La régulation de la phosphorylation de NPM1 n'est pas totalement comprise, toutefois nous savons qu'elle peut être modulée par CKII (Casein Kinase II), PLK2 (Polo-Like Kinase 2) ou par le complexe cycline E/CDK2 (Cyclin-dependent kinase 2) (Okuda et al., 2000). La phosphorylation de NPM1 semble réguler les différentes fonctions de NPM1, son oligomérisation, sa localisation nucléaire, son implication dans la progression de la mitose et la biogenèse des ribosomes (Negi and Olson, 2006). Cette phosphorylation est notamment nécessaire, pour interagir avec les protéines basiques et pour la décondensation de l'ADN (Bañuelos et al., 2003). L'acétylation de la protéine conduit à une plus grande affinité de NPM1 pour les histones acétylées (Swaminathan et al., 2005). NPM1 peut être également dé-ubiquitinée par l'enzyme USP36 (ubiquitin specific peptidase 36), stabilisant NPM1 et augmentant ses fonctions dans le nucléole (Endo et al., 2009).

c. Localisation cellulaire

Bien que NPM1 soit une protéine navette entre le noyau et le cytoplasme, elle est majoritairement nucléaire, et plus précisément nucléolaire (Zirwes et al., 1997) (Figure 8). De plus, au sein du nucléole, la protéine NPM1 est enrichie au niveau des composants granulaires qui sont associés aux particules de maturation pré-ribosomales (Spector et al., 1984). Cependant, une petite fraction de NPM1 peut être observée dans le nucléoplasme. De manière intéressante, NPM1 peut être également détectée au niveau périchromatinien et au niveau des fuseaux mitotiques au cours de la mitose cellulaire (Valdez et al., 1994). Ces différences de localisation cellulaire de NPM1 sont essentielles à ses différentes fonctions et les anomalies de localisation de NPM1 pourraient être directement associées à son impact dans la leucémogénèse.

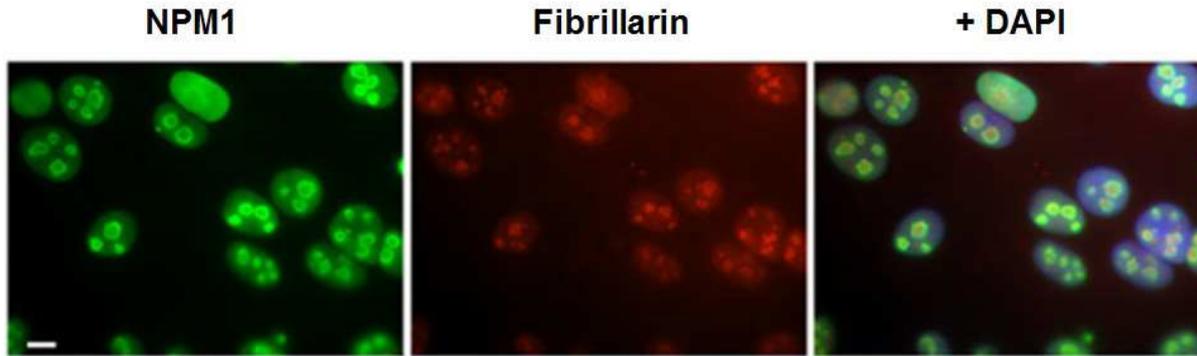


Figure 8 : Localisation nucléolaire de NPM1. Marquage immunofluorescent de NPM1 (vert) dans les nucléoles. NPM1 co-localise avec la Fibrillarin (rouge), une protéine nucléolaire (Adapté de Holmberg Olausson et al., 2014).

d. Fonctions

Ces dernières années, de nombreuses études ont mis en avant l'importance fonctionnelle de NPM1 intervenant dans de nombreux mécanismes moléculaires et processus physiologiques potentiellement impliqués et dérégulés au cours de la cancérogenèse : la biogenèse des ribosomes, l'assemblage des histones, la réparation de l'ADN, la régulation de la transcription, l'apoptose et la duplication des centrosomes.

i. Biogenèse des Ribosomes

Initialement, NPM1 a été considérée comme une protéine chaperone des ribosomes, ou comme un facteur d'assemblage du fait de son association aux particules préribosomales. En effet, NPM1 a été décrite comme nécessaire à la maturation des sous-unités ribosomales 60S, et à l'export et l'assemblage des ribosomes 80S (Yu et al., 2006) (Figure 9). La mutation du domaine d'export nucléaire ou du domaine d'interaction aux acides nucléiques de NPM1 empêche l'export des ribosomes vers le cytoplasme, provoquant ainsi une diminution de la croissance cellulaire (Maggi et al., 2008). *In vitro*, NPM1 augmente le clivage des ARNs ribosomiaux (ARNr), agit comme une endoribonucléase dans la maturation des transcrits ARNr, et la délétion de *NPM1* prévient notamment la maturation de l'ARNr 28S (Savkur and Olson, 1998)(Itahana et al., 2003). De plus, il semblerait que NPM1 puisse inhiber l'activité endoribonucléase médiée par APE1 sur les ARNr (Tell et al., 2010). NPM1 interagit également directement avec les protéines ribosomales comme RPL5 (Ribosomal Protein L5), RPL23 (Ribosomal Protein L23) et RPS9 (Ribosomal Protein S9), confortant donc l'impact de la nucléophosmine dans la biogenèse des ribosomes et par conséquent dans la prolifération et la croissance cellulaire, fortement perturbées dans les cancers (Lindström and Zhang, 2008)(Wanzel et al., 2008).

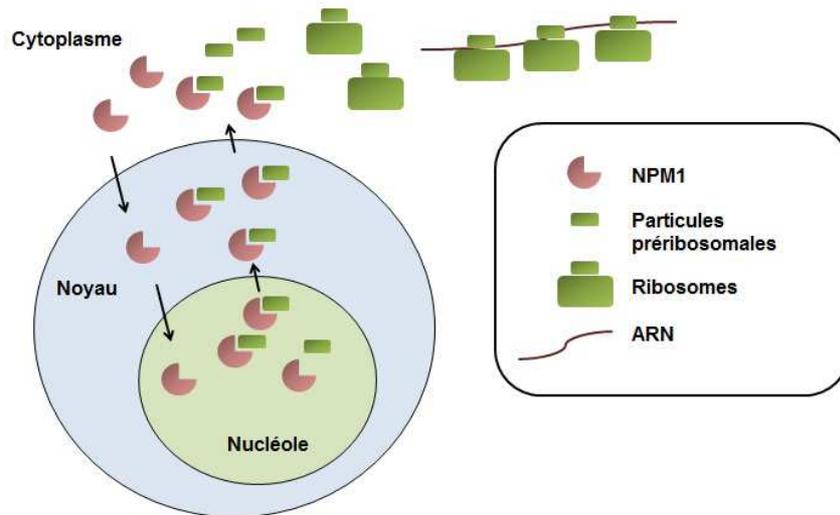


Figure 9 : Représentation schématique de l'impact de NPM1 dans la biogenèse des ribosomes. Dans le nucléole, NPM1 interagit avec les particules préribosomales et les transportent du noyau au cytoplasme afin de permettre leur maturation en ribosome.

ii. Chaperone d'histones

La deuxième grande fonction de NPM1 est de réguler l'assemblage des histones et des nucléosomes. Les histones peuvent être divisées en deux catégories, les histones cœurs formant le nucléosome, et les linker-histones H1 stabilisant la structure nucléosomes/ADN. Les histones cœurs H2A, H2B, H3 et H4 s'associent en dimères H2A/H2B et tétramères H3/H4 pour former l'octamère d'histones autour duquel s'enroule l'ADN pour former le nucléosome (Figure 10).

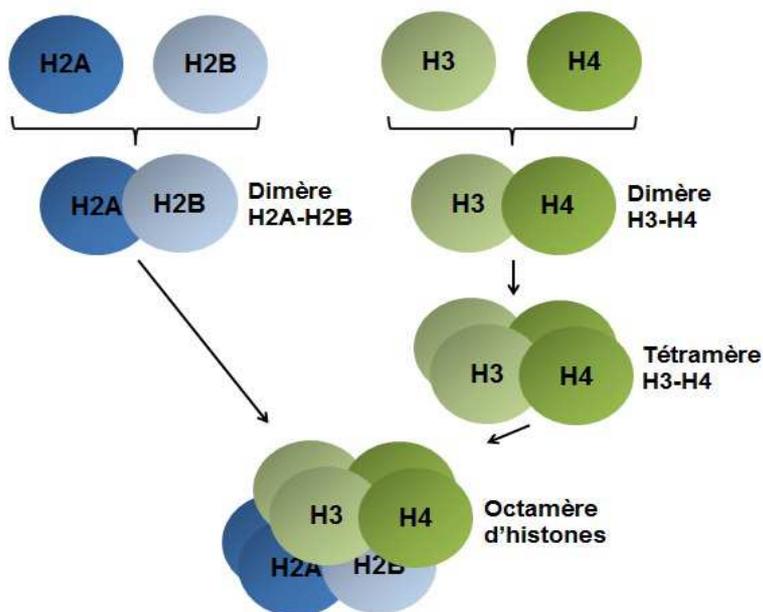


Figure 10 : Représentation schématique de l'assemblage des histones.

L'assemblage des nucléosomes est dépendant de protéines chaperones telle que NPM1 qui stimule l'assemblage et la déposition/éviction des nucléosomes sur l'ADN, ou encore qui facilite l'interaction entre les enzymes de modification des histones et les histones (De Koning et al., 2007). Ce processus est dynamique et intervient dans la régulation de l'expression génique, la réplication et la réparation de l'ADN. Il a été montré que NPM1 interagissait avec les histones cœurs et les histones linker dans le nucléole. Cette association est préférentiellement réalisée avec l'histone H3 et permet l'assemblage des nucléosomes, le transfert des nucléosomes sur l'ADN nu et la décondensation de la chromatine de sperme *in vitro* (Namboodiri et al., 2004)(Okuwaki et al., 2001b)(Okuwaki et al., 2001c) (Figure 11). La délétion de *NPM1* dans des lignées cellulaires humaines provoque notamment un réarrangement de l'hétérochromatine périnucléolaire (Holmberg Olausson et al., 2014). De plus, il semblerait que la modulation des nucléosomes par NPM1 soit dépendante des modifications post-traductionnelles que subit NPM1. L'acétylation de NPM1 et des histones par p300 conduit à une ouverture de la chromatine et à une augmentation de la transcription (Swaminathan et al., 2005). Ces fonctions de chaperone d'histones et de transfert des nucléosomes de NPM1 s'avèrent être d'une importance fondamentale dans la régulation de différents processus biologiques, comme la réparation et le maintien de la stabilité de l'ADN.

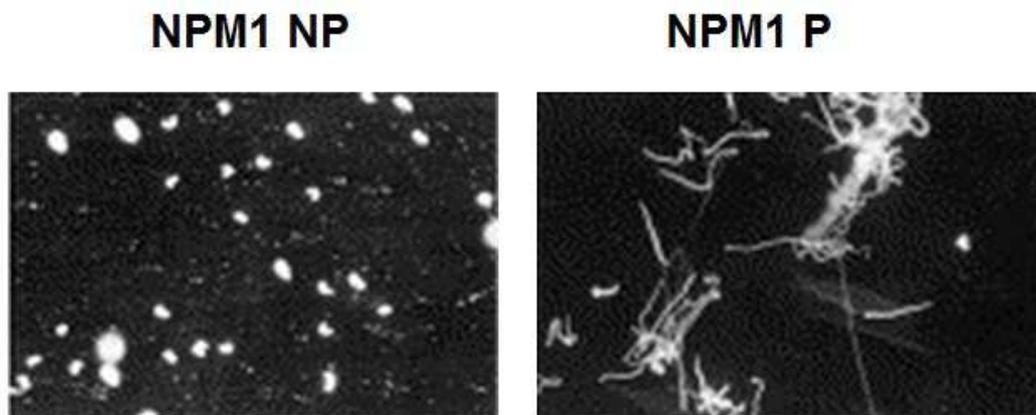


Figure 11 : Impact de NPM1 dans la condensation de la chromatine. En présence de NPM1 non phosphorylée (inactive) la chromatine de sperme est condensée. En présence de NPM1 phosphorylée (active), on observe une décondensation massive de la chromatine (Adapté de Bañuelos et al., 2003).

iii. Duplication des centrosomes

Au cours de la mitose, NPM1 est détectée au niveau des fuseaux mitotiques et plus précisément au niveau des centrosomes non dupliqués en interphase (Zatsepina et al., 1999). Il a été observé que la phosphorylation de NPM1 par cdk2/cycline E permettait le détachement de NPM1 du centrosome et l'initiation de la duplication, suggérant fortement un rôle direct de NPM1 lors de la division cellulaire (Figure 12) (Okuda et al., 2000). Son implication dans la duplication des centrosomes semble dépendante d'interactions entre NPM1, NEK2A (NIMA-related kinase 2), BRCA2 et ROCK2 et dépendante de la phosphorylation de NPM1 par PLK2 et PLK1 (Wang et al., 2011a)(Krause and Hoffmann, 2010). L'inhibition de la phosphorylation de NPM1 par PLK2 conduit notamment à une diminution de la duplication des centrosomes et à un blocage en phase S des cellules. De plus, Foltz *et al.* ont montré que NPM1 interagissait avec CENPA (Centromere Protein A), une histone H3 particulière présente dans des nucléosomes spécifiques des centromères (Foltz et al., 2006). Sachant que les centromères sont les régions chromosomiques centrales contrôlant le comportement des chromosomes lors de la mitose, il a été suggéré que NPM1 était importante pour le maintien de la stabilité génomique lors de la mitose. En effet, les cellules NPM1^{-/-} présentent des anomalies de duplication des centrosomes et une augmentation significative de l'instabilité génomique (Grisendi et al., 2005). La perte de fonction de NPM1 induit donc un accroissement de la susceptibilité qui mène à la transformation oncogénique.

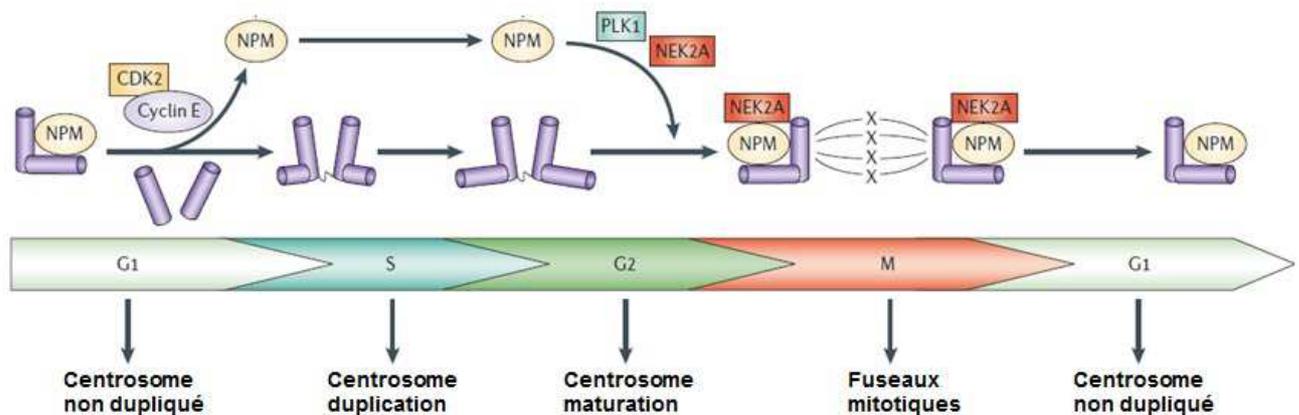


Figure 12 : Représentation schématique des mécanismes potentiels impliquant NPM1 dans la duplication des centrosomes. En phase G1 du cycle cellulaire, NPM1 est associée au centrosome non dupliqué. La phosphorylation de NPM1 induite par cdk2/cycline E permet le détachement de NPM1 et la duplication des centrosomes. Au cours de la phase G2 du cycle cellulaire, NEK2A redirige NPM1 vers les centrosomes dupliqués et permet la mise en place des fuseaux mitotiques (potentiellement via son interaction avec CENPA) pour la cytokinèse (Adapté de Grisendi et al., 2006).

iv. Réparation de l'ADN

L'impact de NPM1 sur la stabilité et le maintien de l'intégrité génomique a également été associé à la réparation de l'ADN. En cas de dommages à l'ADN induits par radiations ionisantes, NPM1 est surexprimée et délocalisée du nucléole vers la chromatine (Lee et al., 2005) et un pool de NPM1 phosphorylées est recruté au niveau des sites de cassures (Koike et al., 2010). Il a été montré par Wu *et al.* que l'irradiation aux UV de fibroblastes murins NIH 3T3 induisait une surexpression rapide de NPM1 (Wu et al., 2002). De plus, ces auteurs ont également montré que cette surexpression de NPM1 *in vitro* était associée à une augmentation de la réparation des dommages à l'ADN de type NER (Nucleotide Excision Repair) et à l'inhibition de la mort cellulaire induite par les UV. Ainsi, il a été suggéré que NPM1 pourrait agir comme chaperone des histones intervenant dans la réparation de ce type de cassures. Colombo *et al.* ont par ailleurs mis en évidence dans un modèle murin de knock-out de *NPM1* que les cellules *NPM1*^{-/-} présentaient une augmentation de la phosphorylation de l'histone H2AFX, une cible des kinases ATM (Ataxia Telangiectasia Mutated) et ATR (Ataxia Telangiectasia RAD3 related) intervenant dans la réparation de l'ADN (Colombo et al., 2005). L'augmentation de la phosphorylation de cette histone spécifique dans ce modèle d'étude a également été associée à l'augmentation de la formation de foci de réparation de l'ADN H2AFX et ATM positives (Wu et al., 2002).

v. Régulation de la transcription

NPM1 peut indirectement influencer la transcription via son activité sur les histones mais NPM1 est également impliquée directement dans la régulation de la transcription. Tout d'abord, NPM1 peut interagir avec des facteurs de transcriptions tel que YY1 (Yin Yang 1), IRF1 (interferon regulatory factor-1), p53, NF- κ B ou MYC, et agir comme un corépresseur ou co-activateur transcriptionnel sur les promoteurs des gènes cibles (Li et al., 2008). NPM1 peut par exemple interagir avec le facteur de transcription AP2 α et agir comme un corépresseur de la transcription en recrutant des histones déacétylases (Liu et al., 2007). De plus, NPM1 peut interagir et inhiber HEXIM1 (hexamethylene bis-acetamide inducible 1), un régulateur négatif de la transcription par l'ARN polymérase II (ARN pol II) (Gurumurthy et al., 2008). Enfin, NPM1 est impliquée dans la régulation de la transcription des ARNr effectuée par l'ARN polymérase de type I dans le nucléole, en recrutant le facteur de transcription spécifique de ce type de transcription TAF(I)48 (TATA box binding protein (TBP)-associated factor), et en réorganisant la chromatine (Bergstralh et al., 2007)(Murano et al., 2008).

vi. Régulation de l'apoptose

Les différents mécanismes moléculaires dans lesquels est impliquée NPM1 ont des conséquences sur la prolifération cellulaire (biogenèse des ribosomes, régulation de facteurs de transcription clés...) mais aussi sur le processus d'apoptose. NPM1 semble avoir un rôle double dans la régulation de l'apoptose comme un acteur pro ou anti-apoptotique, suivant le contexte cellulaire et son niveau d'expression.

Il a pu être observé que le niveau d'expression de NPM1 diminuait dans les cellules sortant du cycle ainsi que dans les cellules s'engageant dans l'apoptose (You et al., 1999). La surexpression de NPM1 induite dans la lignée LAM HL-60 diminue la sensibilité des cellules à l'apoptose, et à l'inverse les cellules NPM1 déficientes sont moins résistantes à l'apoptose. De manière intéressante, l'inhibition de fonction de navette noyau/cytoplasme de NPM1 conduit également à une plus grande sensibilité à l'apoptose (Hsu and Yung, 2000)(Brady et al., 2004). L'induction de l'apoptose par irradiations aux UV de cellules fibroblastiques dans lesquelles NPM1 est surexprimée, a mis en évidence que la résistance à l'apoptose est provoquée par une réduction de l'activité de IRF1 (Interferon regulatory factor 1) induite par NPM1. IRF1 est notamment un facteur de transcription tumeur suppresseur impliqué dans l'apoptose induite après dommages à l'ADN, et dans l'arrêt du cycle cellulaire (Kondo et al., 1997). De plus, NPM1 peut également former dans le noyau un complexe avec PIP (Phosphatidylinositol (3,4,5)-triphosphate) en réponse à des facteurs anti-apoptotiques. En effet, NPM1 est un effecteur en aval de la voie PI3 Kinases en fonctionnant comme un récepteur à PIP, pour moduler l'effet anti-apoptotique de NGF (Nerve Growth Factor) en inhibant la fragmentation de l'ADN par CAD (Caspase-activated DNase) (Ahn et al., 2005).

D'autres études ont montré un effet opposé de NPM1 dans la régulation de l'apoptose, plus particulièrement via ARF et p53. ARF est une protéine nucléolaire qui est impliquée dans le déclenchement de l'arrêt du cycle cellulaire et du programme d'apoptose en réponse à des signaux apoptotiques. Dans ce contexte, ARF inhibe le régulateur négatif de p53, MDM2, en le délocalisant dans le nucléole, ce qui stabilise p53 et permet la réponse transcriptionnelle adaptée. Il a été montré que la surexpression de NPM1 augmentait significativement la stabilité d'ARF, alors que l'inhibition de NPM1 déstabilise ARF (Kuo et al., 2004). En cas de stress cellulaire, NPM1 peut directement inhiber MDM2, augmentant *in fine* la stabilité de p53 (Figure 13). De manière intéressante, ARF peut également réguler l'activité de NPM1. En effet, la surexpression de ARF bloque l'activité de navette de NPM1, augmente la dégradation de NPM1 et en définitive réduit la maturation des ribosomes (Brady et al., 2004).

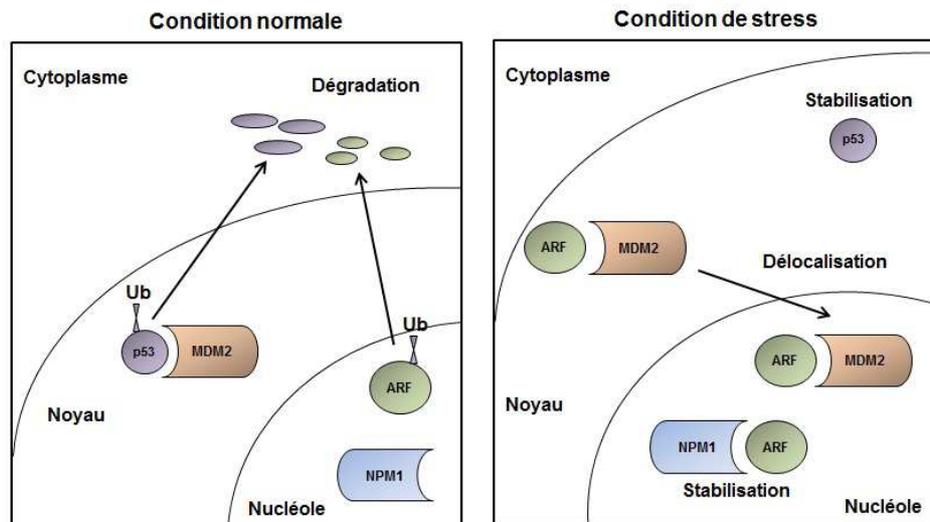


Figure 13 : Représentation schématique du mécanisme de régulation de NPM1 sur la voie ARF/p53. En condition normale, MDM2 interagit avec p53 dans le noyau et conduit à sa dégradation. NPM1 interagit peu avec ARF et ne le stabilise pas. En condition de stress, NPM1 stabilise ARF et ce-dernier délocalise MDM2 dans le nucléole, stabilisant p53.

Ainsi, il apparaît que NPM1 remplit des fonctions essentielles dans le contrôle de la croissance, le maintien de la stabilité génomique et la modulation de la réponse apoptotique. Ses fonctions sont principalement dépendantes de son activité de chaperone d'histones, de ses interactions avec de nombreux partenaires protéiques, de l'activation de programmes transcriptomiques spécifiques, et de sa localisation cellulaire. La perte de fonction de NPM1 ou la dérégulation de son expression conduisent potentiellement à un fort pouvoir transformant.

II) NPM1 dans les cancers

NPM1 est fréquemment impliquée dans les cancers, soit par le biais de la dérégulation de son expression, soit par le biais de mutations. Selon le contexte cellulaire, NPM1 peut agir en tant qu'oncogène ou en tant que suppresseur de tumeur.

a. NPM1 en tant qu'oncogène

On observe des surexpressions de NPM1 dans différents cancers solides tels que le cancer du foie, de la prostate, du côlon ou le carcinome ovarien. La surexpression de NPM1 dans les cellules tumorales est corrélée à des caractéristiques de transformation néoplasique telles que l'augmentation de la prolifération et l'inhibition de l'apoptose.

L'impact de NPM1 dans la prolifération cancéreuse semblerait dépendre de ses fonctions dans la biogenèse des ribosomes, de la réparation de l'ADN et de la duplication des centrosomes. En effet, Boon *et al.* ont mis en évidence dans des cellules de neuroblastomes que MYC régulaait positivement

l'expression de NPM1 ainsi que la machinerie de synthèse protéique (Boon et al., 2001). De plus, il a été montré que l'inhibition de NPM1 par ARN interférence dans la lignée cellulaire Hela, retardait la progression du cycle cellulaire et l'entrée en mitose des cellules, potentiellement par sa fonction dans la régulation des fuseaux mitotiques (Jiang and Yung, 1999).

La capacité de NPM1 à inhiber l'apoptose pourrait jouer un rôle important dans le développement des tumeurs. En effet, la surexpression de NPM1 dans la lignée fibroblastique murine NIH 3T3 conduit à une transformation maligne de ces cellules (Kondo et al., 1997). Les auteurs de ce travail ont suggéré que cet effet transformant serait notamment la conséquence de l'inhibition de l'activité de IRF1 par NPM1. Par ailleurs, la croissance des cellules cancéreuses est dépendante de la réponse à l'hypoxie au cours de la progression tumorale. Li *et al.* ont montré que la surexpression de NPM1 protégeait les cellules de la mort cellulaire induite par l'hypoxie, en inhibant l'activité de p53 (Li et al., 2004). En effet, la surexpression de NPM1 dans différentes lignées cancéreuses humaines en condition hypoxique provoque une augmentation de la survie cellulaire et une inhibition de voies pro-apoptotiques. De manière intéressante, cet effet n'est pas observé dans les lignées p53 négatives surexprimant NPM1, suggérant que l'effet anti-apoptotique de NPM1 est dépendant de sa capacité à inhiber p53.

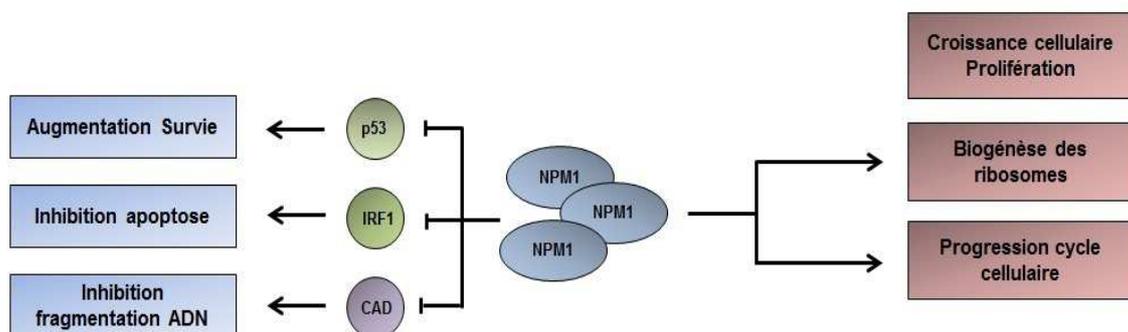


Figure 14 : Résumé schématique des conséquences de la surexpression de NPM1 dans les cancers. La surexpression de NPM1 peut conduire à l'augmentation de la croissance cellulaire et de la prolifération, via une hyperactivité de la machinerie des ribosomes et l'impact de NPM1 dans la progression du cycle cellulaire. La surexpression de NPM1 peut conduire également à l'inhibition de l'apoptose et à l'augmentation de la survie en réprimant l'activité de IRF1, p53 et CAD.

b. NPM1 en tant que tumeur suppresseur

À l'inverse des cancers solides, on observe une perte de fonction de NPM1 dans les hémopathies malignes, par le biais de translocations chromosomiques, de délétions ou de mutations. Les lymphomes anaplasiques à grandes cellules (LAGC) peuvent présenter des fusions *NPM-ALK* (Anaplastic Lymphoma Kinase), les leucémies aiguës promyélocyaires peuvent présenter des fusions

NPM1-RAR α , et les LAMs peuvent présenter des fusions *NPM-MLF1* (Myeloid Leukemia Factor 1) (Figure 15) (Naoe et al., 2006). La contribution de *NPM1*, dans le cas de ces translocations, semble être d'activer le potentiel oncogénique de ses partenaires en étant une interface de dimérisation. Il a été montré que la capacité de transformation de la protéine de fusion *NPM-ALK* ne dépendait que de sa capacité à former des oligomères par le biais de *NPM1*. En effet, différentes protéines mutantes *NPM1-ALK*, incapables de se dimériser, ont été décrites comme non tumorigéniques (Bischof et al., 1997). Il a également été mis en évidence que les syndromes myélodysplasiques pouvaient présenter des délétions 5q (contenant *NPM1*). Le rôle tumeur suppresseur de *NPM1* a notamment été mis en évidence dans des modèles murins de knock-out de *NPM1*. La délétion de *NPM1* chez la souris a montré que *NPM1* était nécessaire au maintien de l'intégrité de l'ADN. En effet, les souris *NPM1* haploinsuffisantes développent des pathologies similaires aux myélodysplasies, avec une instabilité génomique (aneuploïdie), une augmentation du nombre de centrosomes et une augmentation de l'activation des points de contrôle des dommages à l'ADN (Grisendi et al., 2005). Mais en 2005, Falini *et al.* ont mis en évidence une nouvelle mutation de *NPM1* présente dans 55% des LAM-CN (Falini et al., 2005).

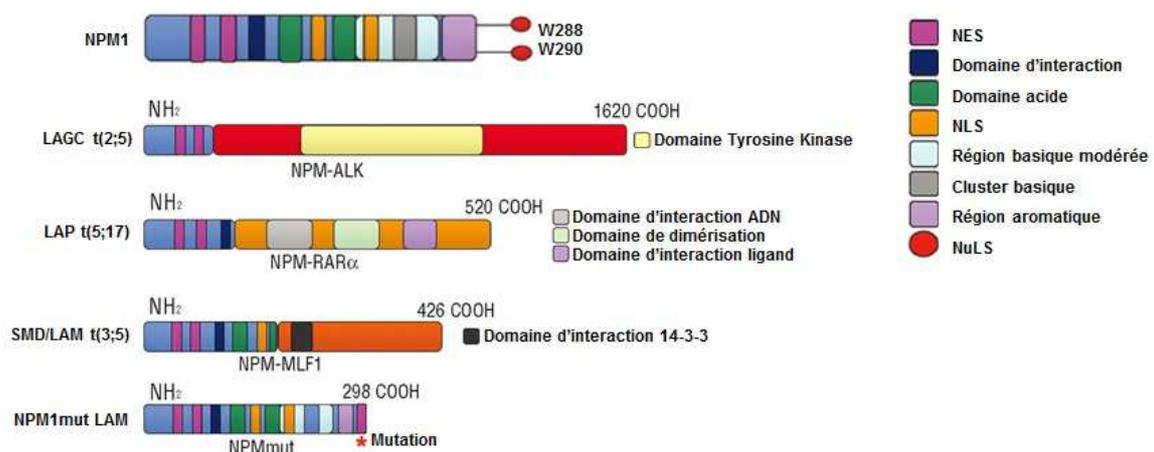


Figure 15 : Domaines structuraux et fonctionnels de *NPM1* sauvage et mutante dans différentes hémopathies malignes. *NPM1* sauvage présente différents motifs nécessaires à sa fonction et sa localisation (Cf Partie II, Chapitre I et sous chapitre a. "Structure"). La fusion *NPM-ALK* est constituée du domaine d'oligomérisation de *NPM1* et du domaine tyrosine kinase de *ALK*. La fusion *NPM1-RAR α* est constituée du domaine d'oligomérisation de *NPM1* et des domaines de dimérisation et d'interaction avec l'ADN et les ligands de *RAR α* . La fusion *NPM-MLF1* est constituée du domaine de dimérisation et d'une partie de la région centrale de *NPM1* et de la protéine complète *MLF1*. *NPM1* muté dans les LAMs consiste en une insertion de 4 nt décalant le cadre de lecture et conduisant à la perte des tryptophanes 288 et 290 et du signal de localisation nucléolaire (Adapté de Falini et al., 2007).

III) La mutation NPM1 dans les LAMs

La mutation du gène *NPM1* est la mutation la plus récurrente dans les LAM-CN (Marcucci et al., 2011). On dénombre une quarantaine de mutations, toujours hétérozygotes, dont 3 mutations présentent dans 90% des LAMs *NPM1* muté. Mis à part quelques rares mutations dans les exons 9 et 11, les mutations sont majoritairement des insertions de 4 nt dans l'exon 12, décalant le cadre de lecture pour coder une protéine NPM1 différente sur les 7 derniers résidus (Figure 16).

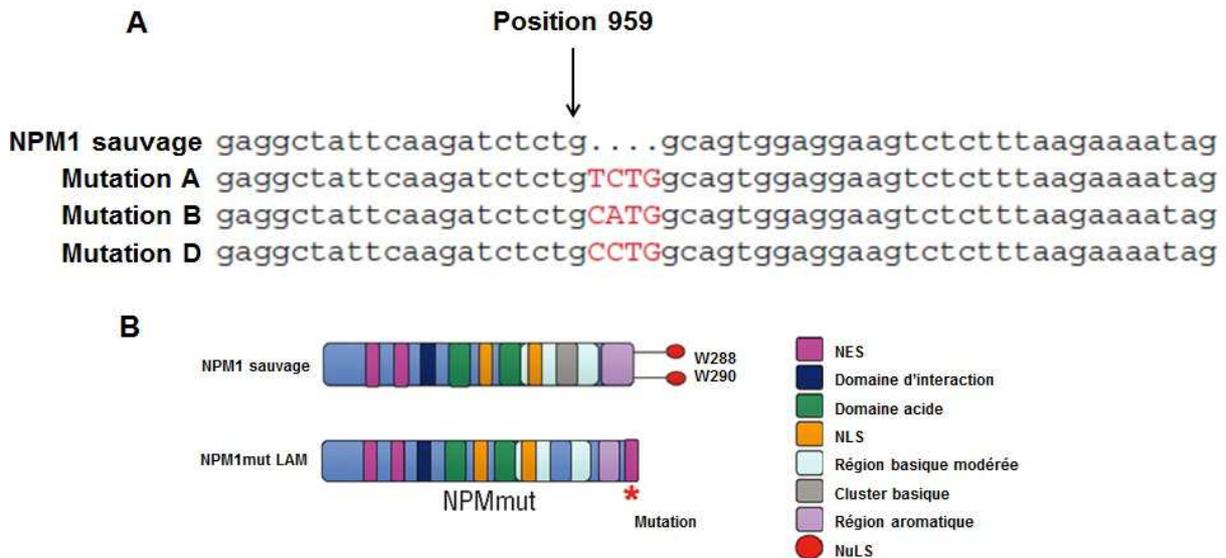


Figure 16 : Mutations *NPM1* dans les LAMs. A. Séquences des 3 mutations par insertion dans l'exon 12 de *NPM1* les plus fréquentes dans les LAMs. B. Impact de ces mutations au niveau protéique. Il y a une perte des tryptophanes en position 288 et 290 et du signal de localisation nucléolaire, ainsi que la création d'un nouveau signal d'export nucléaire (Adapté de Falini et al., 2007).

La mutation principale, la mutation A, consiste en une insertion TCTG en position 959 du gène. Toutes les mutations conduisent à la perte d'au moins un des deux résidus tryptophanes 288 et 290 nécessaires à la localisation nucléolaire, et à la création d'un nouveau signal d'export nucléaire dans la partie N-Terminale de la protéine (Falini et al., 2006a) (Figure 16). La conséquence de cette mutation est la délocalisation du nucléole vers le cytoplasme de NPM1 mutée via le complexe Ram-Crm1 (Figure 17) (Falini et al., 2005). En effet, l'inhibition de Crm1 dans la lignée LAM OCI-AML3 contenant la mutation A de *NPM1*, conduit à un blocage de NPM1 mutée dans le nucléoplasme (Falini et al., 2006b). De plus, NPM1 mutée conserve ses domaines d'oligomérisation conduisant à la formation d'hétérodimères NPM1 mutée/NPM1 sauvage. Ainsi, une grande proportion de NPM1 sauvage est également délocalisée dans le cytoplasme, provoquant une "perte de fonction" de NPM1 sauvage dans le noyau.

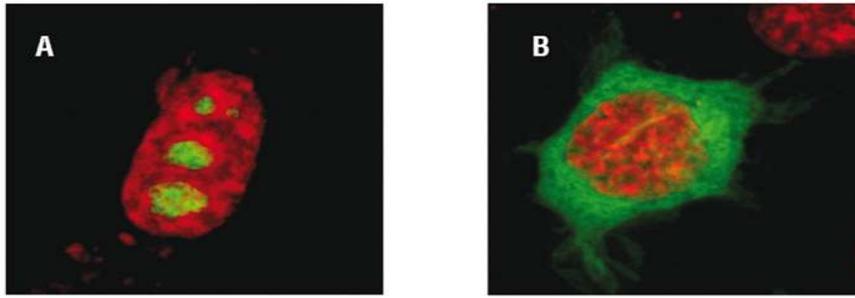


Figure 17 : Analyse par immunofluorescence de la localisation cellulaire de NPM1 dans des fibroblastes murins NIH3T3 transfectés avec *NPM1* sauvage ou muté couplé à eGFP. A. La protéine NPM1 sauvage (vert) s'accumule et est localisée au niveau nucléolaire. B. La protéine NPM1 mutante (vert) est délocalisée du nucléole vers le cytoplasme. Les noyaux sont colorés à l'iodure de propidium (rouge). (Adapté de Falini et al., 2007).

La mutation *NPM1* est associée à des caractéristiques phénotypiques, biologiques et transcriptomiques précises comme l'augmentation du pourcentage de blastes dans la moelle, de lactate déshydrogénase dans le sérum et du nombre de leucocytes. Typiquement, les blastes des LAMs *NPM1* muté expriment fortement le marqueur CD33 mais pas le CD34, et sont ainsi d'un phénotype myéloïde plutôt différencié. En effet, la majorité des LAMs *NPM1* muté présente un FAB de type M4 voire M5. D'un point de vue clinique, ces mutations sont associées à un meilleur pronostic (Schlenk et al., 2008), plus particulièrement en absence de mutation *FLT3-ITD*. Ainsi, le génotype *NPM1* muté/*FLT3* non muté dans les LAM-CN a été ajouté dans la catégorie de risque favorable (Döhner et al., 2010). Cependant, il ne semblerait pas qu'il y ait de différence en terme de pronostic clinique entre les différentes mutations affectant *NPM1* dans les LAMs (Pastore et al., 2014) (Figure 18).

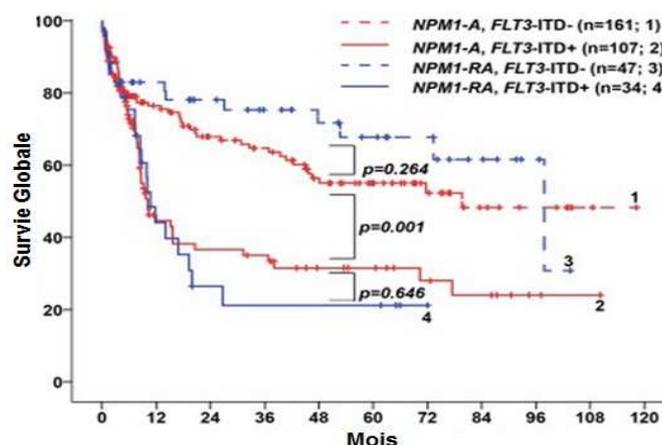


Figure 18 : Impact des mutations du gène *NPM1* en association avec *FLT3-ITD* dans la LAM sur la survie globale des patients. Il n'y a pas de différence sur la survie entre les mutations majoritaires (*NPM1-A*) et les mutations rares (*NPM1-RA*) de *NPM1*, en présence ou absence de *FLT3-ITD* (Adapté de Pastore et al., 2014).

Comme les mutations de *NPM1* sont stables au cours de la maladie, elles sont d'excellents marqueurs pour suivre la maladie résiduelle. En effet, plusieurs études ont montré l'intérêt d'évaluer l'évolution de l'expression de *NPM1* muté par RT-qPCR chez les patients après rémission, et Schnittger *et al.* ont montré que ce niveau d'expression était clairement associé au pronostic des patients, notamment de la rechute (Schnittger *et al.*, 2009).

D'un point de vue transcriptomique, il a été décrit de nombreuses signatures associées spécifiquement à la présence de *NPM1* muté dans les LAMs. On peut citer la surexpression des gènes *HOX*, *MEIS1*, *PBX3* (pre-B-cell leukemia homeobox 3) (Mullighan *et al.*, 2007)(Verhaak *et al.*, 2005), la surexpression de miR-10a et b, miR-29a, b et c, ainsi que la sous expression du miR-204 (ciblant *HOXA10* et *MEIS1*) et miR-128a (Garzon *et al.*, 2008). Très récemment, Garzon *et al.* ont également mis en évidence des profils d'expression de lncRNAs particuliers associés à différentes mutations dont *NPM1* (Garzon *et al.*, 2014).

L'ensemble de ces caractéristiques a conduit à considérer les LAMs *NPM1* muté comme des entités particulières, d'où leur inclusion dans le classement OMS de 2008 en tant qu'entité provisoire.

D'un point de vue mécanistique, le mode d'action de *NPM1* mutée dans la leucémogénèse n'est pas clairement défini. Leong *et al.* ont mis en évidence que *NPM1* mutée exerçait un effet anti-apoptotique et inhibait la différenciation myéloïde en transduisant *NPM1* muté dans des cellules humaines CD34⁺ de sang de cordon (Leong *et al.*, 2010). A l'inverse, l'inhibition de *NPM1* mutée par l'utilisation de la Deguelin (une drogue ciblant *NPM1* mutée) ou par ARN interférence dans la lignée LAM OCI AML3, restaure la capacité de différenciation des cellules hématopoïétiques et la sensibilité à l'apoptose (Yi *et al.*, 2015a). Le changement de localisation de *NPM1* du noyau au cytoplasme semble être l'élément clé de la transformation tumorale dans les LAMs *NPM1* muté. Mais savoir si les effets observés sont la conséquence d'un gain de fonction de *NPM1* mutée dans le cytoplasme et/ou l'haploinsuffisance de *NPM1* sauvage dans le noyau, reste à déterminer. On peut supposer que la délocalisation de *NPM1* dans le cytoplasme conduit également à la délocalisation de ses nombreux partenaires (ARF, HEXIM1, APEX1, NF-κB, etc...). En effet, la transfection de *NPM1* muté dans la lignée NIH 3T3 conduit à une délocalisation de *NPM1* mutée, *NPM1* sauvage et ARF du noyau au cytoplasme (Figure 19) (Bolli *et al.*, 2009). Par ailleurs, l'utilisation de différents modèles murins a permis de mieux comprendre le rôle oncogénique de *NPM1* mutée. Les modèles de knock-in murin avec *NPM1* muté humain (la forme la plus représentée mutation A) conduisent à l'apparition de lésions initiatrices de LAM, à une surexpression des gènes *HOX*, et un tiers des souris développent tardivement des LAMs (Vassiliou *et al.*, 2011).

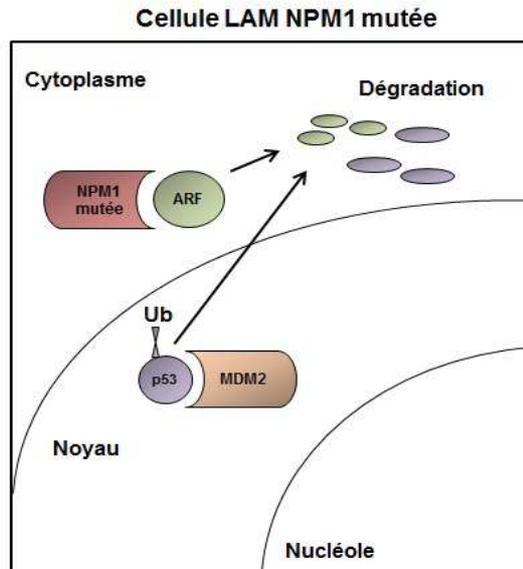


Figure 19 : Impact de NPM1 mutée sur la voie ARF/p53. Dans les cellules normales, NPM1 est localisée dans le nucléole et est requise pour la stabilité d'ARF. En condition de stress, ARF peut inactiver MDM2 conduisant à la diminution de la dégradation de p53. Dans les blastes LAM *NPM1* muté, NPM1 mutée délocalise ARF dans le cytoplasme où ARF sera dégradée. Par conséquent, MDM2 peut induire la réaction d'ubiquitination et la dégradation de p53 même en cas de stress cellulaire.

La plupart des mécanismes oncogéniques décrits se fondent sur la capacité de NPM1 à interagir avec des partenaires intervenant dans le contrôle de la prolifération et de l'apoptose. Cependant, les autres fonctions de NPM1 telles que la biogenèse des ribosomes, la duplication des centrosomes et son activité de chaperone d'histones, n'ont à ce jour jamais été décrites dans le processus de leucémogénèse. De plus, on ne peut pas exclure l'intervention d'autres partenaires non décrits de NPM1 participant au développement de la LAM, tels que les lncRNAs.

I) Historique de la "matière noire" du génome

Il y a maintenant une cinquantaine d'années que Jacob et Monod ont établi le rôle central de la molécule d'ARN en tant qu'ARN messager, intermédiaire entre l'ADN et la protéine (Jacob and Monod, 1961). Depuis, plusieurs études ont mis en avant la présence d'ARNs non codants composant une portion importante du transcriptome total (Weinberg and Penman, 1968)(Paul and Duerksen, 1975)(Salditt-Georgieff and Darnell, 1982). Une partie de ces ARNs non codants est expliquée par la forte présence d'ARNs non codants impliqués dans la machinerie de traduction et sa régulation (ARNr, ARNt, RNase P, SRP-7S ARNs). La plupart des chercheurs ont alors supposé que le reste des ARNs non traduits n'avaient pas de fonction et constituaient uniquement un "bruit de fond" génomique. A partir des années 90, ces séquences non codantes ont peu à peu été associées par des chercheurs tels que John Mattick à de potentiels rôles régulateurs dans le processus d'évolution et dans le développement (Mattick, 1994). Mais les progrès plus récents en termes de séquençage haut débit, l'automatisation du séquençage Sanger pour la cartographie des ESTs dans les années 90 (Schuler et al., 1996), l'avènement des Microarrays et bien évidemment l'explosion du RNA-sequencing lors de la dernière décennie, ont montré que 70 à 90% du génome était transcrit (Birney, 2007)(Carninci et al., 2005)(Kapranov et al., 2005) et était majoritairement non codant (Djebali et al., 2012).

Certaines classes d'ARNs non codants sont bien caractérisées moléculairement et fonctionnellement comme les ARNr, les ARNt, les snARNs, les snoARNs, et plus récemment les microARNs. De manière tout à fait intéressante, la communauté scientifique a pu mettre en évidence que la majorité des ARNs non codants était constituée d'ARNs relativement longs et structurellement très hétérogènes, appelés longs ARNs non codants ou lncRNAs. Ces ARNs sont définis de manière arbitraire comme des ARNs de plus de 200nt et n'ayant à priori aucun potentiel codant. Ce seuil de 200nt permet essentiellement de faire la distinction avec les autres classes d'ARNs non codants généralement en deçà de ce seuil (microARNs, snARNs, etc...). Bien qu'ils constituent la plus grande part du transcriptome, cela n'est en aucun cas une preuve suffisante démontrant leur fonctionnalité biologique. L'argument mis en avant pour justifier une transcription massive de lncRNAs non fonctionnels est que celle-ci pourrait résulter de l'imperfection de la machinerie de transcription. En effet, l'ARN polymérase II (ARN pol II) peut initier une transcription non spécifique chez les eucaryotes conduisant à 90% de "bruit génomique" (Struhl, 2007). De plus, il semblerait que la transcription puisse être bidirectionnelle à partir du promoteur, ou dépasser le transcrit "légitime" pour conduire à la fuite de produits de transcription non spécifiques (Ebisuya et al., 2008).

Mais ce "bruit" est constitué d'exemples de plus en plus nombreux de lncRNAs ayant un impact fonctionnel. Les premiers à avoir été décrits étaient les lncRNAs *H19*, impliqué dans l'embryogenèse en 1990 (Brannan et al., 1990), et *XIST* (X-inactive specific transcript), impliqué dans l'inactivation du chromosome X en 1992 (Brown et al., 1992). Depuis lors, la quantité de lncRNAs annotés chez l'homme a considérablement augmenté grâce, aux nouvelles technologies de séquençage (notamment les protocoles brin-spécifique à partir de 2008), le séquençage du génome humain en 2000, ainsi qu'au projet ENCODE (Encyclopedia of DNA Elements). Mais le nombre de lncRNA annotés est variable selon les études et les bases de données utilisées (Harrow et al., 2006). A l'heure actuelle, on dénombre chez l'homme plus de 55 000 gènes transcrits en plus de 95 000 lncRNAs selon la base de données NONCODE V4.0 (Xie et al., 2014), la grande majorité de ces lncRNAs restant encore très peu décrite et caractérisée.

II) Structure et conservation des lncRNAs

a. Structure des lncRNAs

i. Généralités

Comme cité précédemment, les lncRNAs sont définis comme des ARNs d'une longueur supérieure à 200nt et n'ayant pas de potentiel codant. Mis à part ces caractéristiques, il n'y a pour l'heure aucun archétype commun ou séquence consensus permettant de classifier les lncRNAs. Il semble donc qu'ils constituent un groupe très hétérogène d'ARNs sur le plan structural. Par ailleurs, il n'y a globalement pas non plus de différence majeure entre les lncRNAs et leurs pendants codants puisque les lncRNAs peuvent être cappés, polyadénylés, épissés, édités, multi-exoniques et soumis aux mêmes régulations transcriptionnelles et post-transcriptionnelles que les ARNm. Il a aussi été observé des similarités entre les lncRNAs et les 3'UTR (Untranslated Region) des ARNm en terme de séquence (Pourcentage GC faible) et de structure secondaire (structures moins stables par rapport aux ORFs et aux 5'UTR des ARNm) (Niazi and Valadkhan, 2012).

Cependant, quelques différences sont à noter :

Avec un panel de plus de 8000 lncRNAs, Cabili *et al.* ont pu montrer qu'en moyenne un lncRNA était plus petit et possédait moins d'exons qu'un mRNA (Cabili et al., 2011), même si certains macroARNs dépassent les 200k nt comme le lncRNA *XACT* (X Active Coating Transcript) (Vallot et al., 2013) (Figure 20). Cette étude a également mis en évidence une tendance des lncRNAs à contenir en moyenne 2 exons et à être soumis à au moins un épissage alternatif. Ces résultats ont été confirmés par une étude en 2012 sur le catalogue GENCODE montrant que 42% des lncRNAs présentent 2 exons alors que seulement 6% des ARNm présentent 2 exons (les ARNm présentant généralement plus que 2 exons) (Derrien et al., 2012). Cependant, il semblerait que les exons et introns des lncRNAs

soient plus longs que ceux des ARNm. Les auteurs de ce travail ont également montré que 98% des lncRNAs pourraient être épissés et que plus de 25% des lncRNAs présenteraient un épissage alternatif avec au moins 2 isoformes. En effet, la majorité des introns sont flanqués des sites canoniques d'épissage (GT/AG). Le lncRNA *PCBP1-AS1* (Poly(rC) binding protein 1-Anti Sens 1) par exemple présente jusqu'à 40 isoformes. Cette analyse a également mis en évidence que 39% des lncRNAs contenaient au moins 1 des 6 motifs poly(A) communs alors que 51% des ARNm contenaient ces motifs. Il semblerait ainsi qu'une grande proportion des lncRNAs soit transcrite par l'ARN pol II, bien qu'il ait été montré que la transcription de certains lncRNAs pouvait dépendre de l'ARN pol III (Pagano et al., 2007) ou encore de l'ARN pol I, comme les IGS RNAs (Intergenic Spacer RNAs) situés dans les régions de gènes ribosomiaux (Jacob et al., 2012). Les demi-vies sont également très variables (de 30min à 16h) et fortement dépendantes de la localisation cellulaire des lncRNAs. De plus, la stabilité des lncRNAs paraît être en moyenne légèrement inférieure à celle des ARNm (Clark et al., 2012).

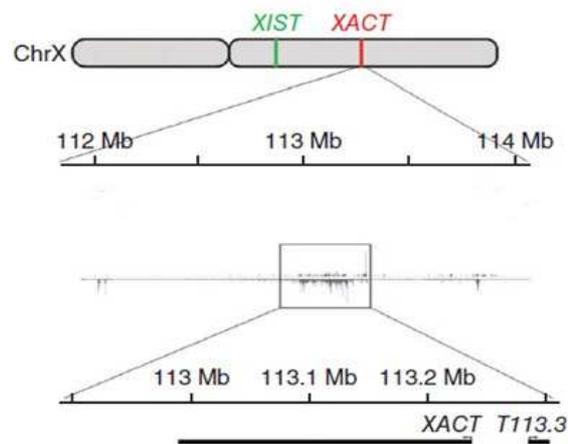


Figure 20 : Représentation de *XACT* sur le chromosome X par rapport au lncRNA *XIST*. Les données RNA-seq présentent une transcription continue sur plus de 250kb (Adapté de Vallot et al., 2013).

Comme pour les protéines, la structure d'un ARN se distingue en 3 catégories d'organisation : primaire, secondaire et tertiaire.

ii. Structure primaire

La structure primaire fait référence à la séquence de l'acide nucléique. Il n'y a pas de différence majeure entre les lncRNAs et les ARNm du point de vue de la séquence hormis l'absence d'ORF. Cependant, des lncRNAs aux séquences particulières ont été récemment mis en évidence : ce sont des lncRNAs introniques situés entre des snoARNs (sno-lncRNAs). L'expression de ces lncRNAs est dépendante de la machinerie de maturation des snoARNs qui ne processerait pas l'intron (le lncRNA)

générant un lncRNA. Ce dernier présente les séquences de deux snoARNs à ses extrémités 3' et 5' (Yin et al., 2012) (Figure 21).

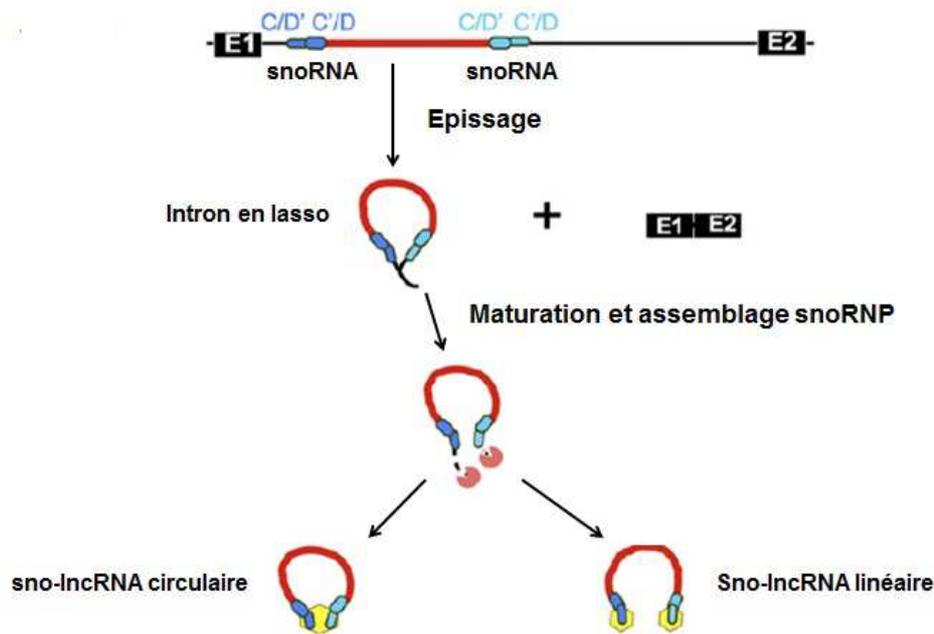


Figure 21 : Modèle de processing snor-lncRNAs. Les snoARNs sont processés à partir du gène hôte mais la séquence interne n'est pas excisée, conduisant en un lncRNA avec des snoARNs aux extrémités. Cette structure peut se circulariser ou rester linéaire (Adapté de Yin et al., 2012).

Une large portion des séquences exoniques des lncRNAs provient d'éléments génétiques mobiles, représentant par ailleurs 40% du génome humain. En se basant sur une analyse croisée entre des données RNA-seq et GENCODE, Kelley et Rinn ont estimé que 41% des séquences des lncRNAs dériveraient d'éléments transposables, et que 83% des lncRNAs contiendraient au moins un fragment d'élément transposable (Kelley and Rinn, 2012). Par conséquent, la plupart des transcrits lncRNAs contiennent des combinaisons de multiples fragments répétés. Cependant, la forte fréquence de séquences répétées et d'éléments transposables au sein de lncRNAs ne semble pas préjudiciable, puisqu'il a été décrit dans la littérature plusieurs cas de lncRNAs fonctionnels tout en étant riche en séquences répétées (Carrieri et al., 2012). Les régions centromériques sont particulièrement riches en séquences répétées et considérées comme des régions ne possédant pas de gènes codants. Il a tout de même été mis en évidence que ces régions pouvaient être transcrites, notamment en lncRNAs. En effet, il a été récemment montré par Quénet et Dalal l'expression d'un lncRNA en phase G1 du cycle cellulaire, qui s'est avérée être nécessaire à l'organisation de la structure chromatinienne par le recrutement de l'histone centromérique CENP-A (Quénet and Dalal, 2014).

iii. Structure secondaire

La structure secondaire implique des paires de bases Watson-Crick et des régions non appariées, qui sont supposées être la base fonctionnelle des lncRNAs (Wan et al., 2011). Cependant, les données de structures secondaires des lncRNAs en relation avec leurs fonctions sont très peu décrites. Il a été rapporté par exemple que le lncRNA *MALAT1* (metastasis associated lung adenocarcinoma transcript 1) pouvait être processé en un ARN plus petit comportant une structure ARNt-like (Wilusz et al., 2008) (Figure 22A). Une des études les plus complètes sur la structure secondaire d'un lncRNA a été publiée en 2012 sur *SRA RNA* (Steroid receptor RNA activator) (Novikova et al., 2012). Cette étude a révélé une organisation structurale complexe de *SRA* comprenant 4 domaines et plusieurs structures secondaires plus ou moins larges comme des hélices et des tiges boucles (Figure 22B). Le lncRNA *HOTAIR* (HOX Transcript Antisense RNA) semble participer à l'échafaudage de complexes protéiques via des motifs structuraux complexes en 5' et 3' permettant l'interaction directe avec différentes protéines (Tsai et al., 2010a) (Somarowthu et al., 2015). La structure des lncRNAs semblerait donc avoir un rôle important pour les interactions avec d'autres molécules, comme avec des régulateurs de la chromatine et même avec d'autres acides nucléiques (ADN, microARNs, etc...) (Paraskevopoulou et al., 2013). Il existe également des lncRNAs se circularisant comme les circRNAs ou ciRNAs (Circular intronic RNAs), pouvant être dérégulés dans certaines pathologies et impliqués dans divers processus, comme dans la régulation de la transcription et dans la séquestration de microARNs (Memczak et al., 2013)(Qu et al., 2015).

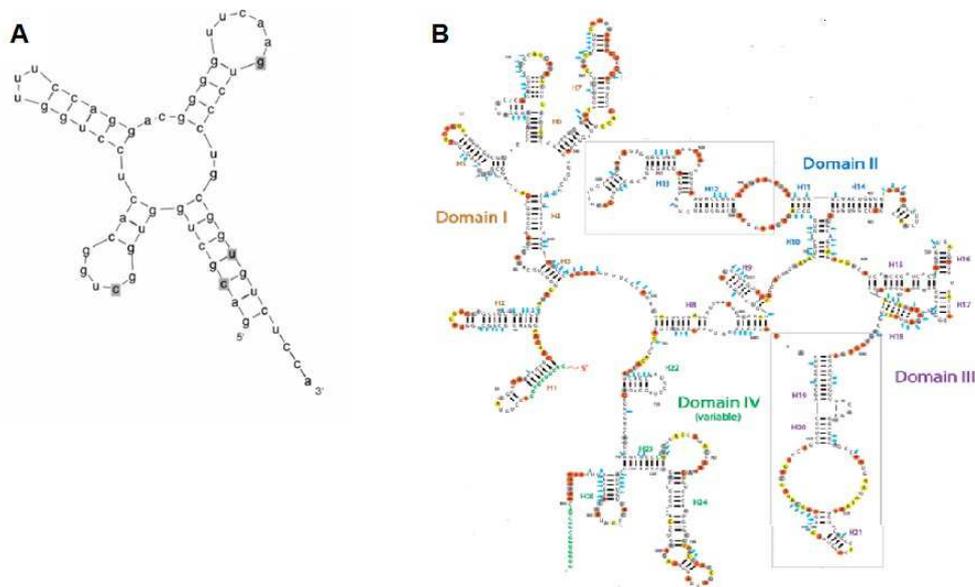


Figure 22 : Structure des lncRNAs. A. Prédiction de structure de *mascRNA* en feuille de trèfle ARNt-like (Adapté de Wilusz et al., 2008). B. Structure complexe de *SRA* déterminée expérimentalement. Ce lncRNA est constitué notamment de 4 domaines distincts et de 25 hélices (Adapté de Novikova et al., 2012).

iv. Structure tertiaire et quaternaire

Il n'y a pour l'heure que très peu de données concernant la structure tertiaire des lncRNAs. Par contre, un exemple intéressant de structure quaternaire a été mis en évidence pour le lncRNA *NEATI* (Nuclear Paraspeckle Assembly Transcript 1). *NEATI* peut former des interactions quaternaires avec différentes protéines nécessaires à la formation des fibres appelées "paraspeckles" dans le noyau, et dont la fonction n'est pas totalement déterminée (Naganuma and Hirose, 2013) (Figure 23). Même si la structure n'a pas clairement été étudiée, il est peu probable que *NEATI* forme des structures semblables aux ribosomes, mais il est suggéré que le lncRNA formerait plutôt des structures proches du complexe télomérase.

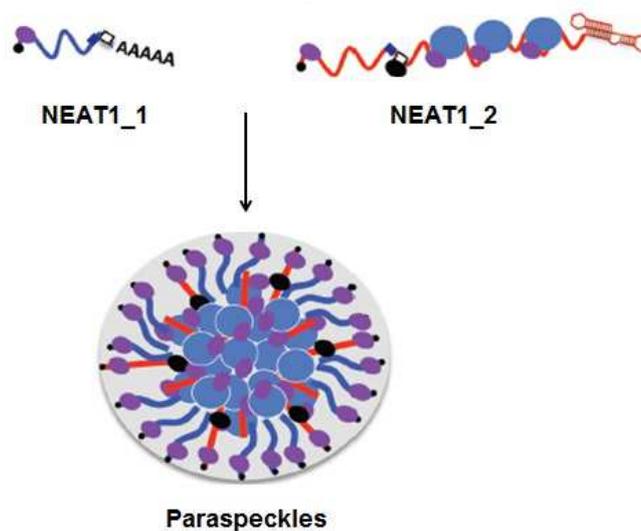


Figure 23 : Formation des paraspeckles nucléaires à partir des deux isoformes de *NEATI*, *NEATI-1* et *NEATI_2* s'associant à différentes protéines (Adapté de Naganuma and Hirose, 2013).

b. Conservation au cours de l'évolution

Plusieurs études ont démontré que les exons des lncRNAs présentaient une plus faible conservation par rapport à celle des ARNm. Cependant, les lncRNAs présentent une meilleure conservation que des séquences aléatoires du génome renforçant leur probable importance fonctionnelle (Marques and Ponting, 2009)(Guttman et al., 2009). Ce faible niveau de conservation pourrait refléter une adaptabilité plus élevée des lncRNAs au cours de l'évolution, contrairement aux séquences codantes sous fortes pressions sélectives. De plus, les promoteurs des lncRNAs sont en moyenne plus conservés que leurs exons et presque autant conservés que ceux des ARNm. L'étude menée par Derrien *et al.* sur le catalogue GENCODE semble confirmer ces observations (Derrien et al., 2012). Les auteurs ont également montré qu'environ 30% des lncRNAs étaient primates spécifiques et moins de 1% était spécifique à l'homme. Ceci suggère que l'unité génique du lncRNA

est relativement bien conservée, mais que les séquences ont évolué trop rapidement pour les analyses classiques d'homologie. D'autres études ont pu identifier des lncRNAs orthologues très conservés et spécifiques du cerveau chez la souris, dans des espèces aussi distantes phylogénétiquement que l'opossum ou la poule (Chodroff et al., 2010). De plus Cabili *et al.* ont identifié 993 lincRNAs sur 8195 ayant un transcrit orthologue dans d'autres espèces, de la souris au poisson, avec une homologie relativement modeste (Cabili et al., 2011). Par ailleurs, certains de ces transcrits chez la souris sont annotés comme codants, reflétant soit la présence de petits peptides cryptiques chez l'homme, soit une évolution d'un transcrit non codant à partir d'un gène codant. Il existe également une classe particulière de lncRNAs appelés T-UCR (Ultra Conserved Region) qui sont presque parfaitement conservés entre les espèces (>95% au sein des mammifères) et plus conservés que les gènes codants (Bejerano et al., 2004). De manière tout à fait intéressante, les multiples alignements réalisés par Derrien *et al.* ont également mis en évidence une forte conservation de certaines régions des lncRNAs, ce qui pourrait signifier une plus forte conservation des structures secondaires potentiellement fonctionnelles (Derrien et al., 2012). En effet, des analyses de conservation entre l'homme et la souris ont montré que des portions du lncRNA *XIST* étaient très conservées, suggérant un maintien des domaines fonctionnels de *XIST* à travers l'évolution. A l'inverse, le lncRNA *GAS5* est un lncRNA producteur de snoARNs à partir de ses introns, et la conservation entre homme et souris est uniquement située au niveau des régions contenant les snoARNs, démontrant potentiellement une conservation se limitant à la fonction (Smith and Steitz, 1998).

III) Organisation génomique et classification des lncRNAs

Etant considérée pour l'heure comme un groupe très hétérogène d'un point de vue structural, la classification des lncRNAs est couramment effectuée selon leur position génomique et leur position relative à un gène codant (Figure 24). Cette classification n'indique pas nécessairement une information quant à la fonction ou à l'origine du lncRNA. On peut donc citer plusieurs catégories relatives à un gène codant:

a. Classification relative à un gène codant

Le lncRNA intronique, lorsque celui-ci est situé dans l'intron d'un gène codant et transcrit en sens ou anti-sens par rapport au gène hôte.

Le lncRNA sens lorsque celui-ci chevauche partiellement au moins un exon d'un gène codant sur le même brin.

Le lncRNA anti-sens si ce chevauchement se situe sur le brin opposé. Ces lncRNAs sont plus couramment appelés NATs pour Natural Anti-sens Transcripts, et il a été rapporté que plus de 70% des transcrits présentaient un transcrit anti-sens (Katayama et al., 2005). Le chevauchement peut être

complet ou incomplet et il semblerait que les NATs soient préférentiellement chevauchants en 5' des transcrits sens. De manière intéressante, plusieurs exemples de lncRNAs anti-sens à un gène codant sont impliqués dans l'empreinte parentale comme *KCNQ1OT1* (*KCNQ1* overlapping transcript 1) sur le locus *KCNQ1* (potassium channel, voltage gated KQT-like subfamily Q, member 1) (Kanduri et al., 2006). La paire ARN sens/anti-sens peut également être composée de deux lncRNAs comme dans le cas de *XIST/TSIX* (*Xist* Antisens RNA), deux lncRNAs impliqués dans l'inactivation du chromosome X (Lee et al., 1999).

Le lncRNA bidirectionnel lorsque sa transcription est initiée à proximité et dans le sens opposé du gène codant voisin.

Les lncRNAs intergéniques ou lincRNAs qui sont des unités transcriptionnelles uniques situées entre 2 gènes codants. Un grand nombre de ces derniers ont été identifiés par l'analyse des signatures chromatinienne par ChiP-seq (Chromatin immunoprecipitation-Sequencing). En effet, ces lncRNAs présentent des marques épigénétiques correspondant à une transcription active comme H3K4me3 au niveau du promoteur, ou H3K36me3 le long du gène transcrit (Guttman et al., 2009). Les lincRNAs sont certainement les plus étudiés et décrits des lncRNAs. Ils sont majoritairement transcrits par l'ARN pol II, polyadénylés, épissés, et présentent plusieurs isoformes avec une longueur de 1kb en moyenne. De plus, une étude se basant sur plus de 8000 lincRNAs a mis en évidence que ces ARNs avaient une tendance à être situés à proximité de gènes codants impliqués dans le développement, et à présenter une corrélation significative entre leurs expressions et celles des gènes codants voisins (Cabali et al., 2011).

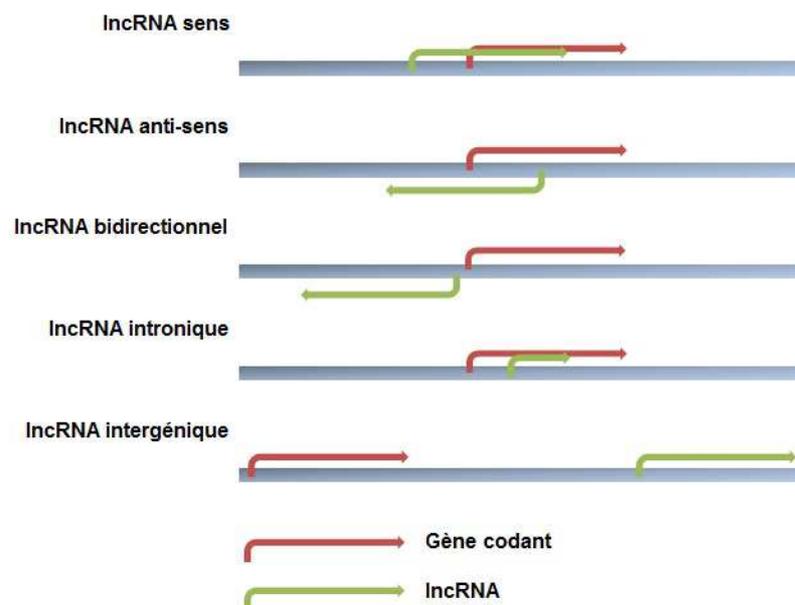


Figure 24 : Classification des lncRNAs selon la position génomique relative aux gènes codants.

b. Classification relative à une région génomique

Il existe deux autres classes de lncRNAs relatives non pas à un gène codant mais à des régions régulatrices : les eRNAs (Enhancer RNA) pour les lncRNAs transcrits à partir d'enhancers, et les pRNAs (Promoter-Associated Long Non Coding RNAs) ou les PROMPTs (PROMoter uPstream Transcripts) pour les lncRNAs transcrits à partir de promoteurs.

Les eRNAs : Les enhancers sont caractérisés essentiellement par la présence d'une mono-méthylation en K4 de l'histone H3 (He et al., 2010). L'association de cette marque avec une acétylation en K27 de l'histone H3 est corroborée avec une activité positive de l'enhancer et à l'inverse, l'association de cette marque avec la tri-méthylation en K27 de l'histone H3 est corroborée avec l'inactivité de l'enhancer (Creyghton et al., 2010). Il a été noté récemment que ces sites présentaient massivement des chevauchements avec les sites de transcription de l'ARN polymérase II, suggérant que les enhancers pourraient être transcrits en ARNs non codants (De Santa et al., 2010). En 2014, Andersson *et al.* ont publié un atlas sur les eRNAs à partir de données de séquençage haut débit et de ChIP-seq sur un large panel de tissus primaires et de lignées cellulaires, conduisant à l'identification de plusieurs milliers de eRNAs (Andersson et al., 2014). Cela a permis d'observer que ces ARNs sont en majorité non polyadénylés, nucléaires et non épissés.

Les PROMPTS : les promoteurs peuvent être bidirectionnels. Ainsi la transcription se réalise dans les deux sens, produisant l'ARNm et un autre ARN en amont, le PROMPT. Ces ARNs non codants partagent les mêmes caractéristiques au niveau du TSS (Transcription Start Site) que le TSS du gène codant associé. Il semblerait que leurs transcriptions permettent d'orienter l'ARN polymérase II dans le sens de l'ARNm, suite à leur dégradation par l'exosome (Ntini et al., 2013) (Figure 25).

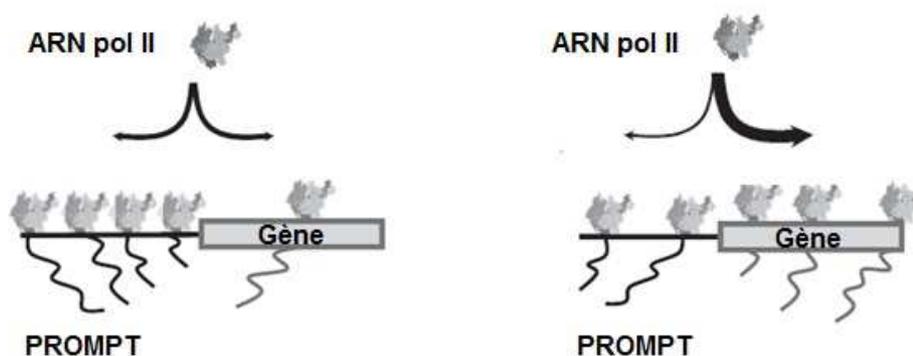


Figure 25 : Mécanisme supposé de transcription des PROMPTs. L'ARN pol II transcrit bidirectionnellement à partir du promoteur l'ARNm et les PROMPTs en amont. L'expression de ces lncRNAs et leur rapide dégradation conduirait à orienter l'ARN pol II majoritairement dans le sens du gène associé.

Les pARNs : une autre classe de lncRNAs moins spécifique est également associée aux promoteurs de gènes codants, les pARNs. Ces lncRNAs ont été décrits comme des ARNs présentant une expression corrélée avec celle du promoteur, et impliqués dans la modulation de l'expression du gène codant associé. En effet, une région riche en îlots CpG du promoteur du gène *SPHK1* (sphingosine kinase 1) peut être transcrite en un ARN non codant intervenant dans la méthylation et la déméthylation du promoteur de *SPHK1*, régulant ainsi la transcription du gène codant (Imamura et al., 2004).

c. Autres classes

Les T-UCRs : dans une classe très spécifique de lncRNAs, on dénombre 481 T-UCRs, associés à des régions ultra-conservées. La plupart de ces lncRNAs sont organisés en cluster et il semblerait que ces éléments soient soumis à une forte sélection évolutive depuis plus de 400 millions d'années, ce qui expliquerait la forte conservation observée du poisson à l'homme. Un tiers de ces T-UCRs est situé dans des "déserts génétiques". Une autre portion de ces ARNs est située à proximité de gènes impliqués dans le développement, ce qui pourrait expliquer leur forte conservation. Par contre, la majorité des T-UCRs sont à cheval sur des exons, dans les introns ou à proximité de gènes codants impliqués dans divers processus, comme la régulation de l'épissage et de la transcription (Bejerano et al., 2004).

Les pseudogènes : la dernière classe relativement particulière de longs ARNs non codants est issue des pseudogènes. Ils sont considérés comme des gènes ancestraux ayant perdu leur potentiel codant par succession de mutations (Balakirev and Ayala, 2003). La plupart de ces gènes sont le produit de duplications en tandem de gènes ou de rétro-transpositions, conduisant à un certain nombre de copies de gènes surnuméraires qui ne sont plus sous de fortes pressions de sélection. Cependant, une petite portion de ces gènes non codants "morts" peut être transcrite. On considère aujourd'hui que ces pseudogènes sont soit en voie de complète extinction (Harrison et al., 2005), soit qu'ils resurgissent en acquérant de nouvelles fonctions par un mécanisme encore inconnu (Bekpen et al., 2009). On suppose par exemple que le lncRNA *XIST* a évolué à partir de la pseudogénisation du gène codant *LNX3* et par l'intégration de séquences répétées transposables (Elisaphenko et al., 2008).

IV) Expression et localisation

a. Généralités

L'expression des lncRNAs est soumise aux mêmes conditions que les ARNm, c'est-à-dire dépendante de facteurs transcriptionnels et épigénétiques. Une analyse à grande échelle a montré que les marques de modifications post-traductionnelles des histones (H3K4me3, H3K9me3, H3K27me3 et H3K36me3) et leur distribution corrélaient avec l'expression de ces lncRNAs, de la même manière que pour les ARNm. De plus, il a été montré qu'indépendamment de l'expression, les TSS associés aux

lncRNAs étaient plus méthylés que ceux des ARNm. De manière intéressante, la méthylation de l'ADN au niveau des TSS ne semble pas intervenir sur l'expression du lncRNA, contrairement aux ARNm (Sati et al., 2012). L'analyse par Derrien *et al.* sur le catalogue GENCODE a mis en évidence que les lncRNAs pouvaient être des unités transcriptionnelles indépendantes, supportant donc le fait qu'il y ait des mécanismes spécifiques de régulation transcriptionnelle (Derrien et al., 2012). Ces marques ont par ailleurs permis d'identifier un grand nombre des lincRNAs dans le travail de Guttman *et al.* Les auteurs ont montré que ces lncRNAs étaient régulés transcriptionnellement par les facteurs de transcriptions clés comme p53, NF- κ B (Nuclear factor of kappa light polypeptide gene enhancer in B-cells), Sox2 (SRY (sex determining region Y)-box 2), POU5F1 (POU class 5 homeobox 1) et Nanog (Guttman et al., 2009)(Huarte et al., 2010).

Globalement, l'expression des lncRNAs est très nettement inférieure à celle des gènes codants dans la plupart des tissus testés, la médiane d'expression maximale étant dix fois inférieure à celle des ARNm (Cabili et al., 2011)(Derrien et al., 2012). Il a quand même été observé une expression relativement élevée des lncRNAs dans les testicules (Ørom et al., 2010). De plus, il existe un certain nombre d'exemples de lncRNAs fortement exprimés, de façon ubiquitaire et à des niveaux similaires aux ARNm, comme les lncRNAs *XIST* chez la femelle placentaire (Masui and Heard, 2006), *MALAT1* ou *NEAT1* (Naganuma and Hirose, 2013).

Ce qui a justifié un peu plus le potentiel fonctionnel des lncRNAs est que leur expression peut être fortement dépendante du contexte cellulaire et/ou du stade de développement, et profondément perturbée dans certaines pathologies. Il a également été noté qu'ils présentaient des variations d'expression entre les différents tissus bien plus importantes que les ARNm. Le travail de Derrien *et al.* a produit plusieurs résultats intéressants concernant l'expression des lncRNAs (Derrien et al., 2012). Les auteurs ont pu observer de fortes corrélations d'expressions entre les lncRNAs et leurs gènes codants voisins. Cela semble corroborer les résultats de précédentes études démontrant l'impact des lncRNAs sur la régulation des gènes codants proximaux, tel que le lncRNA *ANRIL* (antisense non-coding RNA in the INK4 locus) régulant le locus *INK4* (Yap et al., 2010).

b. Expression spécifique au cours du développement

Plusieurs études ont pu montrer un enrichissement de gènes codants impliqués dans le développement localisés à proximité (moins de 10kb) de lncRNAs (Cabili et al., 2011)(Guttman et al., 2009). Ces études ont par ailleurs mis en évidence une régulation de l'expression des lncRNAs par les facteurs de pluripotence et une expression très tissu-spécifique, suggérant ainsi un rôle potentiel de ces lncRNAs dans le développement.

Certains lncRNAs ont été décrits comme intervenant dans la régulation des gènes *HOX* qui sont des facteurs déterminants dans la spécification de l'axe antéro-postérieur et de l'identité

positionnelle des organes chez les animaux. *HOTAIR* est un lncRNA exprimé à partir du locus *HOXC* chez les mammifères et régule négativement l'activité du locus *HOXD* (Rinn et al., 2007a). L'expression de *HOTAIR* au cours du développement chez la souris s'avère très localisée au niveau du tronc postérieur et du bourgeon distal des membres (Rinn et al., 2007b). Il en est de même pour *HOTTIP* (*HOXA* distal transcript antisense RNA), un lncRNA qui est exprimé au niveau du locus *HOXA* et qui régule positivement l'expression de ce locus (Wang et al., 2011b). Il apparaît donc que ces lncRNAs sont importants pour la régulation de l'expression de ces gènes homéotiques cruciaux dans le développement. En effet, les souris knock-out pour *HOTAIR* présentent, d'une part, une perturbation de l'expression des gènes *HOX* dont le locus *HOXD*, et d'autre part des anomalies développementales comme des malformations des membres du squelette (Li et al., 2013a) (Figure 26).



Figure 26 : Impact de *HOTAIR* dans le développement. Malformation des os du poignet dans les modèles de knock-out *HOTAIR* (Coloration Rouge alizarine et bleu alcian). (II-V) os métacarpiens. (1, 2, 3, 4/5) Carpes. (c) élément central. (r) radial. (ra) radius. (u) ulnaire. (ul) ulna. (c-3) et (1-2-c) fusion des os carpiens. (*) Radius manquant (Adapté de Li et al., 2013a).

Au-delà du développement, plusieurs lncRNAs ont été décrits comme nécessaires et essentiels au maintien de la pluripotence des cellules souches et dans la capacité des cellules souches à se différencier. Ainsi, Ng *et al.* ont mis en évidence des lncRNAs qui sont impliqués dans le maintien de la pluripotence des cellules souches embryonnaires humaines en interagissant directement avec le facteur de pluripotence SOX2 (Ng et al., 2012). De la même manière, les auteurs ont identifié des lncRNAs nécessaires à la neurogenèse. En effet, les expériences inhibant l'activité de ces lncRNAs conduisent à un blocage de la différenciation neuronale. A l'inverse, Kretz *et al.* ont démontré que le lncRNA *ANCR* (anti-differentiation ncRNA) permettait le maintien des cellules souches de l'épiderme dans un état indifférencié (Kretz et al., 2012). On peut observer la baisse d'expression de ce lncRNA au cours de la différenciation des cellules souches, et sa perte d'expression conduit à la différenciation de ces cellules sans l'aide de stimuli externes. Une très belle étude, publiée en 2012 dans la revue *Nature*, décrit des expériences de pertes de fonctions de plusieurs lincRNAs exprimés dans les cellules

souches embryonnaires de souris. La perturbation de l'expression de ces lncRNAs a montré que ces ARNs régulaient l'expression d'un large panel de gènes, similaire d'un point de vue de la quantité à un régulateur transcriptionnel clé dans les cellules souches embryonnaires. Par ailleurs, cette régulation est nécessaire au maintien de l'état de pluripotence des cellules avec notamment une régulation de l'expression des facteurs Nanog et POU5F1. Les auteurs ont mis en évidence la répression exercée par les lncRNAs sur les programmes spécifiques de différenciation, impliqués très tôt dans le développement et dans les 3 feuillettes embryonnaires (Guttman et al., 2011).

c. Expression tissu spécifique

Comme cité précédemment, les lncRNAs présentent régulièrement une expression tissu-spécifique, avec des variations d'expressions entre les tissus bien plus importantes que les ARNm. Cabili *et al.* ont estimé que 78% des lincRNAs étaient tissus spécifiques alors que seulement 19% des ARNm présentaient ce type d'expression (Cabili et al., 2011). Ce résultat n'est pas dû à la différence d'expression globale entre lncRNAs et ARNm, puisque le même résultat est obtenu pour des transcrits exprimés à des niveaux similaires. Les auteurs de ce travail ont également estimé qu'environ un tiers des lincRNAs était spécifique aux testicules et qu'il existait des clusters de lincRNAs spécifiques à chacun des 24 tissus et lignées testés (Figure 27). De plus, l'expression de ces clusters de lincRNAs tissus spécifiques était significativement corrélée à l'expression des gènes codants impliqués dans des processus spécifiques de chaque tissu ou dans leurs différenciations.

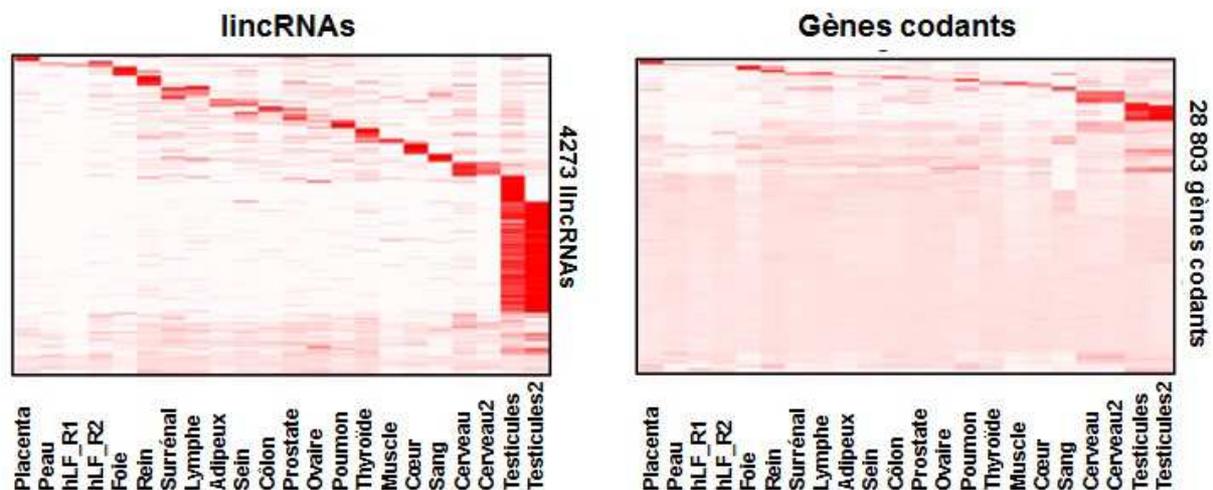


Figure 27 : Expression tissu-spécifique des lncRNAs et des ARNm. Abondance de 4373 lncRNAs (lignes) et 28 802 ARNm (lignes) Dans les différents tissus testés (colonnes) l'intensité de couleur représente le niveau d'expression du gène. Il y a majoritairement plus de lincRNAs exprimés spécifiquement mais leurs expressions sont globalement moins fortes que celles des ARNm. Il y a cependant une forte expression de lncRNAs dans les testicules (Adapté de Cabili et al., 2011).

Le lncRNA *Eyf2* par exemple, est majoritairement exprimé dans le cerveau antérieur en développement et régule l'expression des gènes *Dlx5/6* (distal-less homeobox) (Feng, 2006). Son expression et cette régulation s'avèrent nécessaires au développement du cerveau antérieur et à l'établissement des circuits neuronaux GABA-dépendants de l'hippocampe chez la souris (Feng, 2006). Exprimée de manière plus tissu-spécifique encore, la paire de lncRNAs *HARIA* et *HARIB* (highly accelerated region 1) est uniquement exprimée au niveau des neurones Cajal-Retzius dans le néocortex à la mi-gestation chez les primates. L'expression de ces ARNs s'étend à l'hippocampe, au cortex cérébelleux et au gyrus denté. A l'âge adulte, l'expression n'est plus uniquement cérébrale mais aussi ovarienne et testiculaire (Pollard et al., 2006).

Tout comme dans un contexte normal, l'expression des lncRNAs peut être dépendante d'un contexte pathologique. Il existe des lncRNAs spécifiques du tissu cancéreux comparé au tissu adjacent normal, constituant notamment d'excellents marqueurs néoplasiques. *CCAT1* (Colon Cancer Associated Transcript 1) par exemple, est un lncRNA exprimé uniquement dans le tissu cancéreux du côlon et non dans le tissu sain (Kam et al., 2013). Par ailleurs, Iyer *et al.* ont mis en évidence récemment, par une méta-analyse de 25 études indépendantes, près de 8 000 lncRNAs tissu-spécifique ou tumeur spécifique sur plus de 58 000 lncRNAs identifiés, confirmant plus encore cette caractéristique des lncRNAs (Iyer et al., 2015).

d. Localisation cellulaire

Plusieurs travaux ont montré que la grande majorité des lncRNAs était nucléaire et associée à la chromatine (Djebali et al., 2012)(Mondal et al., 2010) et cela indépendamment de leurs classes (intronique, intergénique, anti-sens) (Derrien et al., 2012). Bien que ce soit en accord avec plusieurs études présentant les lncRNAs comme majoritairement impliqués dans les régulations épigénétiques et associés à la chromatine, cette localisation peut également être le reflet d'une plus grande dégradation par des mécanismes comme le NMD (Nonsense mediated decay). Ces ARNs non codants peuvent présenter des ORFs "de mauvaise qualité" et ainsi être des cibles de ce mécanisme de dégradation cytoplasmique. Malgré cet enrichissement dans le noyau, une partie des lncRNAs peut être enrichie dans le cytoplasme. C'est le cas par exemple de *masrcRNA* (MALAT1-associated Small Cytoplasmic RNA) (Wilusz et al., 2008), de *NRON* (non coding RNA, repressor of NFAT) (Willingham et al., 2005) ou du lncRNA *GAS5* (growth arrest-specific 5) (Kino et al., 2010). Au vue des grandes différences fonctionnelles qui existent entre un lncRNA nucléaire ou cytoplasmique, la question de la localisation des lncRNAs s'avère donc cruciale dans l'établissement du potentiel fonctionnel d'un lncRNA.

V) Fonctions des lncRNAs

Dans tous les aspects décrits ci-dessus, que ce soit structural, au niveau de l'expression ou de la conservation, les lncRNAs sont un groupe très hétérogène. Il en est de même pour leurs fonctions. En effet, malgré la faible proportion de lncRNAs pour lesquels une fonction a été décrite, il y a des exemples montrant qu'ils peuvent intervenir à tous les niveaux biologiques, de la régulation épigénétique à la régulation de mécanismes post-transcriptionnels, jusqu'à la participation à l'assemblage de complexes protéiques. Afin de déterminer les fonctions de tous ces nouveaux lncRNAs, plusieurs méthodes sont utilisées. Les corrélations d'expressions avec des gènes codants suivies d'analyses ontologiques de type GO Term (Gene Ontology) permettent d'associer les lncRNAs de fonctions inconnues à des voies cellulaires ou à des processus particuliers : cette méthode est appelée *Guilt By Association*. Toutefois, il reste nécessaire d'analyser l'impact de leur inhibition et de leur surexpression sur leurs cibles potentielles, ou les processus cellulaires dans lesquels ces lncRNAs sont suspectés d'intervenir. L'inhibition des lncRNAs est par ailleurs un challenge pour ce type de molécules majoritairement nucléaires, pour lesquelles les méthodes basées sur les ARN interférents (siARNs) sont souvent peu ou pas efficaces. Dans ce contexte, des molécules telles que les siARNs modifiés avec LNA (Lock Nucleic Acid) ou BNA (2',4'-Bridged Nucleic Acid) ont été développées pour augmenter l'hybridation sur les lncRNAs cibles. De plus, l'utilisation de GapmeRs permet de s'affranchir du problème de localisation cellulaire. Les GapmeRs sont des petits oligonucléotides anti-sens spécifiques de l'ARN cible, formant un hétéroduplexe ARN/ADN qui sera dégradé par la RNase H endogène. D'un point de vue moléculaire, la caractérisation de la fonction des lncRNAs passe régulièrement par la recherche des partenaires, protéiques ou ARNs, et des gènes cibles. Par contre, les études *in vivo* sont plus rares, notamment à cause de la faible conservation entre les espèces de la majorité des lncRNAs.

Malgré une connaissance encore limitée et constituée à partir de quelques dizaines d'exemples, les modes d'actions des lncRNAs sont actuellement regroupés sous 4 archétypes qui ne sont pas mutuellement exclusifs: signal, leurre, échaffaudage et guide. Globalement, tous ces modes d'action sont la conséquence des informations de la structure primaire des ARNs, et également de leur grande plasticité et complexité structurale. La nature intrinsèque de ces lncRNAs leur confère un rôle dual à la fois en tant que partenaires pour des protéines mais également pour des acides nucléiques par homologie de séquence (ARNs, ADN).

Signal : Ce sont des lncRNAs qui intègrent des informations cellulaires spatio-temporelles, développementales et stimulus-spécifiques. On peut citer le lncRNA *XIST* qui inhibe un des chromosomes X chez la femelle placentaire au cours du développement ou *HOTAIR* régulant l'expression spatio-temporelle de gènes *HOX*.

Leurre : Ce sont des lncRNAs qui ont la capacité de séquestrer des partenaires protéiques mais également des ARNs. Le lncRNA *PANDA* peut par exemple interagir avec le facteur NF-YA et l'empêcher d'accrocher ses cibles. Le transcrite du pseudogène *PTENP1* peut quant à lui séquestrer des microARNs ciblant *PTEN*, inhibant de fait leurs fonctions (Poliseno et al., 2010).

Échaffaudage : Ce sont des lncRNAs qui peuvent servir d'adaptateur ou de plateforme entre deux ou plusieurs protéines pour la formation de complexes. TERC (Telomerase RNA) est un exemple classique de lncRNA adaptateur puisqu'il permet l'assemblage du complexe Télomérase.

Guide : Ce dernier mode d'action fait référence à des lncRNAs qui "guident" ou recrutent des complexes de modification de la chromatine ainsi que d'autres protéines nucléaires à des loci génomiques spécifiques. Il existe de nombreux exemples de lncRNAs présentant ce mode d'action tels que *HOTAIR* ou *XIST*.

Tous ces modes d'action des lncRNAs sont retrouvés à différents niveaux de régulation, transcriptionnel, post-transcriptionnel, traductionnel et épigénétique.

a. Dans la régulation transcriptionnelle

La grande majorité des lncRNAs fonctionnels sont impliqués dans la régulation de l'expression génique en *cis* ou en *trans*. La régulation en *cis* fait référence à une régulation des gènes sur le même chromosome que le gène du lncRNA et à l'inverse, la régulation en *trans* fait référence à une régulation génique entre des chromosomes différents (Figure 28).

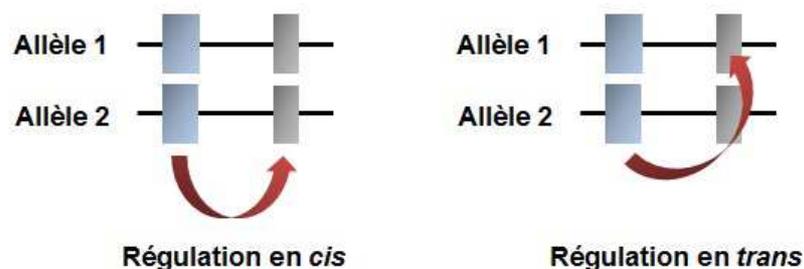


Figure 28 : Représentation schématique des régulations transcriptionnelles en *cis* et *trans*.

Cependant, par abus de langage dans la littérature des lncRNAs, la régulation en *cis* est définie comme une régulation des gènes proximaux du lncRNA sans distinction de l'allèle. Pour plus de cohérence avec la littérature, je m'appuierai sur la deuxième définition pour la suite de cette thèse. La plupart des mécanismes de régulation transcriptionnelle induite par les lncRNAs font intervenir des mécanismes de régulation épigénétique que je décrirai dans le chapitre "Dans la régulation épigénétique". Dans ce chapitre-ci, nous nous concentrerons sur les autres mécanismes de régulation. En effet, la capacité qu'ont les lncRNAs à interagir avec différentes protéines, facteurs et co-facteurs

de transcription, les ARN polymérase et même l'ADN, laisse à supposer qu'il existe plusieurs mécanismes de régulation de l'expression génique dans lesquels interviennent les lncRNAs.

i. Régulation de la machinerie de transcription

Il existe quelques exemples pertinents de lncRNAs intervenant directement sur le processus général de transcription en influençant notamment l'activité de l'ARN pol II. Bien qu'étant à proprement parler un snRNA, *7SK* est un ARN de 331 nucléotides qui réprime l'élongation de la transcription en empêchant la phosphorylation du domaine carboxy-terminal de l'ARN pol II par PTEF β (Peterlin et al., 2012). De la même manière, le snRNA B2 interagit avec l'ARN pol II et empêche la phosphorylation de l'ARN pol II par TFIIF, perturbant ainsi la capacité de la polymérase à lier l'ADN (Yakovchuk et al., 2014).

La régulation *in cis* du gène *DHFR* (dihydrofolate reductase) par des lncRNAs sens de ce gène est également un exemple intéressant qui fait intervenir deux mécanismes distincts (Figure 29). D'une part, il a été montré que ces lncRNAs pouvaient séquestrer les facteurs de transcription SP3, SP1 et TFIIB et altérer de manière concentration-dépendante la formation du complexe de pré-initiation de la transcription (PIC) sur le promoteur majeur de *DHFR* (Blume et al., 2003). Cet effet répressif est uniquement possible lorsque le lncRNA contient la séquence du promoteur majeur de *DHFR* indiquant la spécificité de cette répression. D'autre part, il a également été montré que ces lncRNAs pouvaient contribuer à la répression de *DHFR* par la formation avec le promoteur majeur de structures en triplex ARN/ADN spécifiques et stables *in vitro* (Gee et al., 1992)(Martianov et al., 2007)(Figure 29).

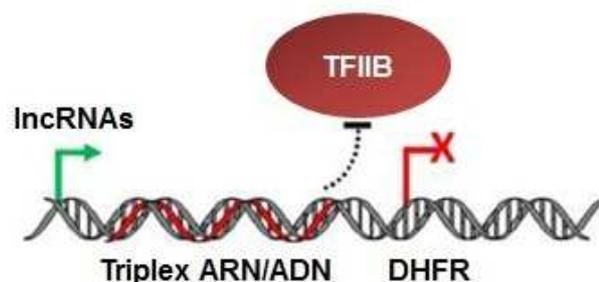


Figure 29 : Représentation schématique du mécanisme de régulation transcriptionnelle du gène *DHFR*. Un des mécanismes fait intervenir un lncRNA impliqué dans la formation d'une structure triplex ARN/ADN au niveau du promoteur de *DHFR*, empêchant ainsi la fixation de facteur de transcription.

Les eRNAs (enhancer RNAs) sont des régulateurs positifs de la transcription *in cis*. Bien que ces ARNs ne soient pas essentiels à l'expression basale des gènes auxquels ils sont associés, ils

semblent néanmoins moduler très finement leurs expressions par plusieurs mécanismes distincts. Un des mécanismes par lequel les eRNAs régulent la transcription se base sur la séquestration de facteurs de transcription tel que NELF (Negative elongation factor). Ce facteur, en coopération avec DSIF (DRB sensitivity-inducing factor), est connu pour interagir avec l'ARN pol II et le transcrit naissant afin de bloquer transitoirement la transcription. Schaukowitch *et al.* ont démontré que les eRNAs déstabilisaient transitoirement le complexe NELF, permettant ainsi l'élongation de la transcription des gènes associés (Schaukowitch et al., 2014).

ii. Séquestration/Recrutement des facteurs de transcription

Dans le cadre de ces régulations, différents travaux semblent montrer que ces régulations sont des régulations positives et généralement en *cis*. Cependant, il existe également différents exemples de lncRNAs régulant négativement l'expression de gènes codants, majoritairement via l'interaction avec des facteurs de transcription. Ce type de mécanisme d'interférence fut mis en évidence la première fois en 2004 chez la levure où la transcription du lncRNA *SRG1* (SER3 regulatory gene 1), à partir de séquences régulatrices du gène codant *SER3*, empêchait l'interaction de facteurs activateurs de transcription sur le promoteur de *SER3* (Martens et al., 2004).

De la même manière, le lncRNA *PANDA* (P21 associated ncRNA DNA damage activated) peut réguler en *trans* l'expression de gènes pro-apoptotiques. *PANDA* est un lncRNA bidirectionnel de 1500nt transcrit à partir du promoteur du gène *CDKN1A* (cyclin-dependent kinase inhibitor 1A). L'expression de *CDKN1A* (p21) et de *PANDA* est induite par p53 en réponse aux dommages à l'ADN afin de réguler le cycle cellulaire et l'apoptose. En effet, il a été montré que *PANDA* pouvait séquestrer le facteur de transcription NF-YA (nuclear transcription factor Y alpha), inhibant ainsi l'expression de gènes tels que *FAS*, *BBC3* (BCL2 binding component 3) et *PMAIP1* (phorbol-12-myristate-13-acetate-induced protein 1) (Hung et al., 2011). Il existe d'autres exemples de lncRNAs impliqués dans l'activation transcriptionnelle via des interactions protéiques, comme PSF Inhibiting RNAs. Identifiés chez l'homme, les PSF Inhibiting RNAs sont au nombre de 5 et ces lncRNAs ont été décrits comme interagissant avec le facteur hPSF (Protein associated splicing factor). Cette protéine interagit avec la région régulatrice en 5' de *GAG6* (G antigen 6), inhibant ainsi la transcription de ce proto-oncogène. Les PSF Inhibiting RNAs séquestrent hPSF, ce qui libère *GAG6*, conduisant à sa transcription et à son effet tumorigénique (Li et al., 2009).

La répression transcriptionnelle peut faire intervenir un autre mécanisme, le lncRNA agissant comme un leurre. En effet, parmi ses nombreuses fonctions, le lncRNA *GAS5* peut se lier aux sites d'interactions à l'ADN des récepteurs aux glucocorticoïdes. Cette interaction empêche la fixation de ces récepteurs aux GREs (Glucocorticoid Response Element) exprimés en réponse à l'activation des récepteurs aux glucocorticoïdes, conférant ainsi à *GAS5* une fonction d'immunosuppresseur (Kino et al., 2010).

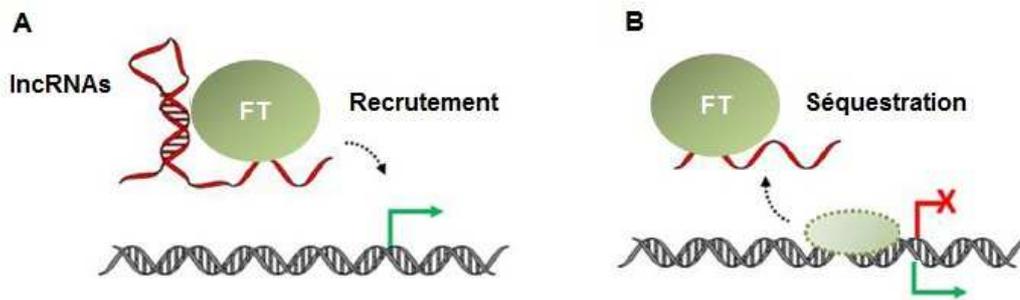


Figure 30 : Représentation schématique des principaux mécanismes de régulation transcriptionnelle.

iii. Régulation de l'activité des facteurs de transcription

Un grand nombre de travaux ont montré que les lncRNAs pouvaient interagir avec des protéines, cependant il n'y a que peu d'études démontrant l'impact de ces interactions sur l'activité protéique. C'est le cas du lncRNA *NRON* (Noncoding RNA Repressor of NFAT) qui régule la localisation cellulaire et *in fine* l'activité du facteur de transcription NFAT (Nuclear Factor of Activated T cells) (Figure 31). En effet, la délétion de *NRON* conduit à une augmentation de l'activité transcriptionnelle de NFAT via la délocalisation de cette protéine du cytoplasme au noyau. De plus, 11 protéines ont été rapportées comme interagissant avec *NRON* par purification d'affinité ARN/Protéine, dont IQGAP1 (IQ motif containing GTPase activating protein 1), le facteur de transport nucléaire KPNB1 (Karyopherin (importin) beta 1), et également trois membres de la superfamille des bêta-importines régulant le transport nucléaire/cytoplasme (Willingham et al., 2005)(Sharma et al., 2011). Ces observations suggèrent fortement un rôle répresseur de *NRON* sur l'activité transcriptionnelle de NFAT par délocalisation de ce facteur du noyau au cytoplasme.

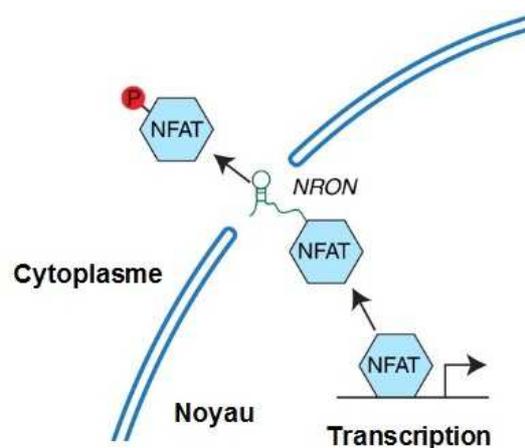


Figure 31 : Représentation schématique de l'impact de *NRON* sur l'activité de NFAT. *NRON* conduit à la localisation cytoplasmique de la forme phosphorylée et inactive de NFAT, potentiellement par le biais de l'importine, régulant ainsi son activité.

Le lncRNA *Inc-DC* régule quant à lui positivement l'activité de STAT3 (Signal transducer and activator of transcription 3). En effet, l'interaction *Inc-DC*/STAT3 empêche l'inactivation de ce facteur de transcription par la déphosphorylation induite par SHP-1 (Src Homology region 2 domain-containing phosphatase-1) (Wang et al., 2014b).

iv. Autres régulations transcriptionnelles

Comme mentionné précédemment, les lncRNAs intervenant dans la régulation transcriptionnelle du gène DHFR forment un triplex avec le promoteur du gène et inhibent sa transcription.

Pour ce qui est des eRNA, le deuxième mécanisme, par lequel les eRNAs régulent l'activité transcriptionnelle des gènes auxquels ils sont associés, fait intervenir la génération et le maintien de boucles chromatiennes entre l'enhancer et le promoteur (Figure 32). Il a été observé que les eRNAs pouvaient interagir avec les sous unités SMC3 et RAD21 du complexe Cohésin, qui a été décrit comme un régulateur des boucles chromatiennes enhancer-promoteur (Kagey et al., 2010). En effet, l'inhibition de eRNAs par ARN interférence diminue la formation de boucles enhancer-promoteurs, ainsi que l'expression des gènes codants associés (Li et al., 2013b).

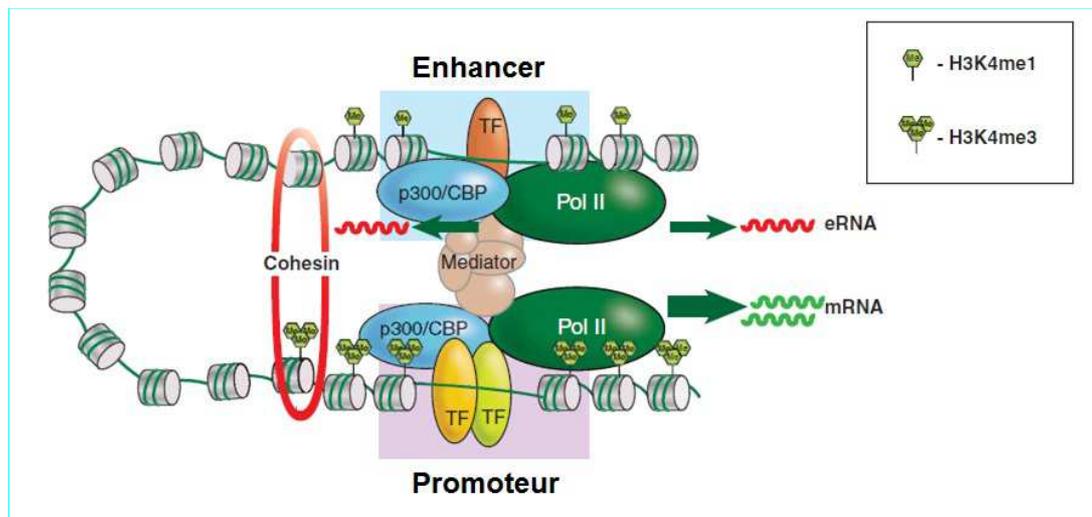


Figure 32 : Représentation schématique des boucles chromatiennes entre enhancer et promoteur dans lesquelles participent les eRNAs. Au cours de l'activation de la transcription, l'ARN pol II et des facteurs co-activateurs (p300/CBP) permettent la transcription bidirectionnelle de eRNAs. La boucle chromatinienne entre l'enhancer et le promoteur va permettre le rapprochement de l'eRNA de son promoteur cible pour coordonner l'activation transcriptionnelle du gène associé. De plus, certains eRNAs interagissent avec le complexe cohesin pour faciliter la formation et la stabilité de ces boucles spécifiques entre promoteur et enhancer, afin d'augmenter l'activité transcriptionnelle (Adapté de Kim et al., 2015).

b. Dans la régulation post-transcriptionnelle

Bien que les lncRNAs aient principalement été décrits comme intervenant dans les régulations transcriptionnelles, plusieurs lncRNAs ont été mis en évidence comme impliqués également dans les modifications post transcriptionnelles des ARNm.

i. Dans l'épissage

Il est estimé qu'au moins 70 % des gènes codants chez l'homme peuvent subir un épissage alternatif, et qu'en moyenne un gène donne naissance à 4 variants issus d'un tel épissage. Un gène transcrit en ARN pré-messager (pré-ARNm) contient à la fois des introns et des exons. Durant l'épissage du pré-ARNm, les exons sont soit conservés dans l'ARNm, soit ciblés et éliminés, menant ainsi à la création de différents d'ARNm à partir d'un seul pré-ARNm. Ce processus d'épissage alternatif de l'ARN fait appel au spliceosome (machinerie complexe faisant intervenir plus de 200 protéines notamment les snRNAs U1, U2, U3, U4/U6 et U5 reconnaissant les sites d'épissages), à des protéines Sérine/Arginine riches (SR) se fixant sur les séquences activatrices de l'ARN, et à des ribonucléoprotéines hnRNP se fixant sur les séquences inhibitrices. Ce processus complexe et dynamique permet d'augmenter la complexité du transcriptome. Il existe deux exemples majeurs de l'impact des lncRNAs sur ce processus :

Le lncRNA *MALAT1* est enrichi dans le noyau et plus précisément dans les granules nucléaires ou "speckles" (Hutchinson et al., 2007). Ces granules ne sont pas un lieu de transcription ni d'épissage mais plutôt un lieu de stockage, de modifications et/ou d'assemblage de différents facteurs d'épissage (Lamond and Spector, 2003). Il a donc été supposé que *MALAT1* pouvait être impliqué dans le métabolisme des pré-ARNs. Ainsi, Tripathi *et al.* ont observé que *MALAT1* pouvait interagir avec les protéines Sérine/Arginine riches SRSF1, SRSF2, SRSF3 et plus faiblement avec SRSF5. Les interactions entre le lncRNAs et ces différents facteurs permettent notamment de moduler l'association des facteurs d'épissage aux granules nucléaires. Des séquences de *MALAT1* influencent à la fois la distribution de *MALAT1* dans ces granules nucléaires et le recrutement des facteurs d'épissage. La délétion nucléaire de *MALAT1 in vitro* a démontré qu'il pouvait moduler le niveau de phosphorylation des facteurs SR et les changements d'épissage alternatif d'un set d'ARNm endogènes (Figure 33). De plus, une nouvelle méthode d'étude des interactions ARN-ARN (RAP-RNA), développée récemment par Engreitz *et al.*, a mis en évidence l'interaction entre *MALAT1* et des pré-ARNm par le biais de protéines intermédiaires (Engreitz et al., 2014). Cela suggère une implication probable de *MALAT1* dans la modulation du phénomène d'épissage alternatif global (Tripathi et al., 2010)(Bernard et al., 2010).

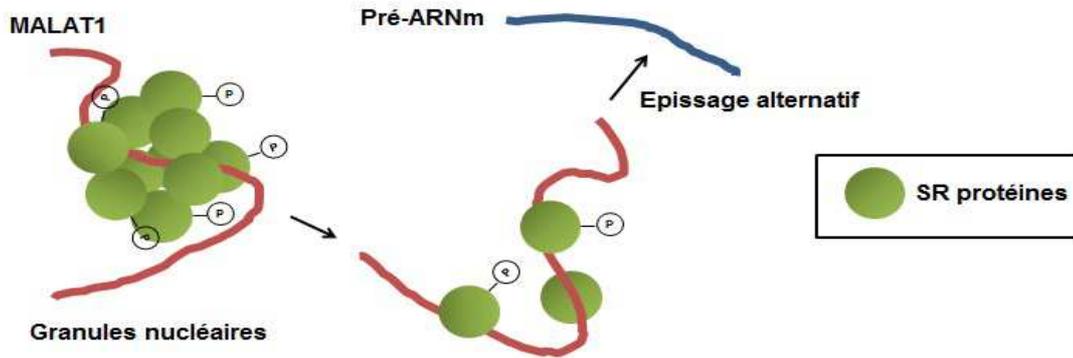


Figure 33 : Représentation schématique du potentiel impact de *MALAT1* dans la régulation de l'épissage alternatif. *MALAT1* peut interagir avec les facteurs d'épissage SR, moduler leur phosphorylation et ainsi réguler l'épissage alternatif des ARNm.

Le deuxième exemple de lncRNA impliqué dans ce processus est le lncRNA *ZEB2NAT* (*ZEB2* Natural Anti sense), affectant le processus d'épissage de manière beaucoup plus spécifique. Ce lncRNA est l'anti sens du gène *ZEB2*, un répresseur transcriptionnel de l'E-Cadherin. Il a été montré que la traduction de *ZEB2* nécessitait la rétention d'un de ces introns en 5'UTR contenant un IRES (Internal Ribosome Entry Site), et que ce mécanisme dépendait du lncRNA. En effet, *ZEB2NAT* s'hybride à l'ARNm de *ZEB2*, masque le site d'épissage en 5' de l'intron, et conduit à la rétention de l'intron et de l'IRES dans le produit final, permettant ainsi la traduction de *ZEB2* (Beltran et al., 2008) (Figure 34).

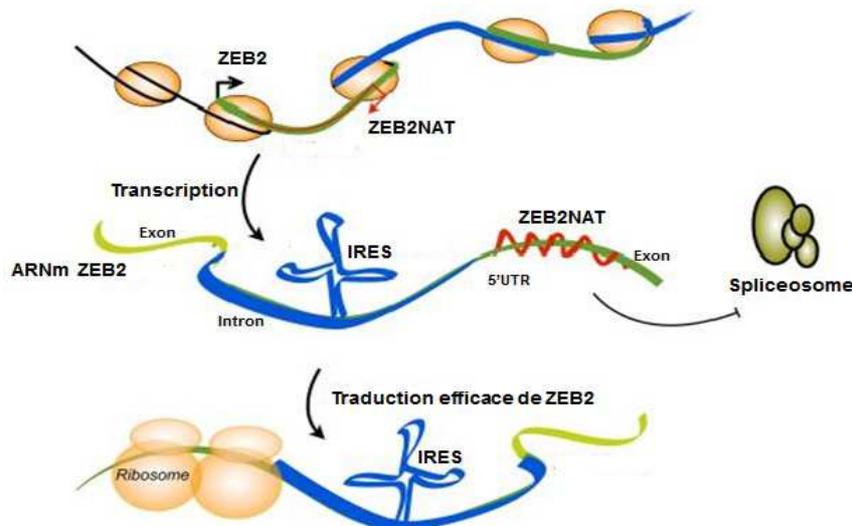


Figure 34 : Représentation schématique de la régulation de l'épissage de *ZEB2* par *ZEB2NAT*. *ZEB2NAT* interagit avec l'ARNm de *ZEB2*, empêchant la fixation du spliceosome. Cela conduit à la rétention de l'intron contenant l'IRES, ce qui permet la traduction de *ZEB2*.

ii. Dans l'édition

L'édition est un processus de modification post-transcriptionnelle faisant intervenir des enzymes ayant pour substrat l'ARN et qui catalysent la réaction de conversion des ribonucléotides. Ce mécanisme implique des déaminases comme la cytidine déaminase catalysant la conversion de la cytidine en uracile ou, majoritairement, les enzymes ADARs (Adenosine Deaminase Acting on RNA) catalysant à partir d'un ARN double brin la conversion d'une adénosine en inosine (reconnu comme une guanine lors de la traduction) (Figure 35). Sachant que les lncRNAs peuvent s'apparier aux ARNm, cela suggère que l'ARN double brin formé (ARNm/lncRNA) pourrait être soumis à ce type de modifications. À l'heure actuelle, il n'y a que peu d'évidences quant à cette implication des lncRNAs dans ce processus. En effet, une seule étude récente a mis en évidence l'implication du lncRNA *PCA3* (prostate cancer associated 3) dans la régulation post-transcriptionnelle d'ARNm via l'édition. Les auteurs ont identifié le gène codant sens de *PCA3*, *PRUNE2* comme étant régulé négativement par *PCA3*. Cette régulation fait intervenir un mécanisme impliquant la formation d'un duplex entre *PCA3* et le pré-ARNm *PRUNE2* qui sera édité par les enzymes ADARs. Les ARNs contenant trop d'inosines sont reconnus par la protéine P54^{NRB}, qui est impliquée dans la rétention nucléaire d'ARNs "défectueux". Ainsi, la reconnaissance de *PRUNE2* hyper-édité par P54^{NRB}, conduit finalement à la réduction de la production de *PRUNE2* (Salameh et al., 2015).

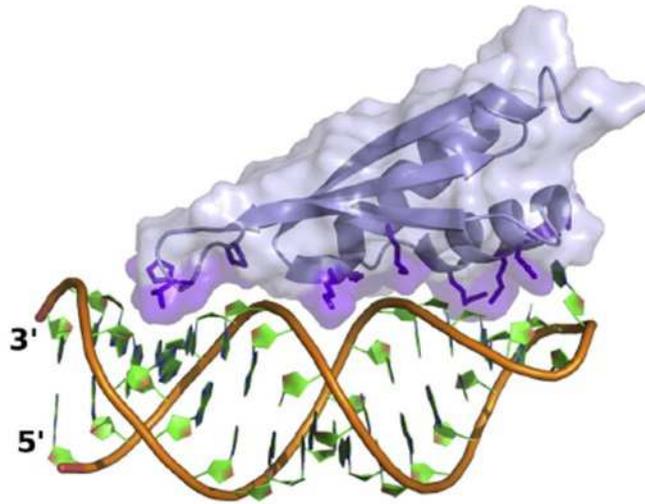


Figure 35 : Représentation 3D de l'interaction entre les enzymes ADARs et la structure double brin de l'ARN. Cette structure peut être formée par lncRNA/ARNm, ce qui modulera l'édition de l'ARNm (Adapté de Blythe et al., 2015).

iii. Dans la dégradation/stabilité des ARNm

La capacité qu'ont les lncRNAs de pouvoir s'hybrider à d'autres ARNs suggère qu'ils pourraient intervenir dans la dégradation/stabilité de ces derniers.

La dégradation des ARNm peut être médiée par plusieurs voies comme la voie NMD (nonsense-mediated mRNA decay), mais également la voie SMD (Staufen 1 ou STAU1-mediated Messenger RNA Decay). Ce processus fait intervenir la protéine STAU1 (staufen double-stranded RNA binding protein 1), qui peut interagir avec la partie 3' des ARNm présentant des structures doubles brins, et conduire au recrutement de la protéine UPF1 pour déclencher la dégradation de l'ARN (Park and Maquat, 2013). Une étude a montré qu'un lncRNA pouvait participer à la dégradation de plusieurs cibles du SMD ajoutant un tout nouveau niveau de régulation de la part de ce type de lncRNAs, appelés *1/2-sbsRNAs* pour half-STAU1-binding site RNAs. Le travail de Gong *et al.* a permis de montrer que les sites d'interactions de STAU1 sur ses cibles pouvaient être constitués d'un appariement au moins partiel entre les séquences *Alu* en 3' de la cible du SMD et les séquences *Alu* de lncRNAs (Gong and Maquat, 2011). Cet appariement conduit à l'activation de la voie SMD et la dégradation de l'ARNm ciblé (Figure 36A).

Hormis l'appariement partiel entre les lncRNAs et les ARNm par le biais de séquences *Alu*, les lncRNAs, notamment les lncRNAs anti-sens, peuvent former des duplex ARN/ARN par appariement de base. Cela suggère donc qu'ils peuvent potentiellement réguler les ARNm sens en altérant la stabilité ou la traduction. En effet, plusieurs exemples de ces anti-sens ont été mis en évidence avec un impact sur la stabilité des ARNm auxquels ils s'associent. Le lncRNA *BACE1-AS* est l'anti-sens du gène codant *BACE1* (β -secretase-1). Il a ainsi été montré que ce lncRNA pouvait former un duplex avec le transcrit sens *BACE1*, et conduire à une augmentation de la stabilité de l'ARNm de *BACE1* (Liu et al., 2014) (Figure 36B).

De manière intéressante, ces interactions lncRNA/ARNm ont soulevé la question de l'interaction potentielle des lncRNAs avec les microARNs. En effet, les lncRNAs peuvent agir comme des éponges à microARNs, les classant de ce fait comme des lncRNAs compétiteurs/séquestreurs ou ceRNAs (Competing Endogenous RNAs) (Figure 36C). Plusieurs études ont mis en évidence des exemples de ceRNAs interagissant avec des microARNs, dé-réprimant ainsi l'expression des ARNm cibles. Par exemple, *lincMD1* séquestre miR-133 et miR-135, des régulateurs de facteurs de transcription activant l'expression des gènes spécifiques impliqués dans la différenciation du tissu musculaire (Cesana et al., 2011). *H19*, qui est également un lncRNA producteur de microARNs, contient 4 sites d'interactions avec le microARN let-7. Ainsi, il semblerait que *H19* régule la différenciation du tissu musculaire en modulant la disponibilité de let-7 (Kallen et al., 2013). Un autre exemple intéressant est le lncRNA *RoR*, qui a été montré comme étant un ceRNA nécessaire au

maintien du phénotype de cellules souches embryonnaires, en séquestrant des microARNs inhibiteurs des facteurs de pluripotence clés Nanog, POU5F1 et SOX2 (Wang et al., 2013).

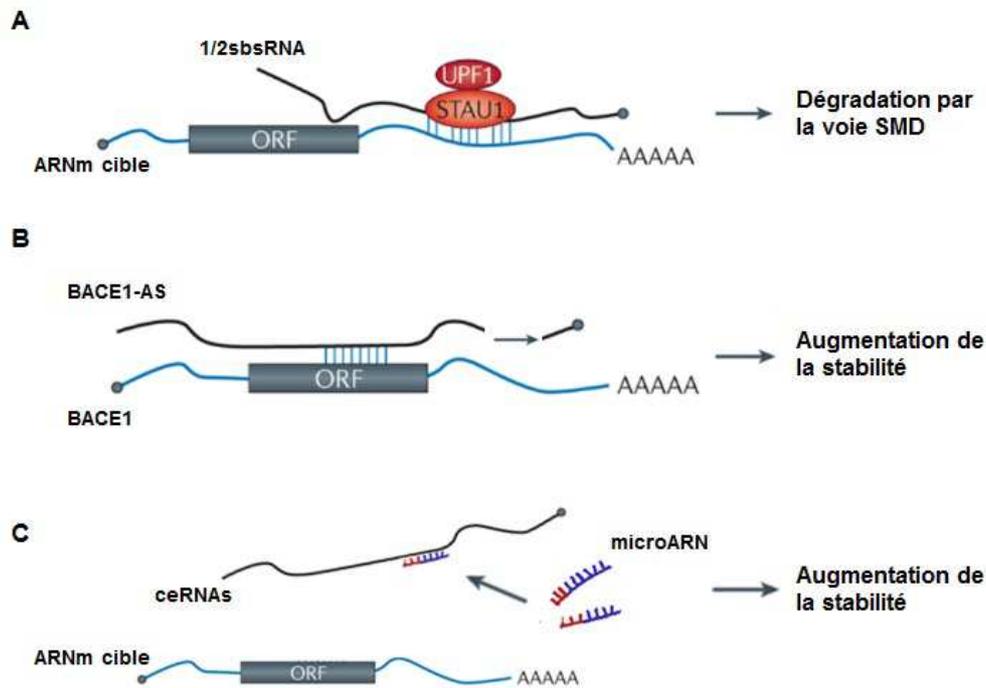


Figure 36 : Représentation schématique des mécanismes de régulation de la dégradation des ARNs impliquant les lncRNAs. A. Les lncRNAs *1/2sbsRNAs* peuvent s'apparier partiellement (via des séquences *Alu*) avec les ARNm. Cet appariement peut être reconnu par STAU1 qui recrutera UPF1 pour induire la dégradation de l'ARNm par la voie SMD. B. Le lncRNA *BACE1-AS* peut s'hybrider à l'ARNm sens *BACE1* et augmenter sa stabilité. C. Les lncRNAs peuvent également séquestrer les microARNs, empêchant ainsi la dégradation ou le blocage de traduction des ARNm cibles des microARNs. (Adapté de Fatica and Bozzoni, 2014).

c. Dans la régulation traductionnelle

Van Heesch *et al.* ont pu mettre en évidence récemment par fractionnement cellulaire, un fort enrichissement en lncRNA dans le cytoplasme et plus particulièrement dans les fractions ribosomales (van Heesch et al., 2014). Sachant que ces lncRNAs sont non codants, cela suggère qu'une portion importante des lncRNAs pourrait intervenir dans la régulation traductionnelle. En effet, la littérature décrit plusieurs exemples pertinents de l'intervention des lncRNAs dans la régulation du processus de traduction d'ARNm cibles. Cette régulation peut être négative comme dans le cas du lncRNA *UCA1* (urothelial carcinoma-associated 1). Ce lncRNA peut interagir et séquestrer hnRNP I, une ribonucléoprotéine qui peut augmenter la traduction du facteur *p27*. Ainsi, cette interaction entre *UCA1* et hnRNP I a mis en évidence que ce lncRNA se comportait comme un compétiteur en empêchant l'interaction entre hnRNP I et l'ARNm de *p27*, diminuant ainsi la traduction de l'ARNm de *p27* (Huang

et al., 2014). A l'inverse, cette régulation de la traduction peut être positive. Carrieri *et al.* ont découvert un lncRNA anti-sens du gène *UCHL1* (Ubiquitin Carboxy-Terminal Hydrolase L1), qui est délocalisé du noyau au cytoplasme en condition de stress pour augmenter la capacité d'association de l'ARNm d'*UCHL1* avec les polysomes (Carrieri et al., 2012).

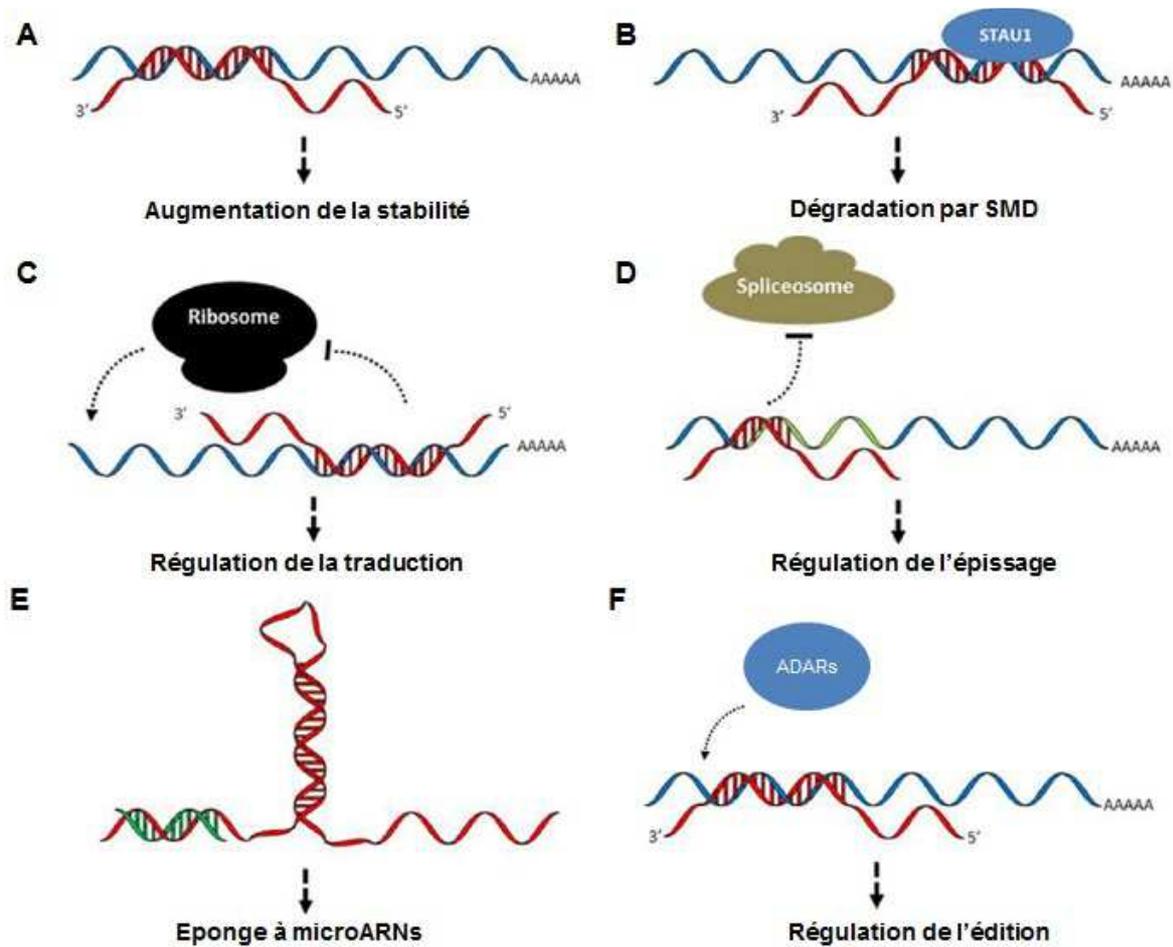


Figure 37 : Représentation schématique des mécanismes de régulation post-transcriptionnelle dans lesquels interviennent les lncRNAs. A. Les lncRNAs s'hybrident aux ARNm et augmentent leur stabilité. B. Les lncRNAs peuvent induire la dégradation par le SMD en s'hybridant partiellement aux ARNm. C. Les lncRNAs peuvent également inhiber ou augmenter la traduction des ARNm. D. Les lncRNAs peuvent empêcher la fixation du spliceosome et moduler l'épissage des ARNm. E. Les lncRNAs peuvent séquestrer des microARNs et empêcher la dégradation de leurs cibles. F. Les lncRNAs peuvent s'hybrider aux ARNm et former un substrat ARN double brin reconnu par les enzymes de l'édition. (Adapté de Angrand et al., 2015).

d. Dans la régulation épigénétique

La fonction la plus abondamment décrite dans la littérature sur les lncRNAs à l'heure actuelle est certainement la régulation transcriptionnelle de gènes cibles par des mécanismes de régulation épigénétiques. Cette régulation intervient principalement par le biais du recrutement de facteurs de modifications post-traductionnelles des histones ou de méthylation de l'ADN au niveau des régions régulatrices. Les 1^{ers} lncRNAs décrits comme étant impliqués dans des régulations épigénétiques sont des lncRNAs participant à l'établissement de l'empreinte parentale.

i. Dans l'empreinte parentale

L'empreinte parentale est un mécanisme épigénétique qui permet l'expression mono-allélique d'un gène, en fonction de l'origine parentale du chromosome qui le porte. On connaît aujourd'hui une centaine de gènes soumis à empreinte parentale, qui sont généralement regroupés en clusters. Le maintien de l'expression différentielle selon l'origine parentale de l'allèle est une information transmissible aux cellules filles après division cellulaire. Il est intéressant de noter que ce phénomène est un processus dynamique et que tous les loci soumis à l'empreinte contiennent des lncRNAs. Un déséquilibre dans l'expression de ces gènes, notamment par perte de l'empreinte ou LOI (Loss Of Imprinting), peut entraîner des pathologies révélant leur spectre fonctionnel, au cours de l'embryogenèse, dans la placentation et le développement, mais également à l'âge adulte dans la régulation de la fonction neuronale, le comportement ou le métabolisme.

Le lncRNA *H19* est un des premiers lncRNAs à avoir été caractérisé comme un régulateur du maintien de l'empreinte parentale du locus *H19/IGF2*. *H19* est exprimé maternellement pour réprimer *IGF2* à partir du stade blastocyste jusqu'à la mise en place des 3 feuilletts embryonnaires, mais uniquement dans le tissu squelettique chez l'adulte (Bartolomei et al., 1991). Les modèles de knock-out murins de *H19* sont viables, fertiles et présentent un surpoids probablement dû à l'expression biallélique du locus *Igf2* (Ripoche et al., 1997). *AIR* (Acute insulin response) est un lncRNA de 108k nt, non épissé et anti-sens du locus d'*IGF2R* (Insulin-like growth factor type 2 receptor). Au cours des stades précoces du développement, *AIR* est maternellement exprimé dans le placenta pour réprimer l'expression paternelle de gènes comme *IGF2R*, *SLC22A2* et *SLC22A3* (solute carrier family 22) (Sleutels et al., 2002). L'empreinte parentale est perdue dans des modèles murins de délétion de *Air*, conduisant à la surexpression de ces gènes et à des souris présentant un poids plus faible à la naissance (Yotova et al., 2008). *KCNQ1OT1* est un autre exemple de lncRNA impliqué dans la régulation de l'empreinte parentale. Exprimé par l'allèle paternel au niveau du locus *KCNQ1*, cet ARN est responsable de l'extinction de plusieurs loci : les loci *KCNQ1*, *CDKN1C* (cyclin-dependent kinase inhibitor 1C) et *PHLDA2* (pleckstrin homology-like domain, family A, member 2) de manière ubiquitaire, et les loci *OSBPL5* (oxysterol binding protein-like 5), *TSSC4* (tumor suppressing

subtransferable candidate 4) et *ASCL2* (achaete-scute family bHLH transcription factor 2) dans les tissus extra-embryonnaires (Mohammad et al., 2009) (Figure 38). Ces régulations se produisent très tôt dans le développement (E3.5 à 5.5) et certaines de ces régulations sont lignages-dépendants démontrant ainsi le rôle important dans l'établissement d'une expression parentale spécifique (Kanduri, 2011).

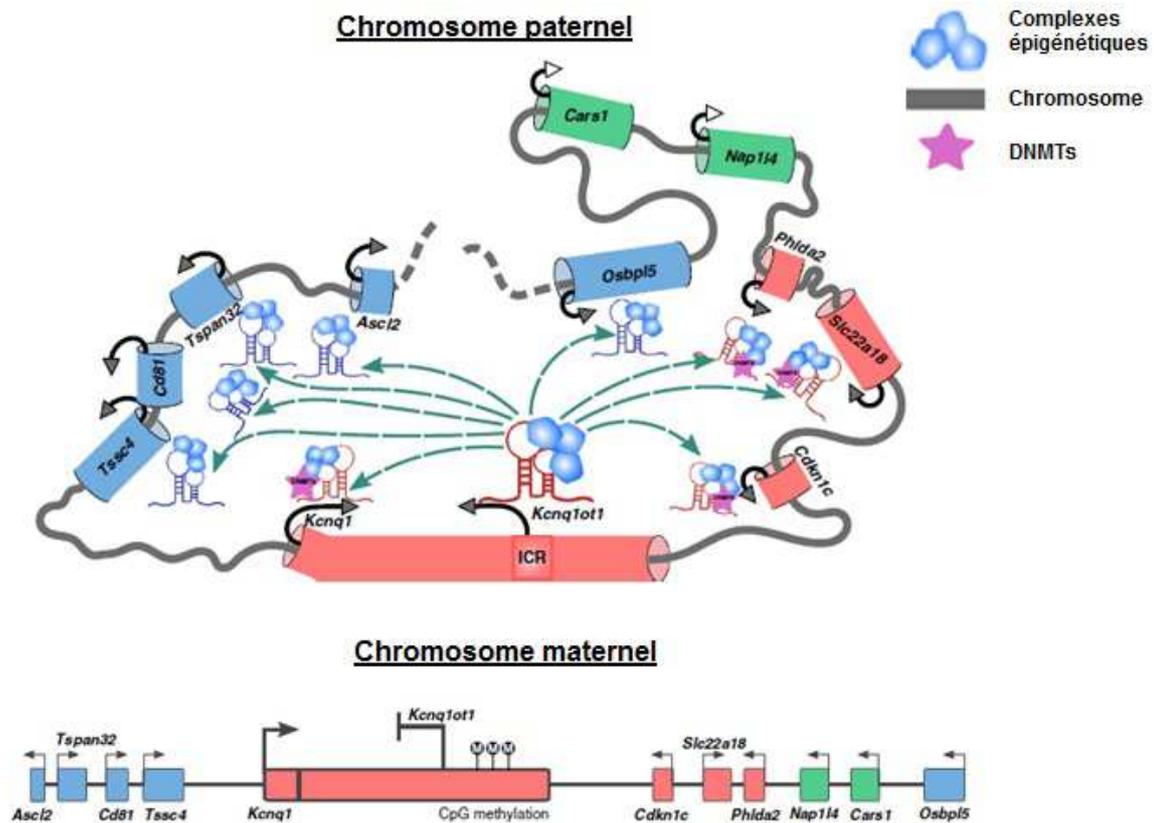


Figure 38 : Représentation schématique du mécanisme d'empreinte parentale induit par le lncRNA *KCNQ1OT1*. *KCNQ1OT1* est exprimé paternellement à partir du locus *KCNQ1* et va recruter différents complexes de régulation épigénétique pour maintenir la répression de différents locus. Sur le chromosome maternel, *KCNQ1OT1* n'est pas exprimé, ce qui permet l'expression des différents gènes des différents locus sur ce chromosome (Adapté de Kanduri, 2015).

Toutes ces régulations modulées par les lncRNAs sont réalisées notamment par le biais du recrutement de différents complexes de régulation épigénétique. En effet, les lncRNAs *AIR* et *KCNQ1OT1* peuvent recruter spécifiquement l'histone méthyltransférase G9a pour induire la triméthylation H3K9 et inhiber transcriptionnellement les locus *IGF2R* et *KCNQ1*, respectivement (Nagano et al., 2008)(Pandey et al., 2008). *H19* participe également à la répression de l'expression du locus *IGF2* en recrutant MBD1 (methyl-CpG-binding domain protein) (Monnier et al., 2013). Cette protéine, membre des méthyl-CpG-binding domain proteins, interagit avec l'ADN méthylé et permet

le recrutement d'histone déacetylases et d'histone méthyltransferases pour le maintien de la répression transcriptionnelle.

Ces différents exemples de lncRNAs impliqués dans la mise en place de l'empreinte parentale illustrent relativement bien la gamme des différentes régulations épigénétiques dans lesquelles interviennent les lncRNAs, de la modification post-traductionnelle des histones à la méthylation de l'ADN.

ii. Modifications post-traductionnelles des histones

Les lncRNAs ne sont pas uniquement impliqués dans la régulation épigénétique des régions soumises à l'empreinte parentale. L'équipe de John Rinn a évalué qu'environ 20% des lincRNAs exprimés dans différents tissus interagissaient avec la sous unité SUZ12 (Suppressor of zeste 12) et/ou EZH2 (Enhancer of Zeste Homolog 2) du complexe PRC2 (Polycomb Repressive Complex 2) (Khalil et al., 2009). Le complexe PRC2 est composé de EZH2, SUZ12, EED (Embryonic ectoderm development) et RbAp46/48 (Figure 39). La sous unité EZH2 possède un domaine SET nécessaire pour l'activité histone lysine méthyltransferase sur la lysine 27 de l'histone H3 (H3K27). EED a un rôle d'échaffaudage pour permettre l'interaction entre PRC2 et son substrat histone H3 et la reconnaissance de résidus lysine méthylés, pour augmenter l'activité de PRC2. Enfin, SUZ12 est nécessaire pour la stabilité de EZH2 et RbAp46/48, et il est connu pour interagir avec les histones H3 et H4, augmentant ainsi l'interaction de PRC2 à son substrat.

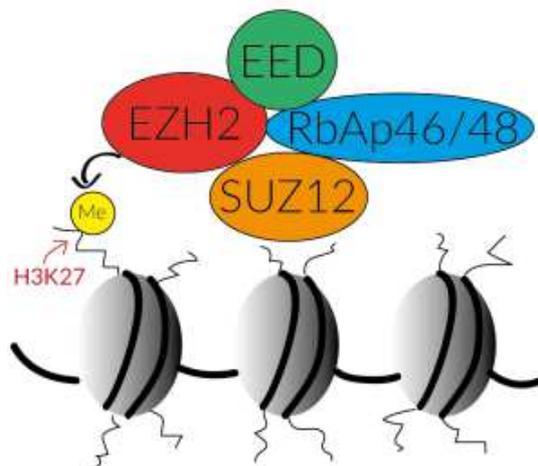


Figure 39: Représentation schématique du complexe PRC2 et de son mécanisme d'action.

A l'heure actuelle, la majorité des lncRNA étudiés sont des lncRNAs "guides" pour les complexes de modifications de la chromatine.

Le lncRNA "guide" le plus étudié est certainement *XIST*, qui a été décrit comme essentiel au cours du développement pour l'inactivation du deuxième chromosome X. Ce lncRNA est exprimé

uniquement chez la femelle à partir du chromosome X qui sera inactivé (Brown et al., 1992). L'expression de *XIST* permet de recruter le complexe de répression épigénétique PRC2 pour inactiver le futur chromosome Xi (inactive X) via la triméthylation H3K27 (Zhao et al., 2008). Le recrutement de PRC2 est réalisé grâce à un domaine conservé dans l'évolution de *XIST*, RepA (Repeat A), qui peut interagir avec la sous-unité méthyltransférase EZH2 de PRC2 (Maenner et al., 2010). Il y a un contrôle très fin du dosage pour l'inactivation du X dans les stades précoces du développement. En effet, les études de pertes de fonction chez l'animal ont démontré l'importance développementale de *XIST* puisque les femelles murines possédant un *XIST* muté d'origine paternelle décédaient *in utero*. Ces décès sont attribués à l'activation des deux chromosomes X dans les tissus embryonnaires, empêchant le développement de l'embryon (Marahrens et al., 1997). Il existe de nombreux exemples de lncRNA régulant l'expression génique en *cis* comme *XIST*. Le lncRNA *ANRIL* recrute le complexe PRC1 (Polycomb Repressive Complex 1) via la sous-unité CBX7 (chromobox homolog 7), pour maintenir la répression transcriptionnelle du locus *INK4* par une action de mono-ubiquitination de H2AK119 (Yap et al., 2010). À l'inverse, *HOTTIP* a été montré comme interagissant avec la protéine WDR5 (WD repeat domain 5), une sous-unité composant le complexe MLL/Trx (Trithorax) qui permet d'induire les modifications activatrices H3K4 triméthylation du locus *HOXA* (Wang et al., 2011b).

Le recrutement de complexes épigénétiques peut s'effectuer soit en *cis* soit en *trans*. Par exemple, le lncRNA *HOTAIR* est exprimé à partir du locus *HOXC* sur le chromosome 12, et recrute PRC2 pour la régulation négative en *trans* du locus *HOXD* sur le chromosome 2 (Rinn et al., 2007a). À l'inverse, le lncRNA *HOTTIP* régule en *cis* le locus *HOXA* (Wang et al., 2011b).

iii. Modifications de l'ADN

Ces régulations épigénétiques induites par les lncRNAs ne sont pas que la conséquence de modifications post-traductionnelles des histones. Il existe également des lncRNAs intervenant dans les modifications de l'état de méthylation de l'ADN et des îlots CpG par le biais de recrutement d'ADN méthyltransférases. Les pRNAs (Promoter-associated RNAs) des gènes ribosomaux répriment par ce type de mécanisme l'expression des ARNr. En effet, il a été montré que ces transcrits complémentaires des promoteurs des ARNr peuvent former un triplex ARN/ADN avec ces régions régulatrices. Ce triplex est reconnu par l'ADN méthyltransférase DNMT3b qui va catalyser la réaction de méthylation de l'ADN, réprimant ainsi la transcription des ARNr (Schmitz et al., 2010). Il a également été montré que *KCNQ1OT1* pouvait maintenir l'extinction allèle spécifique des gènes soumis à l'empreinte parentale via le recrutement de DNMT1 (Mohammad et al., 2010). À l'inverse, le lncRNA *Evf2* a été décrit comme un inhibiteur de la méthylation des îlots CpG de l'enhancer de *Dlx5/6* en séquestrant MECP2 (Methyl-CpG binding protein 2), permettant ainsi la modulation de l'expression de ces gènes homéotiques (Berghoff et al., 2013).

iv. Mécanisme de ciblage?

Les lncRNAs servent donc de "guide" ou "d'adaptateurs" pour recruter un ou même plusieurs complexes protéiques (Khalil et al., 2009). *HOTAIR* peut recruter PRC2 grâce à un domaine spécifique en 5' mais aussi le complexe LSD1 (Lysine-Specific Histone Demethylase)-CoREST avec un autre domaine en 3'. Cette combinaison augmente la répression des gènes cibles avec la triméthylation H3K27 réalisée par PRC2, et la déméthylation de H3K4m2 catalysée par LSD1-CoREST (Tsai et al., 2010b) (Figure 40A). *KCNQ1OT1* peut recruter à la fois PRC2 et G9a pour promouvoir deux marques de répression H3K27me3 et H3K9me3 (Pandey et al., 2008), et *ANRIL* interagit avec PRC1 mais également avec PRC2 pour la régulation de *INK4* (Kotake et al., 2011) (Figure 40B). On peut ainsi concevoir que certains lncRNAs pourraient recruter et assembler différents facteurs de modifications de la chromatine pour permettre un ciblage génomique spécifique.

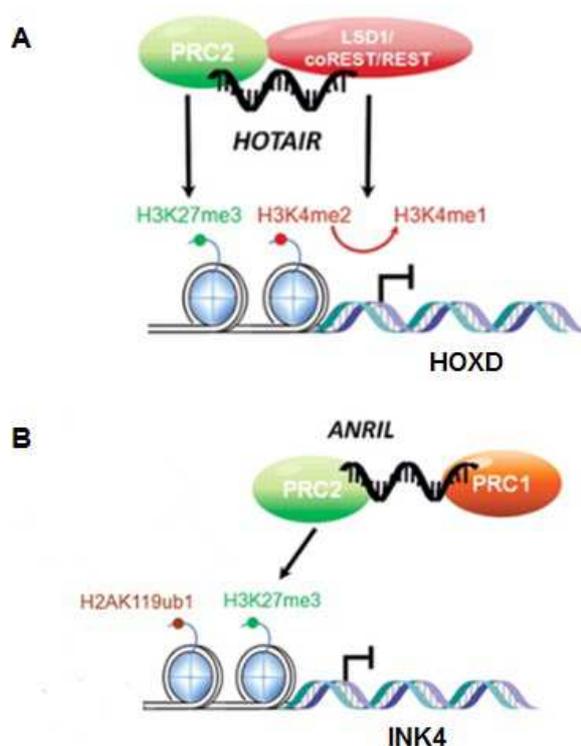


Figure 40 : Représentation schématique du recrutement de complexes épigénétiques par les lncRNAs. A. Le lncRNA *HOTAIR* peut recruter à la fois PRC2 et LSD1-CoREST pour permettre la triméthylation H3K27 et la déméthylation de H3K4m2. B. Le lncRNA *ANRIL* peut recruter à la fois PRC2 et PRC1 pour promouvoir les marques répressives H3K27me3 et la mono-ubiquitination de H2AK119 (Adapté de Kaikkonen et al., 2011).

En plus de cette capacité à lier plusieurs partenaires, les cibles des lncRNAs peuvent être également nombreuses. Afin de pouvoir documenter ces sites dans le génome, la technique de ChIRP a été développée. Ce type d'expérience fait intervenir des sondes biotinylées spécifiques du lncRNA,

permettant l'immunoprécipitation du lncRNA impliqué dans un complexe avec des protéines et l'ADN (Figure 41). Le séquençage de l'ADN immunoprécipité permet ainsi de mettre en évidence toutes les cibles des lncRNAs. De cette manière, on a pu observer l'occupation de *TERC* (Telomerase RNA component) au niveau des télomères, et que *HOTAIR* pouvait recruter PRC2 sur des centaines de sites dans le génome humain (Chu et al., 2011)(Chu et al., 2012).

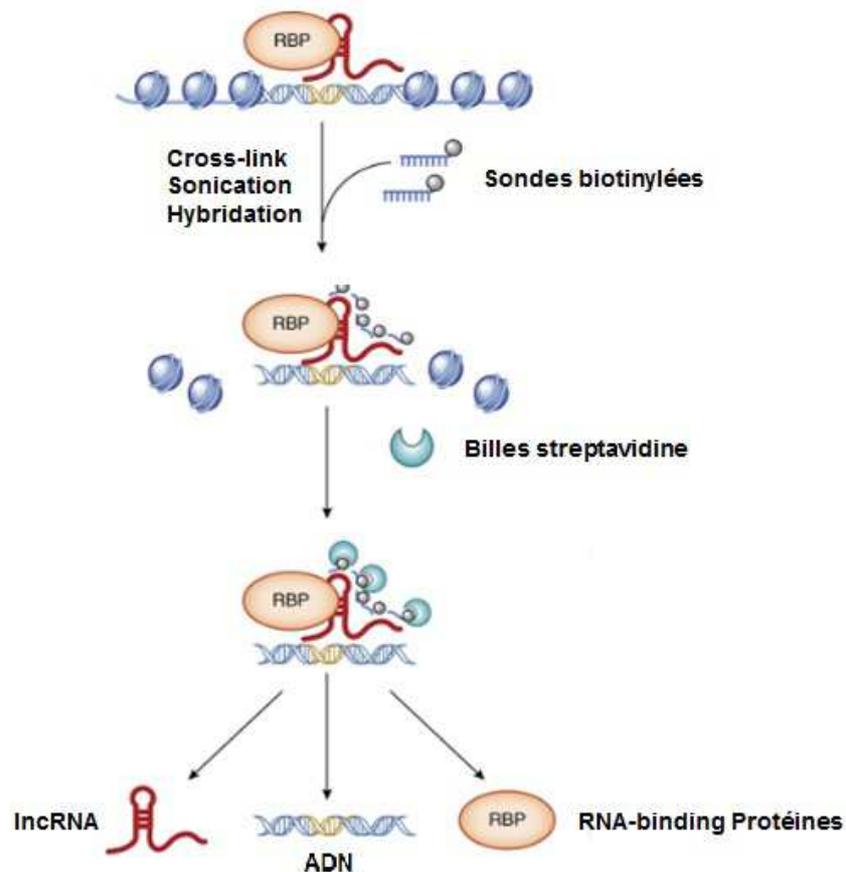


Figure 41 : Représentation schématique du protocole de ChIRP pour déterminer les régions génomiques avec lesquelles interagit un lncRNA. Après cross-linking par glutaraldéhyde et sonication, des sondes biotinylées sont hybridées au lncRNA ciblé. Le complexe lncRNA/ADN/protéines est immunoprécipité avec l'utilisation de billes magnétiques streptavidine et les sites cibles du lncRNA sont déterminés par séquençage. Les protéines partenaires peuvent être déterminées par western blot ou spectrométrie de masse (Adapté de Chu et al., 2012).

Comment ces lncRNAs recrutent-ils des complexes de régulation épigénétique sur des régions spécifiques du génome? Cette question demeure sans réponse précise à l'heure actuelle, même s'il existe quelques pistes intéressantes. Une partie de la réponse réside dans l'homologie de séquence entre un lncRNA et sa cible, comme dans le cas des pRNAs, mais cela n'explique pas les multiples cibles génomiques pour certains de ces lncRNAs. Il a alors été suggéré d'autres mécanismes. Par exemple, la capacité de *XIST* à se localiser sur le chromosome Xi dépend en partie de la capacité de sa

région RepC (Repeat C) à interagir avec le facteur de transcription séquence-spécifique YY1. En effet, l'insertion ectopique dans le génome de motif de reconnaissance pour YY1 permet de mobiliser *XIST* à partir du Xi vers ces sites ectopiques (Jeon and Lee, 2011). Les lncRNAs peuvent également agir comme un échafaudage rapprochant un complexe additionnel pour potentiellement déterminer la spécificité comme *HOTAIR* qui fait interagir PRC2 et LSD1 (Tsai et al., 2010b). Un autre modèle d'interaction fait intervenir des boucles chromatinienne, comme dans le cas de *HOTTIP* régulant en *cis* l'expression du locus *HOXA* (Figure 42). Lorsque *HOTTIP* est transcrit de manière ectopique dans le génome, il n'est pas capable de réguler l'expression du locus *HOXA*, suggérant un mécanisme indépendant de la séquence cible mais dépendant de la proximité spatiale. Ainsi, il a été montré que l'ARN endogène naissant *HOTTIP* est ramené à ses gènes cibles proximales via la formation d'une boucle chromatinienne, permettant ainsi le recrutement de complexe de régulation épigénétique (Wang et al., 2011b).

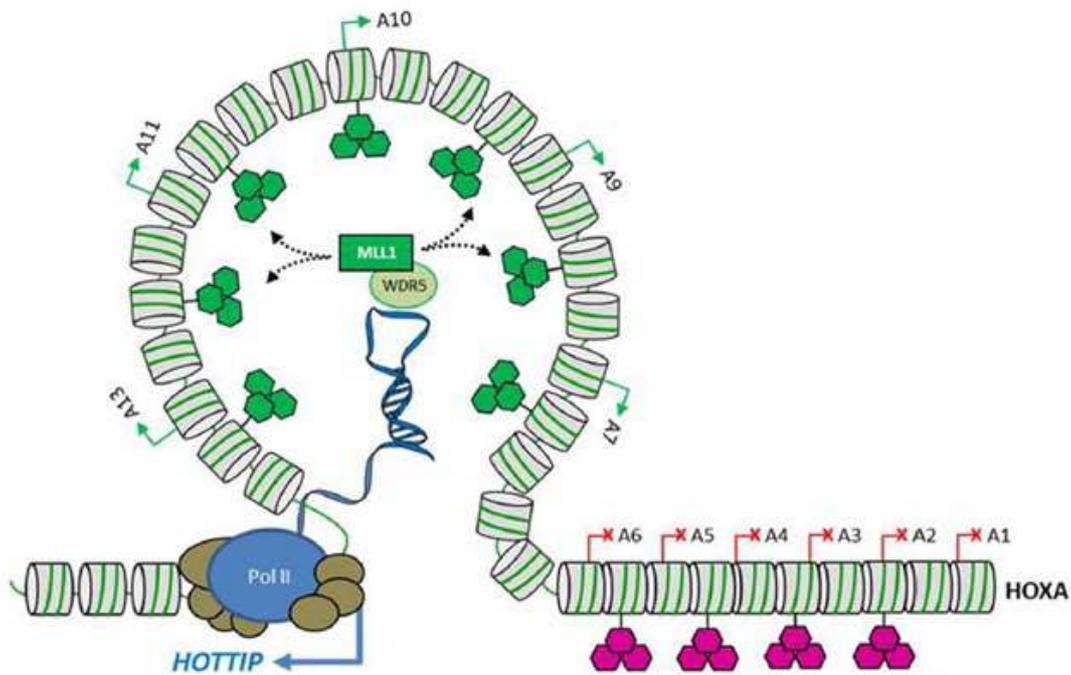


Figure 42 : Représentation schématique de la régulation du locus *HOXA* par le lncRNA *HOTTIP*. *HOTTIP* active la transcription génique, en recrutant le complexe de régulation épigénétique WDR5/MLL responsable de la méthylation H3K4me3, et en modulant la formation d'une boucle de chromatine à l'extrémité du locus *HOXA*. Hexagone violet=H3K27me3 (répression) ; Hexagone vert=H3K4me3 (Activation) (Adapté de Angrand et al., 2015).

e. En tant que réservoirs à petits ARNs non codants

Bien que n'étant pas une fonction à proprement parler et dépendante des 4 archétypes de mode d'action décrits précédemment, plusieurs travaux ont suggéré et démontré qu'une partie des lncRNAs pouvait être également processée en petits ARNs non codants et ainsi être des réservoirs à petits ARNs tels que les microARNs ou les snoARNs (Jalali et al., 2012). Il est important de noter que cette biogenèse de petits ARNs peut être la résultante de différents mécanismes: expression colinéaire du petit ARN et du lncRNA hôte à partir d'un promoteur commun, ou expression dépendante du mécanisme d'épissage. De plus, l'expression du petit ARN peut être également dépendante d'un promoteur interne et donc indépendante de l'expression du lncRNA.

i. Réservoirs à microARNs

Les plus décrits dans la littérature sont les lncRNAs contenant des microARNs. En effet, comme le clivage du microARN primaire induit par l'enzyme RNase III Drosha conduit à au moins trois fragments d'ARNs, les microARNs sont rarement localisés dans des gènes codants. Au contraire, ils sont plus généralement localisés dans les introns et exons de gènes non codants (Cullen, 2004). Dans ce contexte, le lncRNA est un gène dit hôte. Il existe différents exemples et différents modèles de biogenèse de microARNs à partir de gène hôte (Figure 43). Le lncRNA peut être considéré comme le pri-miRNA qui sera ensuite classiquement clivé par Drosha. Dans un autre modèle, c'est l'intron contenant le microARN qui est épissé par le spliceosome, et reconnu comme substrat directement par l'enzyme Dicer. On appelle ce type de microARN, les mirtrons. De manière un peu plus particulière, la biogenèse de certains microARNs introniques ne dépend ni de l'épissage, ni de la voie canonique de biogenèse des microARNs, mais reste dépendante du clivage réalisé par Drosha. On appelle ce type de microARN, les simtrons (Havens et al., 2012).

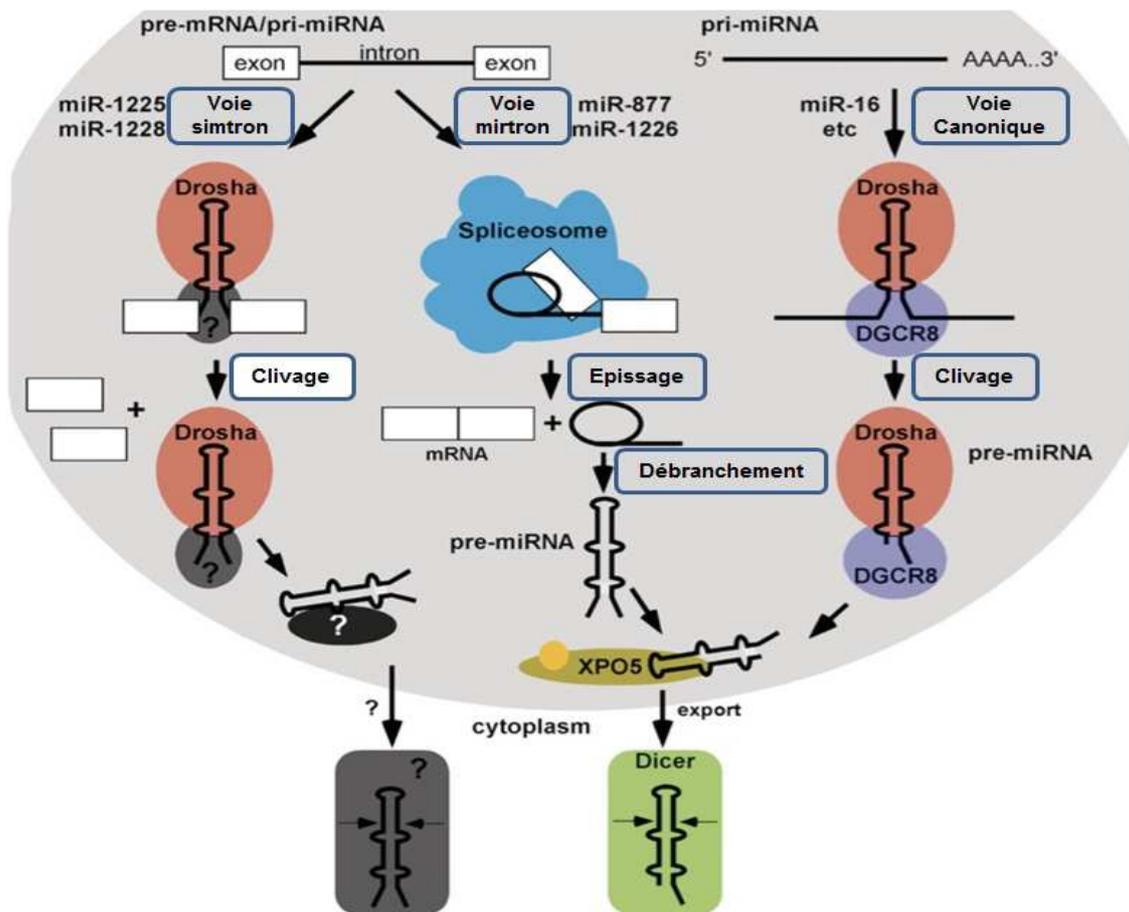


Figure 43 : Modèles de biogénèse des microARNs. **Voie simtron** : l'excision des microARNs issus de l'intron du gène hôte est dépendante de Drosha et d'un potentiel partenaire encore inconnu. Les simtrons sont ensuite processés et exportés par un facteur inconnu et introduits dans le complexe RISC. **Voie mirtron** : Les mirtrons sont excisés du gène hôte par le spliceosome, débranchés, exportés par l'exportin 5, clivés par Dicer et entrent finalement dans le complexe RISC. **Voie canonique** : Les microARNs sont processés par Drosha et DGCR8, exportés par l'exportin 5, clivés par Dicer et entrent dans le complexe RISC. Les 3 voies produisent des microARNs fonctionnels (Adapté de Havens et al., 2012).

DLEU2 est un bon exemple de lncRNA en tant que gène hôte de microARNs. Ce lncRNA est reconnu comme le pri-miRNA par Drosha pour l'excision des microARNs miR15a/miR16-1. Par ailleurs Lerner *et al.* ont montré que l'expression de ces microARNs est entièrement dépendante de l'expression de *DLEU2* (Lerner et al., 2009). Le lncRNA *H19* est quant à lui reconnu comme le pri-miRNA du miR-675. En effet, ce dernier est généré à partir du premier exon de *H19* vraisemblablement suite au clivage par Drosha (Cai and Cullen, 2007). Le microARN et la structure de l'exon de *H19* sont conservés dans l'évolution et tous les deux sont notamment impliqués dans l'ostéoarthrite (Steck et al., 2012). Comme cité précédemment, *H19* a également un rôle dans le maintien de l'empreinte parentale en tant que guide pour des complexes protéiques. Ceci suggère donc

qu'un lncRNA peut avoir une fonction moléculaire, indépendamment de son rôle en tant que gène hôte à petits ARNs non codants.

ii. Réservoirs à snoARNs

La majorité des snoARNs sont exprimés à partir d'introns de gènes hôtes, l'expression des snoARNs est donc dépendante de l'expression du gène hôte et de l'épissage consécutif. Il a ainsi été mis en évidence que ce gène hôte pouvait être un gène non codant tel que *UHG* (U22 Host Gene), un SnHG (Small Nucleolar RNA Host Gene) générant 8 snoARNs à partir de ses introns (U22 et U25 à U31) (Figure 44). Ce lncRNA d'environ 1kb ne présente qu'une conservation très limitée de ces exons, à l'inverse de ces introns, suggérant son implication biologique qu'en tant que lncRNA producteur de snoARNs (Tycowski et al., 1996). Plusieurs autres SnHGs sont décrits dans la littérature comme des lncRNAs contenant des snoARNs comme *U17HG* et *U19HG*. *GAS5* est un autre lncRNA de la classe des SnHGs qui a été initialement mis en évidence en 1988 comme un gène tumeur suppresseur surexprimé lors de l'arrêt de la croissance cellulaire (Schneider et al., 1988). Ce lncRNA contient 10 snoARNs chez l'homme et l'équivalent murin de *GAS5* contient 9 snoARNs, et sa fonction suppresseur de tumeur étant indépendante de sa fonction de producteur de snoRNAs (Sun et al., 2014b).

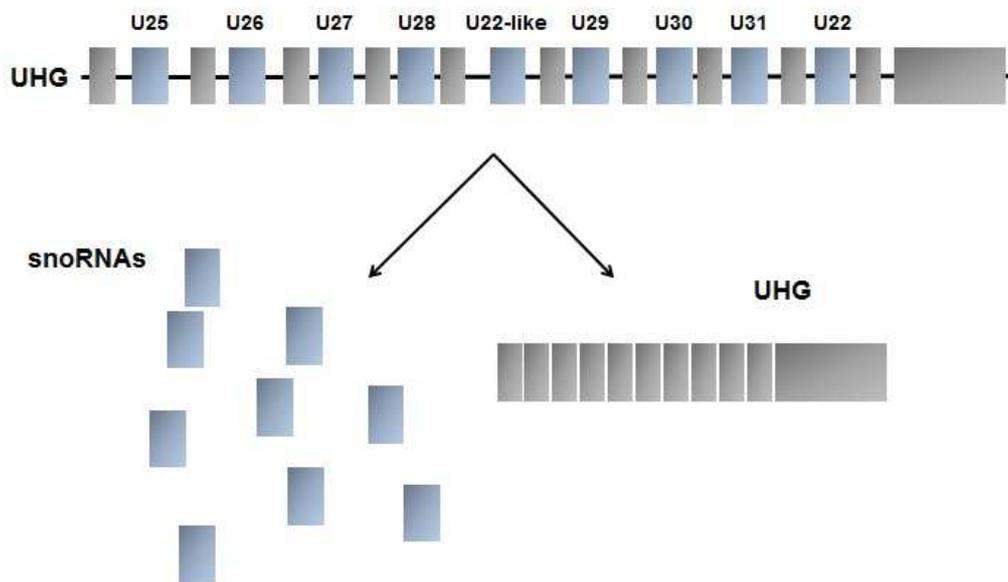


Figure 44 : Exemple de lncRNA producteur de snoARNs. Le gène UHG conduit à la production de 8 snoARNs après épissage de ses introns.

iii. Autres petits ARNs non codants

Un exemple tout à fait particulier d'autres petits ARNs non codants produits par des lncRNAs est produit par le lncRNA *MALATI*. En effet, *MALATI* peut être clivé en deux ARNs non codants, un lncRNA qui est principalement nucléaire et un petit ARN non codant de 61nt, cytoplasmique, arborant

une structure semblable à un ARNt appelé *mascRNA* (MALAT1-associated Small Cytoplasmic RNA) (Figure 45). La fonction de ce petit ARN n'est pas encore clairement définie mais il est suggéré qu'il pourrait servir de censeur à la production de *MALAT1* ou de mimic à ARNt (Wilusz et al., 2008).

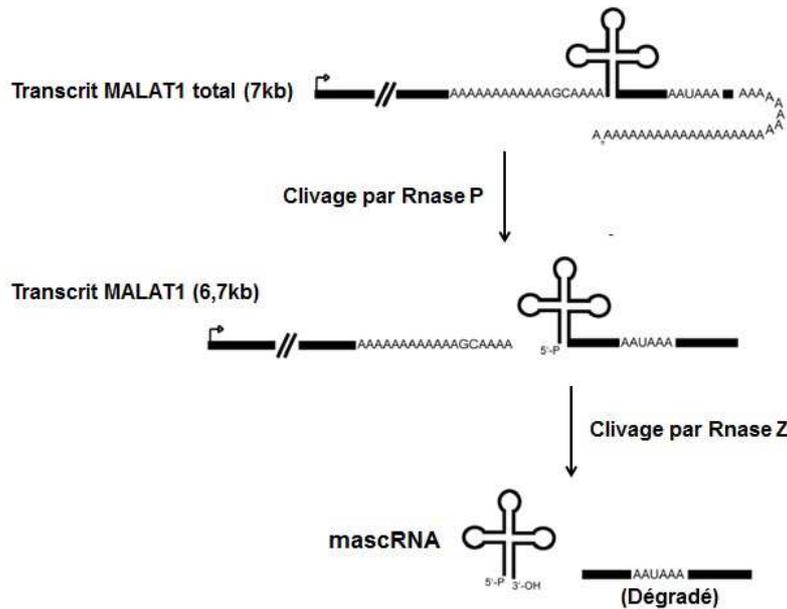


Figure 45 : Biogenèse de MALAT1 mature et *mascRNA*. Le transcrit initial MALAT1 de 7kb est clivé par la RNase P en deux transcrits : MALAT1 mature de 6,7kb et un autre transcrit plus petit. Ce dernier est clivé par la RNase Z en *mascRNA* mature (Adapté de Wilusz et al., 2008).

f. Dans l'assemblage de structures nucléaires

Mis à part toutes ces régulations moléculaires au niveau transcriptionnel, post-transcriptionnel, traductionnel et épigénétique, les lncRNAs sont également impliqués dans des fonctions un peu plus particulières. Par exemple, certains lncRNAs ont été impliqués dans la formation et le maintien de structures ou compartiments nucléaires spécifiques. J'ai cité précédemment les lncRNAs *XIST*, *NEAT1* et *MALAT1*. Chez la femelle placentaire, le "recouvrement" de *XIST* sur Xi conduit à la compaction de ce chromosome, qui est reconnu comme le corpuscule de Barr au niveau de l'enveloppe nucléaire. J'ai également mentionné la formation de structures particulières appelées paraspeckles par le lncRNA *NEAT1* organisé sous forme de fibre. Le rôle de ces granules n'est pas clairement défini, mais ils pourraient jouer un rôle dans la rétention de certains ARNm (Bond and Fox, 2009). *MALAT1* est quant à lui impliqué dans la formation des speckles nucléaires, des granules de taille variable considérées comme des sites de stockage et de modification de facteurs d'épissage des ARNm.

En résumé, nos connaissances sur les lncRNAs sont à l'heure actuelle encore insuffisantes, conduisant à la description d'une classe d'ARN très hétérogènes à la fois sur le plan de leur séquence/structure mais surtout concernant leurs mécanismes d'action. La figure 46 est une synthèse des différentes fonctions des lncRNAs décrites précédemment.

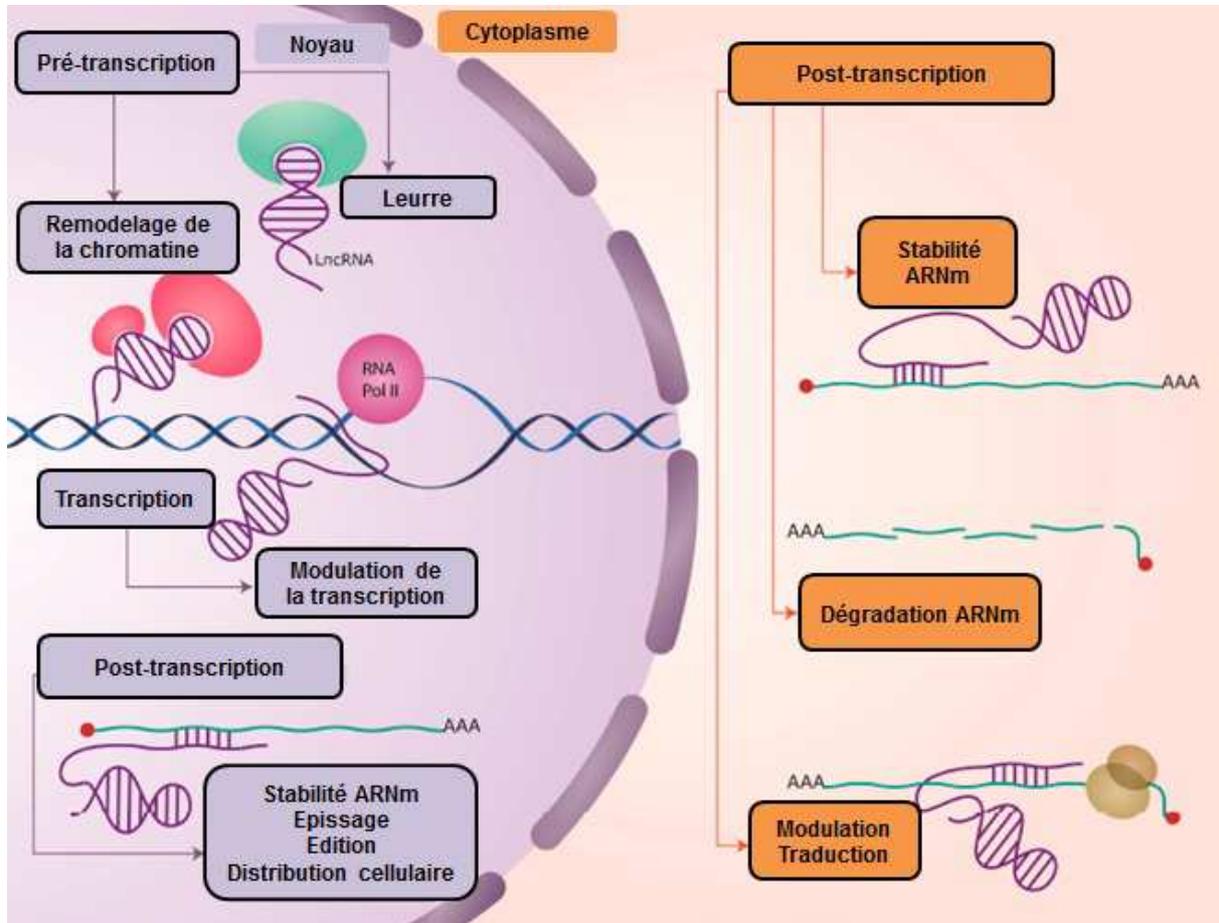


Figure 46 : Représentation schématique des mécanismes de régulation impliquant les lncRNAs (Adapté de Villegas and Zaphiropoulos, 2015).

Bien que la caractérisation fonctionnelle de dizaine de lncRNAs ait permis de mettre en évidence l'implication de ces molécules dans un champ très large de régulations moléculaires (de la régulation épigénétique à la régulation traductionnelle), une vaste majorité de ces transcrits reste encore méconnue. Il est également intéressant, voire fascinant, de spéculer que la vaste diversité de mécanismes de régulations épigénétiques, transcriptionnelles et post-transcriptionnelles, ne soit que le sommet de l'iceberg et que d'autres voies de modulation d'expression puissent être découvertes dans le futur. De plus, des lncRNAs tels que *H19* ou *GAS5* ont montré que les lncRNAs ne sont pas restreints à une seule fonction mais peuvent être bi- voire multi-fonctionnels. *H19* par exemple, est à la fois un producteur de microARN, un séquestreur de microARN, et un guide pour des complexes de modification de la chromatine. *HOTAIR* intervient dans la régulation épigénétique, et également dans

la régulation de la stabilité protéique, en facilitant l'ubiquitinylation de certaines protéines lors de la sénescence (Yoon et al., 2013).

VI) Implications des lncRNAs dans les pathologies humaines

Au vue des multiples processus moléculaires dans lesquels sont impliqués les lncRNAs, il peut paraître logique de les voir impliqués et dérégulés dans les pathologies humaines, que ce soit dans les maladies génétiques, les infections ou les cancers.

De par leur importance dans le développement, les loci soumis à l'empreinte parentale sont impliqués dans un certain nombre de maladies et de syndromes chez l'homme. L'extinction d'*IGF2* ou l'expression biallélique du lncRNA *H19* conduisent au syndrome de Silver-Russell. Les personnes touchées présentent un retard de croissance intra-utérin et post-natal. De plus, la méthylation aberrante de *H19* est impliquée dans l'augmentation du risque de cancer, la présence d'anomalies à la naissance et dans le développement du syndrome de Beckwith-Wiedemann (DeBaun et al., 2002). Ce même défaut de méthylation du promoteur du lncRNA *KCNQ1OT1* intervient également dans le syndrome Beckwith-Wiedemann (Higashimoto et al., 2006). Les patients atteints de cette pathologie présentent une macroglossie, un excès de croissance pré- et/ou post-natal, et des prédispositions au développement de tumeurs embryonnaires.

Des analyses comparatives de transcriptome ont montré une expression différentielle de plus de 500 lncRNAs durant la réponse immunitaire à des infections virales, suggérant un rôle important de la part des lncRNAs dans la régulation de l'immunité (Peng et al., 2010). *Gas5* a par exemple été lié à la susceptibilité de développer un lupus érythémateux disséminé dans des modèles murins, probablement à travers son effet immunosuppresseur de la réponse aux glucocorticoïdes en tant que leurre (Haywood et al., 2006). Le lncRNA *lncRNA-CD244* a quant à lui été impliqué dans la réponse immunitaire à l'infection par la tuberculose. Ce lncRNA, surexprimé chez les patients infectés par la tuberculose, inhibe l'expression d'IFN- γ /TNF- α par le recrutement de PRC2 au niveau de leurs promoteurs dans les cellules T CD8⁺. De plus, le transfert de cellules T CD8⁺ déléetées du lncRNA dans des souris infectées par *Mycobacterium Tuberculosis* a montré une plus grande résistance à l'infection de ces animaux, démontrant ainsi l'impact de l'expression de *lncRNA-CD244* dans la réponse immunitaire à la tuberculose (Wang et al., 2015b).

Le lncRNA *HI-LNC25* a été identifié comme spécifiquement exprimé dans les îlots du pancréas et significativement sous exprimé dans le diabète de Type 2. De plus, il semblerait que ce lncRNA régule l'expression du gène *GLIS3*, un facteur de transcription essentiel à la fonction normale des cellules bêta chez l'adulte (Morán et al., 2012). *H19* semblerait être dérégulé également dans la régulation du métabolisme du glucose des muscles chez les patients atteints de diabète de type 2 (Gao et al., 2014). D'un point de vue génétique, le lncRNA *ANRIL* a été décrit comme un hot spot à

mutation par GWAS (Genome-Wide Association Studies) dans le diabète de type 2, suggérant son implication dans cette pathologie (Pasmant et al., 2011).

Un certain nombre de lncRNAs ont été impliqués dans des pathologies neurodégénératives. Le lncRNA *BACE1-AS* est surexprimé dans la maladie d'Alzheimer et régule la stabilité de l'ARNm de *BACE1* (Liu et al., 2014). L'anti-sens du gène *UCHL1* est également impliqué dans la maladie de Parkinson et d'Alzheimer (Carrieri et al., 2012). Le lncRNA *Gomafu* est sous-exprimé dans les neurones corticaux chez la souris après dépolarisation induite par le chlorure de potassium. Il a été montré par Barry *et al.* en 2014 que ce transcrite régule l'épissage alternatif des ARNm de *DISC1* (Disrupted in schizophrenia 1) et *ERBB4* (Erb-b2 receptor tyrosine kinase 4) (Barry et al., 2014). Il est intéressant de noter que ces changements d'épissage, notamment de *DISC1*, ont été associés à des troubles neurologiques comme la schizophrénie et la dépression (Thomson et al., 2013).

VII) Implications des lncRNAs dans les cancers

Le cancer est une maladie complexe, impliquant de vastes changements de l'expression génique pouvant participer à différents pans du développement cancéreux tels que la prolifération cellulaire, l'apoptose, la formation de métastases ou même l'angiogenèse. Les lncRNAs ont fourni une nouvelle vision de la biologie du cancer. C'est certainement dans les cancers que l'impact des lncRNAs est le plus décrit. Par ailleurs, une grande partie de ceux décrits dans la littérature ont été mis en évidence par des études transcriptomiques comparatives entre tissus sains et cancéreux. Bien que les mécanismes d'action de la plupart des lncRNAs dérégulés dans les cancers n'aient pas été complètement élucidés, la communauté scientifique a commencé récemment à poser des questions de plus en plus complexes telles que : est-ce que les lncRNAs peuvent être la cause sous-jacente des cancers? Est-ce qu'ils sont la conséquence de la maladie elle-même? Est-ce que la dérégulation de leur expression affecte la différenciation cellulaire, la croissance ou le potentiel métastatique? De manière intéressante, la majorité des lncRNAs fonctionnels semblent être impliqués dans la modulation de l'épigénome. Or les cancers sont clairement associés à des profils épigénétiques très perturbés par rapport aux tissus contrôles sains. Il existe des exemples de lncRNAs impliqués dans tous types de cancers, comme tumeurs suppresseurs ou oncogènes, certains étant même de bons biomarqueurs associés à différentes caractéristiques cliniques. De plus, on peut citer des exemples de lncRNAs impliqués dans tous les processus liés aux cancers décrits par Hanahan et Weinberg en 2000, renforçant l'importance de l'étude de ces molécules dans ces pathologies (Hanahan and Weinberg, 2000).

a. Prolifération

En 2011, Prensner *et al.* ont identifié plus d'une centaine de lncRNAs dérégulés dans le cancer de la prostate appelés *PCATs* (Prostate Cancer-Associated ncRNA Transcripts), et le lncRNA *PCAT-1*

a été caractérisé plus fonctionnellement. En effet, les auteurs ont montré qu'il était surexprimé dans les cancers de la prostate de haut grade et métastasés, qu'il était régulé par PRC2, et qu'il modulait l'expression de plus de 370 gènes incluant *BRCA2* (Breast cancer 2). De plus, les expériences de surexpression et d'inhibition de ce lncRNA ont mis en évidence son effet pro-prolifératif drastique *in vitro* (Prensner et al., 2011)(Prensner et al., 2014). *H19* est également dérégulé dans les cancers avec des surexpressions reportées dans plusieurs types de cancers comme le cancer de la vessie, du côlon ou du sein. Il a par ailleurs été montré que l'expression de *H19* avait un effet positif sur la croissance tumorale et Luo *et al.* ont démontré qu'il augmentait la prolifération des cellules cancéreuses de la vessie en inhibant l'expression d'*ID2* (Inhibitor of DNA binding/Differentiation 2) (Luo et al., 2013).

b. Maintien de la croissance

Un des mécanismes pour maintenir le taux de croissance élevé des cellules cancéreuses, est d'inhiber des suppresseurs de tumeur. Les lncRNAs PSF Inhibiting RNAs qui permettent la transcription du prot-oncogène *GAG6* ont été observés comme surexprimés dans 9 lignées tumorales humaines, suggérant ainsi que l'expression de ces lncRNAs contribuent à la tumorigénicité (Song et al., 2004). *ANRIL* utilise un tout autre mécanisme pour inhiber les suppresseurs de tumeurs. En effet, ce lncRNA réprime l'expression de *p15* (INK4B), un suppresseur de tumeur bien connu, et la déplétion de *ANRIL in vitro* permet l'expression de *p15* et l'inhibition de la prolifération (Kotake et al., 2011). Une autre étude a mis en évidence que *ANRIL* était surexprimé dans le cancer de la prostate et qu'il pouvait également recruter le complexe PRC1 pour réprimer *p16* et *p14* (Yap et al., 2010). Ce lncRNA est par ailleurs associé à un mauvais pronostic dans d'autres cancers comme le carcinome hépatique (Huang et al., 2015), le cancer du poumon (Lin et al., 2015) et dans le cancer ovarien séreux (Qiu et al., 2015).

A l'inverse, *GAS5* est un lncRNA se comportant comme un suppresseur de tumeur, notamment via l'inhibition de la réponse aux glucocorticoïdes rendant les cellules plus sensibles à l'apoptose (Kino et al., 2010). En effet, l'expression de *GAS5* induit l'arrêt de croissance et l'apoptose dans des lignées tumorales de la prostate et du sein, et une sous-expression de *GAS5* a été notée dans des cancers du sein (Mourtada-Maarabouni et al., 2009).

c. Apoptose et sénescence

Associées à la prolifération excessive, les cellules cancéreuses présentent également un défaut dans la régulation de la mort cellulaire, et plus particulièrement dans les mécanismes d'échappement au processus d'apoptose. En 2000, Srikantan *et al.* ont découvert un lncRNA surexprimé dans le cancer de la prostate, *PCGEM1* (Prostate-specific transcript 1) (Srikantan et al., 2000). Les études suivantes ont mis en évidence son effet anti-apoptotique. En effet, la surexpression de *PCGEM1* dans une lignée de cancer prostatique a montré une diminution de l'apoptose induite par la doxorubicine associée à un

retard de l'induction de p53 et p21, et une réduction du clivage de caspase 7 et PARP (Poly(ADP-ribose) polymerase). De plus, il semblerait que ce mécanisme soit androgène-dépendant puisque inexistant dans les lignées cancéreuses prostatiques sans récepteurs aux androgènes (Fu et al., 2006)(Petrovics et al., 2004). Dans une autre étude, les auteurs ont comparé les transcriptomes de lignées de mélanomes avec des mélanocytes et kératinocytes normaux, et mis en évidence la surexpression dans le mélanome du lncRNA *SPRY4-IT1*. Ce lncRNA, dérivé de l'intron du gène *SPRY4* (Sprouty RTK signaling antagonist 4), est principalement nucléaire, et les expériences d'inhibition de ce lncRNA ont montré une augmentation de l'apoptose et une diminution de l'invasion (Khaitan et al., 2011). Cependant, les mécanismes impliqués dans son effet anti-apoptotique ne sont pas encore connus. *ANRIL* est également impliqué dans la sénescence cellulaire. En effet, il a été montré dans des modèles de cancer gastrique qu'il régulait l'expression de miR-99a et miR-449a par répression épigénétique (Zhang et al., 2014a). Le miR-99a est un inhibiteur de la prolifération dans le carcinome hépatocellulaire (Li et al., 2011), et miR-499a induit l'arrêt du cycle cellulaire et la sénescence dans les cellules cancéreuses de la prostate (Noonan et al., 2010).

A l'inverse, le lncRNA *PINT* (p53 induced transcript), régulé par p53 et sous-exprimé dans les cancers colorectaux, est responsable de l'arrêt du cycle cellulaire et promeut l'apoptose indépendamment des dommages à l'ADN via des régulations épigénétiques dans lesquelles intervient PRC2 (Marín-Béjar et al., 2013).

d. Angiogenèse

Ce processus est nécessaire à la croissance, à l'augmentation de la masse et de la taille des tumeurs. L'angiogenèse est généralement régulée par le facteur clé HIF α (Hypoxia Induced Factor Alpha) qui est nécessaire à l'activation de gènes induits par l'hypoxie et la stabilisation de p53 au cours de l'hypoxie. Il a été mis en évidence un lncRNA anti-sens de HIF α et complémentaire de sa région 3'UTR, *aHIF* (Thrash-Bingham and Tartof, 1999). Plusieurs études ont montré que l'hypoxie, et plus précisément l'hypoxie prolongée (>12h), induisait l'expression de *aHIF* (par les protéines HIF) pour diminuer HIF α en augmentant la dégradation de l'ARNm (Rossignol et al., 2002)(Uchida et al., 2004). Cette régulation forme donc une boucle de rétrocontrôle modulant la réponse à l'hypoxie. Le lncRNA *aHIF* est exprimé dans plusieurs tissus sains et dérégulé dans plusieurs types de cancers (Thrash-Bingham and Tartof, 1999)(Rossignol et al., 2002). De plus, ce lncRNA peut être associé à un mauvais pronostic dans le cancer du sein (Cayre et al., 2003).

Le lncRNA-*LET* ('low expression in tumor') est sous-exprimé dans les carcinomes hépatiques et pulmonaires et dans le cancer colorectal. Il a été montré que l'expression de HDAC3, induite par l'hypoxie, réprimait l'expression de *lncRNA-LET* dans ces différents cancers. De plus, l'expression ectopique de *lncRNA-LET* a démontré qu'il inhibait la formation de métastases alors que sa déplétion

augmentait l'invasion dans des modèles de xénogreffes (Yang et al., 2013). La diminution d'expression de ce lncRNA a également été associée à la stabilisation de NF90 (Nuclear Factor 90), qui est nécessaire à l'induction de l'hypoxie des cellules cancéreuses, suggérant fortement le rôle de *lncRNA-LET* en tant que régulateur dans la signalisation de l'hypoxie (Tominaga-Yamanaka et al., 2012).

e. Invasion et Métastases

MALAT1 (Metastasis associated lung adenocarcinoma transcript 1) est surexprimé dans plusieurs cancers comme le cancer du poumon (Ji et al., 2003) ou le sarcome endométrial (Yamada et al., 2006). Tano *et al.* ont mis en évidence l'implication de *MALAT1* dans la formation de métastases dans les cancers. En effet, la surexpression de ce lncRNA augmente la motilité cellulaire de lignées cancéreuses pulmonaires par la régulation transcriptionnelle et post-transcriptionnelle de gènes comme *CCT4* (Chaperonin containing TCP1, subunit 4), *PTBP3* (polypyrimidine tract binding protein 3), et de gènes impliqués dans la régulation du cytosquelette et de la matrice extracellulaire (Tano et al., 2010). De plus, l'extinction de *MALAT1* dans une lignée de cancer cervical réprime la prolifération et la capacité de migration de ces cellules avec une surexpression de Caspase 8 et 3, BAX (BCL2-associated X protein), et une sous-expression de Bcl-2 (B-cell CLL/lymphoma 2) et Bcl-xL (Guo et al., 2010).

HOTAIR est également impliqué dans ce processus. A la lumière de son importance dans la régulation transcriptionnelle et épigénétique, il n'est pas étonnant de le voir dérégulé dans divers cancers comme le carcinome hépatique (Yang et al., 2011), le cancer colorectal (Kogo et al., 2011) ou le cancer du sein (Gupta et al., 2010). Dans le cancer du sein, l'expression de *HOTAIR* est associée à la progression tumorale ainsi qu'au stade métastatique. À l'inverse, l'extinction de *HOTAIR* inhibe l'invasion cellulaire des cellules cancéreuses épithéliales. Les expériences de ChIRP ont montré qu'il y avait parmi les nombreuses cibles de *HOTAIR*, des gènes impliqués dans le processus métastatique (Chu et al., 2012).

f. Immortalité répllicative

Le potentiel de division presque illimité des cellules cancéreuses est associé au maintien des télomères afin d'empêcher la sénescence cellulaire. Ce processus est dépendant en grande partie de l'activité de la télomérase, exprimée et active dans la majorité des cancers (Bryan et al., 1997).

La télomérase est constituée de deux unités : la reverse transcriptase TERT (Telomerase Reverse Transcriptase) et un ARN non codant, *TERC* (Telomerase RNA Component). L'immortalisation des cellules par la synthèse des télomères est dépendante de *TERC*, qui sert de matrice à TERT pour la transcription reverse. Une amplification de l'expression de *TERC* en plus de l'expression de TERT ont été reportées dans plusieurs cancers, dont la leucémie et le carcinome

cervical (Nowak et al., 2006)(Andersson et al., 2006). De manière intéressante, il a été montré que les régions subtéломériques et téломériques peuvent être transcrites en ARNs non codants appelés *TERRAs* (Telomeric repeat-containing RNA). Ces lncRNAs interagissent, d'une part avec la composante ARN de la télomérase, et avec la sous unité reverse transcriptase de la télomérase TERT d'autre part. Ainsi, ces lncRNAs *TERRAs* agissent par compétition avec l'ADN téломérique et inhibent la fonction de la télomérase (Figure 47)(Redon et al., 2010). De plus, l'inhibition de l'expression de ces lncRNAs est nécessaire à l'activité de la télomérase pour le maintien des téломères et l'immortalité répliquative (Ng et al., 2009).

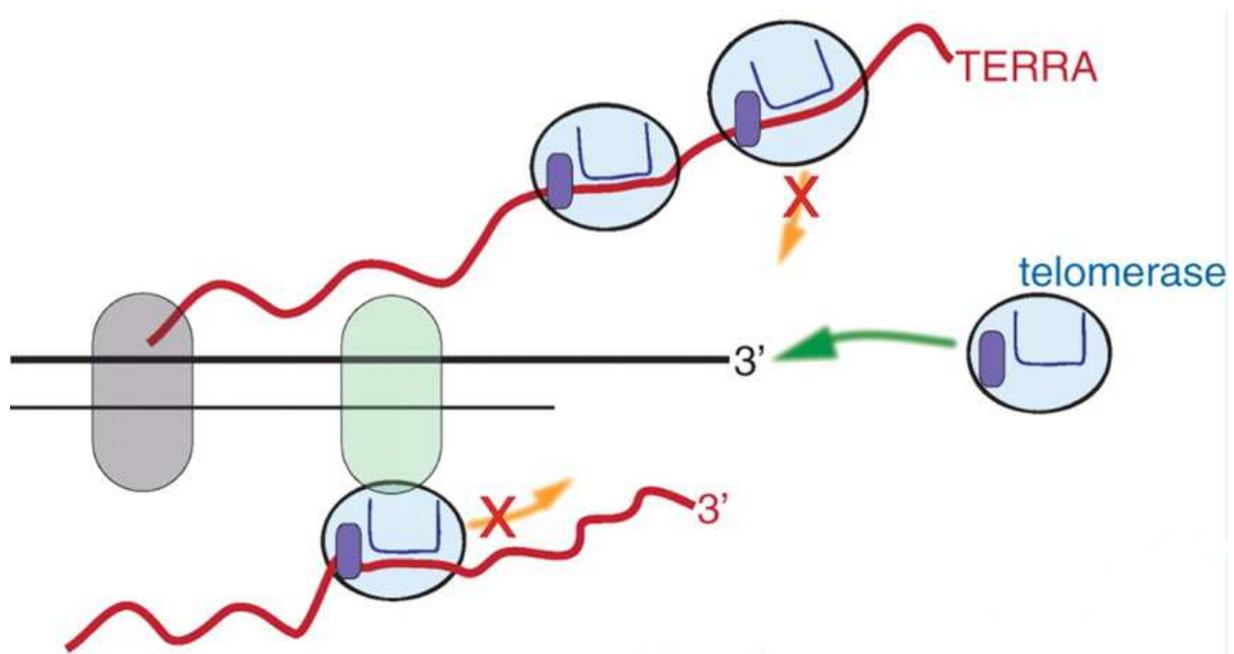


Figure 47 : Représentation schématique de l'inhibition de l'activité Télomérase induite par les lncRNAs TERRA.

Pour l'heure, il n'y a pas de preuve pertinente quant à la dérégulation potentielle, l'extinction ou la mutation des lncRNAs *TERRA* dans le cancer. Nous savons en revanche que ces lncRNAs sont fortement exprimés dans les tissus adultes sains avec peu ou pas d'activité télomérase, et faiblement exprimés dans les lignées cancéreuses comme la lignée Hela (Azzalin et al., 2007)(Schoeftner and Blasco, 2008).

VIII) Mécanismes de dérégulations des lncRNAs dans les cancers

Il n'y a que peu de données décrivant des mécanismes intervenant dans la dérégulation de l'expression des lncRNAs dans les cancers. La majorité des dérégulations des lncRNAs, associées aux cancers et qui ont été caractérisées, sont la conséquence de dérégulations transcriptionnelles et épigénétiques. Il existe également quelques travaux qui ont mis en avant les événements génétiques pouvant directement affecter l'expression des lncRNAs dans le cancer. Un de ces rares exemples est l'amplification du lncRNA *PVT-1*. Ce lncRNA est situé sur le chromosome 8q24 et de fréquentes amplifications de cette région ont été rapportées et associées à une survie diminuée (Popescu and Zimonjic, 2002). Le gène *PVT-1* est donc fréquemment amplifié et surexprimé dans le cancer du sein et de l'ovaire (Guan et al., 2007), ainsi que dans le cancer colorectal où cette surexpression est associée à une inhibition de l'apoptose et à un mauvais pronostic (Takahashi et al., 2014). Des altérations du nombre de copies d'une région du chromosome 3q13.31, contenant les lncRNAs *BC040587* et *LOC285194*, ont également été rapportées dans des cas d'ostéosarcomes. Il est intéressant de noter que le lncRNA *LOC285194* a été décrit comme un suppresseur de tumeur en inhibant la prolifération, et sa délétion est associée à un mauvais pronostic chez les patients atteints d'ostéosarcomes (Pasic et al., 2010). De manière intéressante, il existe des translocations chromosomiques impliquant les lncRNAs. Nakamura *et al.* ont ainsi mis en évidence la translocation t(1;3)(q25;q27) générant la fusion du gène *BCL6* (B-cell CLL/Lymphoma 6) avec *GAS5* chez un patient atteint de lymphome diffus à grandes cellules B (Nakamura et al., 2008).

IX) Biomarqueurs et potentiel clinique

Le fait que les lncRNAs puissent présenter une expression plus tissu-spécifique avec des variations d'expressions très larges entre les tissus par rapport aux gènes codants, suggère qu'ils sont potentiellement de bons candidats en tant que biomarqueurs ou de cibles. Un certain nombre d'entre eux sont associés à des paramètres cliniques, comme la réponse aux traitements, la survie, la rechute, etc... A titre d'exemples, *XIST* est fortement associé à la DFS (Disease Free Survival) chez les patients avec un cancer ovarien traité au Taxol (Huang et al., 2002), et *MALAT1* est un bon marqueur pronostique indépendant de la survie dans les premiers stades de l'adénocarcinome du poumon (Ji et al., 2003b). Certains lncRNAs sont particulièrement spécifiques du tissu néoplasique comme *CCATI*, uniquement exprimé par les cellules cancéreuses du côlon, suggérant que les lncRNAs pourraient être de bons marqueurs diagnostiques. Une méthode élégante de détection *in vitro*, *ex vivo* et surtout *in situ* a été développée par Kam *et al.*. En effet, les auteurs ont mis au point une sonde beacon spécifique de *CCATI* lié à un PNA (Peptid Nucleic Acid) pour mettre en évidence le tissu tumoral par rapport au tissu sain à partir d'échantillons sur lame (Kam et al., 2013) (Figure 48). *HULC* (Hepatocellular carcinoma up-regulated long non-coding RNA) est un lncRNA compétiteur du miR-372, fortement exprimé dans les tumeurs primaires du foie et dans les métastases hépatiques du cancer colorectal,

mais n'est pas exprimé dans le cancer colorectal primaire ni dans les métastases non hépatiques (Matouk et al., 2009)(Panzitt et al., 2007). Une autre particularité de *HULC* est qu'il peut être facilement détecté dans le sang des patients présentant un carcinome hépatocellulaire par RT-qPCR. Mais l'exemple le plus pertinent à l'heure actuelle est le lncRNA *PCA3*. Ce lncRNA est exclusivement exprimé et associé au cancer de la prostate. Par ailleurs, *PCA3* est sécrété dans les urines des patients (de Kok et al., 2002). Ces caractéristiques particulièrement intéressantes ont mené au développement d'un test urinaire quantifiant le niveau d'expression de *PCA3* pour une utilisation clinique : Progensis™ *PCA3* urine Test (Durand et al., 2011). C'est un avantage non négligeable au diagnostic du cancer de la prostate, et qui permet d'éviter la répétition des biopsies.

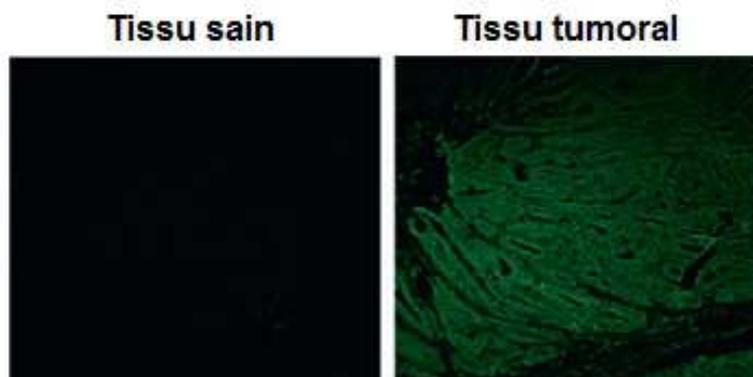


Figure 48 : Détection de *CCAT1* dans les tissus sains et tumoraux en paraffine à l'aide d'une sonde beacon. *CCAT1* n'est détecté que dans l'adénocarcinome colorectal (photo de droite) dans pratiquement toutes les cellules tumorales alors que l'on n'observe qu'un faible bruit de fond de fluorescence dans le tissu sain (photo de gauche) (Adapté de Kam et al., 2013).

Il est bien évidemment très prématuré de parler de ciblage des lncRNAs dans le cancer, même si des stratégies d'inhibition via l'utilisation de molécules inhibitrices d'interactions avec certaines protéines sont envisagées. Mais l'utilisation de l'expression tissu-spécifique des lncRNA a été mise à profit dans des traitements génétiques plus ciblés. En effet, l'expression du lncRNA *H19* est fortement augmentée dans une grande variété de cancers, avec de grandes différences d'expressions par rapport au tissu sain adjacent. Le plasmide BC-819 (DTA-H19) a alors été développé avec un gène codant pour la sous-unité A de la toxine diphtérique sous la régulation du promoteur de *H19* (Amit and Hochberg, 2010). Le plasmide peut être utilisé par injections intra-tumoral ou même en aérosol, induisant l'expression de la toxine diphtérique spécifiquement dans la tumeur. Il y a déjà eu quelques essais chez l'homme avec des résultats encourageants dans différents types de cancer comme le cancer du poumon ou de la vessie (Hasenpusch et al., 2011)(Sidi et al., 2008).

Nos connaissances concernant la fonction et l'impact de la dérégulation des lncRNAs se sont considérablement étendues ces dernières années. Cependant, comme l'a montré cet état de la

littérature, le nombre de lncRNAs caractérisés reste très faible par rapport au nombre total de transcrits annotés. De plus, un grand nombre de mécanismes d'actions restent incomplets mais leur intérêt en tant qu'acteur, biomarqueur, diagnostic ou pronostic est certain. Toutes les études d'analyses de profils d'expression de lncRNAs ont mis en évidence des lncRNAs dérégulés dans de nombreuses pathologies, notamment dans les cancers. Mais il n'y a pour l'heure que très peu de données dans les cancers hématologiques, et plus spécialement chez les patients atteints de leucémies aiguës myéloblastiques.

X) LncRNAs et hématopoïèse normale

L'hématopoïèse est le processus par lequel sont produits et renouvelés tous les éléments figurés du sang. Ce processus s'effectue au sein de la moelle osseuse par étapes successives à partir de la cellule souche hématopoïétique. En réponse à des signaux exogènes, ces cellules multi-potentes peuvent se différencier en érythrocytes (globules rouges), en cellules myéloïdes ou lymphoïdes (globules blancs) ou en thrombocytes (plaquettes). On distingue quatre compartiments le long de la différenciation hématopoïétique : les cellules souches hématopoïétiques, les progéniteurs s'engageant dans l'une ou l'autre lignée, les précurseurs qui se divisent et mûrissent, et les cellules matures différenciées et fonctionnelles. Ce processus nécessite un microenvironnement médullaire particulier avec l'intervention de facteurs de croissance pour s'engager dans les différents lignages comme l'IL-7 (Interleukine 7), l'EPO (Erythropoïétine), le TPO (Thrombopoïétine) et le GM-CSF. De plus, cette différenciation est sous la dépendance de facteurs de transcription spécifiques et nécessaire comme PAX5, PU-1 ou GATA-1.

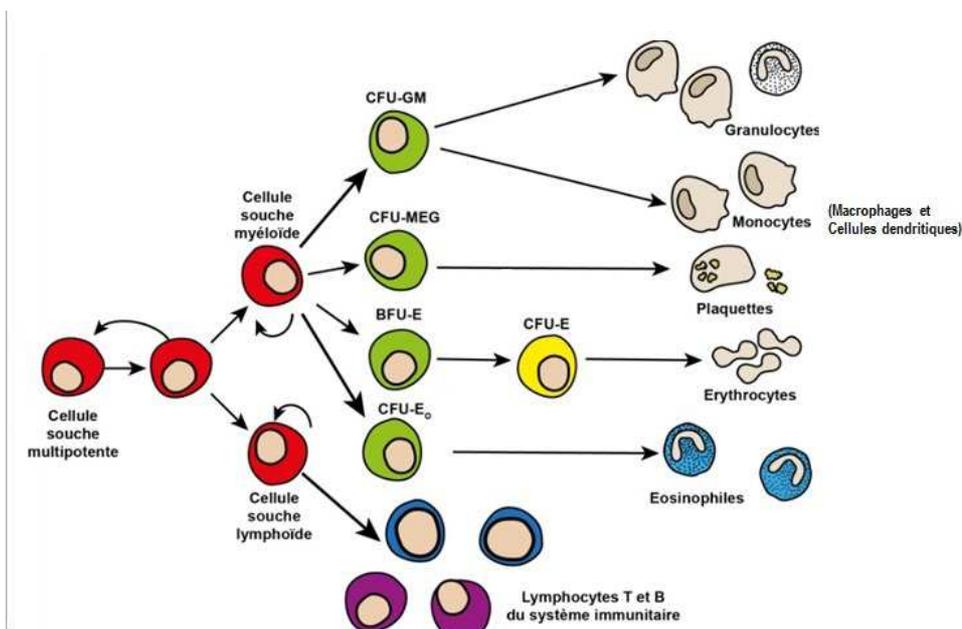


Figure 49 : Représentation schématique de l'hématopoïèse.

A ce jour, il y a très peu d'exemples de lncRNAs intervenant dans l'hématopoïèse (Tableau). Je vais donc détailler les exemples les plus pertinents dans les différents lignages hématopoïétiques.

Lignages	lncRNAs impliqués		
Cellules souches hématopoïétiques	<i>H19</i> (Venkatraman et al., 2013)	<i>lncHSC-1</i> et 2 (Luo et al., 2015)	
Granulocytaire	<i>EGO</i> (Wagner et al., 2007)	<i>HOXA-AS</i> (Zhao et al., 2013)	<i>HOTAIRMI</i> (Zhang et al., 2009)
Dendritique	<i>lnc-DC</i> (Wang et al., 2014)		
Erythrocyte	<i>EPS</i> (Hu et al., 2011)	<i>EC7</i> (Alvarez-Dominguez et al., 2014)	
Lymphoïde	<i>linc-MAF-4</i> (Ranzani et al., 2015)	<i>GATA3-AS1</i> (TH2) (Zhang et al., 2013)	<i>NEST</i> (TH1) (Collier et al., 2014)

Tableau 5 : Liste des lncRNAs impliqués dans les différents lignages hématopoïétiques.

Dans les LT-HSCs (Long Term Hematopoietic Stem Cell), *H19* est fortement exprimé puis son expression diminue dès le stade ST-HSCs (Short Term Hematopoietic Stem Cell). La perte de l'expression maternelle de *H19 in vivo* conduit à une diminution des LT-HSCs et une augmentation des ST-HSCs suggérant un impact de ce lncRNA dans le maintien de ces cellules hématopoïétiques. Les cellules souches hématopoïétiques déficientes en *H19* entrent plus facilement en cycle, provoquant une augmentation de la prolifération et une perte de l'auto-renouveaulement (Venkatraman et al., 2013). Ainsi *H19* semble nécessaire à la quiescence des cellules souches hématopoïétiques. Plus récemment, Luo *et al.* ont mis en évidence que les lncRNAs *lncHSC-1* et 2 étaient impliqués dans l'auto-renouveaulement et l'engagement dans les différents lignages hématopoïétiques par des mécanismes de régulation épigénétique, impliquant des facteurs clés de l'hématopoïèse tel que E2A (Luo et al., 2015b).

Il existe plusieurs exemples de lncRNA impliqués dans la différenciation érythroïde. En 2011, Hu *et al.* ont identifié chez la souris le lncRNA *EPS* (Erythroid ProSurvival lncRNA) régulant positivement la différenciation terminale érythroïde. Ce lncRNA n'est pas exprimé dans les autres lignages hématopoïétiques et sa surexpression promeut la survie des érythrocytes en inhibant l'expression de gènes pro-apoptotiques *Picard* et *Cideb* chez la souris (Hu et al., 2011). D'autres études ont identifié des lncRNAs spécifiques et potentiellement régulateurs du lignage érythroïde chez la souris. Certains sont par ailleurs régulés par les facteurs de transcription érythroïdes comme GATA1,

TAL1 (T-cell acute lymphocytic leukemia 1) et KLF1 (Kruppel-like factor 1). On peut citer le lncRNA *EC7* transcrit à partir de la région enhancer du gène *SLC4A1*, un transporteur membranaire des anions dans les globules rouges impliqués dans l'anémie hémolytique. La déplétion de *EC7* inhibe l'expression de *SLC4A1*, la maturation des érythrocytes, la diminution de la taille cellulaire et l'énucléation *ex vivo* (Alvarez-Dominguez et al., 2014). Une deuxième étude similaire a identifié des lncRNAs spécifiques des érythroblastes humains (dont 4 en communs avec l'étude précédente), des mégacaryocytes, et des précurseurs mégacaryocyte/érythroïdes murins. Quelques-uns de ces lncRNAs sont également régulés par les facteurs de transcription, GATA1 et TAL1, et nécessaires à la différenciation terminale des érythrocytes. Cependant, seulement 15% des lncRNAs identifiés dans les érythrocytes murins sont conservés et exprimés chez l'homme (Paralkar et al., 2014).

Les seuls exemples de lncRNAs impliqués dans la différenciation myéloïde granulocytaire sont *EGO*, *HOXA-AS2* et *HOTAIRM1*. *EGO* (Granule Ontogeny lncRNA) est un lncRNA anti-sens intronique du gène *ITPR1* (Inositol 1,4,5-TrisPhosphate Receptor type 1), qui est exprimé à partir des progéniteurs hématopoïétiques CD34⁺ jusqu'aux cellules éosinophiles matures. Ce lncRNA permet l'expression de MBP (Major Basic Protein) et d'EDN (Eosinophil Derived Neurotoxin), des protéines nécessaires au bon fonctionnement des cellules éosinophiles (Wagner et al., 2007). *HOXA-AS2* est un lncRNA surexprimé lors de la différenciation myéloïde de la lignée promyélocytaire NB4 induite par l'acide rétinoïque. Au cours de ce processus, il semble que ce lncRNA joue un rôle d'inhibiteur de l'apoptose (Zhao et al., 2013). *HOTAIRM1* (HOXA transcript antisens RNA, Myeloid specific 1) est également un lncRNA anti-sens du cluster *HOXA*. Ce lncRNA a été identifié comme étant surexprimé lors de la différenciation granulocytaire induite par l'acide rétinoïque de lignées cellulaires LAP et de progéniteurs myéloïdes. La délétion de *HOTAIRM1* dans la lignée LAP NB4 conduit à la sous expression de gènes *HOX* tels que *HOXA1* et *HOXA4* et de gènes intervenant dans la différenciation granulocytaire tels que *CD11b* et *CD18* (Zhang et al., 2009). De plus, la sous expression de *HOTAIRM1* provoque un retard de la différenciation induite par l'acide rétinoïque *in vitro*, et une diminution des gènes associés à l'activation granulocytaire, la défense immunitaire et la maturation (Zhang et al., 2014).

Des lncRNAs ont également été identifiés dans la différenciation des monocytes en cellules dendritiques (DC). Le *lnc-DC* est un lncRNA cytoplasmique régulé par PU.1, qui active le facteur de transcription nécessaire à la différenciation dendritique STAT3 (Signal transducer and activator of transcription 3). Dans ce processus, *lnc-DC* interagit directement avec STAT3 et empêche son inactivation par déphosphorylation médiée par SHP-1 (Src Homology region 2 domain-containing phosphatase-1). De plus, son inactivation conduit à un défaut de différenciation des cellules dendritiques *in vitro*, associé à la perte d'expression de facteurs nécessaires à leur fonction d'activateur des lymphocytes T comme les facteurs CD40, CD80, CD86, et HLA-DR pour la présentation de l'antigène (Wang et al., 2014b).

L'exemple le plus pertinent de lncRNA intervenant dans la lignée lymphocytaire est le lncRNA *linc-MAF-4* impliqué dans la balance entre différenciation lymphocytaire TH1/TH2 (T Helper). Ce lncRNA est spécifiquement exprimé dans les lymphocytes TH1 dans lesquelles il semble réguler négativement les facteurs de transcription lymphocyte TH1-spécifiques MAF par le biais du recrutement de LSD1 et PRC2. En effet, l'inhibition de *linc-MAF-4* au cours de la différenciation lymphocytaire conduit à une différenciation biaisée vers un phénotype TH2 (Ranzani et al., 2015).

De manière intéressante, très peu de ces lncRNAs décrits comme intervenant dans le processus normal d'hématopoïèse ont été associés aux hémopathies.

XI) LncRNAs et hémopathies malignes

Comme dans l'hématopoïèse normale, il y a peu d'exemples de lncRNAs impliqués dans les hémopathies malignes, je m'attacherai donc à décrire les exemples les plus pertinents.

Hémopathies	lncRNAs impliqués		
Leucémie lymphoïde chronique	DLEU2 (Garding et al., 2013)		
Leucémie lymphoïde aiguë T	LUNAR (Trimarchi et al., 2014)	T-ALL-R-LncR1 (Zhang et al., 2014)	
Leucémie lymphoïde aiguë B	BARL2 (Fernando et al., 2015)		
Lymphome	FAS-AS1 (Sehgal et al., 2014)	PEG10 (Peng et al., 2015)	GAS5 (Mourtada-Maarabouni and Williams, 2014)
Leucémie myéloïde chronique	H19 (Guo et al., 2014)	BGL3 (Guo et al., 2015)	
Myélodysplasie	XIST (Yildirim et al., 2013)		

Tableau 6 : Liste des lncRNAs impliqués dans les différentes hémopathies malignes.

Dans la leucémie aiguë lymphoblastique T (LAL-T), la plupart des événements oncogéniques conduisent à une sur-activation de la voie Notch. Le lncRNA *LUNARI* (Leukemia-induced Non-coding Activator RNA) est un lncRNA oncogénique qui promeut la croissance cellulaire dans cette pathologie, en augmentant l'expression d'*IGF1R* et de la voie de signalisation d'*IGF1 in vitro*. Dans ce contexte, *LUNARI* est activé par la voie Notch, et agit comme un enhancer ARN recrutant le complexe Mediator au promoteur d'*IGF1R*. De plus, les cellules LAL-T avec une faible expression de

LUNARI présentent une capacité de prolifération diminuée en xénogreffes (Trimarchi et al., 2014). Récemment, l'analyse d'expression des lncRNAs dans la leucémie aiguë lymphoblastique B (LAL-B) a mis en évidence l'impact de l'expression du lncRNA *BARL-2* (B-ALL-associated long RNAs-2) qui était corrélée à un mauvais pronostic. Ce lncRNA oncogénique augmente la croissance cellulaire, diminue l'apoptose et la réponse aux traitements utilisés dans ce type d'hémopathie (Fernando et al., 2015).

H19 impliqué dans l'hématopoïèse normale est aussi impliqué dans les hémopathies et plus particulièrement dans la leucémie myéloïde chronique (LMC) avec translocation *BCR-ABL*. Dans des lignées de LMC, *H19* est régulé positivement par *MYC* et la sous expression de *H19* diminue la prolifération et la formation de tumeurs en xénogreffe. Il a été suggéré que cette caractéristique oncogénique de *H19* était la conséquence de sa fonction en tant qu'ARN compétiteur de l'oncosuppresseur microARN let-7 (Guo et al., 2014). À l'inverse, le lncRNA *BGL3* (Beta Globin Locus 3) est un lncRNA suppresseur de tumeur dans la LMC avec translocation *BCR-ABL*. En effet, il a été montré qu'il régulait positivement l'expression du suppresseur de tumeur *PTEN* en agissant comme un ARN compétiteur des microARNs ciblant *PTEN*, miR-13 et miR-20a. De plus, l'expression de ce lncRNA dans une lignée cellulaire *BCR-ABL* positive conduit à une diminution de la capacité de survie et à une augmentation de l'apoptose induite par l'Imatimib, un inhibiteur de tyrosine kinase utilisé dans cette pathologie (Guo et al., 2015).

En 2013, Yildirim *et al.* ont montré l'impact de *Xist* dans le cancer à l'aide de modèles murins de délétion conditionnelle de *Xist* dans les cellules hématopoïétiques. Les auteurs ont pu observer chez ces animaux le développement d'organomégalies lymphoïdes massives, d'infiltrats lymphoïdes et de leucémies myéloïdes létales. De plus, la moelle osseuse de ces animaux présente des défauts comme une myélofibrose et une dysplasie multi-lignage, récapitulant la progression de syndromes myéloprolifératifs humains. Les cellules souches hématopoïétiques présentent également des marqueurs aberrants de maturation et d'auto-renouvellement, et leurs transplantations dans un environnement sauvage conduit au transfert de la maladie. D'un point de vue mécanistique, la déficience de *Xist* dans les cellules souches hématopoïétiques conduit à une réactivation du chromosome X inactif et à la dérégulation de facteurs impliqués dans les leucémies myéloïdes comme *GATA1*, *FLT3*, *MYB* et *CSF1* (Colony stimulating factor 1) (Yildirim et al., 2013).

XII) LncRNAs dans la leucémie aiguë myéloblastique

Les 1^{ers} lncRNAs qui ont été associés aux LAMs sont *H19* et *MEG3*. Une étude sur 32 cas de LAMs a rapporté une sous expression de *H19* et une autre étude a montré une hyperméthylation du promoteur de *MEG3* associées à un mauvais pronostic. Mais aucune étude n'est réellement allée plus loin quant à leurs implications dans cette pathologie (Tessema et al., 2005)(Benetatos et al., 2010). Par

ailleurs, il n'existe que très peu d'études établissant le lien entre lncRNAs et LAMs et celles-ci sont généralement centrées sur des entités très spécifiques.

Récemment, une nouvelle technique de R3C (RNA-guided Chromatin Conformation Capture) a été développée pour permettre d'identifier les lncRNAs interagissant avec une région chromosomique donnée. L'ARN est rétro-transcrit en ADNc biotinylé (en présence de dCTP-biotine) et ligué à l'ADN génomique adjacent après digestion enzymatique. Le complexe ADNc/ADN est purifié par l'utilisation de billes streptavidine et analysé par PCR (Zhang et al., 2014b). Par cette technique, deux lncRNAs ont été identifiés dans des loci impliqués dans les leucémies, *RUNXOR* et *IRAIN*.

Le lncRNA *RUNXOR* (*RUNX1* overlapping RNA) a été identifié comme étant transcrit à proximité du gène *RUNX1*, recouvrant le gène codant et pouvant interagir avec le promoteur de ce dernier. Ce lncRNA intervient dans la mise en place de boucles intrachromatiniennes entre l'enhancer et le promoteur de *RUNX1*. Les auteurs de ce travail ont observé que *RUNXOR* était surexprimé dans les lignées cellulaires LAM et chez les patients. De manière surprenante, il ne semble pas qu'il y ait une corrélation d'expression entre *RUNX1* et *RUNXOR*. Afin d'identifier les régions génomiques avec lesquelles le lncRNA pouvait interagir, Wang *et al.* ont développé la RAT (Reverse Transcription-associated Trap). Cette méthode permet d'identifier les régions génomiques cibles des lncRNAs par immunoprécipitation de l'ADNc biotinylé en complexe avec l'ADN génomique (Figure 50).

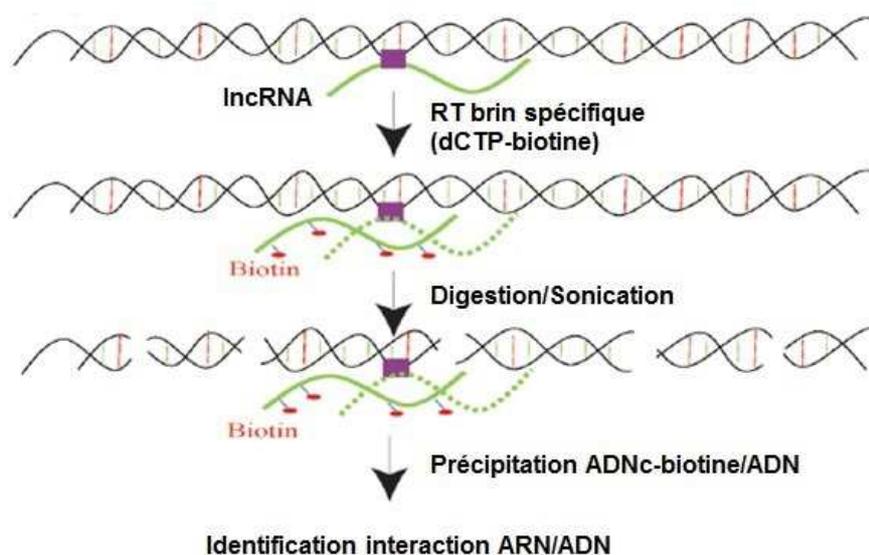


Figure 50 : Représentation schématique du protocole de RAT pour déterminer les interactions ARN/ADN. Le lncRNA ciblé est rétro-transcrit de façon spécifique en présence de dCTP biotinylés. Le complexe ADNc-biotine/ADN génomique est digéré et immunoprécipité à l'aide billes magnétiques streptavidine. La région génomique avec laquelle interagit le lncRNA est identifiée par PCR ou séquençage (Adapté de Sun et al., 2014a).

Par cette méthode, il a été mis en évidence que *RUNXOR* interagissait avec des régions fréquemment associées aux translocations avec *RUNX1*. C'est un travail particulièrement intéressant, puisque cela suggère que *RUNXOR* intervient dans la formation de boucles interchromatiniennes potentiellement impliquées dans la génération de translocations de *RUNX1* retrouvées dans les LAMs (Wang et al., 2014a) (Figure 51). Le lncRNA *IRAIN* a également été identifié par cette méthode de R3C. *IRAIN* est un lncRNA anti-sens exprimé à partir d'un promoteur intronique du locus *IGFRI*, un locus impliqué dans l'augmentation de la croissance cellulaire. Ce lncRNA est surexprimé dans des lignées cellulaires LAMs et chez les patients LAMs avec caryotypes complexes associés à un mauvais pronostic. Tout comme pour *RUNXOR*, *IRAIN* permet la formation de boucles chromatinienne entre le promoteur de *IGFRI* et un enhancer distant. Cependant, il n'y a pas de donnée quant à la corrélation d'expression entre *IRAIN* et *IGFRI* (Sun et al., 2014a).

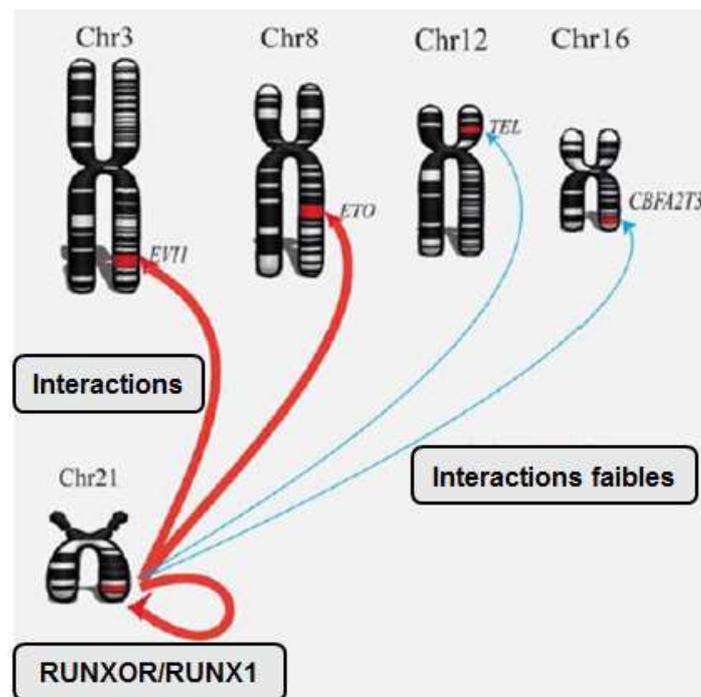


Figure 51 : Représentation schématique des interactions entre le lncRNA RUNXOR et différentes régions du génome fréquemment impliquées dans des translocations avec le gène RUNX1 (Adapté de Wang et al., 2014a).

Emmrich *et al.* ont étudié les deux lncRNAs hôtes des microARNs miR-99a ~ 125b-2 et miR-100 ~ 125b-1, *miR100HG* et *MONC* (ou *MIR99AHG*) (Emmrich et al., 2014). Ces microARNs ont clairement été associés à la promotion de la leucémie aiguë mégacaryoblastique. Dans cette hémopathie maligne, ces lncRNAs sont surexprimés et leurs expressions sont corrélées à l'expression de leurs microARNs introniques respectifs. De plus, *MONC* participe au blocage de la différenciation myéloïde indépendamment de sa fonction de producteur des microARNs. En effet, il a été montré qu'indépendamment de l'expression des microARNs, *MONC* perturbait l'hématopoïèse et augmentait

la prolifération des progéniteurs érythroïdes, lorsqu'il était surexprimé dans les cellules souches et progénitrices de sang de cordon (Emmrich et al., 2014).

Très récemment, un lncRNA a été associé aux LAMs avec mutation du gène *CEBPα*. En effet, Hughes *et al.* ont montré que la forme sauvage de *CEBPα* se fixait au promoteur du lncRNA oncogénique *UCA1* (Urothelial Carcinoma Associated 1) et réprimait son expression. A l'inverse, la forme mutée de *CEBPα* permet l'expression du lncRNA. Dans les LAMs avec la mutation de ce gène, on observe effectivement une surexpression de *UCA1* et il semblerait que ce dernier augmente la prolifération des blastes myéloïdes en réprimant l'expression du régulateur du cycle cellulaire p27kip1 (Hughes et al., 2015).

Les lncRNAs peuvent également être dérégulés par le biais de translocations chromosomiques. Récemment, des fusions du gène codant *NSMCE2* (Non-SMC element 2) ont été identifiées avec deux lncRNAs, *PVT1-NSMCE2* chez un patient atteint de LAM et *BF104016-NSMCE2* dans la lignée LAM HL-60 (Chinen et al., 2014). *PVT1* est un lncRNA oncogénique, producteur de microARNs, régulé par MYC et p53 et inhibant l'apoptose (Carramusa et al., 2007). *BF104016* est un lncRNA composé de l'exon 4 de *CCDC26*, un lncRNA impliqué dans l'inhibition de l'apoptose et l'augmentation de la survie. Le gène codant *NSMCE2* est quant à lui impliqué dans l'instabilité chromosomique et l'augmentation de la fréquence des cassures de l'ADN chez la levure (Rai et al., 2011). Il n'y a cependant pas plus de données quant à l'implication de ces translocations dans la LAM.

Dans le paragraphe consacré à *NPM1* dans les LAMs, j'ai cité des travaux ayant montré un profil d'expression de lncRNAs particulier dans les LAMs présentant la mutation *NPM1*. En effet, dans ce travail publié fin 2014, Garzon *et al.* ont étudié le profil d'expression des lncRNAs dans les LAMs à caryotype normal à l'aide de microarrays. Les auteurs ont pu mettre en évidence différentes signatures associées aux mutations dans les gènes *NPM1*, *CEBPα* (dont le lncRNA *UCA1*), *IDH2*, *ASXL1*, *RUNX1* et *FLT3-ITD*. De plus, cette étude a pu corrélérer l'expression de lncRNA à la survie et à la réponse aux traitements (Garzon et al., 2014). Cependant, l'analyse a été réalisée microarrays, il n'y a donc pas d'identification de nouveaux transcrits. De plus, il n'y a pas de caractérisation fonctionnelle de lncRNAs dans ces LAMs particulières.

Comme on peut l'observer dans cet état de la littérature, les différents travaux impliquant les lncRNAs dans la LAM représentent des exemples ponctuels, associés à des LAMs spécifiques (LAPs, leucémie aiguës mégacaryoblastiques, LAMs *CEBPα* muté), ou associés à des translocations chromosomiques. Il n'y a donc pas d'analyse à grande échelle de l'expression des lncRNAs dans les LAMs, et plus spécifiquement dans les LAMs à caryotype normal.

Objectifs de la thèse

Les données de la littérature suggèrent depuis plusieurs années que les lncRNAs sont impliqués dans le processus de cancérogenèse par divers mécanismes moléculaires impliquant pratiquement tous les niveaux de régulation moléculaire. Au début de ma thèse, il n'y avait que très peu de données impliquant les lncRNAs dans les hémopathies malignes et moins encore dans la leucémie aiguë myéloblastique. La mise en évidence de lncRNAs dérégulés dans ces pathologies pourrait nous permettre de mieux comprendre leurs fonctions d'un point de vue fondamental, mais également leurs rôles dans les processus oncogéniques. Les objectifs de ma thèse ont été :

- Déterminer le profil d'expression des lncRNAs dans les LAMs présentant un caryotype normal et corréler les données transcriptomiques aux données clinico-biologiques de ces patients. Les LAMs résultent de l'acquisition de mutations dans les cellules souches ou dans les progéniteurs hématopoïétiques. Elles peuvent être divisées selon des critères cytogénétiques. On y retrouve des patients avec des anomalies chromosomiques récurrentes, des caryotypes complexes, des translocations chromosomiques rares, mais la moitié des cas est représentée par des caryotypes normaux pour lesquels il existe peu de facteurs pronostiques, car les événements oncogéniques sont inconnus. Nous avons donc choisi de nous concentrer sur ces patients à caryotypes normaux. La cohorte de patients sélectionnés pour cette étude comprend 21 cas de LAMs à caryotype normal avec des mutations de la tyrosine kinase *FLT3* et/ou de la nucléophosmine *NPM1* (mutations récurrentes dans les LAMs), et 19 cas de LAMs à caryotype normal sans mutation *FLT3* et *NPM1*. L'analyse du profil d'expression des lncRNAs a été réalisée par RNA-sequencing (Illumina) en collaboration avec le Beijing Genomics Institute (BGI, Hong Kong).

- Déterminer l'impact fonctionnel de la dérégulation de lncRNAs d'intérêt dans la LAM et les mécanismes moléculaires dans lesquels ceux-ci sont impliqués. Leur dérégulation pourrait alors avoir un impact dans la transformation tumorale, c'est pourquoi nous avons testé les conséquences de leur inactivation et/ou surexpression dans nos différentes lignées cellulaires. Afin de pouvoir étudier la fonction de ces lncRNAs, nous les avons caractérisés au préalable pour mettre en place des modèles d'étude. En effet, des approches de biologie moléculaire (RACE), cellulaire (fractionnement cellulaire) et bioinformatique, nous ont permis de déterminer la séquence exacte des transcrits, s'ils sont cappés, polyadénylés, leur localisation (nucléaire ou cytoplasmique) et leur conservation entre espèces. Nous avons utilisé différentes lignées cellulaires myéloïdes humaines : la lignée NB4 promyélocytaire et *NPM1* non muté, la lignée OCI AML3 avec la mutation *NPM1* et une lignée HL60ER *NPM1* non muté, présentant un récepteur écotropique murin utilisable pour la transduction par des rétrovirus murins.

Ce projet s'inscrit dans l'idée récente que ces ARNs ont une véritable fonction physiologique et que leurs dérégulations ont des conséquences pathologiques. Cette étude, ciblant les lncRNAs présentant une expression inadaptée dans le contexte leucémique, pourrait nous permettre

d'approfondir nos connaissances fondamentales sur la fonction des lncRNAs, et aussi d'étudier leurs rôles dans les processus oncogéniques. A travers l'étude des lncRNAs dans les hémopathies malignes, nous visons une plus grande connaissance du monde méconnu des ARNs non codants et une meilleure compréhension de ce nouveau niveau de complexité dans les régulations géniques. Ces lncRNAs pourraient rapidement constituer des outils moléculaires de choix pour de futures thérapies, pour de nouveaux tests diagnostiques, et ainsi aider au pronostic.

*Article en cours de
rédaction*

Résumé :

Au cours de ce travail, nous avons évalué les profils d'expression des lncRNAs par RNA-sequencing sur 40 patients atteints de leucémies aiguës myéloblastiques avec un caryotype normal (LAM-CN). Cette approche a mené à la découverte de plus de 8000 lncRNAs non identifiés à ce jour. Par analyse non supervisée, nous avons observé une signature spécifique de lncRNAs dépendante du statut mutationnel de *NPM1*. L'utilisation de la méthode mathématique Sparse Partial Least Squares Discriminant Analysis (Sparse PLS-DA) nous a permis d'identifier un set minimal de 19 lncRNAs capable de discriminer les patients *NPM1* muté des patients *NPM1* non muté. Ces lncRNAs pourraient être de bons candidats d'études fonctionnelles car ils pourraient être impliqués dans le développement de la leucémie et plus particulièrement dans les LAMs *NPM1* muté.

En parallèle, l'analyse de l'association entre l'expression des lncRNAs précédemment sélectionnés et les paramètres clinico-biologiques des patients, nous a permis d'identifier une combinaison de deux lncRNAs, XLOC_010909 et XLOC_028779 qui présentent des profils d'expression associés au pronostic clinique des patients LAM-CN.

Ces résultats ont permis d'une part, d'identifier des lncRNAs candidats potentiellement impliqués dans le processus de leucémogénèse des LAM-CN *NPM1* muté, et d'autre part, cette étude suggère que les lncRNAs pourraient être également de bons biomarqueurs pronostiques dans cette pathologie.

Long non-coding RNA expression in cytogenetically normal AML identifies a distinct signature associated with NPM1 mutations

Etienne De Clara¹, Ma Hanjing², François Vergez¹⁻³, Cathy Quelen¹, Sébastien Dejean⁴, Cécile Demur¹⁻³, Eric Delabesse¹⁻³, Christian Recher¹⁻³, Christian Touriol¹, Maria Paola Martelli⁵, Brunangelo Falini⁵, Pierre Brousset¹ and Marina Bousquet^{1,6}

¹Centre de Recherches en Cancérologie de Toulouse (CRCT), UMR1037 Inserm/Université Toulouse III Paul Sabatier, ERL5294 CNRS, Oncopole de Toulouse, 2 avenue Hubert Curien CS 53717, 31037 TOULOUSE CEDEX 1, France

²BGI, 11F, Main Building, Beishan Industrial Zone, Yantian District, Shenzhen, 518083, China

³Laboratoire et service d'Hématologie, Centre Hospitalier Universitaire de Toulouse, Institut Universitaire du Cancer, 31059 Toulouse, France.

⁴Institut de Mathématiques de Toulouse, UMR 5219 Université de Toulouse / CNRS Université Paul Sabatier, 118 route de Narbonne 31062 Toulouse Cedex 9

⁵Institute of Hematology, University of Perugia, Ospedale S. Maria della Misericordia, Perugia, Italy

⁶Corresponding Author: Marina Bousquet

Centre de Recherches en Cancérologie de Toulouse (CRCT)
UMR1037 Inserm/Université Toulouse III Paul Sabatier
ERL5294 CNRS
Oncopole de Toulouse
2 avenue Hubert Curien
CS 53717
31037 TOULOUSE CEDEX 1 – France

Phone: +33582741671

Fax: +33582741685

E-mail: bousquetmarina@gmail.com

Conflict-of-interest disclosure: The authors declare no competing financial interests.

The analysis of 200 acute myeloid leukemia (AML) genomes in the context of the Atlas Genome Project highlighted the genetic and epigenetic architecture of cytogenetically normal AML (CN-AML)(1), provided new important information on genetic alterations involved in AML development and also identified potential tools for risk stratification of AML and new diagnostic and therapeutic targets. However, these studies mainly focused on genetic alterations that either impact on protein coding genes or microRNAs. The discovery of lncRNAs provides a unique opportunity to identify new biomarkers and key players in leukemogenesis. Accumulating evidence points towards an important role for lncRNAs in cancer initiation, development, and progression (2). A recent publication showed by a custom microarray strategy that lncRNAs expression is associated with common mutations in AML and that some lncRNAs correlate with survival (3). Despite this, our overall knowledge of lncRNAs in cancer, including leukemia, remains very limited (4-6).

In this study we sought to determine lncRNA expression profiles in CN-AML. RNA sequencing of forty CN-AML patients (Table S1, Cohort 1) allowed us to identify 11036 expressed lncRNAs. Among them, we identified 8526 new lncRNAs, not previously reported in the RefSeq, UCSC and ENCODE databases.

We performed an unsupervised analysis with the 11036 CN-AML lncRNAs using hierarchical clustering. As shown in Figure 1A (and Figure S1A), clustering segregated the samples into two main groups. Different criteria were evaluated to define the two groups, namely FAB status, age and mutational status (presence/absence of NPM1 and/or FLT3-ITD mutations). The main observation was that NPM1-mutated patients were largely enriched in the first group (75%) compared to the second group (8%) ($p=6 \times 10^{-6}$). This result suggests that there is a lncRNA expression profile that is specific to NPM1-mutated patients.

To identify the differentially-expressed lncRNAs associated with NPM1 mutations in the 40 patients with CN-AML, we compared NPM1-mutated AML patients (n=14) with NPM1-wild type AML samples (n=26) using the EdgeR package. We focused our study on highly-expressed lncRNAs by selecting lncRNAs that had at least 10 cpm (counts per million) in at least 14 of the 40 patients. Among the 5333 lncRNAs selected, 236 lncRNAs were significantly differentially-expressed ($FDR < 10^{-3}$) between the two groups. Of the 236 differentially-expressed lncRNAs, 107 lncRNAs had a fold change higher than 4, 26 of which were up-regulated in NPM1-mutated patients and 81 were down-regulated.

To validate our strategy we applied the same approach to mRNAs and compared NPM1-mutated with NPM1-wild type AML patients. Among the 71 most differentially-expressed mRNAs, the HOX genes were up-regulated and the CD34, BALCC and MN1 genes were down-regulated. These have already been described in several previous studies and validate our lncRNA findings (Table S2) (7, 8).

Among the 107 lncRNAs that were found to be differentially-expressed in NPM1-mutated patients, 74 are located within introns or exons of protein-coding genes and 33 are intergenic or anti-sense lncRNAs. To avoid quantification of pre-mRNA by RT-qPCR, we focused our attention on the 33 intergenic and anti-sense lncRNAs (Table S3). Hierarchical clustering using these 33 lncRNAs clearly separated NPM1-mutated from NPM1-wild type AML patients, as shown in Figure S1B ($p = 1.2 \times 10^{-9}$). RNAseq data were validated by RT-qPCR (Fluidigm). Quantification was done only on 32 of the 33 lncRNAs previously selected due to technical reasons. We observed a large positive correlation between the two techniques (Table S4). Hierarchical clustering with Fluidigm data confirmed the separation of the NPM1-mutated and NPM1-wild type groups ($p = 1.6 \times 10^{-8}$) (Figure S1C). In order to validate this lncRNA signature, a validation set (n=50) composed of 33 new NPM1-mutated AML patients and 17 NPM1-wild type was used (Table S1, Cohort 2). Among the 50 patients, we included 8

patients with no recurrent cytogenetic abnormalities that had been classified within the intermediate risk group to increase the cohort (Table S5). As shown in Figure S1D, hierarchical clustering allowed us again to discriminate between NPM1-mutated and -wild type patients ($p=0.002$).

In order to reduce the number of discriminating lncRNAs we used the Sparse partial least squares discriminant analysis (Sparse PLS-DA) approach on the 33 lncRNAs identified from the RNA sequencing data, which allowed us to identify a minimal set of 19 lncRNAs that are able to discriminate NPM1-mutated patients from NPM1-wild type patients in the first ($p=5.2 \times 10^{-9}$) and validation cohort ($p=8.0 \times 10^{-5}$) (leave one out error rate 5% and 10% respectively) (Figures 1B, 1C, S2 and S3). Of note, this lncRNA signature was also observed in the validation set using only CN-AML patients ($n=42$, $p=1.9 \times 10^{-4}$; Figure S4).

To determine the potential function of the 33 lncRNAs associated with the NPM1-mutated genotype we analyzed for each lncRNA its correlation with mRNA expression using the Pearson Correlation (Table S6). We selected mRNAs with positive ($r>0.5$; $p<0.001$) and negative correlation ($r<-0.5$; $p<0.001$) and performed Gene Ontology (GO) analysis. Among the 33 lncRNAs, 14 (42%) were not significantly associated with any GO Terms. Four lncRNAs (12%), XLOC_010909, XLOC_034440, XLOC_034443 and XLOC_087120, were positively correlated with mRNAs belonging to the HOX family (despite their different localization) (Table S6). This finding is consistent with previous gene expression studies highlighting overexpression of HOX genes in NPM1-mutated patients (7, 8). When taken together with other features of NPM1 mutations, such as their mutual exclusivity with other AML recurrent cytogenetic abnormalities, high specificity for AML, stability at relapse and association with unique gene expression and microRNA profiles (9), our finding that NPM1-mutated AML displays a distinct lncRNA signature further supports the view that this leukemia represents a distinct disease entity. This has obvious biological and clinical impact

and also implications for the up-coming update of the WHO classification. Moreover, 40% of these lncRNAs were found to correlate with mRNAs involved in pathways including those of ribonucleoproteins, cell cycle, centrosome duplication, genomic stability and DNA repair, nuclear pore formation and export and protein chaperoning (Figure S5 and Table S6). These pathways are known to be highly associated with NPM1 functions (10).

We noticed that some NPM1 mutated patients did not cluster with the majority of NPM1 mutated patients and we wondered to decipher if some clinical parameters could explain this phenomenon so we next asked whether these differentially-expressed lncRNAs have any impact on prognosis. To answer this question we pooled the first and second cohorts and selected a homogeneous cohort of 58 patients under 70 years old that had achieved complete remission (CR) after treatment. Hierarchical clustering using the 32 lncRNAs defined two groups: group 1 (n=36) and group 2 (n=22) (Figure S6A). Kaplan Meier curves of overall survival (OS) and disease free survival (DFS) with these two groups showed significant prognostic differences (Figure S6B) as patients in group 2 had significantly shorter OS and DFS compared to patients in group 1. From a multivariate analysis, the signature of these 32 lncRNAs (Group 1 vs Group 2) was the only independent prognostic marker for OS and DFS when considered alongside covariates such as age, NPM1 and FLT3-ITD mutational status, WBC count and karyotype (Table S7). This result can be explained by the relative small number of patients in our cohort or by the differences of lncRNA expression among the NPM1 mutated patients.

The Sparse PLS-DA approach on the 32 lncRNAs allowed us to identify the two most discriminative lncRNAs: XLOC_010909 and XLOC_028779 (Figure S6C and D). Hierarchical clustering using these two lncRNAs resulted in 6 patients (5/6 are NPM1 mutated patients) shifting from group 1 (n=30) to group 2 (n=28) (Figure 2A). Kaplan Meier curves of these two new groups showed a higher prognostic significance using only these two

markers (Figure 2B and S7). Using a multivariate analysis, the XLOC_010909 and XLOC_028779 lncRNA signature was again the only independent prognostic marker for OS and DFS (Table S7). XLOC_010909 and XLOC_028779 expression signature therefore show potential as good prognosis marker for post-CR CN-AML patients as their combined overexpression seems to be associated with a better prognosis. Further studies in larger cohort will help to determine if lncRNA signature could be used to better sub-classify NPM1mutated +/- FLT3-ITD AML.

A major finding of this study is the discovery that NPM1-mutated AML patients show a distinct lncRNA expression profile. This finding was reinforced by the recent publication of Garzon *et al.* identifying by microarray 180 known lncRNAs associated with NPM1 mutational status (3). Our study enhanced our overall knowledge of lncRNAs profiles in AML as we used a RNA-seq approach, leading to the identification of more than 8000 new lncRNAs. Of note, among the 33 identified lncRNAs associated with NPM1, 31 are new lncRNAs. Furthermore we also reduced this signature to a minimal set of 19 lncRNAs able to discriminate NPM1 mutated and NPM1 wild type patients. We also provide evidence that lncRNA signatures may play a role in risk stratification of patients with CN-AML in accordance with Garzon *et al.* showing a 48 lncRNAs signature associated with older patients outcome in CN-AML (3). Altogether, these data suggest that lncRNAs should be considered as key players in the pathogenesis of acute myeloid leukemias. The functional role of these lncRNAs is of great interest and will have to be determined in the future.

ACKNOWLEDGEMENTS

M.B. was supported by a fellowship from ARC (Association pour la Recherche sur le Cancer). E.D.C. was supported by a fellowship from LABEX. This work was also supported by ARC grants, the Institut Universitaire de France (P.B.) and by the Associazione Italiana Ricerca Cancro (AIRC, IG 2013 n.14595). English proofreading was performed by Scientific Scripts (<http://scientificscripts.com>).

Conflict-of-interest disclosure: The authors declare no competing financial interests.

AUTHORSHIP CONTRIBUTIONS

M.H. performed RNA sequencing and the identification of new lncRNAs. M.B., E.D.C., C.Q. and C.T. performed in vitro experiments. F.V., C.D., E.D., C.R., M.P.M. and B.F. provided patient samples and clinical information. E.D.C and M.B analyzed the RNA sequencing data using the EdgeR package. E.D.C., S.D., F.V. and M.B. performed statistical analyses. E.D.C., B.F., P.B and M.B. wrote the manuscript. P.B. and M.B designed the study.

REFERENCES

- 1 Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med* 2013; **368**: 2059–74.
- 2 Yang G, Lu X, Yuan L. LncRNA: A link between RNA and cancer. *Biochim Biophys Acta* 2014; **1839**: 1097–1109.
- 3 Garzon R, Volinia S, Papaioannou D, Nicolet D, Kohlschmidt J, Yan PS *et al.* Expression and prognostic impact of lncRNAs in acute myeloid leukemia. *Proc Natl Acad Sci U S A* 2014; **111**: 18679–84.
- 4 Trimarchi T, Bilal E, Ntziachristos P, Fabbri G, Dalla-Favera R, Tsirigos A *et al.* Genome-wide Mapping and Characterization of Notch-Regulated Long Noncoding RNAs in Acute Leukemia. *Cell* 2014; **158**: 593–606.
- 5 Emmrich S, Streltsov A, Schmidt F, Thangapandi VR, Reinhardt D, Klusmann J-H. LincRNAs MONC and MIR100HG act as oncogenes in acute megakaryoblastic leukemia. *Mol Cancer* 2014; **13**: 171.
- 6 Alvarez-Dominguez JR, Hu W, Gromatzky AA, Lodish HF. Long noncoding RNAs during normal and malignant hematopoiesis. *Int J Hematol* 2014; **99**: 531–41.

- 7 Alcalay M, Tiacci E, Bergomas R, Bigerna B, Venturini E, Minardi SP *et al.* Acute myeloid leukemia bearing cytoplasmic nucleophosmin (NPMc+ AML) shows a distinct gene expression profile characterized by up-regulation of genes involved in stem-cell maintenance. *Blood* 2005; **106**: 899–902.
- 8 Becker H, Marcucci G, Maharry K, Radmacher MD, Mrózek K, Margeson D *et al.* Favorable prognostic impact of NPM1 mutations in older patients with cytogenetically normal de novo acute myeloid leukemia and associated gene- and microRNA-expression signatures: a Cancer and Leukemia Group B study. *J Clin Oncol* 2010; **28**: 596–604.
- 9 Falini B, Martelli MP, Bolli N, Sportoletti P, Liso A, Tiacci E *et al.* Acute myeloid leukemia with mutated nucleophosmin (NPM1): is it a distinct entity? *Blood* 2011; **117**: 1109–20.
- 10 Grisendi S, Mecucci C, Falini B, Pandolfi PP. Nucleophosmin and cancer. *Nat Rev Cancer* 2006; **6**: 493–505.

Figure 1: **Specific LncRNA expression profile in *NPM1*-mutated AML patients with normal cytogenetics.** **A** Unsupervised hierarchical clustering analysis of 40 patients with CN-AML (Cohort 1) using 11 065 lncRNAs (RNA-seq data). **B** Hierarchical clustering of 40 CN-AML (Cohort 1) and associated heatmap of the 19 lncRNAs identified by Sparse PLS-DA comparing *NPM1*-mutated patients (n=14) with *NPM1*-wild type patients (n=26) (RNA-seq data). Heatmap depicts high expression (red, +1) and low expression (green, -1). **C** Hierarchical clustering of 50 AML patients with intermediate risk (Cohort 2) and associated heatmap of the 19 lncRNAs identified by Sparse-PLS-DA (Fluidigm data).

Figure 2: **Combined expression of XLOC_010909 and XLOC_28779 lncRNAs as a prognosis marker.** **A** Hierarchical clustering and associated heatmap of XLOC_010909 and XLOC_28779 lncRNAs identified by Sparse PLS-DA. **B** Overall Survival (OS) and Disease Free Survival (DFS) Kaplan Meier curves comparing groups 1 and 2. LogRank test for OS: $p= 0.0332$, hazard ratio (HR): 2.510, 95% CI: 1.076 to 5.856 and for DFS $p=0.0080$, HR: 2.972, 95% CI: 1.329 to 6.646.

Figure 1

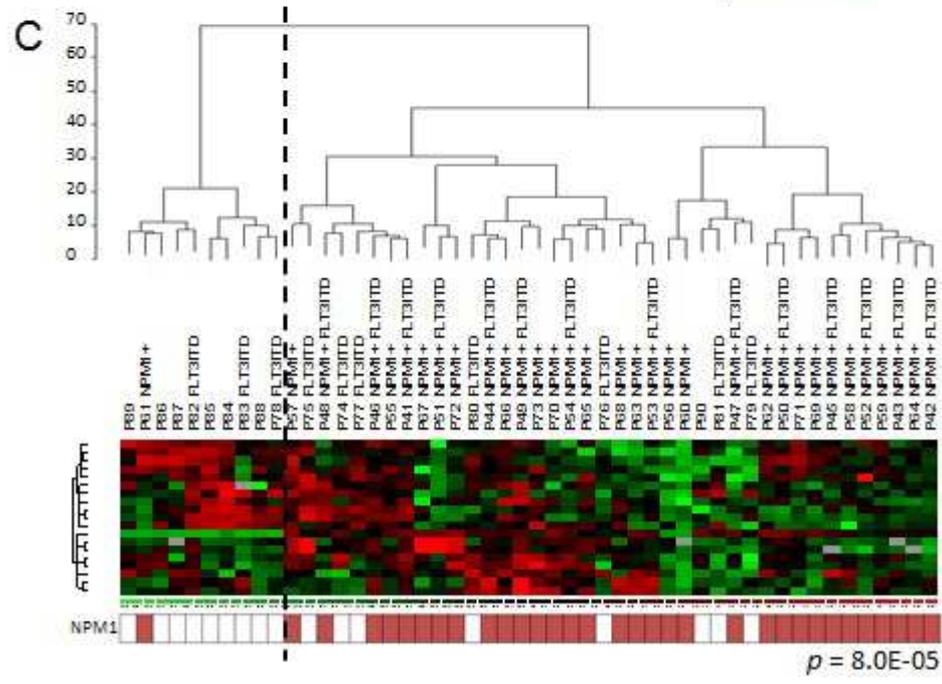
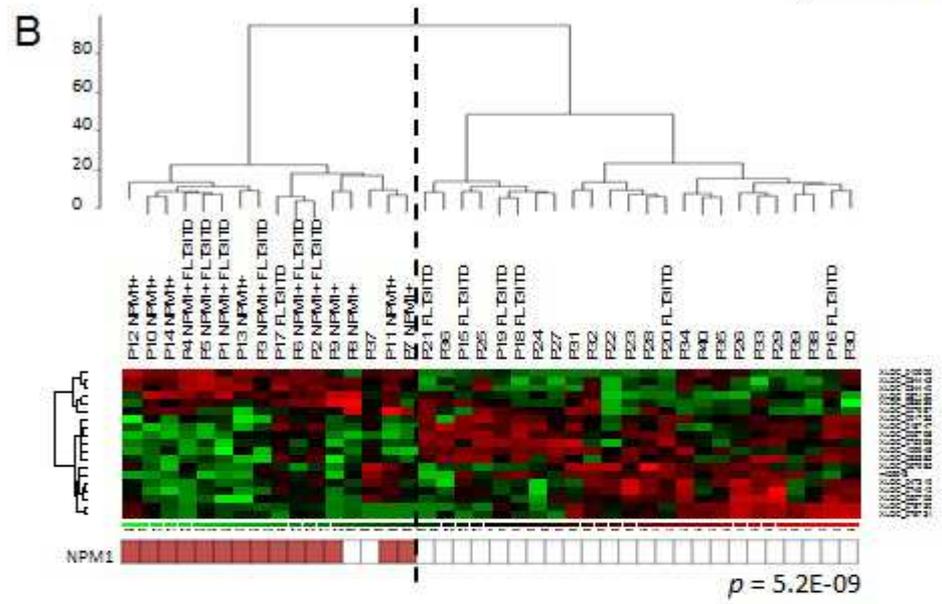
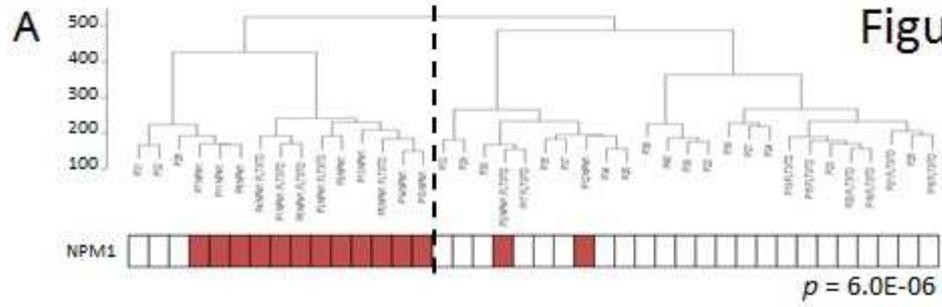
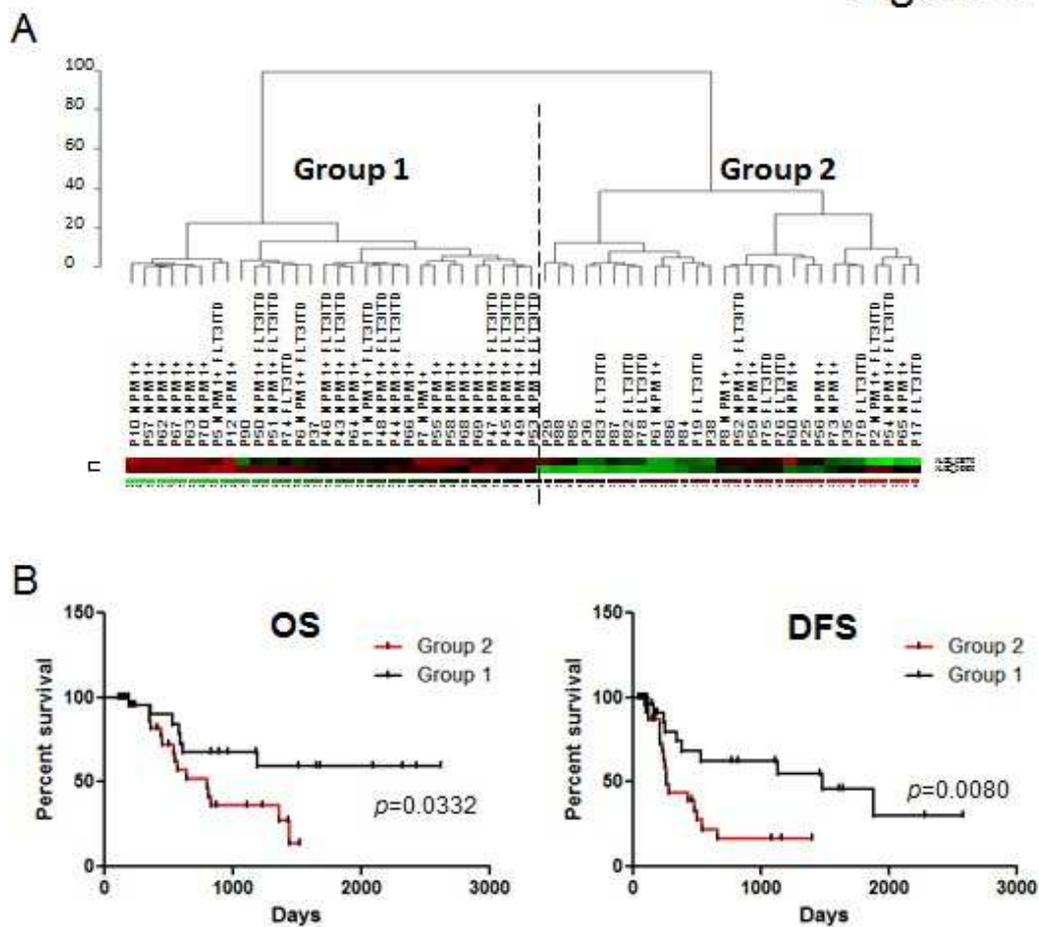


Figure 2



Supporting Information

SUPPLEMENTAL METHODS

AML samples

Forty AML samples were collected from patients registered at the HIMIP (Hémopathies INSERM Midi-Pyrénées) collection for the RNA-sequencing experiment (5µg of RNA). For the validation cohort, forty-two AML samples were obtained from the HIMIP collection and 8 AML samples from the Hematology Institute of Perugia University (1µg of RNA). According to French law, the HIMIP collection has been declared to the Ministry of Higher Education and Research (DC 2008-307 collection 1) and a transfer agreement has been obtained (AC 2008-129) after approbation by ethical committees (Comité de Protection des Personnes Sud-Ouest et Outremer II and APHP ethical committee). Clinical and biological annotations have been declared to the CNIL (Comité National Informatique et Libertés). Fresh and thawed samples were obtained after informed consent and the study was conducted in accordance with the Declaration of Helsinki and approved by local ethic committee. The patients received intensive Cytarabine/Daunorubicin-based therapy.

For prognosis analysis, we pooled the two cohorts (n=90) and we discarded patients that did not receive intensive chemotherapy, older than 70 years old, with no OS (overall survival) and DFS (Disease free survival) information and patients who died within 1 month after treatment.

RNA isolation and cDNA synthesis

Total RNAs were extracted with TRIzol reagent (Ambion, Austin, TX, USA). RNA integrity was evaluated using the RNA 6000 Nano Chip kit (Agilent Technologies, Massy, France). Only RNA extracts with RNA integrity number values ≥ 7 underwent further reverse transcription. All samples were reverse transcribed using the Superscript II reverse transcription kit (Invitrogen, Beijing, China) according to the manufacturer's protocol.

RNA-Sequencing Library Preparation, Read Generation and Mapping

RNA-sequencing was performed at the BGI (Hong Kong). rRNA depletion was performed from total RNA with Ribo-Zero™ rRNA Removal Kits (Epicentre, Madison, WI, USA). Paired-end, strand-specific reads of 91 nt were generated on an Illumina HiSeq™2000. Alignment and mapping were performed using Tophat (1) against the hg19 genome and the

mapped reads were assembled by Cufflinks 2.0.2 (2). The Cuffcompare program was used to merge the RefSeq, ENCODE and UCSC human known genes into one gene annotation set for comparison with the assembled transcripts (3). For novel lncRNA prediction, transcripts with a long exonic length (>200 nt) were extracted, according to the definition of lncRNA. Multi-exon (≥ 2 exons) transcripts and single-exon transcripts which recurred in multiple samples (≥ 3 samples) were then extracted. Assemblies that had long (≥ 300 nt) putative open reading frames (ORFs) were excluded. In order to filter out the known lncRNAs, the filtered assembly transcripts were BLASTed to find known lncRNAs from the ENCODE, Refseq and UCSC databases, according to the threshold: identity >0.9 and coverage >0.8. Transcripts which did not match known lncRNAs were then used for the protein annotation. The transcripts were aligned to four protein databases: kegg, nr, cog and swissprot, according to the threshold: identity >0.9 and coverage >0.8. The coding potential of extracted transcripts which did not align with protein databases was assessed using CPC based on six biologically meaningful sequence features such as ORF score, coverage, integrity, number of hits, hit score and frame score (4). Known lncRNAs and predicted novel lncRNAs were then combined as the reference, and reads were aligned to this reference by SOAP2 (5). Among the 35046 putative lncRNA transcripts expressed in our samples, we selected only lncRNAs that were at least 90% identical to the reference transcript in at least two samples. Gene expression levels were calculated using the RPKM (Reads per Kb per Million) method (6). The raw and processed RNA-sequencing data have been deposited at the National Center for Biotechnology Information Gene Expression Omnibus (repository number GSE62852).

Fluidigm 96.96 Dynamic ArrayIFCs (Integrated Fluidic Circuits)

lncRNA expression analysis was performed using the BioMark 96×96 gene expression platform (Fluidigm) according to the manufacturer's instructions. Primer pairs were designed using the Primer3 - PCR primer design tool (Steve Rozen, Helen J. Skaletsky (1998) Primer3), using an amplicon length of 80–120 bp. Table S1 shows the list of primers used in the experiment. RNaseP, 5s rRNA and MLN51 genes were used as reference genes for RT-qPCR data normalization. The Biomark System was used to run the chips and data were collected with Fluidigm Real-Time PCR Analysis Software. Primers having more than one peak were eliminated from further analysis. Linear derivative and automatic detector methods were set for baseline correction and Ct calculation. The $-\Delta Ct$ (threshold cycle) values were used for the clustering analysis and heatmaps generation.

Statistical analysis

All unsupervised clustering and heatmap generation was performed using the «hclust» function in R and the Ward's agglomeration method. Differential expression analysis was performed with the EdgeR package on R and cutoffs were established at a fold change ≥ 4 and False Discovery Rate (FDR) <0.001 (7). The principal component analysis (PCA) and the sparse partial least squares discriminant analysis (Sparse PLS-DA) methods were used using the R mixOmics package. The “*perf*” method, from the “*Mixomics*” package, was used to estimate the error rate of the model using leave-one-out cross-validation. **Bootstrapping cluster analysis** was carried out using the R ClassDiscovery package (1000-fold re-sampling). Expression correlation between lncRNA and mRNA was assessed using the Pearson correlation. mRNAs with a Pearson correlation coefficient of r over 0.5 or under -0.5 ($p<0.001$) were selected and analyzed using the Database for Annotation, Visualization and Integrated Discovery (DAVID) (DAVID;<http://www.david.abcc.ncifcrf.gov>). The significance of related GO cellular components and molecular pathways were identified based on FDR <0.05 . The Fisher test and the unpaired t test were used to assess the statistical significance of the differences between groups. Overall survival (OS) time was calculated from the date of diagnosis until death. Disease-Free-Survival (DFS) was calculated from the date of complete remission (CR) until relapse or death. Surviving patients were censored at the date of last follow-up. For patients who had undergone allogeneic transplantation ($n=17$), OS and DFS time were censored at the time of transplantation. The Kaplan-Meier method was used to generate survival curves, and curves were compared using a log-rank test. A multivariate Cox regression using Stata 13 software was performed to adjust for other covariates. Even if covariates do not show statistical significance on the univariate analysis in our cohort, we included them in the multivariate analysis since they have been shown to have an impact on the prognosis in the literature. The multivariate model used is a backward Cox Regression model with variable selection threshold set to $p<0.1$. For all analyses, a p -value <0.05 was considered significant.

Figure S2: **A minimal set of 19 lncRNAs is able to discriminate *NPM1*-mutated from *NPM1*-wild type AML patients.** **A** Sparse PLS-DA plot of *NPM1*-mutated versus *NPM1*-wild type patients based on 19 discriminating lncRNAs. **B** Variable plot of the 19 discriminative lncRNAs. **C** Hierarchical clustering of 40 CN-AML patients (Cohort 1) and associated heatmap of the 19 lncRNAs identified by Sparse-PLS-DA (Fluidigm data). **D** PCA plot of AML patients (Cohort 2; n=50) with 19 lncRNAs identified by Sparse PLS-DA.

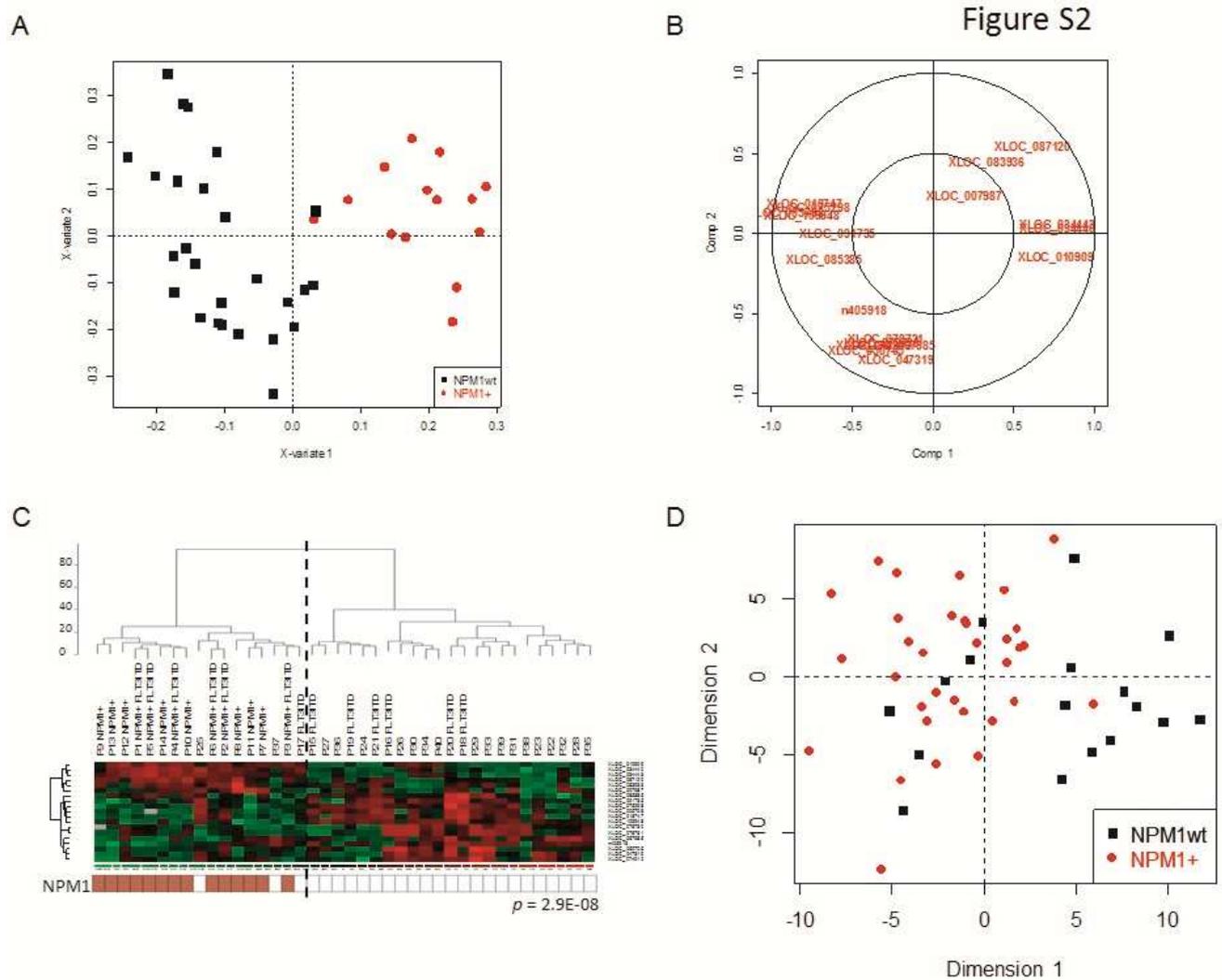


Figure S3: **Bootstrapping cluster analysis showing the stability of the groups based on the 19 lncRNA signature.** Bootstrapping analysis results using the 19 lncRNA signature with the first (A) and second (B) cohorts (1000-fold re-sampling). Hierarchical clustering was used, with Euclidean distance and Ward's method for agglomeration. The purity of the color corresponds to the stability of the sample to be in a specific group. Analysis was carried out in R using the ClassDiscovery package.

Figure S3

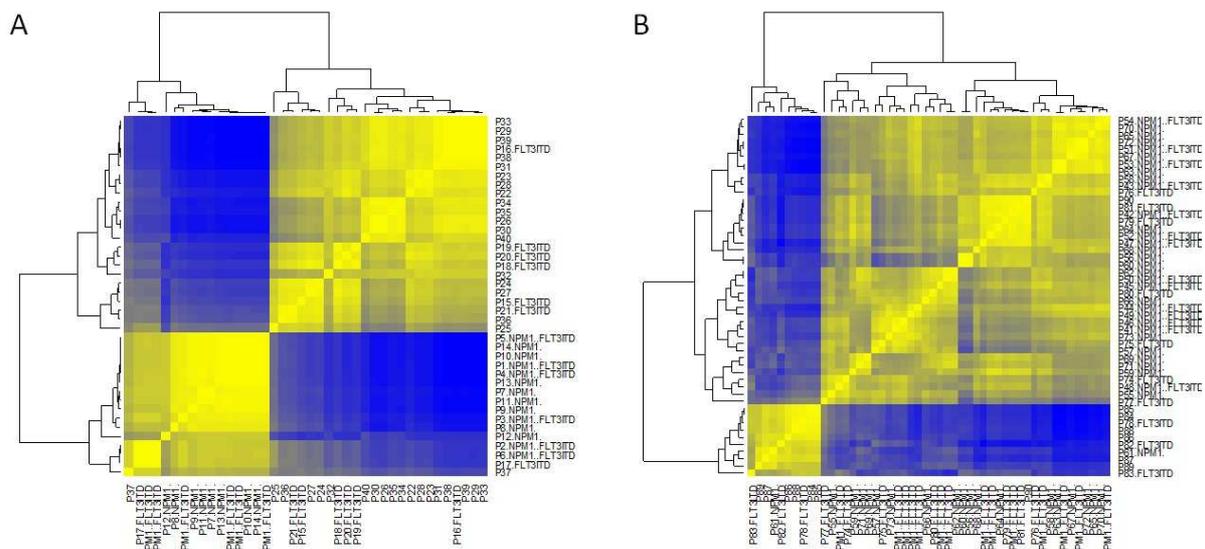


Figure S4: **Validation of the 19 lncRNA signature in the second cohort with only CN-AML patients.** Hierarchical clustering and associated heatmap of Fluidigm data from CN-AML patients (Cohort 2; n=42) with 19 lncRNAs identified by Sparse PLS-DA.

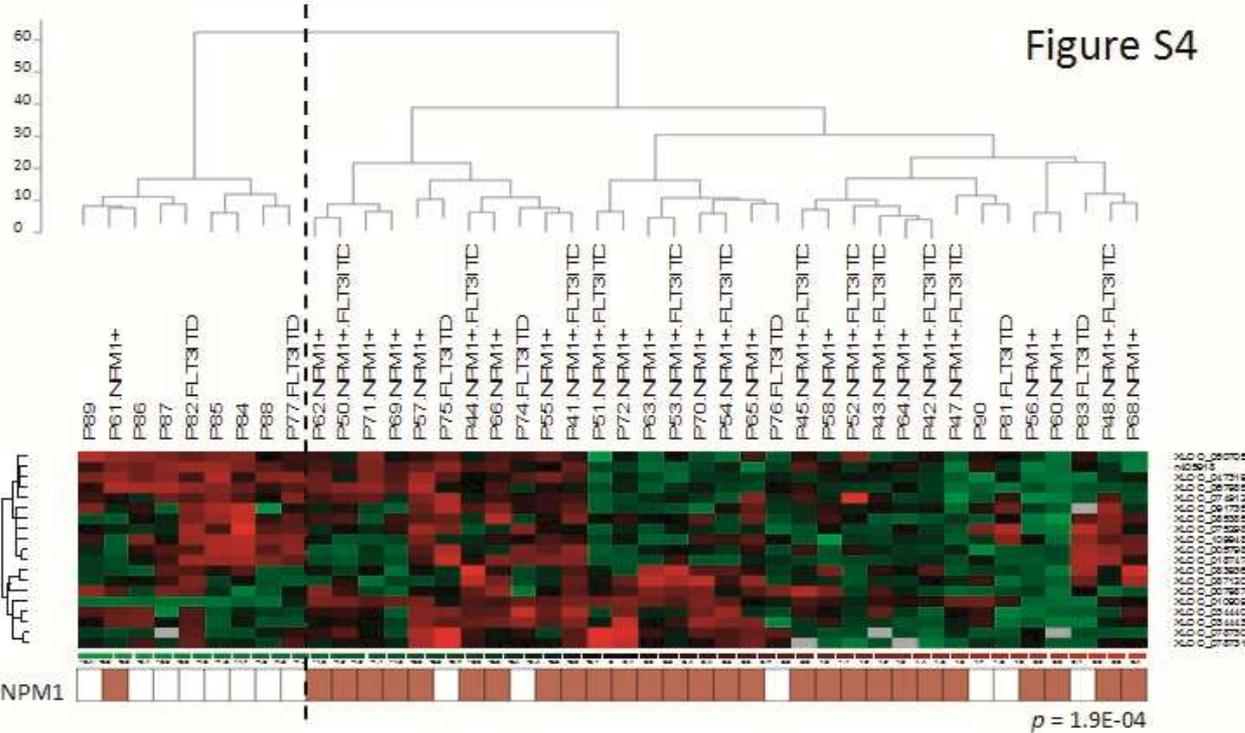


Figure S5: **Gene ontology analysis results with correlated and anti-correlated mRNAs for the 33 differentially expressed lncRNAs.** Of note, 39% of lncRNAs are associated with pathways related to NPM1 functions.

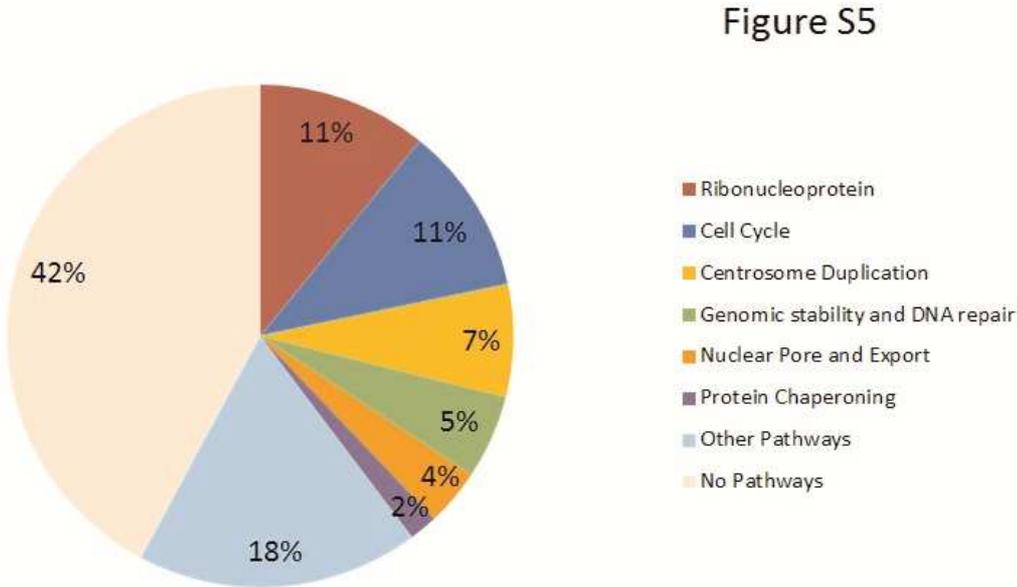


Figure S6: **LncRNA expression as a prognosis marker** **A** Hierarchical clustering and associated heatmap of Fluidigm data from 58 AML patients under 70 years old that had achieved complete remission after treatment, with 32 previously identified lncRNAs. **B** Overall Survival (OS) and Disease Free Survival (DFS) Kaplan Meier curves comparing groups 1 and 2. (OS: $p=0.0419$, hazard ratio (HR): 2.456, 95% confidence interval (CI): 1.033 to 5.838; DFS: $p=0.0205$, HR: 2.687, 95% CI: 1.165 to 6.198). **C** Sparse PLS-DA plot of group 1 (n=36) versus group 2 (n=22) based on two discriminating lncRNAs: XLOC_010909 and XLOC_28779. **D** Variable plot of the two discriminating lncRNAs: XLOC_010909 and XLOC_28779.

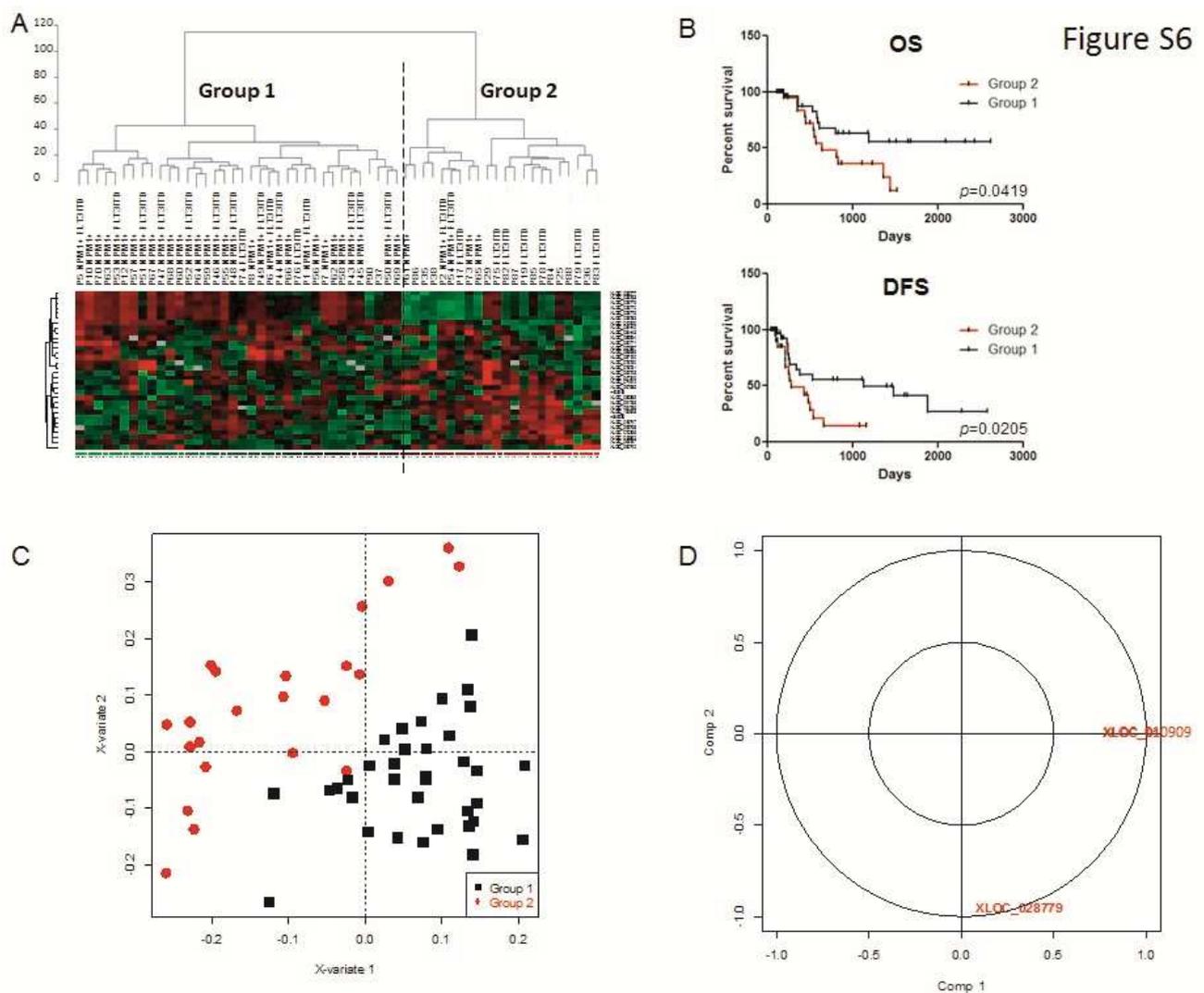


Figure S7: **Kaplan Meier curves comparing group 1 (n=27) and group 2 (n=24) based on XLOC_010909 and XLOC_28779 lncRNA signature with only CN-AML patients.** (OS: $p=0.0759$, hazard ratio (HR): 2.239, 95% confidence interval (CI): 0.919 to 5.454; DFS: $p=0.0232$, HR: 2.641, 95% CI: 1.142 to 6.105).

Figure S7

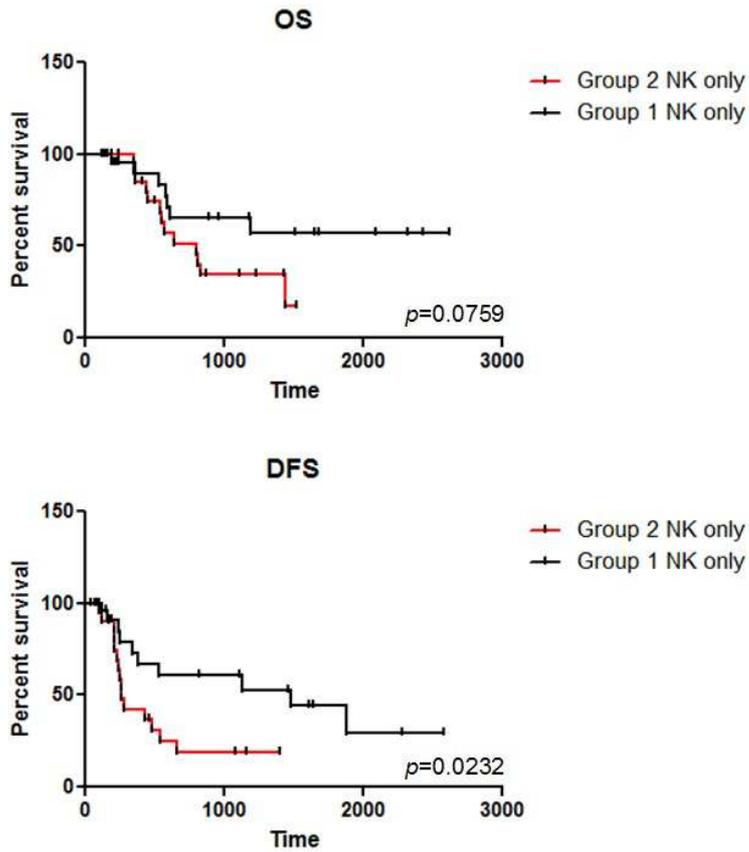


Table S1 : **Patients Characteristics**

Characteristics	Cohort 1 (n=40)	Cohort 2 (n=50)
Median age at diagnosis	67.5	53
Male:female; ratio	21:19; 1.1	20:30; 0.67
White Blood Cells Count (x10⁹ cells/L)		
WBC<50	22 (55%)	25 (50%)
WBC>50	14 (35%)	25 (50%)
NC	4 (10%)	0 (0%)
French American British classification (FAB)		
M0	0 (0%)	2 (4%)
M1	14 (35%)	17 (34%)
M2	9 (22.5%)	9 (18%)
M4	13 (32.5%)	12 (24%)
M5	3 (7.5%)	9 (18%)
M6	1 (2.5%)	0 (0%)
NC	0	1 (2%)
Mutational Status		
NPM1 mutation	14 (35%)	33 (66%)
FLT3-ITD mutation	13 (32.5%)	24 (48%)
Median Bone Marrow Blast Count (%)	79.5	85.5
Cytogenetics		
Normal	40 (100%)	42 (84%)
No recurrent cytogenetic abnormalities (intermediate risk)	0 (0%)	8 (16%)
Chemotherapy		
Intensive	21 (52.5%)	50 (100%)
NC	8 (20%)	0 (0%)

NC: Not Communicated

Table S2: mRNAs differentially-expressed between *NPM1*-mutated and *NPM1*-wild type patients from RNA sequencing data (EdgeR analysis). Highlighted genes are genes known to be deregulated between *NPM1*-mutated versus –wild type AML patients. logFC (log Fold Change), log CPM (log Count per million), FDR (False Discovery Rate).

GeneID	logFC	logCPM	PValue	FDR
Over-expressed in NPM1-mutated patients				
FOXC1	3.7992385	6.617073	1.12E-07	2.49E-05
SERPINB2	3.7624086	7.241341	1.89E-07	3.66E-05
HOXA5	3.2648033	5.568959	8.13E-10	4.35E-07
HOXA6	3.0653511	4.56229	9.94E-08	2.38E-05
PTPN14	2.9203656	5.288403	8.63E-08	2.31E-05
PTX3	2.9068023	5.085375	8.41E-09	3.07E-06
HOXB3	2.8705499	5.989047	7.27E-10	4.14E-07
KRT18	2.8525907	4.451534	2.07E-06	2.35E-04
CD300E	2.6436774	6.685434	9.94E-05	3.71E-03
HOXB4	2.6059775	5.452031	9.03E-10	4.57E-07
IFIT2	2.4890158	6.243663	2.03E-05	1.24E-03
NEU4	2.4444477	4.846467	4.57E-05	2.16E-03
TP53INP2	2.3964926	5.808486	7.70E-11	9.70E-08
PHLDA1	2.2744287	5.596325	7.67E-06	6.47E-04
HOXA7	2.2621114	3.595158	3.46E-05	1.81E-03
ETV5	2.2126576	3.909572	8.32E-09	3.07E-06
IRAK2	2.1879898	4.334432	1.27E-05	9.31E-04
PBX3	2.1392841	6.041945	2.35E-13	7.15E-10
MEIS1	2.1127907	6.818388	7.39E-07	1.04E-04
HOXA9	2.0020864	6.987189	4.75E-05	2.20E-03
Down-regulated in NPM1-mutated patients				
HGF	-2.0269719	5.906225	3.09E-07	5.21E-05
FZD6	-2.0329842	3.641024	9.55E-05	3.64E-03
FAM105A	-2.0421482	5.466201	3.93E-06	3.73E-04
C18orf1	-2.0764253	5.281759	2.28E-06	2.54E-04
CEACAM4	-2.1238841	4.692292	1.54E-06	1.90E-04
MYEF2	-2.1815986	4.084038	1.59E-05	1.07E-03
STK32B	-2.1820237	3.221524	1.72E-05	1.15E-03
LY9	-2.1917438	3.591488	1.79E-06	2.07E-04
GATM	-2.2189544	3.506641	1.86E-05	1.20E-03
SLAIN1	-2.2403403	4.0648	2.80E-06	2.93E-04
TRPS1	-2.2711756	6.504062	3.57E-08	1.08E-05
TNXB	-2.2879998	5.496351	1.01E-04	3.72E-03
GNAI1	-2.3396894	4.894691	3.23E-04	8.04E-03
COBLL1	-2.3414112	4.747695	4.49E-07	7.05E-05
NRP1	-2.4106922	3.987752	1.51E-04	4.78E-03

SPRY1	-2.4227937	4.780269	2.53E-07	4.44E-05
PFKFB2	-2.4260788	5.28147	5.28E-09	2.30E-06
IGJ	-2.524673	7.655813	2.65E-05	1.50E-03
PROM1	-2.5341643	6.820754	5.94E-05	2.57E-03
LTF	-2.5680995	9.211821	1.05E-04	3.78E-03
HLA-DQA1	-2.5988382	5.944873	3.85E-04	9.13E-03
HIST1H3G	-2.6152367	7.499084	4.40E-08	1.25E-05
C10orf10	-2.6196339	4.432897	2.90E-06	2.98E-04
MN1	-2.683685	4.580429	1.02E-05	7.83E-04
STARD9	-2.7103892	5.638785	5.71E-09	2.36E-06
SPARC	-2.7172767	6.100833	6.78E-12	1.54E-08
FAM111B	-2.7939696	3.918043	1.13E-09	5.41E-07
AKT3	-2.8172513	5.861471	6.18E-10	3.75E-07
SETBP1	-2.820028	4.219782	2.62E-07	4.50E-05
NAV1	-2.8628071	4.082467	1.37E-08	4.46E-06
RASD1	-2.8813058	4.833077	5.73E-06	5.28E-04
PRR5L	-2.9081594	4.977676	2.42E-06	2.66E-04
KLHL13	-2.9215933	4.252025	3.73E-08	1.10E-05
ADAMTS10	-3.0293413	4.612616	1.74E-07	3.44E-05
CD200	-3.0689633	4.101284	2.20E-06	2.47E-04
APP	-3.1129035	6.956252	1.10E-07	2.49E-05
ACY3	-3.1149827	4.464958	4.57E-05	2.16E-03
IKZF2	-3.1627802	5.301738	5.29E-09	2.30E-06
TMIGD2	-3.2453597	3.919937	3.44E-10	2.41E-07
SLCO5A1	-3.2615407	4.123672	2.48E-06	2.66E-04
HLA-DOA	-3.3002241	5.611109	1.80E-10	1.36E-07
MLLT3	-3.667135	5.52831	1.76E-20	1.61E-16
L3MBTL4	-3.6796474	3.832075	9.27E-08	2.38E-05
CRHBP	-3.7111283	4.927968	9.51E-08	2.38E-05
CD34	-3.8018409	5.23215	2.31E-07	4.28E-05
ATP10A	-3.9279846	5.308236	4.23E-10	2.75E-07
BAALC	-4.1531125	7.245319	5.02E-07	7.62E-05
BEND4	-4.6850487	4.415333	1.20E-10	9.93E-08
LRP6	-4.7783951	4.287007	4.10E-11	7.47E-08
SDK2	-5.290874	4.833749	8.71E-18	3.97E-14
TRH	-6.0396973	4.54351	8.84E-11	9.70E-08

Table S3: Sequences and localization of the 33 lncRNAs differentially expressed between NPM1-mutated and NPM1-wild type patients

GeneID	localisation	strand	length	exons
n346526	chr9:139,534,589-139,541,618	-	1856	3
n405918	chr9:124,217,319-124,262,306	-	1755	7
XLOC_005798	chr1:25,188,258-25,189,524	-	825	2
XLOC_007987	chr1:151,915,378-151,918,279	-	2902	1
XLOC_010909	chr10:61,810,320-61,845,161	+	1745	5
XLOC_013842	chr10:54,246,152-54,246,835	-	684	1
XLOC_018747	chr11:58,756,982-58,767,516	-	4013	3
XLOC_028763	chr13:110,303,472-110,305,351	-	1880	1
XLOC_028770	chr13:110,324,428-110,325,930	-	1503	1
XLOC_028776	chr13:110,347,864-110,349,260	-	1397	1
XLOC_028777	chr13:110,359,142-110,360,361	-	1220	1
XLOC_028779	chr13:110,364,295-110,365,134	-	840	1
XLOC_028783	chr13:110,377,963-110,379,023	-	1061	1
XLOC_034440	chr15:50,078,505-50,080,055	-	1551	1
XLOC_034443	chr15:50,091,315-50,094,732	-	3418	1
XLOC_047319	chr19:42,119,856-42,121,317	-	1462	1
XLOC_050705	chr2:166,666,979-166,668,466	+	1488	1
XLOC_051554	chr2:201,659,264-201,659,916	+	653	1
XLOC_060742	chr21:21,616,713-21,631,054	-	2794	2
XLOC_063787	chr3:18,315,929-18,319,889	+	3961	1
XLOC_067985	chr3:46,475,919-46,477,217	-	1299	1
XLOC_074912	chr4:70,586,917-70,589,152	-	2236	1
XLOC_075298	chr4:89,336,836-89,376,492	-	866	6
XLOC_078730	chr5:120,504,656-120,505,782	+	1127	1
XLOC_078731	chr5:120,513,301-120,514,526	+	1226	1
XLOC_080771	chr5:56,021,002-56,021,770	-	769	1
XLOC_083936	chr6:27,863,983-27,864,675	+	693	1
XLOC_085385	chr6:119,685,366-119,687,007	+	1642	1
XLOC_085994	chr6:147,235,025-147,238,497	+	3473	1
XLOC_087120	chr6:26,112,403-26,114,712	-	2310	1
XLOC_091735	chr7:24,732,006-24,732,506	+	501	1
XLOC_091736	chr7:24,735,012-24,736,269	+	1258	1
XLOC_109948	chrX:45,688,377-45,689,937	-	1561	1

Table S4: **Validation of RNA-seq results by Fluidigm.** Pearson Correlation r values for each of the 32 lncRNAs comparing RNA-seq and Fluidigm data using the first CN-AML cohort (n=40).

LncRNAs	r value (Pearson Correlation)
XLOC_010909	0.953424872033776
XLOC_007987	0.95329551883069
XLOC_063787	0.94221125910222
XLOC_034443	0.931598622694955
XLOC_067985	0.928830951055749
XLOC_078730	0.927306803540744
XLOC_028776	0.920909079582142
XLOC_028779	0.917401908886419
n405918	0.903424586826947
XLOC_028783	0.89542084109703
XLOC_087120	0.891354935775359
XLOC_080771	0.88818994847133
XLOC_005798	0.884350078636733
XLOC_078731	0.884107622032436
n346526	0.881161931600738
XLOC_028763	0.870957128610975
XLOC_109948	0.866576692659733
XLOC_028770	0.866199780415936
XLOC_075298	0.864919262191128
XLOC_091735	0.852034130723094
XLOC_034440	0.843624082630362
XLOC_050705	0.837092293447742
XLOC_074912	0.835569852789351
XLOC_083936	0.816751303888042
XLOC_018747	0.809607920913865
XLOC_028777	0.808276575737014
XLOC_013842	0.786279797612319
XLOC_047319	0.743913795686985
XLOC_060742	0.716237293782074
XLOC_091736	0.709462935977276
XLOC_085994	0.422500692928231
XLOC_085385	0.21123986212753

Table S5: Karyotype of patients with no recurrent cytogenetic abnormalities

<i>Patients</i>	<i>Karyotype</i>
<i>Patient 46</i>	<i>46,XX, add(6)(p25)[2]; 46,XX[18]</i>
<i>Patient 49</i>	<i>47,XX,+8[11]; 46,XX [2]</i>
<i>Patient 59</i>	<i>46,XX, del(14); (q22q31)[4]; 46,XX[6]</i>
<i>Patient 67</i>	<i>46,X, del(X)(q21q24)<5>; 46,XX<15></i>
<i>Patient 73</i>	<i>46, XX, der(7)t(4;7)(q21;q31)[11]; 46, XX[4]</i>
<i>Patient 78</i>	<i>46,XY[18]; 45,X,-Y[2]</i>
<i>Patient 79</i>	<i>51,XY, +8,+8, +8, +10, +13[10]</i>
<i>Patient 80</i>	<i>46,XX, add(17)(q12)[19] - 46,XX[1]</i>

Table S7: Univariate and multivariate Analysis

Univariate analysis was performed using the LogRank test. The multivariate model is a backward Cox Regression model with variable selection threshold set to $p < 0.1$. The two lncRNA signatures were tested with covariates such as age (< 60 years vs ≥ 60 years), *NPM1* and *FLT3-ITD* mutational status, WBC count ($< 50 \times 10^9$ cells/L¹ vs $\geq 50 \times 10^9$ cells/L) and karyotype (normal or not recurrent cytogenetic abnormalities).

Multivariate analysis with 32 lncRNAs, age, NPM1, FLT3, WBC and karyotype variables

OS	Variables	LogRank test	Cox regression	Hazard ratio	95%IC
	32 lncRNAs	0.04	0.048	0.42	0.17-0.99
	Age < 60y	0.19	0.31		
	NPM1 mut	0.09	0.79		
	FLT3-ITD mut	0.25	0.33		
	WBC > 50G/L	0.99	0.86		
	karyotype	0.65	0.84		

DFS	Variables	LogRank test	Cox regression	Hazard ratio	95%IC
	clust 33	0.02	0.025	0.40	0.17-0.89
	Age < 60y	0.28	0.42		
	NPM1 mut	0.07	0.78		
	FLT3-ITD mut	0.33	0.55		
	WBC > 50G/L	0.54	0.49		
	karyotype	0.88	0.83		

Multivariate analysis with 2 lncRNAs, age, NPM1, FLT3, WBC and karyotype variables

OS	Variables	LogRank test	Cox regression	Hazard ratio	95%IC
----	-----------	--------------	----------------	--------------	-------

	2 lncRNAs	0.03	0.04	0.39	0.15-0.96
	Age < 60y	0.19	0.25		
	NPM1 mut	0.09	0.84		
	FLT3-ITD mut	0.25	0.33		
	WBC > 50G/L	0.99	0.99		
	karyotype	0.65	0.74		
DFS	Variables	LogRank test	Cox regression	Hazard ratio	95%IC
	2 lnc	0.008	0.011	0.33	0.14-0.78
	Age < 60y	0.28	0.41		
	NPM1 mut	0.07	0.80		
	FLT3-ITD mut	0.33	0.45		
	WBC > 50G/L	0.54	0.82		
	karyotype	0.88	0.86		

Multivariate analysis with age, NPM1 , FLT3 , WBC and karyotype variables

OS	Variables	LogRank test	Cox regression	Hazard ratio	95%IC
	Age < 60y	0.19	0.22		
	NPM1 mut	0.09	0.09	2.04	0.87-4.78
	FLT3-ITD mut	0.25	0.32		
	WBC > 50G/L	0.99	0.97		
	karyotype	0.65	0.86		

DFS	Variables	LogRank test	Cox regression	Hazard ratio	95%IC
	Age < 60y	0.28	0.42		
	NPM1 mut	0.07	0.07	2.04	0.94-4.45
	FLT3-ITD mut	0.33	0.39		
	WBC > 50G/L	0.54	0.54		
	karyotype	0.88	0.81		

Table S8: Primer list

lncRNAs	Forward Primer	Reverse Primer
n346526	GGAGAGCTTGGGTGAGGACCAC	AAGGGGGCTCAGCAGGTGATAA
n405918	AGCTTCGGCTATGGGACTGGTT	CGCATGATCTCTCAGTCCAGGTC
XLOC_005798	GGCAGTCAGGAAGGAAAACCTGGA	CGGTGGCTGTTTCTCCACTGTC
XLOC_007987	AAACCCACCAACGTGACCAGAA	GCATCCTGGCAAATGAGGACAT
XLOC_010909	GACAAGGGGGAGACTACAAGAAGCA	CTGCGCTCAAGCTCGAAGAGTAA
XLOC_013842	GGTTTGGAGTGAAAGGCACATGA	CCAGGCAGCACTATCTGGGAAA
XLOC_018747	GCCTGGACAGATTACCAGATCATCA	AGCAAGCTGGAAGCTGAGCTCTT
XLOC_028763	CAGGCACCAGGAGAGGAGACAC	CACGGGGACCCCTGAACTTAAA
XLOC_028770	TCCGGTCATTGGAAAAGTGCAT	AGGGTTTCCCTGGCCAACCTGAG
XLOC_028776	CTTTTGGTGGCCAAATCCTGTG	ACGGTTTCCACCCCTAGTGCAT
XLOC_028777	TTTGCTGCTCACCTGCTGTGAC	TCTAGGCTGGAGGCAGTGAGGA
XLOC_028779	TGAGCTTCCCCACTGCATGAAT	AGGCCGAAATGATTGTGATGTGA
XLOC_028783	CCTGGGCCAATCACTGATGTTC	GGGCAGGAAAGCTGTCAGTCAC
XLOC_034440	TGCTTAGTGAAAGATGCCAGACAAAA	TCCCCAGGCAGTAACCAAACCTG
XLOC_034443	TGCACCGACTGACAGAATGAGG	TCCAGGGTTCCAAGATAGCGAAA
XLOC_047319	TCATCAGCCTCCTGCTCCACTC	TGATTACCACGCAAGGCTGACA
XLOC_050705	TTCCCGGTGGGTGTCATATCAG	CTAGCAGCCCCCTCCCTCCATC
XLOC_060742	GGGAATGGAGAATTACCCTCTGGA	AAACAGGCCATTGACAGGTAGACAA
XLOC_063787	GCCAATGCCAACTTCACTCCAG	ACGCAGCCTCATACACCCATTG
XLOC_067985	GCAAATGGCCTCCAATCAAAA	AGCCACAGATCCATTGCTGACC
XLOC_074912	CACAGGTGAATCTCCGCTGTCA	ATCCTTTGCTGGGGTTGGAAA
XLOC_075298	TTGGGACTACCACCCAAAGCAG	GGTGTGAGCCATTCTGCTGGAA
XLOC_078730	GCATGCTAAGAGCCCATGAAGC	AGAAGCTGGCCCAAGTGTGTGT
XLOC_078731	GGCCAGGTGCAGTAGTTCATGC	AAACAAGGTTCCCACGGCTGAT
XLOC_080771	TACATGTGCCATGGTGGTTTGC	TAGGGGGCTAGGGGAGGGATAG
XLOC_083936	GCAATCTCGGCTCACTGCAACT	TGTGGCAGGTGCCTGTCTGTAG
XLOC_085385	TCCATTTTGGAGGCTGGGAAGT	TAGGGAGTTGGCCATCTGCAAG
XLOC_085994	GACCGAGATGAGGCAAAGAGCA	GAGCAATGGGGATAGTGCCACA
XLOC_087120	CCATGCACCAGGCACTGTAATG	TGGGTAGCACCTCTCCATGAGC
XLOC_091735	GCTGGTGGAAAGGTCTGTTCTGC	TGGCAGCAAGGATTTGCTAAGG
XLOC_091736	AAGAACCCGTGTCAGGCTTGTG	CCTCTGACAGCCAGCAAGGAAA
XLOC_109948	CCTGCCTTCTTGTCTCCACCT	GGCCAGGAAACAAATTGCAACA
MLN51	TAATCCCAGTTACCCTTATGCTCCA	GTTATAGTAGGTCACTCCTCCATATACCTGT
5S	ACGCGCCCGATCTCGTCTGAT	GCCTACAGCACCCGGTATTCCC

SUPPLEMENTAL LEGEND OF ADDITIONAL TABLE

Table S6: **Correlation analysis between lncRNAs and mRNAs.** Sheet 1, Pearson Correlation (r values) for the 33 lncRNAs (columns) differentially-expressed between *NPM1*-mutated ($n=14$) and *NPM1*-wild type patients ($n=26$) for each mRNA (rows, $n=11328$). Sheet 2, Gene ontology analysis results for each lncRNA using positively correlated ($r>0.5$; $p<0.001$) and anti-correlated ($r<-0.5$; $p<0.001$) mRNAs.

SUPPLEMENTAL REFERENCES

- 1 Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 2009; **25**: 1105–11.
- 2 Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 2010; **28**: 511–5.
- 3 Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* 2012; **7**: 562–78.
- 4 Kong L, Zhang Y, Ye Z-Q, Liu X-Q, Zhao S-Q, Wei L *et al.* CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res* 2007; **35**: W345–9.
- 5 Li R, Yu C, Li Y, Lam T-W, Yiu S-M, Kristiansen K *et al.* SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* 2009; **25**: 1966–7.
- 6 Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 2008; **5**: 621–8.
- 7 Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010; **26**: 139–40.

Résultats

Supplémentaires

Grâce à la technique de RNA-seq, nous avons pu identifier 8526 nouveaux lncRNAs et observer des profils d'expressions particuliers de ces lncRNAs au sein des LAMs à caryotype normal. Par ailleurs, nous avons pu mettre en évidence, par l'analyse biostatistique de données RNA-seq et la validation par RT-qPCR, 33 lncRNAs associés à la présence de la mutation *NPM1* dans ces LAM-CN. Sur ces 33 lncRNAs, 2 seulement ont été décrits mais non caractérisés, laissant donc un grand champ d'investigations dans l'analyse de l'impact de ces lncRNAs dans cette pathologie.

Afin de déterminer quelles pourraient être les fonctions de ces lncRNAs, nous avons dans un premier temps utilisé la méthode de Guilt By Association. Cette technique consiste à identifier les ARNm dont l'expression est corrélée (positivement ou négativement) à celle des lncRNAs. L'analyse GOTerm de ces gènes permet alors de définir dans quelles voies moléculaires et cellulaires ces lncRNAs pourraient intervenir. Nous avons pu observer que 40% des lncRNAs identifiés dans cette signature sont associés à des voies et des gènes impliqués dans les fonctions de NPM1, telles que la biogenèse des ribosomes, le cycle cellulaire, la duplication des centrosomes, la stabilité génomique et la réparation de l'ADN, les pores nucléaires et l'export, ou des gènes codants pour des chaperones de protéines. De plus, 4 lncRNAs ont été associés à l'expression de gènes homéotiques, bien connus comme étant associés à la mutation *NPM1* : les lncRNAs XLOC_010909, XLOC_034440, XLOC_034443 et XLOC_087120. Ces résultats permettent de valider la sélection de nos lncRNAs, mais ne démontrent en aucun cas leurs fonctions.

Les mécanismes de transformations cellulaires, liés à la présence de NPM1 mutée, restent encore obscurs. L'identification de nouveaux acteurs dans le processus oncogénique liée à NPM1 mutée est donc très importante. Les données de la littérature et nos résultats permettent d'envisager que ces nouveaux acteurs pourraient être des lncRNAs.

Dans cette étude nous avons utilisé différents modèles cellulaires, dont 3 lignées cellulaires LAMs : la lignée OCI AML3 contenant la mutation *NPM1* et deux lignées de LAM promyélocytaire NB4 et HL60ER, non porteuses de la mutation *NPM1*. Les techniques utilisées sont décrites dans la partie "Matériels et méthodes".

1) Quel est l'impact de la mutation *NPM1* sur l'expression des lncRNAs

Au vue de l'association des lncRNAs avec la mutation de *NPM1*, la première question a été de déterminer quel pouvait être l'impact de la présence de la mutation *NPM1* sur l'expression de ces lncRNAs. Pour cela, nous avons transduit la lignée LAM HL60ER (exprimant le récepteur écotropique permettant la transduction par des retrovirus murins) avec le vecteur rétroviral XZ contenant *NPM1* muté ou le vecteur vide. Le vecteur rétroviral XZ possède une séquence IRES-GFP, ce qui permet de trier les cellules GFP⁺ transduites. La mutation introduite dans *NPM1* est la mutation

A présente dans la majorité des patients LAM *NPM1* muté (mutation consistant en une insertion TCTG en position 956 du gène). A l'aide d'amorces spécifiques de la mutation A, nous avons pu confirmer par RT-qPCR l'expression de la forme *NPM1* muté dans nos cellules transduites et triées (Figure 1A). Ces résultats ont également été validés par western blotting avec un anticorps spécifique de *NPM1* mutée (fourni gracieusement par le Pr Brunangelo Falini, Pérouse, Italie) (Figure 1B). Il est important de noter que le niveau d'expression de l'ARNm *NPM1* muté introduit est presque similaire à celui de *NPM1* sauvage, comme observé chez les patients LAM *NPM1* muté (données non montrées). On remarque aussi une légère augmentation de l'expression de *NPM1* sauvage après introduction de *NPM1* muté dans cette lignée (Figure 1A).

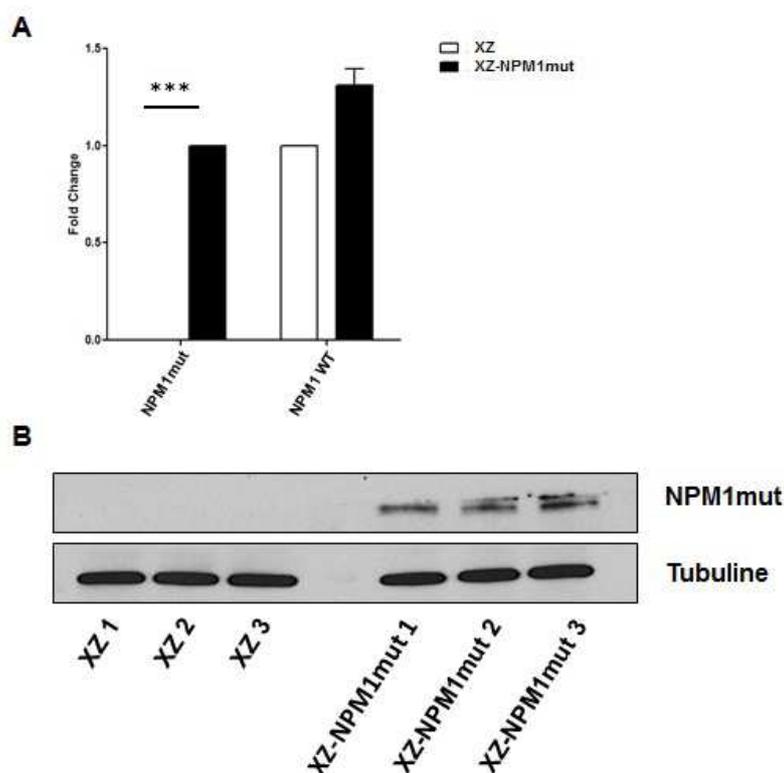


Figure 1: Expression de *NPM1* dans la lignée HL60ER transduite avec XZ-NPM1mut. A. Représentation graphique du niveau d'expression des transcrits *NPM1* muté et *NPM1* sauvage dans les cellules HL60ER-XZ et HL60ER-NPM1mut (triées GFP⁺) évalué par RT-qPCR à partir de trois transductions indépendantes; ***= p valeurs<0.001. B. Expression de la protéine *NPM1* mutée évaluée par western blot avec un anticorps spécifique de *NPM1* mutée. Trois transductions indépendantes, suivies d'un tri GFP⁺, ont été réalisées pour HL60ER-XZ et HL60ER-NPM1mut. Normalisation avec la Tubuline.

A partir de ces lignées, nous avons pu comparer le niveau d'expression de chacun des lncRNAs de la signature entre les cellules HL60-XZ et HL60-NPM1mut (Figure 2). A noter que parmi les 33 lncRNAs de la signature, seulement 22 lncRNAs sont détectables par RT-qPCR dans cette

lignée. La majorité des lncRNAs ne semble pas être affectée par l'introduction de la mutation *NPM1* dans ce modèle. Seuls les lncRNAs XLOC_074912 et XLOC_091736 sont significativement diminués en présence de la mutation *NPM1* et ces derniers étaient également sous exprimés chez les patients porteurs de la mutation *NPM1* (Figure S1 "Article"). On observe aussi une sous expression des lncRNAs XLOC_028763, XLOC_028770 et XLOC_028776. Ces lncRNAs sont regroupés en cluster au niveau du bras long du chromosome 13 que nous avons appelé cluster 028, et qui sont normalement surexprimés chez les patients portant la mutation *NPM1* (Figure S1 "Article").

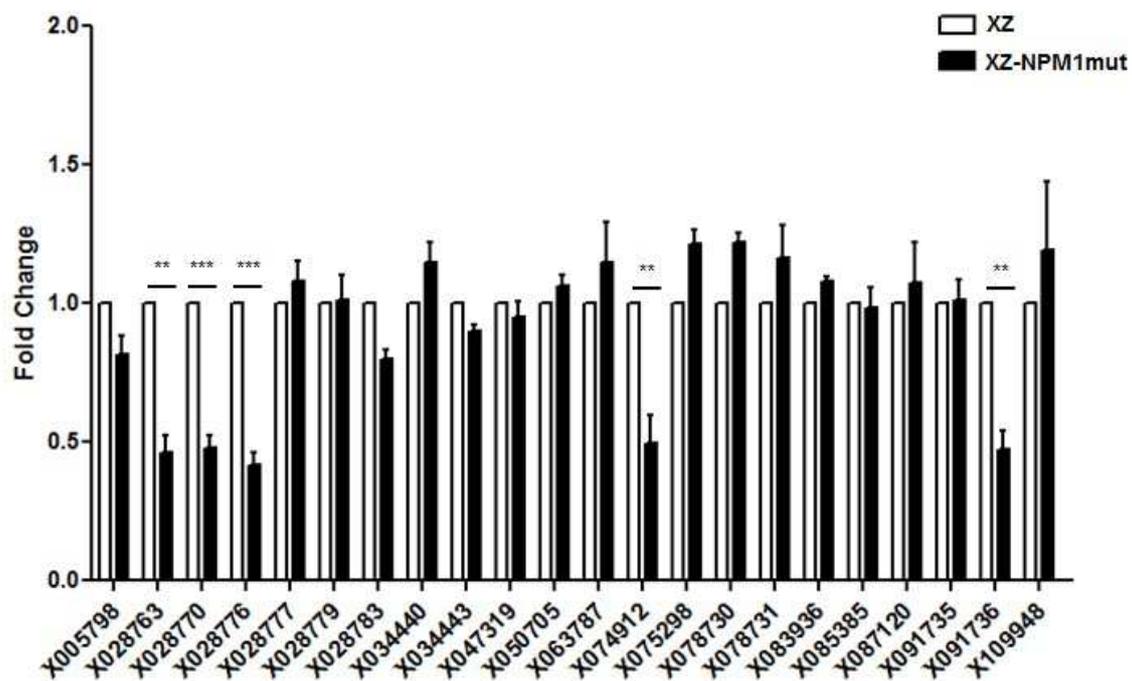


Figure 2: Impact de la mutation *NPM1* sur l'expression des lncRNAs. Expression relative de 22 sur 33 des lncRNAs de la signature *NPM1* évaluée par RT-qPCR dans les HL60ER-XZ et HL60ER-NPM1mut; n=3 ; ***= p valeurs<0.001 ; **= p valeurs<0.01.

Afin de déterminer si la présence de la mutation *NPM1* avait un impact sur le niveau d'expression des lncRNAs sélectionnés, nous avons utilisé une autre stratégie consistant à diminuer l'expression de la forme *NPM1* mutée dans la lignée OCI-AML3 (*NPM1* muté). Pour ce faire, nous avons utilisé une approche par GapmeRs. Les GapmeRs sont des petits oligonucléotides de 16nt s'hybridant spécifiquement avec un ARN cible. Cette méthode présente quelques avantages majeurs par rapport aux méthodes se basant sur l'ARN interférence. D'une part, cela fait intervenir une dégradation par la RNase H endogène qui permet une meilleure dégradation des ARNs nucléaires. D'autre part, les GapmeRs peuvent entrer par simple diffusion passive sans agent de transfection dans les cellules, augmentant ainsi l'efficacité de transfection, et la durée de l'inhibition si le GapmeR est présent dans le milieu. Par ailleurs, l'utilisation d'une amorce plus petite qu'un siARN permet dans ce

cas de diminuer les hybridations partielles sur l'ARN *NPM1* sauvage (4nt de différence entre *NPM1* muté et sauvage). Nous avons donc transfecté transitoirement les cellules OCI AML3 avec le GapmeR ciblant *NPM1* muté ou le GapmeR contrôle, et analysé les résultats à 48h, 96h et jusqu'à 6 jours post-transfection. En RT-qPCR, on observe une baisse très nette de l'ARNm de *NPM1* muté (environ 75%), mais également une baisse significative de *NPM1* sauvage (environ 50%) (Figure 3A). On obtient la même efficacité d'inhibition entre 48h et 6 jours (données non montrées). Ce résultat est confirmé au niveau protéique par western blotting, avec une baisse de NPM1 mutée (figure 3B).

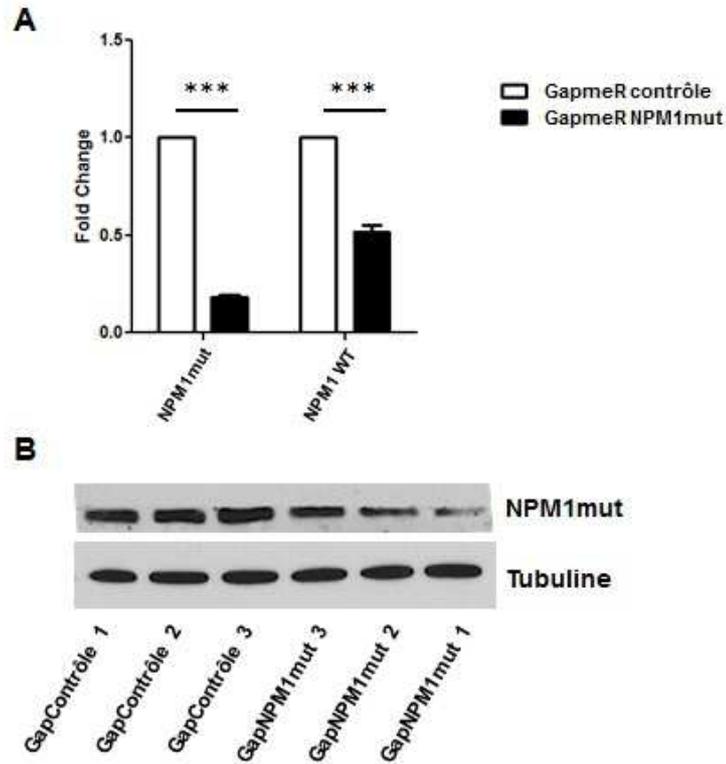


Figure 3: Expression de NPM1 après inhibition par transfection de GapmeR dans OCI AML3. A. Niveaux d'expression de, *NPM1* muté et *NPM1* sauvage dans les OCI AML3 96h après transfection des GapmeRs ciblant *NPM1* muté ou contrôle, évalués par RT-qPCR. n=3 ; ***= p valeurs<0.001. B. Niveau d'expression de la protéine NPM1 mutée dans les OCI AML3 des 3 expériences 96h post-transfection des GapmeRs ciblant *NPM1* muté ou contrôle, par Western Blot. Normalisation avec la Tubuline.

Dans ce modèle, seule l'expression de deux lncRNAs (XLOC_080771 et XLOC_007987) varie en présence du GapmeR *NPM1* muté par rapport au GapmeR contrôle (Figure 4). A noter que la diminution de l'expression de XLOC_007987 est en accord avec les résultats obtenus chez les patients (Figure S1 "Article"). A l'inverse, l'augmentation de l'expression de XLOC_080771 est contradictoire avec les résultats obtenus chez les patients (Figure S1 "Article"). Dans la lignée HL60ER, ces deux lncRNAs ne sont pas exprimés, ce qui ne permet pas de conclure. De plus, nous n'observons pas de

différence d'expression pour les lncRNAs dérégulés par l'introduction de *NPM1* muté dans les HL60ER, suggérant une régulation spécifique de ces lncRNAs dépendante du contexte cellulaire.

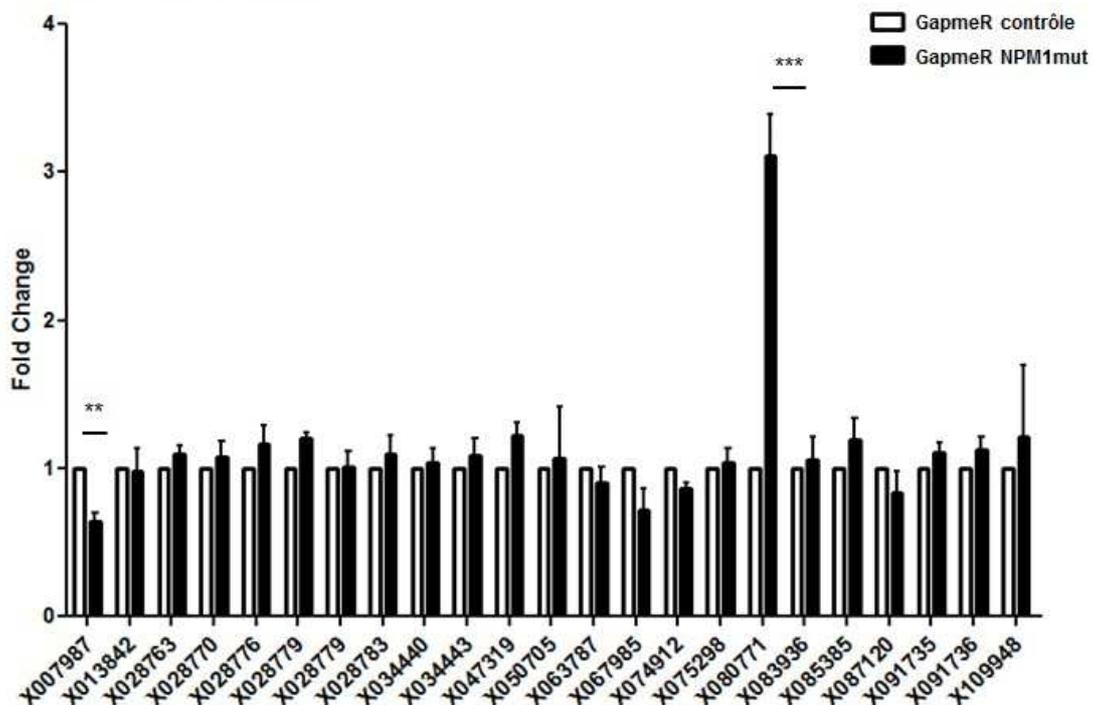


Figure 4: Impact de l'inhibition de *NPM1* mutée sur l'expression des lncRNAs. Expression relative de 23 sur 33 des lncRNAs de la signature *NPM1* évaluée par RT-qPCR dans les OCI AML3 transfectées avec les GapmeRs *NPM1* muté ou contrôle, 96h post-transfection. n=3 ; ***= p valeurs<0.001 ; **= p valeurs<0.01.

Ainsi, il apparait que la mutation de *NPM1* pourrait influencer l'expression de certains des lncRNAs que nous avons observés comme étant associés à la mutation *NPM1* chez les patients. Cependant, pour la majorité d'entre eux, il n'y a pas de dérégulation d'expression. L'introduction de la mutation *NPM1* dans la lignée HL60ER ou la diminution de *NPM1* muté dans la lignée OCI AML3 ne sont peut-être pas suffisantes pour reproduire le phénotype observé *in vivo*.

II) Quelle est la localisation cellulaire des lncRNAs associés à la mutation *NPM1*?

Comme cité dans le chapitre "localisation cellulaire" de la partie III de l'introduction, la localisation cellulaire des lncRNAs est une question importante, permettant de nous guider vers l'identification de leurs fonctions. Pour répondre à cette question, nous avons utilisé la technique de fractionnement cellulaire permettant de séparer le noyau et le cytoplasme. Cette technique a été utilisée sur la lignée NB4 (présentant globalement les plus fortes expressions de lncRNAs et le plus

grand nombre de lncRNAs détectables), et nous avons analysé l'abondance de chaque transcrite de la signature dans le noyau et le cytoplasme. La validation du fractionnement a été évaluée avec l'ARNm de la *GAPDH*, majoritairement cytoplasmique, et le snoARN *snord44*, majoritairement nucléaire. Cette expérience nous a permis d'observer que la majorité des lncRNAs est clairement enrichie dans la fraction nucléaire, en accord avec la littérature (Figure 5). Sur les 26 lncRNAs testés, le n405918, XLOC_075298 et à moindre échelle les XLOC_087120 et XLOC_091735, sont plus abondants dans le cytoplasme (Figure 5).

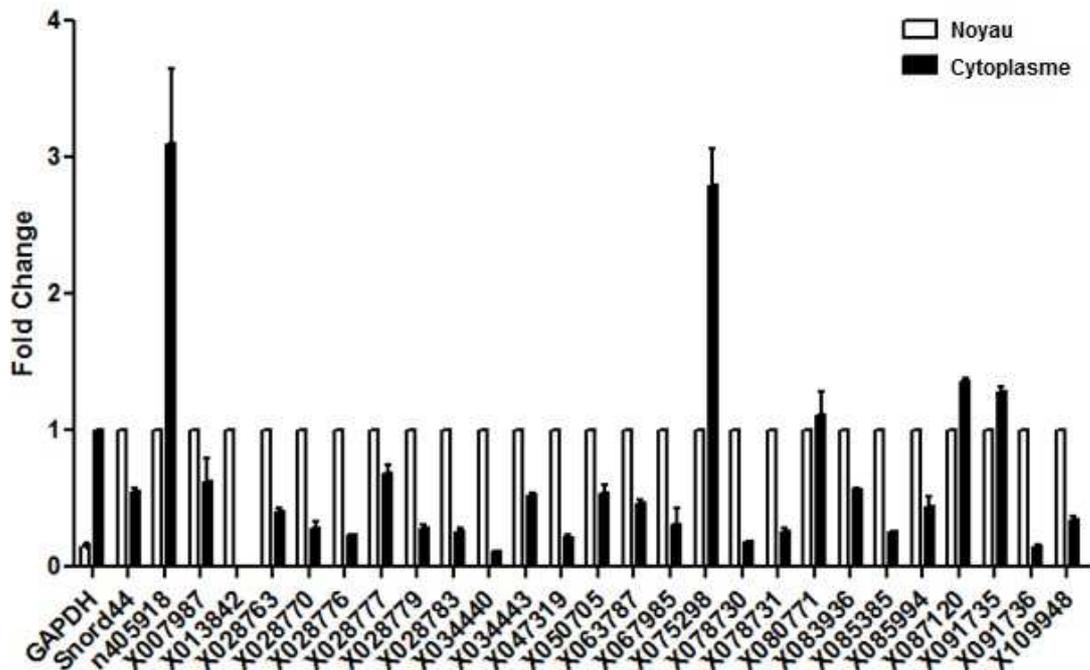


Figure 5: Localisation cellulaire des lncRNAs dans les cellules NB4. Quantification relative de 26 sur 33 des lncRNAs de la signature NPM1 dans les fractions nucléaires et cytoplasmiques de la lignée NB4, par RT-qPCR. *GAPDH* est utilisé comme contrôle positif de la fraction cytoplasmique et *Snord44* comme contrôle positif de la fraction nucléaire. Normalisation par rapport au noyau ; n=3.

III) Quelles peuvent être les fonctions de ces lncRNAs?

Sachant que la majorité des lncRNAs sont nucléaires et que, d'après la littérature, une grande proportion des lncRNAs est impliquée dans des régulations épigénétiques, nous avons voulu déterminer lesquels de ces transcrits pouvaient potentiellement interagir avec des complexes de régulation épigénétique. Pour cela, nous avons utilisé un algorithme de prédiction d'interaction appelé "catRAPID". Ce programme permet d'établir un score d'interaction potentielle entre un ARN et toutes les protéines d'une base de données. Cette analyse nous a permis d'observer que 19 lncRNAs sur les 33 lncRNAs de la signature (57%) présentaient des sites potentiels d'interaction avec une des sous-unités du complexe PRC2 (SUZ12 ou EZH2). Parmi ces 19 lncRNAs, 5 lncRNAs contenaient même des

sites pour les deux sous unités SUZ12 et EZH2, les 14 autres présentant uniquement des sites potentiels d'interaction avec SUZ12 (Figure 6A). Le complexe PRC2 est un complexe répresseur permettant la triméthylation en lysine 27 de l'histone H3 (H3K27m3) entraînant une extinction génique. Comme nous l'avons vu dans l'introduction, les lncRNAs semblent être des acteurs clés dans ce mécanisme. En effet, plusieurs études ont montré que les lncRNAs étaient capables de recruter le complexe PRC2 pour réguler négativement l'expression de gènes en *cis* ou en *trans*. Nous avons donc recherché parmi les 19 lncRNAs prédits comme interagissant avec le PRC2, quels étaient ceux dont l'expression était anti-corrélée avec celle de gènes codants. Pour chacun de ces lncRNAs, nous avons réalisé une analyse GOMTerm des ARNm anti-corrélés. Ainsi, nous avons observé que, parmi les 19 lncRNAs pouvant potentiellement interagir avec PRC2, 10 lncRNAs étaient négativement corrélés à l'expression de gènes enrichis dans des GO Term tels que les complexes ribonucléoprotéiques, le cycle cellulaire, l'organisation de la chromatine, l'épissage, la traduction ou la réponse immunitaire (Figures 6B). Chacune de ces voies a déjà été décrite comme faisant intervenir des lncRNAs. Ces résultats sont à prendre avec précaution, puisqu'ils sont basés uniquement sur de l'anti-corrélation d'expression, mais cela permet d'ouvrir de nouvelles pistes dans la recherche de la fonction des lncRNAs que nous avons sélectionnés.

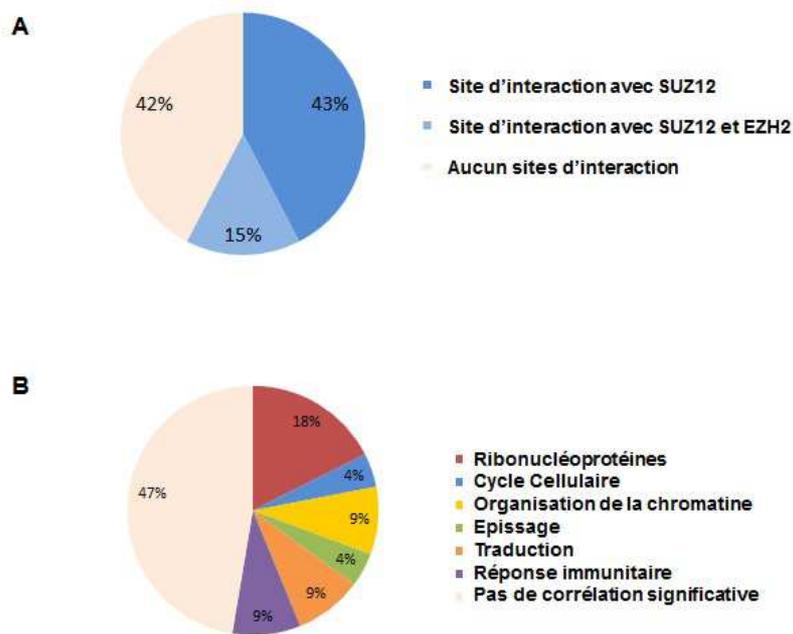


Figure 6: Les lncRNAs associés à la mutation *NPM1* prédits pour interagir avec le complexe PRC2. A. Diagramme en secteur représentant la distribution des 33 lncRNAs associés à la mutation *NPM1* présentant un site d'interaction potentiel avec SUZ12 et/ou EZH2 ou aucun site. B. Diagramme en secteur représentant la distribution des GOMTerm associés aux gènes anti-corrélés à l'expression des lncRNAs présentant au moins un site d'interaction avec le complexe PRC2.

Pour aller plus loin, et sachant que dans la littérature cette régulation semble majoritairement en *cis*, nous avons analysé les corrélations d'expression entre les lncRNAs et leurs gènes codants proximaux dans une fenêtre de 200kb. De manière intéressante, seulement 2 lncRNAs sur les 19 présentent des corrélations négatives significatives avec les gènes adjacents : le lncRNA n405918 anti-corrélé avec le gène *RAB14* ($r=-0.51$; $p<0.001$) et le lncRNA XLOC_087120 anti-corrélé avec 9 gènes codants des histones dans le cluster 1 des histones ($r<-0.3$; $p<0.05$). Le lncRNA n405918 est clairement majoritairement cytoplasmique dans la lignée NB4 et ce lncRNA n'est pas exprimé dans la lignée OCI-AML3. Cela suggère que son association à la chromatine et à des complexes épigénétiques ne pourrait être démontrée de manière pertinente dans ces modèles. Nous ne sommes donc pas allés plus loin dans sa caractérisation. En revanche, le XLOC_087120 semble être particulièrement intéressant si le fait qu'il régule l'expression des histones s'avère exact. En effet, comme XLOC_087120 est surexprimé chez les patients *NPM1* muté et que *NPM1* est une histone chaperone, nous pouvons supposer que ce lncRNA pourrait participer à l'organisation chromatinienne modulée par *NPM1*.

IV) Etude du lncRNA XLOC_087120

Ce lncRNA est localisé en 6p22.2 au sein du cluster principal de gènes codants des histones (Figure 7).

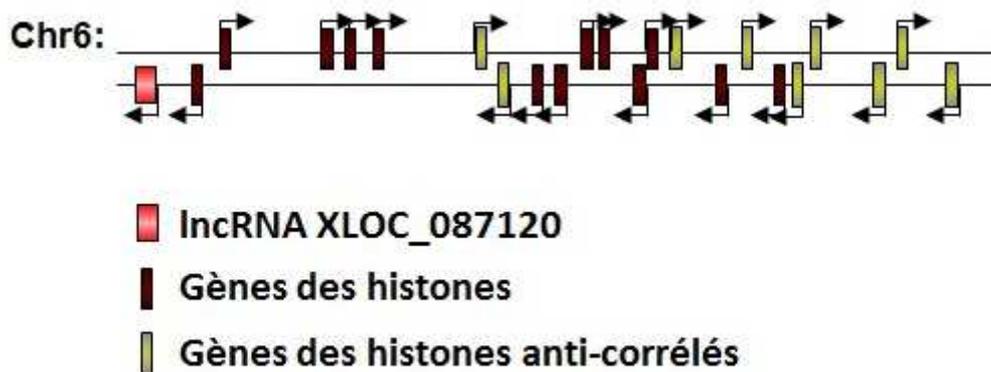


Figure 7 : Représentation schématique du gène XLOC_087120 au sein du cluster 1 des histones.

Chez les patients LAMs portant la mutation *NPM1*, le lncRNA XLOC_087120 est surexprimé (Figure 8A). Par corrélation de Pearson, nous avons pu déterminer que son expression était anti-corrélée avec celles de 9 gènes proximaux codants des histones. Chez les patients, on observe en effet une diminution du niveau d'expression de ces histones en présence de la mutation *NPM1* (Figure 8B).

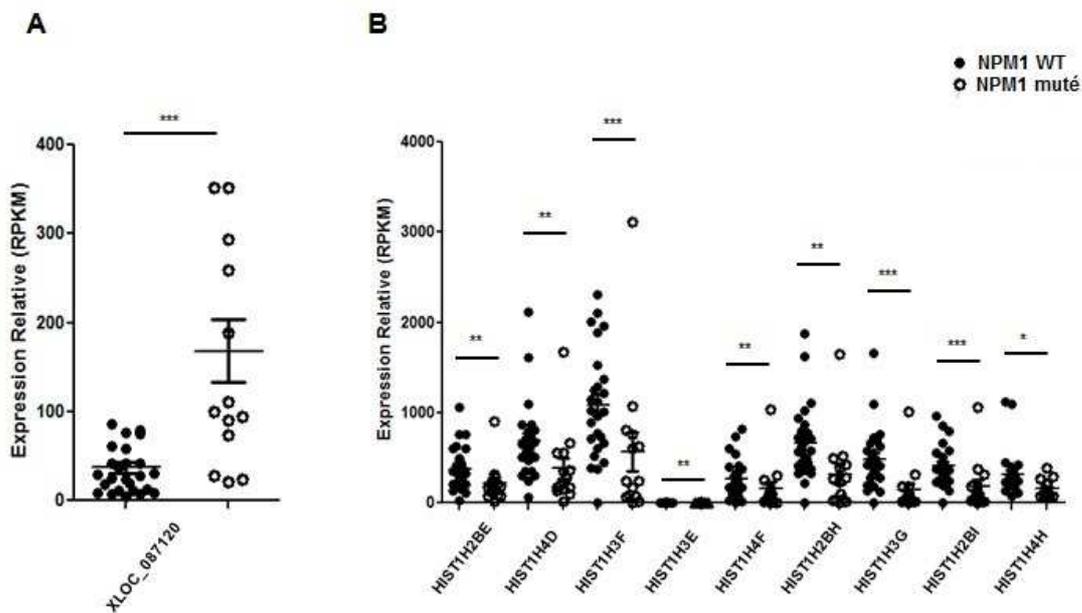


Figure 8 : XLOC_087120 est anti-corrélé à l'expression des gènes proximaux codants les histones. A. Niveau d'expression en RPKM (Reads per Kb per Million) du lncRNA XLOC_087120 entre les patients *NPM1* muté (n=14) et *NPM1* sauvage (n=26). B. Niveau d'expression en RPKM des 9 histones proximales et anti-corrélées à XLOC_087120, entre les patients *NPM1* muté (n=14) et *NPM1* sauvage (n=26). ***= p valeurs < 0.001 ; **= p valeurs < 0.01 ; *= p valeurs < 0.05.

Les gènes des histones dérégulés dans ce cluster appartiennent à la famille des histones canoniques H2B, H3.1 et H4. Sachant qu'il existe plusieurs gènes codant chacun des histones canoniques (17 H2B; 13 H3.1 et 14 H4), nous nous sommes demandés si la diminution du niveau d'expression des histones localisées près de XLOC_087120 avait un réel impact sur le niveau d'expression globale des histones canoniques ou s'il existait des mécanismes de compensation. Nous avons additionné les expressions de chacun des gènes codant la même histone et comme le montre la figure 9, les histones H3.1 et H4 sont significativement sous exprimées en présence de la mutation *NPM1* chez les patients. Pour les histones H2B, la différence n'est pas statistiquement significative, mais elle semble suivre cette tendance. En résumé, nous avons émis l'hypothèse que le lncRNA XLOC_087120 était capable de réguler négativement l'expression de gènes codants des histones par un mécanisme de régulation épigénétique impliquant le complexe PRC2. Cette régulation pourrait être en *cis* avec une régulation des gènes du cluster 1 des histones, mais aussi en *trans* pour une répression plus globale des histones H3.1, H4 et à moindre niveau des histones H2B.

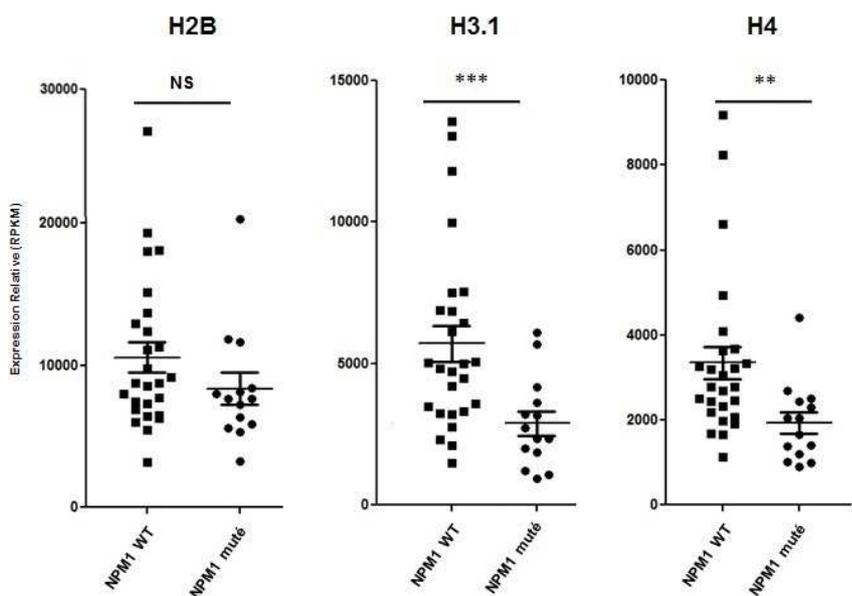


Figure 9 : Expression des histones H2B, H3.1 et H4 chez les patients LAMs. Différences d'expression en RPKM des types d'histones cœurs H2B, H3.1 et H4 entre les patients *NPM1* muté (n=14) et *NPM1* sauvage (n=26). NS= Non Significatif. ***= p valeurs<0.001 ; **= p valeurs<0.01.

v) Quel est l'impact de la dérégulation du XLOC_087120 sur l'expression des histones?

Afin de vérifier notre hypothèse de régulation de l'expression des histones par le lncRNA XLOC_087120, nous avons inhibé ou surexprimé ce lncRNA *in vitro*.

La 1^{ère} étape a été de caractériser ce lncRNA. En effet, il a été identifié et assemblé à partir de données de séquençage haut débit, et ce type d'analyse a déjà montré des faiblesses dans la fiabilité des séquences en 5' et 3' d'un transcrit. Afin de déterminer la séquence exacte de ce lncRNA, nous avons utilisé la technique de RACE PCR (Rapid Amplification of cDNA Ends). Cette approche nous a permis de définir les extrémités 5' et 3' du XLOC_087120 et ainsi de lui attribuer une taille de 1972nt. Ce lncRNA est constitué d'un seul exon et est transcrit à partir du brin - du chromosome 6. La 3'RACE a montré que ce lncRNA était polyadénylé, ce qui a été confirmé par une isolation ARN poly(A)+/poly(A)- montrant un enrichissement significatif de XLOC_087120 dans la fraction poly(A)+, suggérant donc une transcription par l'ARN pol II (Figure supplémentaire 1).

Pour réduire l'expression du XLOC_087120, nous avons dans un premier temps utilisé différents siARNs dirigés contre ce transcrit, qui se sont avérés inefficaces. Nous avons alors choisi d'utiliser la stratégie GapmeR, pour les avantages qu'elle présente par rapport aux siARNs dans le cas d'un ARN nucléaire. La lignée OCI AML3 (*NPM1* muté) a été transfectée transitoirement avec un

GapmeR dirigé contre XLOC_087120 ou un GapmeR contrôle. Cette approche nous a permis d'obtenir une réduction de l'expression du XLOC_087120 d'environ 40% par rapport au contrôle 48h post-transfection (Figure 10A). Par rapport aux histones, la transfection du GapmeR XLOC_087120 induit une augmentation statistiquement significative pour 6 des 9 histones (Figure 10B). Pour les histones H3F et H2BI, la différence n'est pas significative d'un point de vue statistique (liée à la variance entre les expériences), mais montre tout de même une augmentation de l'expression de ces histones en présence du GapmeR XLOC_087120. Seule l'expression de l'histone H3G ne semble pas être affectée par l'inhibition du XLOC_087120. Afin de déterminer l'impact de l'inhibition du XLOC_087120 sur l'expression globale des histones H3.1 et H4, nous avons utilisé des amorces de RT-qPCR capables d'amplifier l'ensemble des transcrits. Les variations de séquence entre les différents gènes codant les histones H2B n'ont pas permis d'utiliser un couple d'amorces spécifique unique. Ainsi, nous avons pu observer une surexpression globale d'au moins 1,5 fois des histones canoniques H3.1 et H4 après inhibition du XLOC_087120 (Figure 10C). Ces résultats suggèrent que le XLOC_087120 pourrait intervenir dans des mécanismes de régulation *in cis* sur les histones adjacentes, mais également *in trans* sur les autres histones afin de réguler l'expression globale des histones H3.1 et H4.

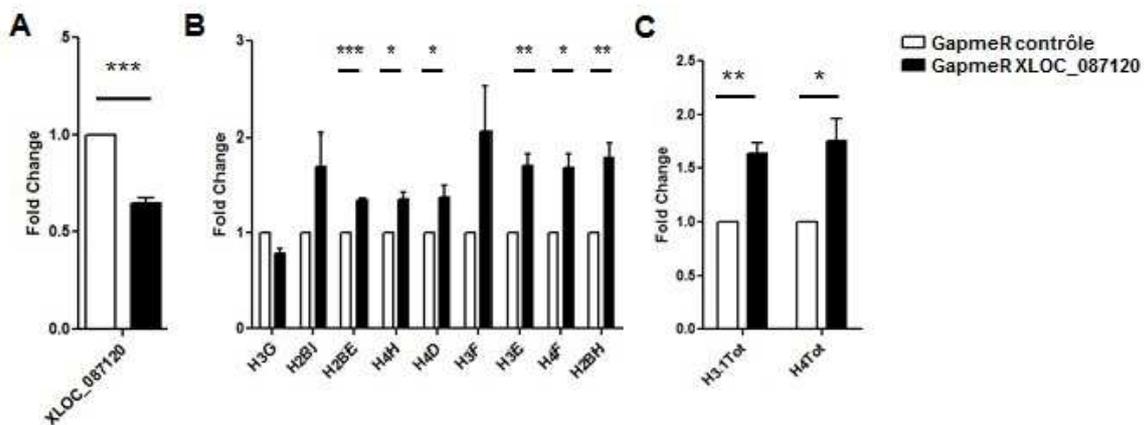


Figure 10 : Impact de l'inhibition de XLOC_087120 sur l'expression des histones. Différences d'expression de XLOC_087120 (A), des 9 histones proximales (B) et des types H3.1 et H4 (C), 48h post transfection des GapmeRs XLOC_087120 ou contrôle dans les OCI AML3, par RT-qPCR. n=3 ; ***= p valeurs < 0.001 ; **= p valeurs < 0.01 ; *= p valeurs < 0.05.

En résumé, nous avons montré que le lncRNA XLOC_087120 est capable de réprimer l'expression d'histones proximales de la famille H2B, H3.1 et H4, et de façon plus globale le niveau d'expression des histones cœurs H3.1 et H4.

VI) Est-ce que XLOC_087120 est capable de recruter PRC2?

Les analyses de prédictions d'interaction ont montré que XLOC_087120 pouvait potentiellement interagir avec le complexe PRC2 via la sous unité SUZ12, et pourrait ainsi recruter ce complexe afin de réprimer les histones comme nous l'avons mis en évidence. Nous avons réalisé des expériences d'immunoprécipitation d'ARN (RIP) avec un anticorps anti- SUZ12 selon le protocole décrit par Rinn *et al.* (Rinn *et al.*, 2007b). Le lncRNA *PINT* a été utilisé comme contrôle positif et l'ARNr *5S* en contrôle négatif. *PINT* est un lncRNA précédemment décrit comme interagissant avec SUZ12 (Marín-Béjar *et al.*, 2013), et l'ARNr *5S* nous permet également d'évaluer le bruit de fond. Par western blotting, nous observons bien la présence de SUZ12 dans la fraction immunoprécipitation de SUZ12 (IP SUZ12), et aucune bande dans la fraction contrôle (Figure 11A). La quantification des transcrits dans ces fractions par RT-qPCR montre effectivement un enrichissement du transcrite de *PINT* dans l'IP SUZ12 et aucun enrichissement de l'ARNr *5S* par rapport au contrôle (Figure 11B). De manière intéressante, on observe également un enrichissement de XLOC_087120 dans la fraction immunoprécipitation de SUZ12. Le lncRNA XLOC_087120 interagit donc avec le complexe PRC2, pour potentiellement réguler l'expression des gènes codant les histones.

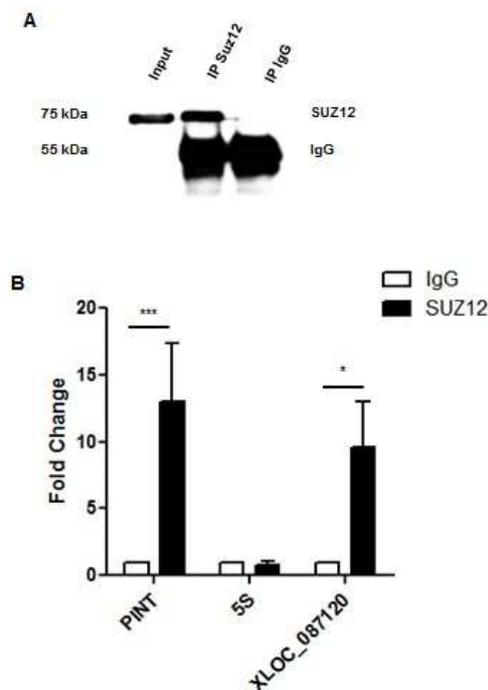


Figure 11 : Interaction entre XLOC_087120 et la sous unité SUZ12 de PRC2. A. Western Blot après RIP (RNA Immunoprecipitation) avec les anticorps anti-SUZ12 ou IgG contrôle. B. Enrichissement du transcrite XLOC_087120 dans la fraction IP de SUZ12 par rapport à la fraction contrôle par RT-qPCR. Le lncRNA *PINT* est utilisé comme contrôle positif et l'ARNr *5S* comme contrôle négatif de la RIP. n=3 ; ***=p valeurs<0.001 ; *=p valeurs<0.05.

VII) Quel pourrait être l'impact de *NPM1* mutée sur *XLOC_087120*?

Chez les patients, la surexpression de *XLOC_087120* est fortement associée à la présence de la mutation *NPM1*. Cependant, aucune variation du niveau d'expression de *XLOC_087120* n'a été mise en évidence dans nos modèles cellulaires d'inhibition ou de surexpression de *NPM1* muté. Sachant que le fractionnement cellulaire dans la lignée NB4 a montré que *XLOC_087120* était plus abondant dans le cytoplasme, et que la forme mutée de *NPM1* conduit à des défauts de localisation, nous avons supposé que la présence de *NPM1* mutée pourrait avoir un impact sur la localisation de *XLOC_087120*. La localisation du *XLOC_087120* a donc été comparée dans les différentes fractions entre la lignée NB4 *NPM1* non muté et la lignée OCI AML3 *NPM1* muté. Nous observons une localisation principalement cytoplasmique de *XLOC_087120* dans les cellules NB4 (Figure 12) et à l'inverse, *XLOC_087120* est préférentiellement et significativement localisé dans le noyau des cellules OCI AML3.

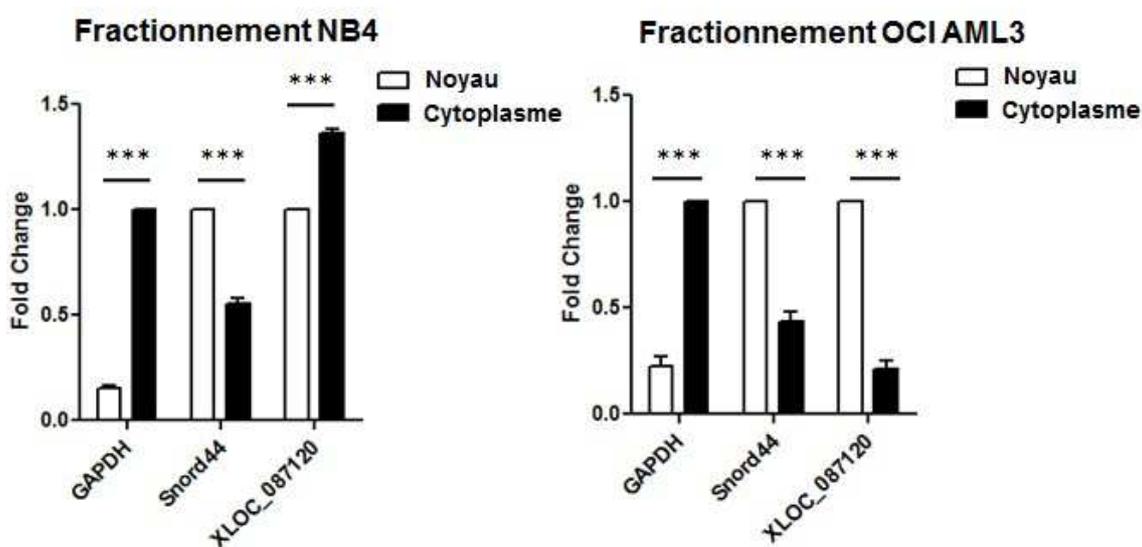


Figure 12 : Localisation cellulaire de *XLOC_087120* dans la lignée NB4 et dans la lignée OCI-AML3. A. Enrichissement de *XLOC_087120* après fractionnement noyau/cytoplasme de la lignée *NPM1* non muté NB4, par RT-qPCR. *GAPDH* est utilisé comme contrôle positif de la fraction cytoplasmique et *Snord44* comme contrôle positif de la fraction nucléaire. B. Enrichissement de *XLOC_087120* après fractionnement noyau/cytoplasme de la lignée *NPM1* muté OCI AML3, par RT-qPCR. *GAPDH* est utilisé comme contrôle positif de la fraction cytoplasmique et *Snord44* comme contrôle positif de la fraction nucléaire. Normalisation par rapport au noyau ; n=3 ; ***=p valeurs<0.001.

Afin de déterminer si c'est la présence de NPM1 mutée qui induit cette différence de localisation et non un autre des nombreux événements génétiques présents dans ces différentes lignées LAMs, nous avons analysé la localisation du XLOC_087120 dans la lignée HL60ER transduite avec NPM1 muté par rapport à la lignée HL60ER contrôle. Dans les HL60ER contrôles, le lncRNA XLOC_087120 est plus abondant dans le cytoplasme (Figure 13A). En revanche dans les HL60ER transduites avec NPM1 muté, on observe une délocalisation du XLOC_087120 du cytoplasme vers le noyau (Figure 13B). Ces résultats sont en accord avec ceux obtenus dans les lignées NB4 NPM1 non muté et OCI AML3 NPM1 muté. Ainsi, il semble que la présence de NPM1 mutée permette une relocalisation nucléaire de XLOC_087120, ce qui pourrait potentiellement augmenter sa fonction de répresseur des histones.

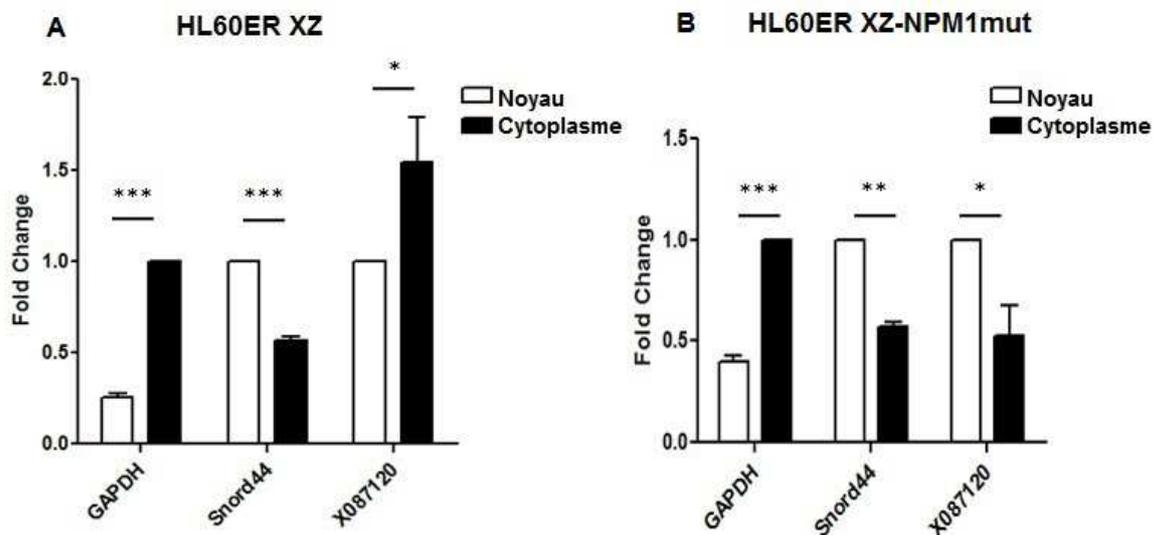


Figure 13 : Impact de la présence de la mutation NPM1 sur la localisation cellulaire de XLOC_087120. A. Enrichissement de XLOC_087120 après fractionnement noyau/cytoplasme de la lignée HL60ER infectée avec le plasmide XZ vide, par RT-qPCR. *GAPDH* est utilisé comme contrôle positif de la fraction cytoplasmique et *Snord44* comme contrôle positif de la fraction nucléaire. B. Enrichissement de XLOC_087120 après fractionnement noyau/cytoplasme de la lignée HL60ER infectée avec le plasmide XZ-NPM1mut, par RT-qPCR. *GAPDH* est utilisé comme contrôle positif de la fraction cytoplasmique et *Snord44* comme contrôle positif de la fraction nucléaire. Normalisation par rapport au noyau ; n=3 ; ***=p valeurs<0.001 ; **=p valeurs<0.01 ; *=p valeurs<0.05.

VIII) Quel est l'impact de la relocalisation de XLOC_087120 induite par NPM1 sur l'expression des histones?

Les différences de localisation de XLOC_087120 dans le noyau ou dans le cytoplasme peuvent potentiellement moduler sa fonction de régulateur de l'expression des histones. Nous avons analysé si le modèle HL60ER exprimant NPM1 mutée présente une différence d'expression des histones proximales à XLOC_087120 et des histones cœurs. Dans ce modèle, seules 5 histones sur 9 sont détectables par RT-qPCR. La présence de NPM1 mutée entraîne une légère diminution du niveau d'expression de ces histones (Figure 14). De plus, l'expression globale des histones H3.1 et H4 présente également une légère diminution, même si celle-ci n'est pas statistiquement significative.

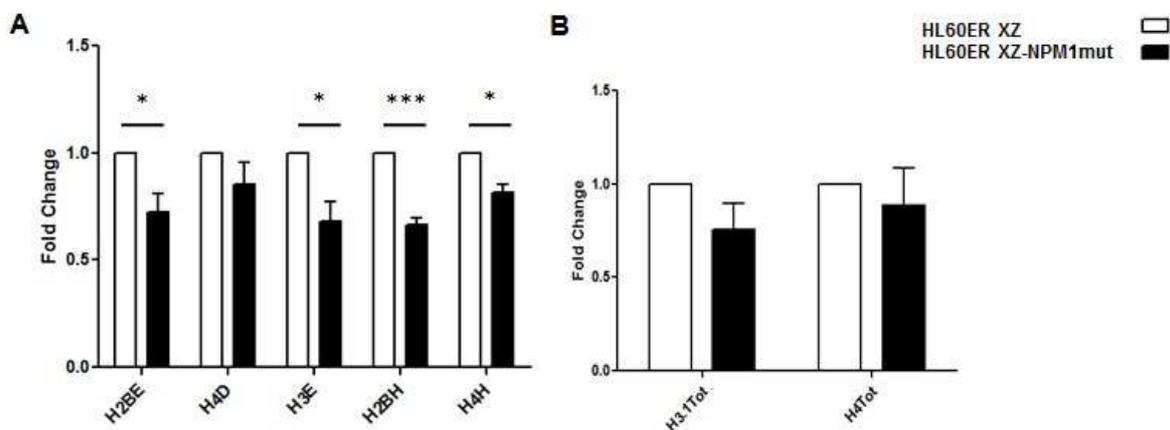


Figure 14 : Impact de la présence de la mutation *NPM1* sur l'expression des histones. Expression relative des histones localisées près de XLOC_087120 (A) et des histones cœurs H3.1 et H4 (B) entre la lignée HL60ER transduite avec XZ-NPM1mut et XZ vide, par RT-qPCR. n=3 ; ***= p valeurs<0.001 ; *= p valeurs<0.05.

En résumé, il semblerait que la mutation *NPM1* ait un impact directement ou indirectement sur l'expression de certains lncRNAs et plus particulièrement sur la localisation du lncRNA XLOC_087120. Ce lncRNA semble être un répresseur transcriptionnel des gènes codant les histones en *cis* et potentiellement en *trans*, pour une répression globale des histones cœurs H3.1 et H4 via le recrutement de PRC2. L'hypothèse qui se dégage de ces résultats est que NPM1 mutée dans les LAMs conduirait à une délocalisation du XLOC_087120 dans le noyau, par un mécanisme qui reste à déterminer. Ainsi, ce lncRNA pourrait alors recruter le complexe PRC2 pour exercer une répression transcriptionnelle sur les gènes codant les histones. L'impact cellulaire de cette dérégulation reste bien évidemment à déterminer, cependant ce travail ouvre un grand nombre de perspectives, que je décrirai dans la discussion.

*Discussion et
Perspectives*

Notre travail nous a permis de mettre en évidence de nouveaux lncRNAs non décrits à ce jour et des dérégulations de lncRNAs dans la leucémie aiguë myéloblastique à caryotype normal associées à la présence de la mutation du gène *NPM1*. De plus, nous avons identifié une signature de lncRNAs associée au pronostic des patients, confirmant donc l'intérêt de l'étude des lncRNAs dans le cancer et dans la leucémogénèse.

Sont-ils vraiment non codants?

La seconde partie de la définition d'un lncRNA est bien évidemment qu'il soit non codant. Le potentiel codant de ces ARNs est évalué selon différents critères. La recherche d'ORF est le premier élément pris en compte. Seuls les ARNs avec des ORF inférieures à 100 acides aminés et avec un contexte codant "pauvre" (déterminé informatiquement) sont sélectionnés. Ces ORF sont alors comparées aux banques de données protéiques (Swissprot, PFAM,...) par analyse d'homologie (BLASTX). Enfin, la méthode de CSF (Codon Substitution Frequency) permet d'évaluer si les codons sont conservés à travers l'évolution, indiquant ainsi la préservation du potentiel codant (Lin et al., 2007). La validation expérimentale est un élément important pour définir si un lncRNA est réellement non codant. Plusieurs approches sont utilisées: la traduction *in vitro*, l'analyse du profil de fixation des ribosomes ou la comparaison avec des données de spectrométries de masse. L'étude de Banfai *et al.* a ainsi permis de déterminer qu'environ 92% des lncRNAs annotés dans le catalogue GENCODE v7 n'étaient pas traduits en protéines en comparant les données de RNA-seq et de spectrométrie de masse sur deux lignées cellulaires (Bánfai et al., 2012). Cette approche reste cependant limitée pour les petites protéines de moins de 50 acides aminés. L'analyse de profil de fixation des ribosomes semble être la méthode la plus fiable à ce jour. Dans une première étude de profil de ribosome en 2011, l'équipe de Weissman avait démontré que la majorité des lncRNAs pouvait interagir avec les ribosomes, suggérant ainsi une traduction de ces lncRNAs (Ingolia et al., 2011). En 2013, cette équipe a déterminé un score d'efficacité de traduction à partir des mêmes données de profil de fixation des ribosomes et de RNA-seq. Les auteurs se sont aperçus que les régions 5'UTR et les différentes classes d'ARNs réellement non codants tels que les microRNAs avaient un score équivalent à ceux des lncRNAs. Ils ont alors analysé ces données différemment, en quantifiant le détachement des ribosomes (comparaison des données de ribosome profiling et de RNA-seq sur les putatives ORF). Cette approche a permis de montrer que les lncRNAs ainsi que les 3'UTR, 5'UTR et les ARNs non codants classiques avaient un profil différent de celui des ARNs codants. La quantification du détachement des ribosomes permettrait donc de distinguer les ARNs codants des ARNs non codants (Guttman et al., 2013). Pour finir, l'équipe de Weissman a mis en évidence en 2014 l'existence de petits peptides produits à partir de "séquences non codantes" par fixation des ribosomes en dehors des cadres ouverts de lecture classiques (Ingolia et al., 2014).

Ainsi, il convient de rester prudent sur le qualificatif "non codant" attribué aux lncRNAs nouvellement identifiés. En effet, certains soit disant lncRNAs ont finalement été décrits comme codant de petits peptides fonctionnels. Par exemple, Anderson *et al.* ont pu mettre en évidence un ARN (reconnu comme ne possédant pas d'ORF) codant le peptide MLN (Motilin). Ce peptide de 46 acides aminés est traduit à partir d'un "lncRNA" spécifique aux muscles squelettiques et est impliqué dans la performance de ces derniers (Anderson *et al.*, 2015). Il apparaît donc nécessaire de suivre les évolutions de la caractérisation du micro-protéome. Cependant, même si certains de ces "lncRNAs" sont capables de produire des petits peptides, cela n'exclut pas qu'ils puissent avoir un rôle fonctionnel en tant qu'ARN. Ces ARNs sont alors qualifiés d'ARNs bifonctionnels. A titre d'exemple, on peut citer l'ARNm de *p53* qui promeut sa propre traduction via la régulation négative de MDM2 (Candeias *et al.*, 2008). Le gène *SRA* est également bifonctionnel mais de manière un peu plus particulière. En effet, suivant l'épissage alternatif et l'isoforme qui en résulte, *SRA* peut être transcrit en un lncRNA fonctionnel ou en un ARN codant la protéine SRAP (Chooniedass-Kothari *et al.*, 2004). Ce mécanisme ajoute un degré de complexité supplémentaire du transcriptome et laisse supposer la présence de nombreux gènes potentiellement bifonctionnels par des mécanismes d'épissage alternatif (Ulveling *et al.*, 2011).

Identification de lncRNAs fonctionnels : le challenge

Le nombre important de nouveaux lncRNAs identifiés (8526 lncRNAs) dans cette étude peut paraître surprenant. Comme il s'agit de la première étude à grande échelle dans les LAMs et qu'il n'existe pas non plus d'étude à grande échelle dans la lignée myéloïde, on peut supposer que ces lncRNAs sont spécifiques de ce lignage ou de cette pathologie. Cette hypothèse est en accord avec les différentes études qui tendent à montrer que les lncRNAs sont exprimés de façon tissu spécifique (Cabili *et al.*, 2011). Une autre explication peut être cette fois technique. En effet la majorité des études par RNA-seq utilise uniquement les ARNs poly(A)⁺. Or, comme mentionné dans l'introduction, un certain nombre de lncRNAs ne possèdent pas de queue poly(A) tout en étant fonctionnels, tels que les eRNAs. Afin d'analyser de manière plus générale et précise le transcriptome des lncRNAs, il paraît primordial d'établir la sélection des ARNs pour le séquençage haut débit sur une base de longueur et non pas sur la présence/absence de queue poly(A). Ainsi, dans ce travail, nous avons choisi de réaliser le RNA-seq à partir des ARNs totaux (déplétés en ARNs ribosomiques) comprenant les ARNs poly(A)⁺ et poly(A)⁻. On peut donc supposer qu'une grande partie des ARNs mis en évidence dans notre étude sont poly(A)⁻. Cette hypothèse est en accord avec les résultats que nous avons obtenus après isolation des ARNs poly(A)⁻/poly(A)⁺ dans les cellules NB4. En effet, parmi les 33 lncRNAs associés à *NPM1* muté, la majorité semble être non polyadénylée (figure supplémentaire 1).

La structure des lncRNAs semble avoir un rôle important pour les interactions avec d'autres molécules, telles que des régulateurs de la chromatine, l'ADN ou les ARNs. *HOTAIR* par exemple possède des motifs structuraux précis et nécessaires pour son interaction avec les complexes LSD1 et PRC2. Déterminer les séquences en lien avec les fonctions pourrait permettre de cibler plus facilement les lncRNAs potentiellement fonctionnels à étudier. Cependant, cela reste un challenge important dans la caractérisation de ces molécules et de leur impact fonctionnel. Ainsi, de nouvelles méthodes ont été développées pour étudier la structure des ARNs comme la SHAPE (Selective 2'-Hydroxyl Acylation analyzed by Primer Extension) (Low and Weeks, 2010) ou la PARS (Parallel Analysis of RNA Structure) (Kertesz et al., 2010). La technique de SHAPE se fonde sur l'utilisation d'agents chimiques induisant préférentiellement l'acylation des acides nucléiques simple brin ou des régions flexibles des ARNs. Les sites de ces modifications sont détectés par la reverse transcription qui est bloquée par ce type de modifications. L'analyse de ces séquences, couplée à des algorithmes de prédiction, s'est révélée être d'une grande précision pour l'étude des structures des ARNs. La PARS quant à elle, combine les méthodes traditionnelles d'ARNs "footprinting" (à l'aide de la RNase V1 ou la nucléase S1) avec le séquençage haut débit pour identifier les régions simple ou double brin de plusieurs ARNs simultanément. Ainsi, cela permettrait d'analyser la structure des lncRNAs à grande échelle, de mettre en évidence des séquences consensus permettant l'interaction avec différents partenaires, et de déterminer les séquences associées aux différentes fonctions potentielles des lncRNAs.

RNA-seq et transcriptomique haut débit : attention aux artefacts

La technique de RNA-seq, bien qu'évidemment puissante pour l'identification de nouveaux transcrits et la quantification globale du transcriptome, génère un certain nombre de biais. La préparation des bibliothèques en premier lieu peut introduire des biais lors de la fragmentation des ARNs et de la transcription reverse, suivie des cycles d'amplification PCR des fragments. Cela peut conduire à la génération d'artefacts et à des erreurs de quantification. De plus, la séquence après l'assemblage de nouveaux transcrits peut être imprécise lorsque la couverture (nombre de reads alignés sur le génome) n'est pas suffisante. Cette identification de nouveaux transcrits nécessite donc des études complémentaires de type RACE et Northern Blot, afin de déterminer les séquences exactes. Corroborer ce type d'analyse avec des résultats de séquençage haut débit se basant sur la partie 5' des transcrits comme la méthode CAGE, pourrait également permettre d'avoir une idée plus précise des TSS des nouveaux transcrits identifiés. Dans notre signature, un certain nombre de lncRNAs sont situés dans un cluster que nous avons appelé cluster 028. Ces gènes sont très proches d'un point de vue génomique et leurs expressions sont très fortement positivement corrélées. Il pourrait s'agir d'un seul et même transcrit ou de portions de transcrits alternatifs, mal assemblés bioinformatiquement. En effet, la dérégulation induite par l'introduction de la mutation *NPM1* dans la lignée HL60ER montre une sous-expression de 3 lncRNAs du cluster 028, alors que les 3 autres ne semblent pas être affectés. Dans ce

contexte, on pourrait penser qu'il s'agisse en réalité de deux transcrits. Seules des expériences de Northern ou de RACE nous permettront de répondre à cette question.

Concernant la quantification des transcrits à partir de donnée RNA-seq, la normalisation actuelle utilise le RPKM ou FPKM (Fragments per kb per million), qui prend en compte à la fois la taille de la librairie et la taille du transcrit. Cependant, utiliser cette unité pour réaliser un différentiel d'expression entre deux conditions introduit un biais important. La puissance du test statistique, basée sur un test de student, prend en considération le facteur taille du gène et conduit ainsi à une p valeur plus petite pour un gène plus grand. C'est pourquoi différentes méthodes de normalisation, plus compliquées d'un point de vue mathématique et légèrement moins précises que les RPKM ou FPKM, ont été utilisées comme les méthodes développées par DESeq ou edgeR. Le choix de l'une ou l'autre de ces méthodes dépend majoritairement du type d'expérience, de la taille des librairies, de la cohorte, etc. Dans notre cas, les résultats du RNA-seq ont été particulièrement reproductibles. En effet, nous avons obtenu de très fortes corrélations entre les données de RNA-seq et de PCR quantitative par Fluidigm sur les mêmes échantillons. Sur les 33 lncRNA sélectionnés, seul deux lncRNAs n'ont pas pu être validés par Fluidigm. L'un d'eux n'a pas pu être quantifié pour des raisons techniques (amplifications aspécifiques malgré l'utilisation de différents couples d'amorces) et le deuxième lncRNA présentait une corrélation positive mais relativement faible entre les deux techniques. Ce résultat peut être lié aux différences techniques et méthodologiques entre le RNA-seq et la quantification relative par PCR. En effet, l'utilisation de gènes supposés de référence pour la normalisation des données de qPCR, ou encore la gamme de détection comprise entre le bruit de fond et la saturation du signal pour la mesure de fluorescence, peuvent introduire des biais de quantification et une limitation dans la quantification de transcrit faiblement exprimé.

Association NPM1 et lncRNAs?

L'analyse d'expression globale non supervisée du transcriptome des lncRNAs sur notre cohorte de patients LAM-CN nous a permis de mettre en avant un profil particulier d'expression de ces lncRNAs en présence de la mutation *NPM1*. Mais nous n'avons pas observé de profils d'expression particuliers associés à d'autres caractéristiques cliniques en analyses non supervisées, comme la mutation du gène *FLT3* ou le statut FAB par exemple. En revanche, la mutation *NPM1* présente un profil transcriptomique bien distinct avec 236 lncRNAs différentiellement exprimés selon nos critères stringents, confortant par ailleurs le statut d'entité à part entière des LAMs *NPM1* muté. La difficulté technique d'étude, ainsi que le risque de n'être qu'un artefact et non un transcrit réel, de certaines classes de lncRNAs (les lncRNAs introniques et dans les UTRs de gènes codants) nous a incité à ne garder que les lncRNAs intergéniques et anti-sens. Nous avons donc pu mettre en évidence une signature de 33 lncRNAs permettant de distinguer significativement les patients *NPM1* muté des patients *NPM1* non muté au sein des LAM-CN. De plus, nous avons noté de fortes corrélations entre

les lncRNAs et les mécanismes moléculaires impliquant NPM1, et entre les lncRNAs et l'expression de gènes spécifiquement associés à cette mutation, tels que les gènes *HOX*. Evidemment, la signature n'a pas d'intérêt d'un point de vue clinique, mais démontre bien que les LAMs présentant la mutation *NPM1* sont des entités en soi avec des profils transcriptomiques particuliers (ARNm, microARNs, lncRNAs), et que les lncRNAs pourraient participer au phénotype de ce type de cancer impliquant NPM1. La signature minimale permettant de regrouper les patients *NPM1* muté étant de 19 lncRNAs et non de un ou deux lncRNAs suggère que l'expression de ces lncRNAs n'est pas uniquement dépendante de la présence de la mutation *NPM1*. En effet, pris indépendamment, aucun des lncRNA de la signature n'est capable de discriminer les patients *NPM1* muté des patients *NPM1* non muté. Cette observation permet deux hypothèses: -soit la dérégulation du lncRNA est un événement secondaire indépendant de NPM1; -soit la présence d'un autre facteur (lié au stade de blocage de différenciation ou la présence d'une autre mutation: *DNMT3a*, *IDH1/2*...) perturbe l'expression du lncRNA. Nous n'avons pas pu mettre en évidence le facteur responsable de cette hétérogénéité, car malheureusement nous n'avons pas toutes les informations cliniques concernant le statut mutationnel pour tous les gènes mutés de façon récurrente dans les LAM-CN. Ces hypothèses pourraient également expliquer les différences entre les résultats obtenus dans les lignées (après inhibition ou surexpression de NPM1) et ce que l'on peut observer chez les patients.

Cette hétérogénéité d'expression des lncRNAs observés chez les patients *NPM1* muté nous a permis de mettre en évidence une signature de deux lncRNAs associée au pronostic des patients. De manière intéressante, cette signature est statistiquement indépendante du statut de *NPM1* et des autres facteurs pronostiques. Cela pourrait être la conséquence de l'hétérogénéité d'expression des lncRNAs et du faible nombre de patients sélectionnés pour les analyses statistiques. Cependant, l'indépendance de cette signature par rapport à *NPM1* pourrait permettre de stratifier un peu plus le risque dans cette pathologie, et plus particulièrement au sein des patients *NPM1* muté indépendamment du statut de *FLT3*. À la différence d'événements génétiques comme les mutations ou translocations représentant un état absolu, muté ou non, l'expression de gènes est soumise à la gradation de leurs expressions. Il est donc important que la gamme d'expression des gènes de la signature soit la plus large possible pour une classification des patients avec le moins de doute possible. La grande variabilité d'expression est un des avantages des lncRNAs qui pourrait faciliter la mise en évidence de signature pronostique forte. En effet, différentes études ont montré que les variations d'expressions entre différentes conditions (tissu sain/tissu malade, tissu indifférencié/tissu différencié, etc...) sont généralement plus importantes pour les lncRNAs que pour les ARNm, ce qui en fait potentiellement de bons marqueurs cliniques et pronostiques. Pour notre signature, nous pouvons observer des différences d'expression de plus de 100 fois entre les groupes de patients. Notre cohorte reste petite, et une confirmation sur une cohorte plus large et multicentrique permettrait de valider et d'évaluer plus précisément l'intérêt clinique de cette signature. Nous avons également remarqué par différentes analyses que la valeur pronostique associée

à l'expression des lncRNAs était régulièrement plus significative dans la rechute que sur la survie globale. De plus, notre signature se base sur des patients ayant atteint au moins une rémission complète, suggérant un impact prédictif principalement dans la rechute. De manière intéressante, il n'y a que très peu de marqueurs de rechute dans la LAM-CN. Une étude plus poussée dans ce sens pourrait permettre de mettre en avant des signatures de lncRNAs avec un fort pouvoir de prédiction de la rechute.

Pendant ma troisième année de thèse, une étude sur les lncRNAs dans les LAM-CN a été publiée par Garzon *et al.*, dans laquelle les auteurs ont identifié des signatures de lncRNAs associées à différentes mutations récurrentes dans les LAMs, dont *NPM1* (Garzon et al., 2014). Ils ont également montré une corrélation entre l'expression de lncRNA et le pronostic de patients LAM-CN en proposant une signature de 48 lncRNAs, validée sur une cohorte de 71 patients. Cette étude confirme l'importance des lncRNAs dans cette pathologie. Il existe cependant de nombreuses différences entre nos deux études. Tout d'abord, la technique de quantification des lncRNAs utilisée est différente puisqu'il s'agit de microarray dans l'étude de Garzon *et al.*. Cette technique ne permettant pas l'identification de nouveaux transcrits, les lncRNAs quantifiés sont ainsi uniquement des lncRNAs connus, répertoriés dans les banques de données. Nous n'avons donc pas de gènes en commun entre nos deux signatures *NPM1* muté, puisque la nôtre est composée de 31 lncRNAs nouvellement identifiés sur 33. Les cohortes de patients LAMs sont également différentes des nôtres puisqu'ils ont sélectionné uniquement des patients de plus de 65 ans. Une autre différence majeure entre nos deux études est que nous avons voulu tester la robustesse de chacune de nos signatures. En effet, notre démarche a été de sélectionner dans un premier temps les gènes différentiellement exprimés entre les patients *NPM1* muté et non muté puis de réaliser un clustering (analyse non supervisée) avec cette liste réduite de gènes. Cette approche nous a permis de valider la séparation des patients *NPM1* muté et non muté. Dans l'étude de Garzon, seul le différentiel d'expression a été réalisé. De plus, nous avons également utilisé une deuxième cohorte de patients (cohorte de validation, n=50) pour confirmer nos résultats. Ces conditions très stringentes sont par ailleurs une des raisons pour lesquelles nous n'avons pas établi de signature de lncRNAs associée à la mutation FLT3-ITD, au contraire de Garzon *et al.*.

Impact de NPM1 mutée sur les lncRNAs?

Expérimentalement, l'introduction de la mutation *NPM1* ou sa délétion ne perturbe pas globalement l'expression des lncRNAs que nous avons sélectionnés. On a pu observer tout de même des dérégulations de l'expression des lncRNAs XLOC_028763, XLOC_028770, XLOC_028776, XLOC_074912 et XLOC_091736 en introduisant la mutation dans la lignée HL60ER et XLOC_007987 et XLOC_080771 en déléant *NPM1* muté transitoirement dans la lignée OCI-AML3. Les différences d'observations entre ces expériences et les patients peuvent être dues à la nécessité d'événements secondaires, au contexte cellulaire et/ou des problèmes techniques. En effet, il existe des

différences inhérentes aux lignées cellulaires présentant des caryotypes complexes et de nombreuses mutations additionnelles, à la différence de nos patients aux caryotypes normaux. Le stade de blocage de différenciation, différent entre les cellules de patients *NPM1* muté et nos lignées, pourrait également avoir un impact sur la régulation des lncRNAs. En effet, les blastes de patient LAM *NPM1* muté sont plutôt différenciés (FAB M4), alors que les lignées LAMs que nous avons utilisées sont plutôt indifférenciées (FAB M2/M3). De plus, nos résultats se basent sur de l'étude *in vitro* qui s'affranchit de tout contexte physiologique tel que le microenvironnement tumoral. Enfin, d'un point de vue technique, l'introduction de la mutation ne compense peut être pas suffisamment le niveau endogène de *NPM1* sauvage. En effet, chez les patients, la mutation *NPM1* est toujours hétérozygote, ce qui conduit à une expression similaire des allèles, muté et sauvage. Or, lorsque nous introduisons *NPM1* muté dans nos lignées, nous observons également une légère augmentation du niveau d'expression de *NPM1* sauvage. A l'inverse, l'inhibition transitoire de *NPM1* muté par GapmeR induit également une diminution de *NPM1* sauvage. Ce dernier effet est dû à de trop fortes similarités de séquences entre *NPM1* muté et sauvage pour un ciblage entièrement spécifique. À la vue de l'importance du niveau d'expression de *NPM1* pour ses différentes fonctions, parfois opposées, nous pouvons supposer également que l'impact de la dérégulation de *NPM1* sauvage peut être à l'origine des dérégulations de lncRNAs en contradiction avec les résultats chez les patients. Bolli *et al.* ont montré que la délocalisation de *NPM1* mutée et de ses partenaires peut être reversée lorsque *NPM1* sauvage est transfecté et exprimé en excès, suggérant ainsi l'importance fonctionnelle de la balance d'expression *NPM1* muté/*NPM1* sauvage (Bolli et al., 2009). La surexpression et l'inhibition de la forme sauvage de *NPM1* pourraient nous permettre d'évaluer l'impact de *NPM1* dans la régulation de l'expression de ces lncRNAs.

Si l'on compare les résultats obtenus sur nos différentes lignées, nous pouvons constater qu'il n'y a pas de correspondance entre les stratégies d'inhibition et d'expression de *NPM1* mutée. En effet, les lncRNAs dérégulés dans ces deux types d'expériences ne sont pas les mêmes. On peut tout d'abord noter des différences d'expression entre les lignées: la lignée HL60ER exprime seulement 22 des 33 lncRNAs sélectionnés, la lignée NB4 26 lncRNAs et la lignée OCI-AML3 23 lncRNAs. D'un point de vue technique, la surexpression de *NPM1* mutée a été obtenue à l'aide de rétrovirus permettant une expression stable de la protéine, alors que pour l'inhibition de *NPM1* muté nous avons utilisé l'approche par GapmeR induisant une inhibition transitoire. L'inconvénient de la transfection transitoire est un temps d'observation court: 6 jours maximum post-transfection. Or, les LAMs sont des maladies avec une progression relativement lente et les études de la mutation *NPM1* chez la souris montrent une apparition tardive du phénotype leucémique. Nous devrions peut-être envisager une extinction stable de *NPM1* muté par sh-RNA ou par knock-out du gène permettant une observation plus tardive. En effet, Wang *et al.* ont transduit un sh-RNA dirigé contre *NPM1* muté dans la lignée OCI AML3, et de façon particulièrement surprenante, cette inhibition de *NPM1* muté par cette

méthode n'a présenté aucun effet sur la forme sauvage de *NPM1* (Wang et al., 2014c). Il en est de même pour nos expériences de surexpression de *NPM1* mutée pour lesquelles nous avons analysé les résultats seulement deux semaines après le tri des cellules GFP positives. Il est intéressant de noter toutefois que les résultats expérimentaux sont en accord avec les observations chez les patients pour les lncRNAs XLOC_074912, XLOC_091736 et XLOC_007987, suggérant un impact nécessaire et suffisant de la mutation *NPM1* sur l'expression de ces lncRNAs. Ces résultats nécessitent bien évidemment une validation par d'autres stratégies et sur d'autres lignées. L'utilisation d'inhibiteurs de l'oligomérisation de *NPM1* tel que NSC34888, diminuant donc ses fonctions, et de la Degueline, une drogue ciblant spécifiquement la forme mutée de *NPM1*, pourraient nous permettre également de mettre en avant l'impact de *NPM1* mutée sur la régulation de l'expression des lncRNAs (Balusu et al., 2011)(Yi et al., 2015b). La Deguelin est une drogue anti-tumorale qui a été décrite comme réprimant spécifiquement *NPM1* mutée en augmentant sa dégradation par le protéasome, mais qui n'a aucun effet sur *NPM1* sauvage. Les raisons expliquant cette différence d'impact entre *NPM1* sauvage et mutée ne sont pas encore connues. Cependant, il est important de prendre en considération que la Deguelin a des effets sur de nombreuses voies de signalisation impliquées dans les cancers comme PI3K/Akt, ce qui pourrait biaiser les résultats. Il serait également intéressant de transfecter transitoirement le GapmeR anti-*NPM1* muté dans les HL60ER XZ-*NPM1*mut afin d'observer si l'on restaure le niveau d'expression des lncRNAs. Dans le but de s'affranchir des différences entre nos modèles d'études et les patients, peut-être pourrions-nous également travailler directement sur les cellules de patients LAM *NPM1* muté. Toutefois, le maintien de ces cellules primaires en culture s'avère particulièrement difficile et limité dans le temps. Enfin, ces différences d'expressions peuvent être dues non pas à une régulation de l'expression, mais plutôt à une différence de stabilité des lncRNAs selon les conditions. D'après plusieurs études, *NPM1* n'aurait pas un rôle direct dans les LAMs sur la régulation de l'expression des ARNm associés à la présence de *NPM1* mutée comme les gènes *HOX*, mais plutôt un rôle indirect via la régulation de l'expression de microARNs ciblant des gènes *HOX* par exemple (De Braekeleer et al., 2014).

D'après nos analyses préliminaires par alignement de séquences des lncRNAs sélectionnés, très peu de ces lncRNAs présentent des séquences homologues chez la souris. Cependant, plusieurs chercheurs s'accordent à dire que les algorithmes pour l'analyse de conservation entre les espèces ne sont pas totalement adaptés pour des séquences évoluant aussi rapidement que celles des lncRNAs. En effet, la recherche de conservation ne doit pas seulement s'attacher à la conservation de séquences sur l'ensemble du transcrit. Il existe d'autres niveaux de conservation à considérer:

- une conservation de petites portions du transcrit responsables de la formation de structures secondaires ou tertiaires nécessaires à la fonctionnalité du lncRNA, comme dans le cas de *XIST*

- une conservation de la localisation génomique: les lncRNAs dont la séquence est différente mais qui sont localisés près des mêmes gènes codants dans différentes espèces
- une conservation de la fonction malgré une faible homologie de séquence entre les différentes espèces telle que *GAS5* (9% d'homologie entre l'homme et la souris mais producteur des mêmes snoARNs)

Ainsi, il serait intéressant de comparer nos résultats avec des résultats RNA-seq sur des souris knock-in *NPM1* muté. Ceci pourrait nous permettre de mettre en évidence des lncRNAs conservés entre les deux espèces et dérégulés par la présence de *NPM1* mutée. Pour finir, il paraît logique d'évaluer également l'impact de *NPM1* sur les lncRNAs à travers d'autres pathologies impliquant la nucléophosmine, comme les LAMs avec la fusion *NPM-MLF1* (Myeloid Leukemia Factor 1), ou les lymphomes anaplasiques à grandes cellules avec la fusion *NPM-ALK* (Anaplastic Lymphoma Kinase).

La mutation de *NPM1* semble également avoir un effet sur la localisation cellulaire des lncRNAs. Nous avons pu mettre en évidence une délocalisation du cytoplasme au noyau du lncRNA XLOC_087120 en présence de la mutation de *NPM1*, mais également une délocalisation du cytoplasme vers le noyau du lncRNA XLOC_091735 en présence de *NPM1* mutée (Cf Figure 5 Résultats supplémentaires et Figure 2 Illustrations supplémentaires). Or, *NPM1* est également une RNA-binding protéine et il a récemment été montré par RIP que *NPM1* était capable de lier le lncRNA Lnc_bc060912 (Luo et al., 2015a). On peut donc imaginer que *NPM1* mutée intervienne dans la délocalisation des lncRNAs dans le contexte des LAMs. Cependant, le sens de délocalisation des lncRNAs (cytoplasme vers noyau) peut apparaître contradictoire, puisque la mutation *NPM1* induit normalement une délocalisation de ses partenaires du noyau vers le cytoplasme. On peut donc supposer que *NPM1* mutée délocalise, inhibe ou diminue l'expression d'une protéine/d'un complexe protéique impliqué dans le transport des lncRNAs. L'utilisation de leptomycin B (un inhibiteur de l'exportine Crm1 impliqué dans le transport de *NPM1* mutée dans le cytoplasme) pourrait nous permettre d'évaluer l'impact de la délocalisation de *NPM1* sur la localisation cellulaire des lncRNAs. De plus, la recherche des partenaires protéiques de ces lncRNAs, par RNA biotin pull down ou par ChIRP couplé à la spectrométrie de masse, pourrait nous permettre d'identifier les acteurs impliqués dans cette délocalisation.

Quelle est la fonction de ces lncRNAs?

Que le génome soit transcrit en grande partie est un fait connu, mais que tous ces transcrits aient une fonction reste à déterminer. Les prédictions de fonctions par analyses bio-informatiques sont encore balbutiantes notamment par manque d'informations liant la séquence et/ou la structure aux fonctions. Une grande partie des fonctions des lncRNAs repose sur des interactions avec des

complexes protéiques. Cependant, la plupart des algorithmes de prédictions d'interactions ARN/protéines présentent de la cross-prédiction entre RNA et DNA-binding protein et sont limités d'un point de vue de la taille des séquences à analyser (Yan et al., 2015). Toutefois, nos prédictions d'interactions entre les protéines et les lncRNAs que nous avons sélectionnés ont généré un grand nombre de résultats intéressants, notamment avec le complexe PRC2 via SUZ12 et/ou EZH2.

Les lncRNAs peuvent également agir en tant qu'éponge à microARNs. Il existe des bases de données de prédictions d'interactions lncRNAs et microARNs. Cependant, ces bases de données ne répertorient que des lncRNAs déjà décrits et ne permettent pas l'utilisation de nouvelles séquences. De manière générale, même si les lncRNAs présentent des sites d'interactions avec des microARNs, il faut être prudent avant de les considérer comme des ceRNAs (competing RNA). Une étude publiée en 2014, traitant de l'effet compétiteur des lncRNAs sur le miR-122 et ses cibles, a conclu qu'il était peu probable que les lncRNAs présentant des sites d'interactions aux microARNs soient des compétiteurs endogènes efficaces (Denzler et al., 2014). En effet, la majorité des lncRNAs n'atteint pas le niveau d'expression nécessaire pour séquestrer les microARNs qui sont exprimés plus fortement. De plus, la plupart des lncRNAs ne contiennent pas de multiples sites d'interactions aux microARNs, une caractéristique nécessaire aux ceRNAs pour cibler de manière efficace les microARNs.

La méthode Guilt By Association permet d'avoir une vue d'ensemble des différentes voies moléculaires et cellulaires dans lesquelles un lncRNA peut être impliqué, mais ce type d'analyse repose sur des corrélations d'expressions qui ne sont en aucun cas synonyme de cause. Les expériences de perte ou gain de fonction sont donc nécessaires mais difficiles. D'une part parce que ces lncRNAs sont généralement peu conservés et peu exprimés, et d'autre part parce que les expériences d'ARN interférents sont difficilement efficaces, surtout pour des molécules associées à la chromatine. Bien que cela ne diminue en rien la fonction du gène proprement dite, le fait qu'un lncRNA puisse aussi être traduit en un petit peptide fonctionnel non présent à l'heure actuelle dans les banques de protéines, rend d'autant plus difficile la caractérisation de la fonction de ces ARNs "non codants" par ce type d'expériences.

Bien que nous nous soyons focalisés sur le lncRNA XLOC_087120, les études fonctionnelles d'autres lncRNAs de la signature pourraient donner des résultats intéressants. Par exemple, les deux lncRNAs XLOC_010909 et XLOC_028779 associés au pronostic des patients LAMs pourraient être impliqués dans des processus cellulaires comme la prolifération, l'apoptose, ou la résistance aux traitements. En effet, l'expression de XLOC_010909 est fortement corrélée à l'expression de gènes homéotiques dérégulés en présence de la mutation *NPM1*. Cependant, nous n'avons pas trouvé pour l'heure de lignée LAMs exprimant XLOC_010909, nous obligeant donc à caractériser et cloner ce lncRNA à partir d'échantillons provenant de patients. L'étude des 3 lncRNAs XLOC_007987, XLOC_074912 et XLOC_091736, qui semblent être dérégulés par la présence de la mutation *NPM1*

dans nos lignées, montre qu'ils pourraient également être impliqués dans le phénotype induit par cette mutation, comme le blocage de la différenciation et la réponse aux traitements dans les LAMs. L'inhibition ou la surexpression de ces lncRNAs au cours de la différenciation des lignées HL60ER et NB4, induite par le DMSO et l'acide rétinoïque respectivement et après traitement chimiothérapeutique à la cytarabine et/ou daunorubicine, pourraient répondre à ces questions. Le lncRNA XLOC_080771 est surexprimé chez les patients *NPM1* muté, et également fortement dérégulé dans nos expériences d'inhibition de *NPM1* mutée, ce qui en fait également un bon candidat d'étude. Ce lncRNA est situé à proximité du gène codant la MAP Kinase *MAP3K1* et son expression est positivement corrélée à l'expression de ce gène. De plus, XLOC_080771 est associé significativement au GoTerm MAPK dans les cancers dans nos analyses ontologiques, il pourrait donc avoir un impact notamment sur la prolifération.

Le lncRNA XLOC_087120 : régulateur négatif des histones?

Nos résultats sur le lncRNA XLOC_087120 sont prometteurs. En effet, nous avons pu mettre en évidence une répression induite par ce lncRNA sur l'expression de différents gènes d'histones à proximité de ce gène, dans le cluster 1 des histones, ainsi que sur l'expression globale des histones cœurs H3.1 et H4. Un certain nombre de questions restent en suspens. Tout d'abord, cette répression via le recrutement de PRC2 reste à confirmer. Nous avons pu montrer que XLOC_087120 interagissait avec la sous unité SUZ12 dans la lignée Hela par RIP. Ce résultat reste lui-aussi à valider, d'une part à partir d'au moins une lignée LAM, ce qui s'est avéré être plus difficile techniquement que dans les Hela: en effet, l'efficacité d'immunoprécipitation de SUZ12 est bien plus faible dans les lignées LAMs, associée par conséquent à une expression des lncRNAs plus difficilement détectable par RT-qPCR. D'autre part, la validation de cette interaction entre le lncRNA et PRC2 pourrait être réalisée par l'immunoprécipitation de l'ARN au lieu de la protéine. Typiquement, cette interaction peut être montrée par Biotin RNA pull down. Cette technique consiste à marquer l'ARN avec la biotine, immunoprécipiter le complexe ARN/protéine avec des billes de streptavidine, et analyser les protéines du complexe par western blot pour les protéines déjà suspectées, ou par spectrométrie de masse pour une identification globale. Cela nous permettrait de valider l'interaction de XLOC_087120 et PRC2 et de déterminer si cette interaction est réalisée avec la sous unité SUZ12 ou EZH2. En effet, la RIP que nous avons réalisée l'a été en conditions non dénaturantes, immunoprécipitant ainsi tout le complexe PRC2. De plus, l'identification de nouveaux partenaires pourrait indiquer d'autres fonctions potentielles, et par exemple nous permettre de déterminer pour quelles raisons XLOC_087120 est délocalisé dans le noyau en présence de la mutation *NPM1*.

Cependant, même si l'interaction avec PRC2 est confirmée, rien ne prouve que ce soit par ce biais que s'effectue la régulation sur les gènes des histones. Comme cité dans l'introduction, les lncRNAs peuvent recruter plusieurs complexes différents, répresseurs ou activateurs de transcription.

L'étude des marques des histones comme H3K27me3, spécifique de PRC2, par ChIP sur les promoteurs des gènes des histones réprimés, permettrait d'identifier le type de régulation et potentiellement le complexe de régulation épigénétique impliqué. Mais il est tout à fait possible aussi que cette répression de XLOC_087120 puisse être indirecte, en régulant l'expression d'un régulateur transcriptionnel des gènes codant les histones. La preuve pourrait être amenée par une expérience de ChIRP (Chromatin Isolation by RNA Purification). Cette technique permettrait d'identifier d'une part les gènes cibles, et d'autre part les partenaires protéiques du lncRNA. L'hypothèse d'une régulation des gènes des histones par le complexe PRC2 a été renforcée récemment par une étude de l'équipe d'Estelle Duprez sur les marques épigénétiques dans les LAMs. En effet, cette étude montre une augmentation des marques épigénétiques répressives spécifiques de PRC2 H3K27me3 dans le cluster 1 des histones chez les patients *NPM1* muté. Ces marques sont associées à une diminution de l'expression de différents gènes dont *HIST1H2BH*, *HIST1H3F* et *HIST1H4F* (Tiberi et al., 2015). De manière intéressante, ces 3 gènes font partie des 9 gènes que nous avons pu montrer comme régulés par XLOC_087120, supportant donc une régulation directe ou indirecte à travers le recrutement de PRC2.

Malgré une régulation qui semblerait proximale en *cis*, nous observons également une perturbation de l'expression des histones cœurs H3.1 et H4 en cas de modulation de l'expression du XLOC_087120. Ces résultats suggèrent un effet plus global du lncRNA sur la régulation de l'expression des histones. On observe notamment des corrélations significatives (négatives ou positives) avec la plupart des histones du cluster 1, cluster le plus important chez l'homme, constitué de 55 gènes d'histones. Mais nous avons également observé des corrélations avec d'autres histones cœurs comme H2B et même des histones plus "exotiques" telles que les histones CENPA et H2AFX, ainsi que des histones des autres clusters 2, 3 et 4, plus petits et sur d'autres chromosomes, suggérant donc un effet *trans* de cette régulation. La difficulté à dessiner des amorces spécifiques pour tous ces gènes très proches du point de vue de leurs séquences, rend difficile l'étude de leurs expressions par RT-qPCR. Il serait donc pertinent d'évaluer l'impact de la dérégulation de XLOC_087120 par RNA-seq après inhibition ou surexpression du lncRNA dans nos modèles. Cette approche, couplée à la ChIRP, nous permettrait également d'identifier de nouvelles cibles du XLOC_087120 autres que les gènes des histones. Il reste également à déterminer si cette régulation de XLOC_087120 sur les histones est spécifique et n'est pas la seule conséquence d'une proximité spatiale. Il apparaît donc comme nécessaire de valider d'une part la répression sur l'expression des histones, et si cet effet est spécifique du lncRNA d'autre part, en sur-exprimant XLOC_087120. Typiquement, nous envisageons une transduction du lncRNA dans des lignées LAMs après clonage du gène.

Impact physiologique de la dérégulation de XLOC_087120?

Nous n'avons pas encore évalué l'impact cellulaire de la dérégulation de XLOC_087120 et par conséquent l'impact de la dérégulation des histones dans les LAMs. En effet, les histones étant impliquées dans la condensation et la décondensation de la chromatine, nous pouvons nous attendre à des effets divers. D'un point de vue général, les histones sont plutôt considérées comme oncogéniques avec des surexpressions de certains types d'histones dans de nombreux cancers comme dans les cancers du côlon, du sein, de l'œsophage, ou encore des poumons (Hua et al., 2008)(Tomonaga et al., 2003)(Graber et al., 1996)(Wu et al., 2012b). Des mutations récurrentes dans des gènes codants des histones H3.1 et H3.3 ont également été retrouvées dans des glioblastomes de l'enfant (Wu et al., 2012a). Mais les régulations induites par les histones sont bien plus complexes qu'il n'y paraît. Elles peuvent dépendre du contexte cellulaire, et certaines histones se sont également avérées être des tumeurs suppresseurs par la régulation négative d'oncogènes (Kapoor et al., 2010).

Le plus évident des effets auxquels nous pourrions nous attendre pourrait être au niveau de la réplication de l'ADN et donc plus généralement dans la prolifération. Mais il pourrait aussi y avoir un effet sur la stabilité et l'intégrité des télomères, dans la régulation transcriptionnelle et épigénétique, la stabilité génomique et la ségrégation des chromosomes (Brickner et al., 2007)(Shia et al., 2006)(Goldberg et al., 2010). Certains gènes d'histones comme les histones H3.1 et H3.2 sont des histones de réplication, synthétisés et nécessaires au passage de la phase S du cycle cellulaire. L'établissement d'une lignée LAM exprimant stablement XLOC_087120 pourrait nous permettre de faire des études de cycle cellulaire. Cela nous permettrait notamment de quantifier le nombre de cellules dans les différentes phases du cycle en présence ou absence du lncRNA, par l'utilisation d'iodure de propidium en cytométrie en flux. L'impact sur la prolifération pourrait être évalué classiquement par l'introduction de CFSE dans les cellules et nous analyserons la diffusion du signal suite à la division cellulaire par cytométrie. La répression des histones induite par XLOC_087120 pourrait ainsi conduire à une diminution de la prolifération, ce qui pourrait être en accord avec le phénotype moins agressif des LAMs présentant la mutation *NPM1*.

Le deuxième effet pourrait être la réponse aux traitements, et plus généralement l'apoptose. La raison pour laquelle cette mutation dans les LAMs est associée à un meilleur pronostic et est plus sensible à la chimiothérapie n'est pas clairement établie : cet effet pourrait être la conséquence de la diminution du métabolisme des ribosomes, ou encore lié à un défaut de réparation de l'ADN. *NPM1* a été impliquée dans la maintenance de l'intégrité génomique et notamment dans la réparation de l'ADN. En effet, il a été montré que *NPM1* a une influence sur la TLS (Translesion DNA Synthesis), un processus de réparation faisant intervenir la polymérase pol η . Dans cette étude, les auteurs ont mis en évidence que *NPM1* régulait ce processus en diminuant la dégradation de cette polymérase. Dans le cas des leucémies avec mutation et délocalisation de *NPM1*, on observe une plus grande dégradation

de pol η et un défaut du mécanisme de TLS. Il pourrait en découler une plus grande instabilité génomique et donc une augmentation de la mort cellulaire après traitement (Ziv et al., 2014). De manière intéressante, certaines histones comme les histones de type H3.3 sont impliquées et nécessaires au bon fonctionnement du processus de réparation de l'ADN comme CENPA et H2AFX, recrutées aux niveaux des sites de lésions. On peut donc imaginer que la répression induite par XLOC_087120 rendrait le mécanisme de réparation moins efficace par son action sur les gènes des histones impliqués dans ce processus, augmentant le nombre de cellules en apoptose. L'implication de lncRNAs dans la réparation de l'ADN a déjà été décrite. Jiao *et al.* ont montré que le lncRNA *LIRRI* (Long intergenic radiation-responsive ncRNAs) était un régulateur de la réparation de l'ADN, induite par irradiations dans la lignée épithéliale bronchique humaine BEAS-2B (Jiao et al., 2015). Ce lncRNA agit notamment en diminuant l'expression de protéines impliquées dans la réparation. Nous testerons l'impact de la dérégulation du XLOC_087120 dans nos lignées sur l'apoptose par cytométrie en flux (marquage Annexin V/7-AAD), après traitement à la daunorubicine.

Enfin, les LAMs étant caractérisées par un blocage de la différenciation myéloïde, nous testerons l'effet de la dérégulation du XLOC_087120 dans ce processus. Pour cela, nous utiliserons les lignées NB4 et HL60ER après surexpression ou inactivation du XLOC_087120 et en présence ou non de *NPM1* muté. Le traitement avec l'acide rétinoïque de la lignée NB4 induit une différenciation granulocytaire. Le traitement avec le DMSO de la lignée HL60ER conduit à une différenciation monocyttaire. L'état de différenciation des cellules sera évalué par cytométrie en flux à l'aide des marqueurs de surfaces CD11b, CD15 et CD14, et par observation morphologique des cellules sur lames après coloration au May Grunwald Giemsa.

Il faut évidemment garder à l'esprit que les lncRNAs peuvent être bi- voir multi-fonctionnels. XLOC_087120 pourrait donc avoir d'autres fonctions et un impact cellulaire indépendamment de son rôle de régulateur d'histones.

Etudes in vivo de XLOC_087120?

Tous ces processus et ces expériences restent des études *in vitro*. L'évaluation de l'impact de XLOC_087120 *ex vivo* et *in vivo* est donc indispensable, afin de confirmer l'importance de la dérégulation de ce lncRNA observée chez les patients LAM-CN. En premier lieu, des études de lignées LAMs surexprimant ou inhibant XLOC_087120 xénotreffées chez la souris, pourraient nous permettre d'évaluer l'impact de la dérégulation de XLOC_087120 *in vivo*. Des études sur des modèles de knock-out ou knock-in murin de XLOC_087120 pourraient s'avérer plus compliquées à mettre en place, étant donné la faible conservation des lncRNAs qui est généralement observée entre l'homme et la souris.

Bien que les clusters d'histones et les différents gènes soient bien conservés dans l'évolution, nos analyses préliminaires ne nous ont pas permis de mettre en évidence des séquences homologues à XLOC_087120 chez la souris. Des expériences de knock-out murin paraissent donc compromises sans analyses plus poussées de la conservation du lncRNA chez la souris. Cependant, nous pourrions tout de même introduire la séquence de XLOC_087120 par recombinaison homologue dans le cluster des histones équivalent au cluster 1 des histones chez l'homme. En effet, nous savons que les mécanismes de régulation épigénétique sont conservés entre les deux espèces, PRC2 étant notamment conservé entre la souris et l'homme. Toutefois, il est nécessaire de prendre en compte que les structures permettant l'interaction entre le lncRNA et PRC2 ne sont peut-être pas les mêmes chez la souris, et que la régulation des histones ne dépend peut être pas de la triméthylation H3K27. Ainsi, il paraît judicieux de déterminer au préalable l'impact de XLOC_087120 humain sur l'expression des histones murines en transduisant des lignées cellulaires murines avec le lncRNA. Nous pourrions également envisager des études à partir de souris humanisées. En effet, nous souhaiterions transplanter des cellules humaines CD34⁺ transduites avec XLOC_087120 ou un sh-RNA dirigé contre le lncRNA dans des souris NSG (NOD/SCID/Gamma). Cela pourrait nous permettre d'observer l'impact du lncRNA sur l'hématopoïèse normale d'une part, son action potentielle sur le blocage de la différenciation et/ou l'augmentation de la prolifération intervenant dans le processus leucémique d'autre part.

Le lien entre NPM1 mutée et XLOC_087120?

Pour le moment, nous n'avons pas pu mettre en évidence *in vitro* un effet de la présence de NPM1 mutée sur le niveau d'expression du XLOC_087120. Il semblerait que la mutation *NPM1* agisse uniquement sur la localisation cellulaire du XLOC_087120. Or, chez les patients *NPM1* muté on observe une surexpression du XLOC_087120. On peut donc supposer que cette surexpression est un événement secondaire qui, associé à la présence de la mutation *NPM1*, entraîne une forte répression des histones. Il n'y a pour l'instant pas d'exemple dans la littérature de la régulation de l'expression des histones par NPM1, bien que NPM1 soit directement impliquée dans la régulation de l'assemblage des histones. Nous pouvons cependant faire un parallèle avec la régulation qu'exerce NPM1 sur la biogenèse des ribosomes. En effet, NPM1 est nécessaire à la maturation et à l'export des ribosomes, mais également à l'expression des ARNr. NPM1 peut s'associer aux promoteurs des ARNr, via des structures G-quadruplexes, et il a été montré que la déplétion de NPM1 ou l'expression d'un dominant négatif comme NPM3, conduit à l'inhibition de la transcription des ARNr (Gadad et al., 2010). Nous pouvons donc supposer que NPM1 présente ces deux niveaux de régulation sur les histones, à la fois en tant que chaperone, mais aussi en tant que régulateur de leur expression via le lncRNA XLOC_087120. Nous pourrions ainsi envisager la transplantation de cellules CD34⁺ humaines co-transduites avec le XLOC_087120 et *NPM1* muté dans les NSG (NOD/SCID/Gamma). Ce modèle de

souris humanisées nous permettrait de déterminer si la dérégulation du XLOC_087120 est un événement secondaire capable d'interférer dans le processus leucémique impliquant NPM1.

Conclusion

Au cours de ce travail de thèse, j'ai pu démontrer par des techniques de transcriptomique quantitative à grande échelle (RNA-seq et Fluidigm) que l'expression des lncRNAs était dérégulée dans la leucémie aiguë myéloblastique à caryotype normal, et plus particulièrement dans les LAM-CN présentant la mutation du gène *NPM1*. J'ai pu également démontrer que le XLOC_087120, un lncRNA nouvellement identifié et surexprimé chez les patients *NPM1* muté, intervenait dans la répression transcriptionnelle des gènes des histones probablement via le recrutement du complexe PRC2. Ainsi, ce travail s'inscrit dans l'idée récente que ces lncRNAs ont une véritable fonction physiologique et que leurs dérégulations ont des conséquences pathologiques. La caractérisation fonctionnelle de ces lncRNAs nous permettra d'approfondir nos connaissances fondamentales sur la fonction des lncRNAs, et aussi de mieux comprendre leurs rôles dans les processus oncogéniques. Ces lncRNAs pourraient rapidement constituer de nouvelles cibles thérapeutiques, et ainsi permettre d'affiner la classification et la stratification du risque dans les cancers. De plus, la suite de ce travail pourrait générer des perspectives intéressantes quant à la compréhension des mécanismes aussi fondamentaux que la régulation de l'expression des histones par des lncRNAs. Cela crée un champ d'investigations qui s'ouvre dans différents domaines, de la biologie fondamentale au développement et à toutes les pathologies impliquant des dérégulations d'expression, la réparation et le maintien de la stabilité génomique.

*Illustrations
supplémentaires*

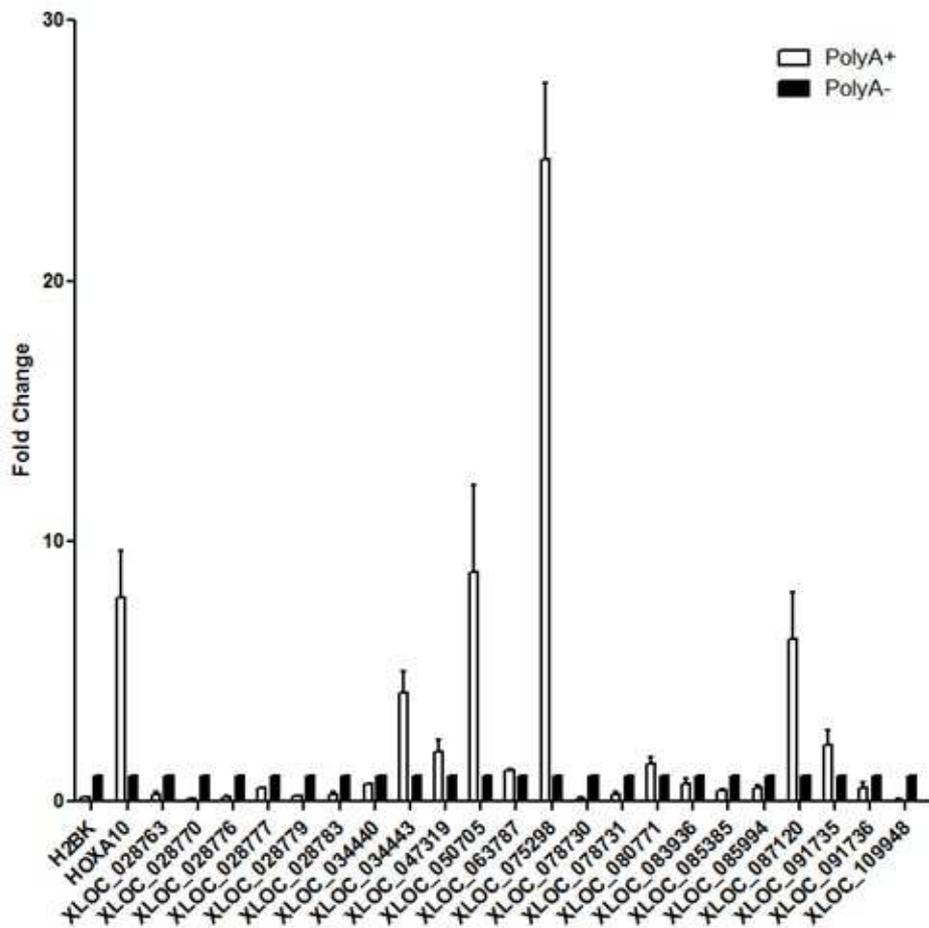


Figure supplémentaire 1: Polyadénylation des lncRNAs associés à la mutation *NPM1* dans les LAMs. Différences d'enrichissement des 22 lncRNAs détectables dans la lignée NB4 après fractionnement poly(A)+/poly(A)-. Le transcrit *H2BK* est utilisé comme contrôle négatif de la polyadénylation et le transcrit *HOXA10* comme contrôle positif. Normalisation par rapport à la fraction poly(A)- ; n=3.

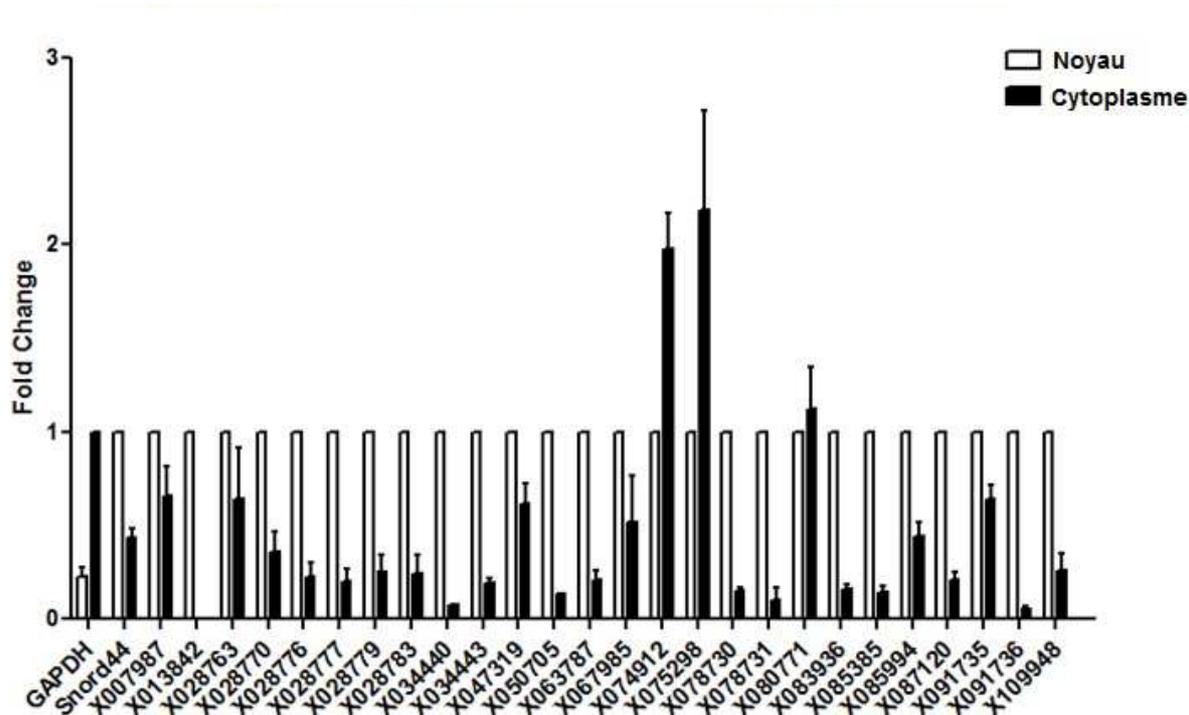


Figure supp 2: Enrichissement des lncRNAs dans le noyau et cytoplasme dans les OCI-AML3. Différences d'abondance de 27 lncRNAs associés à *NPM1* dans les LAMs détectables après fractionnement noyau/cytoplasme de la lignée OCI-AML3, par RT-qPCR. *GAPDH* est utilisé comme contrôle positif de la fraction cytoplasmique et le snoARN *Snord44* comme contrôle positif de la fraction nucléaire. Normalisation par rapport au noyau ; n=3.

Matériels et Méthodes

Prédictions d'interactions protéines/lncRNAs et analyse de GOTerm:

L'algorithme de prédiction catRAPID a été utilisé pour prédire les interactions potentielles entre les lncRNAs sélectionnés et des protéines. Cet algorithme est basé sur la contribution de structures secondaires prédites, les liaisons hydrogènes et van der Waals pour calculer un score d'interaction et la force d'interaction. Nous avons établi des seuils supérieurs à 75% pour la force d'interaction, 70% pour la puissance discriminative, et les partenaires protéiques potentiels ont été sélectionnés selon la propension d'interaction attribuée au Z score (<http://service.tartaglialab.com>) (Agostini et al., 2013). Les corrélations d'expression entre différents gènes ont été déterminées par le calcul du coefficient de Pearson sur Excel. Les gènes présentant un coefficient r de corrélation inférieur à -0,5 ou supérieur à 0,5 ont été sélectionnés et analysés en utilisant DAVID (Database for Annotation, Visualization and Integrated Discovery). La significativité des composants cellulaires et des voies moléculaires GO a été identifiée en se basant sur une FDR (False Discovery Rate) inférieure à 0,05.

Culture cellulaire:

Les lignées cellulaires HeLa et HEK 293T ont été cultivées respectivement en DMEM et IMDM (Sigma) et les lignées cellulaires LAMs HL60 ER (Ecotropic Receptor), NB4 et OCI AML3 ont été cultivées respectivement en IMDM et RPMI, dans une étuve à 37°C et 5% de CO². Chaque milieu a été additionné de 10% de SVF (Life Technologies), 100U/ml de pénicilline (Life Technologies) et 100µg/ml de streptomycine (Life Technologies).

Extraction ARN et synthèse d'ADNc:

Les ARNs totaux ont été extraits par Trizol (Ambion) selon les recommandations du fournisseur. La pureté et la concentration des ARNs ont été déterminées sur le Nanodrop 1000 (Thermo Fisher Scientific). L'intégrité des ARNs extraits de patients a été évaluée par un 2100 Bioanalyzer avec le kit RNA 6000 Nanon Chip (Agilent Technologies). Seuls les échantillons avec un RIN (RNA Integrity Number) > 7 ont été utilisés. De 100ng à 1µg d'ARNs ont été rétrotranscrits en ADNc avec le kit QuantiTect Reverse Transcription (Qiagen), avec un traitement DNase selon les conditions indiquées par le fournisseur.

Quantification par PCR quantitative (qPCR):

L'expression de chaque gène a été quantifiée par qPCR avec des amorces spécifiques déterminées grâce au logiciel en ligne *Primer3-PCR primer design tool*. La réaction de PCR a été réalisée avec le kit SYBR Premix Ex Taq (Takara) sur un thermocycleur StepOnePlus (Applied Biosystems) selon le protocole Fast standard recommandé par le fournisseur. Les gènes TBP (TATA Binding Protein) et Actine ont été utilisés pour la normalisation des données et l'expression a été calculée selon la méthode $2^{-\Delta Ct}$.

Fractionnement Cytoplasme/Noyau:

Entre 7.10^5 et 8.10^5 cellules ont été fractionnées avec le kit Cytoplasmic & Nuclear RNA Purification Kit (Norgen Biotek) selon les recommandations du kit. Brièvement, les cellules sont lysées une première fois et après centrifugation, le surnageant contenant la fraction cytoplasmique est récupéré. Le culot de noyaux est ensuite lysé vigoureusement, centrifugé et le surnageant contenant la fraction cytoplasmique nucléaire est récupéré. Les ARNs sont ensuite directement extraits à partir de chaque fraction selon le protocole du kit avec un rapport volume Cytoplasme/Noyau de 1:1. Les différences d'abondance entre les deux fractions ont ensuite été déterminées par RT-qPCR. L'ARNm de la GAPDH a été utilisé pour évaluer la contamination cytoplasmique dans le noyau et le snoARN SNORD44 a été utilisé pour évaluer la contamination nucléaire dans le cytoplasme.

Fractionnement poly(A)+/poly(A)- :

Les fractions poly(A)+ et poly(A)- ont été isolées à l'aide du kit polyA Spin mRNA Isolation (NEB). Les ARNs totaux extraits de cellules NB4 ont été incubés avec des billes de cellulose oligo(dT), pour isoler les ARNs poly(A)+ accrochés aux billes et les ARNs poly(A)- présents dans l'élution après incubation. Cette sélection a été réalisée deux fois de suite pour assurer la pureté des populations dans chaque fraction. Les différences entre les fractions ont ensuite été évaluées par RT-qPCR. L'ARNm de H2BK a servi de contrôle positif de la fraction poly(A)- et l'ARNm de HOXA10 a servi de contrôle positif pour la fraction poly(A)+.

RNA Immunoprécipitation (RIP):

La RIP a été réalisée selon les modalités précédemment décrites par Rinn *et al.* avec quelques modifications (Rinn et al., 2007b). Les cellules Hela ont été lysées dans un tampon RIPA (Tris-HCl 20 mM, pH 7.4, NaCl 150mM, EDTA 4mM, Triton X-100 1%) additionné d'inhibiteurs de phosphatase (Na₃VO₄ 1mM, NaF 1mM), de phénylméthylsulfonylfluoride (PMSF) et d'un cocktail d'inhibiteurs de protéase (Roche). Environ 5mg d'extrait protéique ont été ensuite incubés avec chaque anticorps (l'anticorps de la protéine à immunoprécipiter SUZ12 (ab12073 Abcam) et l'anticorps contrôle IgG (sc-2027 Santa Cruz)) durant 2h à 4°C. L'extrait protéique plus l'anticorps ont été ensuite incubés avec des billes A pendant 1h à 4°C. Après 4 lavages, les billes ont été centrifugées et suspendues dans le Trizol pour l'extraction ARN et une partie a été analysée par Western Blot.

Western Blot:

30µg d'extrait protéique ou 10% de billes A (RIP) sont fractionnés sur gel polyacrylamide SDS-PAGE (10% ou 15%) et transférés sur membrane de nitrocellulose (Whatman). Les western-blots ont été réalisés en utilisant les anticorps anti, SUZ12 (ab12073 Abcam), NPM1 mutant (fourni par le Pr Brunangelo Falini, Italie), NPM1 totale (ab52644 Abcam), Tubuline β (). Les protéines ont été

visualisées en utilisant des anticorps secondaires couplés à la peroxydase (Jackson ImmunoResearch) et le kit ECL Prime Western Blotting Detection Reagent (Amersham Biosciences).

5'/3' RACE (Rapid Amplification of cDNA Ends):

Les séquences terminales en 5' et en 3' du lncRNA XLOC_087120 ont été déterminées par l'utilisation du kit 5'/3' RACE 2nd Generation (Roche) selon les instructions du fournisseur. Brièvement, 1µg d'ARNs a été rétrotranscrit en ADNc à l'aide d'une amorce oligod(T) présentant une séquence particulière en queue flottante (fournie par le fabricant) pour la 3' RACE, et à l'aide d'une amorce spécifique du lncRNA pour la 5'RACE. Le fragment 3' est amplifié par PCR à l'aide d'une amorce spécifique du lncRNA et d'une amorce spécifique de la fin 3' fournie par le fabricant. Le produit est ensuite visualisé sur gel d'agarose 1% par UV et séquencé (GATC-Biotech). Pour la 5'RACE, l'ADNc est traité par une Terminal Transferase rajoutant une queue poly(A) en 3' (5' du transcrit). Cette queue va pouvoir servir de matrice pour une succession de PCR pseudo-nichées entre la queue poly(A) nouvellement ajoutée et des amorces spécifiques du lncRNAs. Le produit final sera par la suite séquencé pour déterminer la séquence exacte des extrémités du lncRNA XLOC_087120.

Clonages:

La séquence complète du lncRNA XLOC_087120 (déterminée par 3'/5'RACE) a été amplifiée par PCR à partir d'ADNc. Après purification, le fragment a été directement sous cloné dans le vecteur pCRTM-Blunt II-TOPO® (Life) et amplifié dans les bactéries DH5α (NEB). Les sites de restrictions EcoRI entourant le fragment dans le vecteur ont été digérés et le fragment a été inséré par ligation dans le vecteur rétroviral pMSCV PIG (Puro IRES GFP).

L'ADNc de *NPM1* muté (la forme majoritaire A) nous a été fourni par le Pr Brunangelo Falini (Italie). *NPM1* muté a ensuite été amplifié avec des amorces contenant les sites de restrictions pour XhoI et EcoRI, digéré et directement inséré par ligation dans le vecteur rétroviral XZ-IRES-GFP.

Les vecteurs produits ont tous été vérifiés par séquençage et amplifiés pour transfection transitoire ou infection.

Production virale et infection:

Les cellules HEK 293T ont été transfectées avec 15µg du vecteur pCL-Eco et 25µg du vecteur pMSCV PIG-XLOC_087120 ou XZ-NPM1muté, par nucléofection avec le kit TransIT-X2 (Mirus). Le milieu a ensuite été récupéré à 48h, 72h et 96h post-transfection. Après centrifugation, le surnageant contenant les virus a été filtré et congelé à -20°C en attendant l'infection.

Les cellules HL60ER présentent un récepteur écotropique murin pour l'infection par le virus produit précédemment. Trois millions de cellules HL60ER sont suspendues dans 3ml de surnageant, contenant

le rétrovirus d'intérêt ou contrôle supplémenté de 15µg de polybrène, et centrifugées à 2400rpm, à 37°C pendant 2h. Puis les cellules sont cultivées dans leur milieu toute la nuit et réinfectées le lendemain selon le même protocole. Enfin, les cellules GFP positives sont triées par cytométrie en flux 3 jours post-infection.

Transfection Transitoire:

Deux millions de cellules de la lignée OCI AML3 ou NB4 ont été transfectées avec 2µg de plasmide ou avec des LNA-GapmeRs dirigés contre XLOC_087120 (Exiqon), *NPM1* muté (Eurogentec), ou LNA-GapmeRs contrôles (concentration finale: 0,5µM). Le GapmeR contrôle de *NPM1* muté est un GapmeR ayant la même composition et n'ayant aucune cible. Le GapmeR contrôle du XLOC_087120 est un GapmeR "Negative control A" fourni par Exiqon et n'ayant aucune cible. La transfection est réalisée par électroporation avec le système Nucleofector II (Lonza) selon les recommandations du fournisseur, et à l'aide du programme X-001 et des solutions de nucléofection T et V, pour OCI AML3 et NB4, respectivement.

Analyses statistiques:

Les niveaux d'expression entre deux groupes ont été comparés par un test de Student non apparié. La correction de Welch a été appliquée en cas de variances significativement inégales entre les deux groupes d'après un test de Fisher (GraphPad Prism v.5). Une *p* valeur inférieure à 0,05 a été considérée comme significative et toutes les expériences ont été réalisées en triplica. Tous les graphiques ont été réalisés grâce au logiciel GraphPad Prism v.5.

Annexe

Gènes	Forward Primer 5'->3'	Reverse Primer 5'->3'
5S	ACGCGCCCGATCTCGTCTGAT	GCCTACAGCACCCGGTATTCCC
Actine	TCCCTGGAGAAGAGCTACGA	AGGAAGGAAGGCTGGAAGAG
GAPDH	CGGGAAGCTTGTCATCAATGG	GGCAGTGATGGCATGGACTG
NPM1 WT	GAAGAATTGCTTCCGGATGACT	CTTCCTCCACTGCCAGAGATCT
NPM1A	GAAGAATTGCTTCCGGATGACT	CTTCCTCCACTGCCAGACAGA
PINT	GCTGGAACCTTGCCCTGTCTCCT	TCGCTCTGTGCCACCTTCTTCT
Snord44	CCTGGATGATGATAAGCAAATGC	GTCAGTTAGAGCTAATTAAGAC
TBP	GCCTCCCCACCCCCTTCTTT	GCCACACCCTGCAACTCAACATCC
H3.1 Total	CTACCTGGTGGGGCTGTTTGAG	GCGAGCAAGCTGGATGTCCTT
H4 Total	ACAACATCCAGGGCATCACAA	CCGCGAGTCTCCTCGTAGATCA
HIST1H2BE	GTGACCAAGGCGCAGAAGAAGGAC	TTTAGAGGAGATGCCGGTGTGGGG
HIST1H2BH	GCGTAAACGCAGCCGCAAGG	GCCAGTTCCCCAGGCAGCAG
HIST1H2BI	GGGAGATCCAAACGGCTGTGCG	GAGCCTTTGGGTCGTTAGCGCTTT
HIST1H3E	GGACACCAACCTGTGCGCTATT	CAAAACAATTACGCCCTCTCC
HIST1H3F	CTATCCACGCCAAGCGAGTGAC	GGCGCACACCCTCAGTACAACCT
HIST1H3G	GTGCCATCCATGCTAAGCGAGT	TACTGCCCGGAAACCTCTACGC
HIST1H4D	GACGTGGTGTACGCGCTCAAG	CTGAAAAGGGCCGTTGGTTTTG
HIST1H4F	TGGATGTTGTCTACGCGCTCAA	GGGCCTTTTATGGGTGAGCTGT
HIST1H4H	AACGCAAGACCGTGACAGCAAT	AGTGCAGCAAGCAGGAGCCTTA
XLOC_087120	CCATGCACCAGGCACTGTAATG	TGGGTAGCACCTCTCCATGAGC
Gènes	GapmeR spécifique 5'->3'	GapmeR contrôle 5'->3'
XLOC_087120	GAGGTTAATGAGTCA	AACACGTCTATACGC
NPM1A	CTGCCAGACAGAGATC	ATTAGACTGCCACAGT
Gènes Clonage	Forward Primer 5'->3'	Reverse Primer 5'->3'
XLOC_087120	CCGCTCGAGCGGTTAGTGAGGGTCAT TCTTCCCAT	CCGGAATTCCGGTTTACCCACATCATGTTGTT TTTTAT
NPM1A	CCGCTCGAGCGGATGGAAGATTGAT GGACATGG	GAATTCCAGAAATGAAATAAGACG

Bibliographie

- Agostini, F., Zanzoni, A., Klus, P., Marchese, D., Cirillo, D., and Tartaglia, G.G. (2013). *catRAPID omics: a web server for large-scale prediction of protein-RNA interactions*. *Bioinformatics* 29, 2928–2930.
- Ahn, J.-Y., Liu, X., Cheng, D., Peng, J., Chan, P.-K., Wade, P.A., and Ye, K. (2005). *Nucleophosmin/B23, a nuclear PI(3,4,5)P(3) receptor, mediates the antiapoptotic actions of NGF by inhibiting CAD*. *Mol. Cell* 18, 435–445.
- Alvarez-Dominguez, J.R., Hu, W., Yuan, B., Shi, J., Park, S.S., Gromatzky, A.A., van Oudenaarden, A., and Lodish, H.F. (2014). *Global discovery of erythroid long noncoding RNAs reveals novel regulators of red cell maturation*. *Blood* 123, 570–581.
- Amit, D., and Hochberg, A. (2010). *Development of targeted therapy for bladder cancer mediated by a double promoter plasmid expressing diphtheria toxin under the control of H19 and IGF2-P4 regulatory sequences*. *J. Transl. Med.* 8, 134.
- Anderson, D.M., Anderson, K.M., Chang, C.-L., Makarewich, C.A., Nelson, B.R., McAnally, J.R., Kasaragod, P., Shelton, J.M., Liou, J., Bassel-Duby, R., et al. (2015). *A Micropeptide Encoded by a Putative Long Noncoding RNA Regulates Muscle Performance*. *Cell* 160, 595–606.
- Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., et al. (2014). *An atlas of active enhancers across human cell types and tissues*. *Nature* 507, 455–461.
- Andersson, S., Wallin, K.-L., Hellström, A.-C., Morrison, L.E., Hjerpe, A., Auer, G., Ried, T., Larsson, C., and Heselmeyer-Haddad, K. (2006). *Frequent gain of the human telomerase gene TERC at 3q26 in cervical adenocarcinomas*. *Br. J. Cancer* 95, 331–338.
- Angrand, P.-O., Vennin, C., Le Bourhis, X., and Adriaenssens, E. (2015). *The role of long non-coding RNAs in genome formatting and expression*. *Front. Genet.* 6, 165.
- Azzalin, C.M., Reichenbach, P., Khorianti, L., Giulotto, E., and Lingner, J. (2007). *Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends*. *Science* 318, 798–801.
- Balakirev, E.S., and Ayala, F.J. (2003). *Pseudogenes: are they “junk” or functional DNA?* *Annu. Rev. Genet.* 37, 123–151.
- Balusu, R., Fiskus, W., Rao, R., Chong, D.G., Nalluri, S., Mudunuru, U., Ma, H., Chen, L., Venkannagari, S., Ha, K., et al. (2011). *Targeting levels or oligomerization of nucleophosmin 1 induces differentiation and loss of survival of human AML cells with mutant NPM1*. *Blood* 118, 3096–3106.
- Bánfai, B., Jia, H., Khatun, J., Wood, E., Risk, B., Gundling, W.E., Kundaje, A., Gunawardena, H.P., Yu, Y., Xie, L., et al. (2012). *Long noncoding RNAs are rarely translated in two human cell lines*. *Genome Res.* 22, 1646–1657.
- Bañuelos, S., Hierro, A., Arizmendi, J.M., Montoya, G., Prado, A., and Muga, A. (2003). *Activation Mechanism of the Nuclear Chaperone Nucleoplasmin: Role of the Core Domain*. *J. Mol. Biol.* 334, 585–593.
- Barry, G., Briggs, J.A., Vanichkina, D.P., Poth, E.M., Beveridge, N.J., Ratnu, V.S., Nayler, S.P., Nones, K., Hu, J., Bredy, T.W., et al. (2014). *The long non-coding RNA Gomafu is acutely regulated in response to neuronal activation and involved in schizophrenia-associated alternative splicing*. *Mol. Psychiatry* 19, 486–494.
- Bartolomei, M.S., Zemel, S., and Tilghman, S.M. (1991). *Parental imprinting of the mouse H19 gene*. *Nature* 351, 153–155.
- Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W.J., Mattick, J.S., and Haussler, D. (2004). *Ultraconserved elements in the human genome*. *Science* 304, 1321–1325.
- Bekpen, C., Marques-Bonet, T., Alkan, C., Antonacci, F., Leogrande, M.B., Ventura, M., Kidd, J.M., Siswara, P.,

- Howard, J.C., and Eichler, E.E. (2009). Death and resurrection of the human IRGM gene. *PLoS Genet.* 5, e1000403.
- Belson, M., Kingsley, B., and Holmes, A. (2007). Risk factors for acute leukemia in children: a review. *Environ. Health Perspect.* 115, 138–145.
- Beltran, M., Puig, I., Peña, C., García, J.M., Alvarez, A.B., Peña, R., Bonilla, F., and de Herreros, A.G. (2008). A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial-mesenchymal transition. *Genes Dev.* 22, 756–769.
- Benetatos, L., Hatzimichael, E., Dasoula, A., Dranitsaris, G., Tsiara, S., Syrrou, M., Georgiou, I., and Bourantas, K.L. (2010). CpG methylation analysis of the MEG3 and SNRPN imprinted genes in acute myeloid leukemia and myelodysplastic syndromes. *Leuk. Res.* 34, 148–153.
- Bennett, J.M., Catovsky, D., Daniel, M.T., Flandrin, G., Galton, D.A., Gralnick, H.R., and Sultan, C. (1976). Proposals for the classification of the acute leukaemias. French-American-British (FAB) co-operative group. *Br. J. Haematol.* 33, 451–458.
- Berghoff, E.G., Clark, M.F., Chen, S., Cajigas, I., Leib, D.E., and Kohtz, J.D. (2013). Evg2 (Dlx6as) lncRNA regulates ultraconserved enhancer methylation and the differential transcriptional control of adjacent genes. *Development* 140, 4407–4416.
- Bergstralh, D.T., Conti, B.J., Moore, C.B., Brickey, W.J., Taxman, D.J., and Ting, J.P.-Y. (2007). Global functional analysis of nucleophosmin in Taxol response, cancer, chromatin regulation, and ribosomal DNA transcription. *Exp. Cell Res.* 313, 65–76.
- Bernard, D., Prasanth, K. V., Tripathi, V., Colasse, S., Nakamura, T., Xuan, Z., Zhang, M.Q., Sedel, F., Jourden, L., Couplier, F., et al. (2010). A long nuclear-retained non-coding RNA regulates synaptogenesis by modulating gene expression. *EMBO J.* 29, 3082–3093.
- Birney, E. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447, 799–816.
- Bischof, D., Pulford, K., Mason, D.Y., and Morris, S.W. (1997). Role of the nucleophosmin (NPM) portion of the non-Hodgkin's lymphoma-associated NPM-anaplastic lymphoma kinase fusion protein in oncogenesis. *Mol. Cell. Biol.* 17, 2312–2325.
- Blume, S.W., Meng, Z., Shrestha, K., Snyder, R.C., and Emanuel, P.D. (2003). The 5'-untranslated RNA of the human dhfr minor transcript alters transcription pre-initiation complex assembly at the major (core) promoter. *J. Cell. Biochem.* 88, 165–180.
- Blythe, A.J., Fox, A., and Bond, C.S. (2015). The ins and outs of lncRNA structure: How, why and what comes next? *Biochim. Biophys. Acta.*
- Boissel, N., Nibourel, O., Renneville, A., Gardin, C., Reman, O., Contentin, N., Bordessoule, D., Pautas, C., de Revel, T., Quesnel, B., et al. (2010). Prognostic impact of isocitrate dehydrogenase enzyme isoforms 1 and 2 mutations in acute myeloid leukemia: a study by the Acute Leukemia French Association group. *J. Clin. Oncol.* 28, 3717–3723.
- Bolli, N., De Marco, M.F., Martelli, M.P., Bigerna, B., Pucciarini, A., Rossi, R., Mannucci, R., Manes, N., Pettrossi, V., Pileri, S.A., et al. (2009). A dose-dependent tug of war involving the NPM1 leukaemic mutant, nucleophosmin, and ARF. *Leukemia* 23, 501–509.
- Bond, C.S., and Fox, A.H. (2009). Paraspeckles: nuclear bodies built on long noncoding RNA. *J. Cell Biol.* 186, 637–644.
- Boon, K., Caron, H.N., van Asperen, R., Valentijn, L., Hermus, M.-C., van Sluis, P., Roobeek, I., Weis, I., Voûte, P.A., Schwab, M., et al. (2001). N-myc enhances the expression of a large set of genes functioning in ribosome

biogenesis and protein synthesis. *EMBO J.* 20, 1383–1393.

Bousquet, M., Quelen, C., Rosati, R., Mansat-De Mas, V., La Starza, R., Bastard, C., Lippert, E., Talmant, P., Lafage-Pochitaloff, M., Leroux, D., et al. (2008). Myeloid cell differentiation arrest by miR-125b-1 in myelodysplastic syndrome and acute myeloid leukemia with the t(2;11)(p21;q23) translocation. *J. Exp. Med.* 205, 2499–2506.

Bradstock, K.F., Matthews, J.P., Lowenthal, R.M., Baxter, H., Catalano, J., Brighton, T., Gill, D., Eliadis, P., Joshua, D., Cannell, P., et al. (2005). A randomized trial of high-versus conventional-dose cytarabine in consolidation chemotherapy for adult de novo acute myeloid leukemia in first remission after induction therapy containing high-dose cytarabine. *Blood* 105, 481–488.

Brady, S.N., Yu, Y., Maggi, L.B., and Weber, J.D. (2004). ARF Impedes NPM/B23 Shuttling in an Mdm2-Sensitive Tumor Suppressor Pathway. *Mol. Cell. Biol.* 24, 9327–9338.

De Braekeleer, E., Douet-Guilbert, N., Basinko, A., Le Bris, M.-J., Morel, F., and De Braekeleer, M. (2014). *Hox* gene dysregulation in acute myeloid leukemia. *Future Oncol.* 10, 475–495.

Brannan, C.I., Dees, E.C., Ingram, R.S., and Tilghman, S.M. (1990). The product of the *H19* gene may function as an RNA. *Mol. Cell. Biol.* 10, 28–36.

Brickner, D.G., Cajigas, I., Fondufe-Mittendorf, Y., Ahmed, S., Lee, P.-C., Widom, J., and Brickner, J.H. (2007). H2A.Z-mediated localization of genes at the nuclear periphery confers epigenetic memory of previous transcriptional state. *PLoS Biol.* 5, e81.

Brown, C.J., Hendrich, B.D., Rupert, J.L., Lafrenière, R.G., Xing, Y., Lawrence, J., and Willard, H.F. (1992). The human *XIST* gene: Analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* 71, 527–542.

Bryan, T.M., Englezou, A., Dalla-Pozza, L., Dunham, M.A., and Reddel, R.R. (1997). Evidence for an alternative mechanism for maintaining telomere length in human tumors and tumor-derived cell lines. *Nat. Med.* 3, 1271–1274.

Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., and Rinn, J.L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* 25, 1915–1927.

Cai, X., and Cullen, B.R. (2007). The imprinted *H19* noncoding RNA is a primary microRNA precursor. *RNA* 13, 313–316.

Candeias, M.M., Malbert-Colas, L., Powell, D.J., Daskalogianni, C., Maslon, M.M., Naski, N., Bourougaa, K., Calvo, F., and Fähræus, R. (2008). P53 mRNA controls p53 activity by managing Mdm2 functions. *Nat. Cell Biol.* 10, 1098–1105.

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., et al. (2005). The transcriptional landscape of the mammalian genome. *Science* 309, 1559–1563.

Carramusa, L., Contino, F., Ferro, A., Minafra, L., Perconti, G., Giallongo, A., and Feo, S. (2007). The *PVT-1* oncogene is a Myc protein target that is overexpressed in transformed cells. *J. Cell. Physiol.* 213, 511–518.

Carrieri, C., Cimatti, L., Biagioli, M., Beugnet, A., Zucchelli, S., Fedele, S., Pesce, E., Ferrer, I., Collavin, L., Santoro, C., et al. (2012). Long non-coding antisense RNA controls *Uchl1* translation through an embedded *SINEB2* repeat. *Nature* 491, 454–457.

Cayre, A., Rossignol, F., Clottes, E., and Penault-Llorca, F. (2003). *aHIF* but not *HIF-1alpha* transcript is a poor prognostic marker in human breast cancer. *Breast Cancer Res.* 5, R223–R230.

Cesana, M., Cacchiarelli, D., Legnini, I., Santini, T., Sthandier, O., Chinappi, M., Tramontano, A., and Bozzoni,

- I. (2011). A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell* 147, 358–369.
- Challen, G.A., Sun, D., Jeong, M., Luo, M., Jelinek, J., Berg, J.S., Bock, C., Vasanthakumar, A., Gu, H., Xi, Y., et al. (2011). *Dnmt3a* is essential for hematopoietic stem cell differentiation. *Nat. Genet.* 44, 23–31.
- Chang, J.H., Lin, J.Y., Wu, M.H., and Yung, B.Y. (1998). Evidence for the ability of nucleophosmin/B23 to bind ATP. *Biochem. J.* 329 (Pt 3, 539–544.
- Chinen, Y., Sakamoto, N., Nagoshi, H., Taki, T., Maegawa, S., Tatekawa, S., Tsukamoto, T., Mizutani, S., Shimura, Y., Yamamoto-Sugitani, M., et al. (2014). 8q24 amplified segments involve novel fusion genes between *NSMCE2* and long noncoding RNAs in acute myelogenous leukemia. *J. Hematol. Oncol.* 7, 68.
- Chodroff, R.A., Goodstadt, L., Sirey, T.M., Oliver, P.L., Davies, K.E., Green, E.D., Molnár, Z., and Ponting, C.P. (2010). Long noncoding RNA genes: conservation of sequence and brain expression among diverse amniotes. *Genome Biol.* 11, R72.
- Chooniedass-Kothari, S., Emberley, E., Hamedani, M.K., Troup, S., Wang, X., Czosnek, A., Hube, F., Mutawe, M., Watson, P.H., and Leygue, E. (2004). The steroid receptor RNA activator is the first functional RNA encoding a protein. *FEBS Lett.* 566, 43–47.
- 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.
- Chu, C., Quinn, J., and Chang, H.Y. (2012). Chromatin isolation by RNA purification (ChIRP). *J. Vis. Exp.*
- Clark, M.B., Johnston, R.L., Inostroza-Ponta, M., Fox, A.H., Fortini, E., Moscato, P., Dinger, M.E., and Mattick, J.S. (2012). Genome-wide analysis of long noncoding RNA stability. *Genome Res.* 22, 885–898.
- Collier, S.P., Henderson, M.A., Tossberg, J.T., and Aune, T.M. (2014). Regulation of the *Th1* genomic locus from *Ifng* through *Tmevpg1* by T-bet. *J. Immunol.* 193, 3959–3965.
- Colombo, E., Bonetti, P., Lazzerini Denchi, E., Martinelli, P., Zamponi, R., Marine, J.-C., Helin, K., Falini, B., and Pelicci, P.G. (2005). Nucleophosmin is required for DNA integrity and p19Arf protein stability. *Mol. Cell Biol.* 25, 8874–8886.
- Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U. S. A.* 107, 21931–21936.
- Cullen, B.R. (2004). Transcription and processing of human microRNA precursors. *Mol. Cell* 16, 861–865.
- DeBaun, M.R., Niemitz, E.L., McNeil, D.E., Brandenburg, S.A., Lee, M.P., and Feinberg, A.P. (2002). Epigenetic alterations of *H19* and *LIT1* distinguish patients with Beckwith-Wiedemann syndrome with cancer and birth defects. *Am. J. Hum. Genet.* 70, 604–611.
- Delhommeau, F., Dupont, S., Della Valle, V., James, C., Trannoy, S., Massé, A., Kosmider, O., Le Couedic, J.-P., Robert, F., Alberdi, A., et al. (2009). Mutation in *TET2* in myeloid cancers. *N. Engl. J. Med.* 360, 2289–2301.
- Denzler, R., Agarwal, V., Stefano, J., Bartel, D.P., and Stoffel, M. (2014). Assessing the ceRNA Hypothesis with Quantitative Measurements of miRNA and Target Abundance. *Mol. Cell* 54, 766–776.
- Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D.G., et al. (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* 22, 1775–1789.
- Ding, L., Ley, T.J., Larson, D.E., Miller, C.A., Koboldt, D.C., Welch, J.S., Ritchey, J.K., Young, M.A., Lamprecht, T., McLellan, M.D., et al. (2012). Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* 481, 506–510.

- Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., et al. (2012). Landscape of transcription in human cells. *Nature* 489, 101–108.
- Döhner, H., Estey, E.H., Amadori, S., Appelbaum, F.R., Büchner, T., Burnett, A.K., Dombret, H., Fenaux, P., Grimwade, D., Larson, R.A., et al. (2010). Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet. *Blood* 115, 453–474.
- Dombret, H. (2011). Gene mutation and AML pathogenesis. *Blood* 118, 5366–5367.
- Durand, X., Moutereau, S., Xylinas, E., and de la Taille, A. (2011). ProgenesaTM PCA3 test for prostate cancer. *Expert Rev. Mol. Diagn.* 11, 137–144.
- Ebisuya, M., Yamamoto, T., Nakajima, M., and Nishida, E. (2008). Ripples from neighbouring transcription. *Nat. Cell Biol.* 10, 1106–1113.
- Edling, C.E., and Hallberg, B. (2007). c-Kit—a hematopoietic cell essential receptor tyrosine kinase. *Int. J. Biochem. Cell Biol.* 39, 1995–1998.
- Elisaphenko, E.A., Kolesnikov, N.N., Shevchenko, A.I., Rogozin, I.B., Nesterova, T.B., Brockdorff, N., and Zakian, S.M. (2008). A dual origin of the Xist gene from a protein-coding gene and a set of transposable elements. *PLoS One* 3, e2521.
- Emmrich, S., Streltsov, A., Schmidt, F., Thangapandi, V.R., Reinhardt, D., and Klusmann, J.-H. (2014). LincRNAs MONC and MIR100HG act as oncogenes in acute megakaryoblastic leukemia. *Mol. Cancer* 13, 171.
- Endo, A., Matsumoto, M., Inada, T., Yamamoto, A., Nakayama, K.I., Kitamura, N., and Komada, M. (2009). Nucleolar structure and function are regulated by the deubiquitylating enzyme USP36. *J. Cell Sci.* 122, 678–686.
- Engreitz, J.M., Sirokman, K., McDonel, P., Shishkin, A.A., Surka, C., Russell, P., Grossman, S.R., Chow, A.Y., Guttman, M., and Lander, E.S. (2014). RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent Pre-mRNAs and chromatin sites. *Cell* 159, 188–199.
- Erba, H.P. (2010). Has there been progress in the treatment of older patients with acute myeloid leukemia? *Best Pract. Res. Clin. Haematol.* 23, 495–501.
- Ernst, T., Chase, A.J., Score, J., Hidalgo-Curtis, C.E., Bryant, C., Jones, A. V, Waghorn, K., Zoi, K., Ross, F.M., Reiter, A., et al. (2010). Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. *Nat. Genet.* 42, 722–726.
- Estey, E., and Döhner, H. (2006). Acute myeloid leukaemia. *Lancet* 368, 1894–1907.
- Falini, B., Mecucci, C., Tiacci, E., Alcalay, M., Rosati, R., Pasqualucci, L., La Starza, R., Diverio, D., Colombo, E., Santucci, A., et al. (2005). Cytoplasmic nucleophosmin in acute myelogenous leukemia with a normal karyotype. *N. Engl. J. Med.* 352, 254–266.
- Falini, B., Bolli, N., Shan, J., Martelli, M.P., Liso, A., Pucciarini, A., Bigerna, B., Pasqualucci, L., Mannucci, R., Rosati, R., et al. (2006a). Both carboxy-terminus NES motif and mutated tryptophan(s) are crucial for aberrant nuclear export of nucleophosmin leukemic mutants in NPMc+ AML. *Blood* 107, 4514–4523.
- Falini, B., Bolli, N., Shan, J., Martelli, M.P., Liso, A., Pucciarini, A., Bigerna, B., Pasqualucci, L., Mannucci, R., Rosati, R., et al. (2006b). Both carboxy-terminus NES motif and mutated tryptophan(s) are crucial for aberrant nuclear export of nucleophosmin leukemic mutants in NPMc+ AML. *Blood* 107, 4514–4523.
- Falini, B., Nicoletti, I., Bolli, N., Martelli, M.P., Liso, A., Gorello, P., Mandelli, F., Mecucci, C., and Martelli, M.F. (2007). Translocations and mutations involving the nucleophosmin (NPM1) gene in lymphomas and leukemias. *Haematologica* 92, 519–532.

- Fatica, A., and Bozzoni, I. (2014). Long non-coding RNAs: new players in cell differentiation and development. *Nat. Rev. Genet.* 15, 7–21.
- Feng, J. (2006). The *Eyf-2* noncoding RNA is transcribed from the *Dlx-5/6* ultraconserved region and functions as a *Dlx-2* transcriptional coactivator. *Genes Dev.* 20, 1470–1484.
- Fernando, T.R., Rodriguez-Malave, N.I., Waters, E. V, Yan, W., Casero, D., Basso, G., Pigazzi, M., and Rao, D.S. (2015). *LncRNA* Expression Discriminates Karyotype and Predicts Survival in B-Lymphoblastic Leukemia. *Mol. Cancer Res.* 13, 839–851.
- Figuroa, M.E., Abdel-Wahab, O., Lu, C., Ward, P.S., Patel, J., Shih, A., Li, Y., Bhagwat, N., Vasanthakumar, A., Fernandez, H.F., et al. (2010). Leukemic *IDH1* and *IDH2* mutations result in a hypermethylation phenotype, disrupt *TET2* function, and impair hematopoietic differentiation. *Cancer Cell* 18, 553–567.
- Foltz, D.R., Jansen, L.E.T., Black, B.E., Bailey, A.O., Yates, J.R., and Cleveland, D.W. (2006). The human *CENP-A* centromeric nucleosome-associated complex. *Nat. Cell Biol.* 8, 458–469.
- Fu, X., Ravindranath, L., Tran, N., Petrovics, G., and Srivastava, S. (2006). Regulation of apoptosis by a prostate-specific and prostate cancer-associated noncoding gene, *PCGEM1*. *DNA Cell Biol.* 25, 135–141.
- Gadad, S.S., Shandilya, J., Kishore, A.H., and Kundu, T.K. (2010). *NPM3*, a member of the nucleophosmin/nucleoplamin family, enhances activator-dependent transcription. *Biochemistry* 49, 1355–1357.
- Gaidzik, V.I., Paschka, P., Spath, D., Habdank, M., Kohne, C.-H., Germing, U., von Lilienfeld-Toal, M., Held, G., Horst, H.-A., Haase, D., et al. (2012). *TET2* Mutations in Acute Myeloid Leukemia (AML): Results From a Comprehensive Genetic and Clinical Analysis of the AML Study Group. *J. Clin. Oncol.* 30, 1350–1357.
- Gao, Y., Wu, F., Zhou, J., Yan, L., Jurczak, M.J., Lee, H.-Y., Yang, L., Mueller, M., Zhou, X.-B., Dandolo, L., et al. (2014). The *H19/let-7* double-negative feedback loop contributes to glucose metabolism in muscle cells. *Nucleic Acids Res.* 42, 13799–13811.
- Garding, A., Bhattacharya, N., Claus, R., Ruppel, M., Tschuch, C., Filarsky, K., Idler, I., Zucknick, M., Caudron-Herger, M., Oakes, C., et al. (2013). Epigenetic upregulation of *lncRNAs* at *13q14.3* in leukemia is linked to the *In Cis* downregulation of a gene cluster that targets *NF-κB*. *PLoS Genet.* 9, e1003373.
- Garzon, R., Garofalo, M., Martelli, M.P., Briesewitz, R., Wang, L., Fernandez-Cymering, C., Volinia, S., Liu, C.-G., Schnittger, S., Haferlach, T., et al. (2008). Distinctive microRNA signature of acute myeloid leukemia bearing cytoplasmic mutated nucleophosmin. *Proc. Natl. Acad. Sci. U. S. A.* 105, 3945–3950.
- Garzon, R., Volinia, S., Papaioannou, D., Nicolet, D., Kohlschmidt, J., Yan, P.S., Mrózek, K., Bucci, D., Carroll, A.J., Baer, M.R., et al. (2014). Expression and prognostic impact of *lncRNAs* in acute myeloid leukemia. *Proc. Natl. Acad. Sci. U. S. A.* 111, 18679–18684.
- Gee, J.E., Blume, S., Snyder, R.C., Ray, R., and Miller, D.M. (1992). Triplex formation prevents *Sp1* binding to the dihydrofolate reductase promoter. *J. Biol. Chem.* 267, 11163–11167.
- Genovese, G., Kähler, A.K., Handsaker, R.E., Lindberg, J., Rose, S.A., Bakhoum, S.F., Chambert, K., Mick, E., Neale, B.M., Fromer, M., et al. (2014). Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N. Engl. J. Med.* 371, 2477–2487.
- Gilliland, D.G., and Griffin, J.D. (2002). The roles of *FLT3* in hematopoiesis and leukemia. *Blood* 100, 1532–1542.
- Godley, L.A., and Larson, R.A. (2008). Therapy-related myeloid leukemia. *Semin. Oncol.* 35, 418–429.
- Goldberg, A.D., Banaszynski, L.A., Noh, K.-M., Lewis, P.W., Elsaesser, S.J., Stadler, S., Dewell, S., Law, M., Guo, X., Li, X., et al. (2010). Distinct factors control histone variant *H3.3* localization at specific genomic regions. *Cell* 140, 678–691.

- Gong, C., and Maquat, L.E. (2011). lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. *Nature* 470, 284–288.
- Graber, M.W., Schweinfest, C.W., Reed, C.E., Papas, T.S., and Baron, P.L. (1996). Isolation of differentially expressed genes in carcinoma of the esophagus. *Ann. Surg. Oncol.* 3, 192–197.
- Grisendi, S., Bernardi, R., Rossi, M., Cheng, K., Khandker, L., Manova, K., and Pandolfi, P.P. (2005). Role of nucleophosmin in embryonic development and tumorigenesis. *Nature* 437, 147–153.
- Grisendi, S., Mecucci, C., Falini, B., and Pandolfi, P.P. (2006). Nucleophosmin and cancer. *Nat. Rev. Cancer* 6, 493–505.
- Gross, S., Cairns, R.A., Minden, M.D., Driggers, E.M., Bittinger, M.A., Jang, H.G., Sasaki, M., Jin, S., Schenkein, D.P., Su, S.M., et al. (2010). Cancer-associated metabolite 2-hydroxyglutarate accumulates in acute myelogenous leukemia with isocitrate dehydrogenase 1 and 2 mutations. *J. Exp. Med.* 207, 339–344.
- Guan, Y., Kuo, W.-L., Stilwell, J.L., Takano, H., Lapuk, A. V, Fridlyand, J., Mao, J.-H., Yu, M., Miller, M.A., Santos, J.L., et al. (2007). Amplification of PVT1 contributes to the pathophysiology of ovarian and breast cancer. *Clin. Cancer Res.* 13, 5745–5755.
- Guo, F., Li, Y., Liu, Y., Wang, J., Li, Y., and Li, G. (2010). Inhibition of metastasis-associated lung adenocarcinoma transcript 1 in CaSki human cervical cancer cells suppresses cell proliferation and invasion. *Acta Biochim. Biophys. Sin. (Shanghai)*. 42, 224–229.
- Guo, G., Kang, Q., Chen, Q., Chen, Z., Wang, J., Tan, L., and Chen, J.-L. (2014). High expression of long non-coding RNA H19 is required for efficient tumorigenesis induced by Bcr-Abl oncogene. *FEBS Lett.* 588, 1780–1786.
- Guo, G., Kang, Q., Zhu, X., Chen, Q., Wang, X., Chen, Y., Ouyang, J., Zhang, L., Tan, H., Chen, R., et al. (2015). A long noncoding RNA critically regulates Bcr-Abl-mediated cellular transformation by acting as a competitive endogenous RNA. *Oncogene* 34, 1768–1779.
- Gupta, R.A., Shah, N., Wang, K.C., Kim, J., Horlings, H.M., Wong, D.J., Tsai, M.-C., Hung, T., Argani, P., Rinn, J.L., et al. (2010). Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 464, 1071–1076.
- Gurumurthy, M., Tan, C.H., Ng, R., Zeiger, L., Lau, J., Lee, J., Dey, A., Philp, R., Li, Q., Lim, T.M., et al. (2008). Nucleophosmin interacts with HEXIM1 and regulates RNA polymerase II transcription. *J. Mol. Biol.* 378, 302–317.
- Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P., et al. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* 458, 223–227.
- Guttman, M., Donaghey, J., Carey, B.W., Garber, M., Grenier, J.K., Munson, G., Young, G., Lucas, A.B., Ach, R., Bruhn, L., et al. (2011). lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* 477, 295–300.
- Guttman, M., Russell, P., Ingolia, N.T., Weissman, J.S., and Lander, E.S. (2013). Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell* 154, 240–251.
- Hanahan, D., and Weinberg, R.A. (2000). The hallmarks of cancer. *Cell* 100, 57–70.
- Hann, I.M., Stevens, R.F., Goldstone, A.H., Rees, J.K., Wheatley, K., Gray, R.G., and Burnett, A.K. (1997). Randomized comparison of DAT versus ADE as induction chemotherapy in children and younger adults with acute myeloid leukemia. Results of the Medical Research Council's 10th AML trial (MRC AML10). Adult and Childhood Leukaemia Working Parties of the Medica. *Blood* 89, 2311–2318.

- Harrison, P.M., Zheng, D., Zhang, Z., Carriero, N., and Gerstein, M. (2005). Transcribed processed pseudogenes in the human genome: an intermediate form of expressed retrosequence lacking protein-coding ability. *Nucleic Acids Res.* 33, 2374–2383.
- Harrow, J., Denoeud, F., Frankish, A., Reymond, A., Chen, C.-K., Chrast, J., Lagarde, J., Gilbert, J.G.R., Storey, R., Swarbreck, D., et al. (2006). GENCODE: producing a reference annotation for ENCODE. *Genome Biol.* 7 Suppl 1, S4.1–9.
- Hasenpusch, G., Pfeifer, C., Aneja, M.K., Wagner, K., Reinhardt, D., Gilon, M., Ohana, P., Hochberg, A., and Rudolph, C. (2011). Aerosolized BC-819 inhibits primary but not secondary lung cancer growth. *PLoS One* 6, e20760.
- Havens, M.A., Reich, A.A., Duelli, D.M., and Hastings, M.L. (2012). Biogenesis of mammalian microRNAs by a non-canonical processing pathway. *Nucleic Acids Res.* 40, 4626–4640.
- Haywood, M.E.K., Rose, S.J., Horswell, S., Lees, M.J., Fu, G., Walport, M.J., and Morley, B.J. (2006). Overlapping BXS congenic intervals, in combination with microarray gene expression, reveal novel lupus candidate genes. *Genes Immun.* 7, 250–263.
- He, H.H., Meyer, C.A., Shin, H., Bailey, S.T., Wei, G., Wang, Q., Zhang, Y., Xu, K., Ni, M., Lupien, M., et al. (2010). Nucleosome dynamics define transcriptional enhancers. *Nat. Genet.* 42, 343–347.
- van Heesch, S., van Iterson, M., Jacobi, J., Boymans, S., Essers, P.B., de Bruijn, E., Hao, W., Macinnes, A.W., Cuppen, E., and Simonis, M. (2014). Extensive localization of long noncoding RNAs to the cytosol and mono- and polyribosomal complexes. *Genome Biol.* 15, R6.
- Higashimoto, K., Soejima, H., Saito, T., Okumura, K., and Mukai, T. (2006). Imprinting disruption of the CDKN1C/KCNQ1OT1 domain: the molecular mechanisms causing Beckwith-Wiedemann syndrome and cancer. *Cytogenet. Genome Res.* 113, 306–312.
- Hingorani, K., Szebeni, A., and Olson, M.O. (2000). Mapping the functional domains of nucleolar protein B23. *J. Biol. Chem.* 275, 24451–24457.
- Holmberg Olausson, K., Nistér, M., and Lindström, M.S. (2014). Loss of nucleolar histone chaperone NPM1 triggers rearrangement of heterochromatin and synergizes with a deficiency in DNA methyltransferase DNMT3A to drive ribosomal DNA transcription. *J. Biol. Chem.* 289, 34601–34619.
- Hsu, C.Y., and Yung, B.Y. (2000). Over-expression of nucleophosmin/B23 decreases the susceptibility of human leukemia HL-60 cells to retinoic acid-induced differentiation and apoptosis. *Int. J. Cancer* 88, 392–400.
- Hu, W., Yuan, B., Flygare, J., and Lodish, H.F. (2011). Long noncoding RNA-mediated anti-apoptotic activity in murine erythroid terminal differentiation. *Genes Dev.* 25, 2573–2578.
- Hua, S., Kallen, C.B., Dhar, R., Baquero, M.T., Mason, C.E., Russell, B.A., Shah, P.K., Liu, J., Khramtsov, A., Tretiakova, M.S., et al. (2008). Genomic analysis of estrogen cascade reveals histone variant H2A.Z associated with breast cancer progression. *Mol. Syst. Biol.* 4, 188.
- Huang, J., Zhou, N., Watabe, K., Lu, Z., Wu, F., Xu, M., and Mo, Y.-Y. (2014). Long non-coding RNA UCA1 promotes breast tumor growth by suppression of p27 (Kip1). *Cell Death Dis.* 5, e1008.
- Huang, K.-C., Rao, P.H., Lau, C.C., Heard, E., Ng, S.-K., Brown, C., Mok, S.C., Berkowitz, R.S., and Ng, S.-W. (2002). Relationship of XIST expression and responses of ovarian cancer to chemotherapy. *Mol. Cancer Ther.* 1, 769–776.
- Huang, M., Chen, W., Qi, F., Xia, R., Sun, M., Xu, T., Yin, L., Zhang, E., De, W., and Shu, Y. (2015). Long non-coding RNA ANRIL is upregulated in hepatocellular carcinoma and regulates cell apoptosis by epigenetic silencing of KLF2. *J. Hematol. Oncol.* 8, 50.

- Huarte, M., Guttman, M., Feldser, D., Garber, M., Koziol, M.J., Kenzelmann-Broz, D., Khalil, A.M., Zuk, O., Amit, I., Rabani, M., et al. (2010). A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* 142, 409–419.
- Hughes, J.M., Legnini, I., Salvatori, B., Masciarelli, S., Marchioni, M., Fazi, F., Morlando, M., Bozzoni, I., and Fatica, A. (2015). C/EBP α -p30 protein induces expression of the oncogenic long non-coding RNA UCA1 in acute myeloid leukemia. *Oncotarget*.
- Hung, T., Wang, Y., Lin, M.F., Koegel, A.K., Kotake, Y., Grant, G.D., Horlings, H.M., Shah, N., Umbricht, C., Wang, P., et al. (2011). Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nat. Genet.* 43, 621–629.
- Hutchinson, J.N., Ensminger, A.W., Clemson, C.M., Lynch, C.R., Lawrence, J.B., and Chess, A. (2007). A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. *BMC Genomics* 8, 39.
- Imamura, T., Yamamoto, S., Ohgane, J., Hattori, N., Tanaka, S., and Shiota, K. (2004). Non-coding RNA directed DNA demethylation of Sphk1 CpG island. *Biochem. Biophys. Res. Commun.* 322, 593–600.
- Ingolia, N.T., Lareau, L.F., and Weissman, J.S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* 147, 789–802.
- Ingolia, N.T., Brar, G.A., Stern-Ginossar, N., Harris, M.S., Talhouarne, G.J.S., Jackson, S.E., Wills, M.R., and Weissman, J.S. (2014). Ribosome Profiling Reveals Pervasive Translation Outside of Annotated Protein-Coding Genes. *Cell Rep.* 8, 1365–1379.
- Itahana, K., Bhat, K.P., Jin, A., Itahana, Y., Hawke, D., Kobayashi, R., and Zhang, Y. (2003). Tumor suppressor ARF degrades B23, a nucleolar protein involved in ribosome biogenesis and cell proliferation. *Mol. Cell* 12, 1151–1164.
- Iyer, M.K., Niknafs, Y.S., Malik, R., Singhal, U., Sahu, A., Hosono, Y., Barrette, T.R., Prensner, J.R., Evans, J.R., Zhao, S., et al. (2015). The landscape of long noncoding RNAs in the human transcriptome. *Nat. Genet.* 47, 199–208.
- Jacob, F., and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* 3, 318–356.
- Jacob, M.D., Audas, T.E., Mullineux, S.-T., and Lee, S. (2012). Where no RNA polymerase has gone before: novel functional transcripts derived from the ribosomal intergenic spacer. *Nucleus* 3, 315–319.
- Jalali, S., Jayaraj, G.G., and Scaria, V. (2012). Integrative transcriptome analysis suggest processing of a subset of long non-coding RNAs to small RNAs. *Biol. Direct* 7, 25.
- Jan, M., Snyder, T.M., Corces-Zimmerman, M.R., Vyas, P., Weissman, I.L., Quake, S.R., and Majeti, R. (2012). Clonal evolution of preleukemic hematopoietic stem cells precedes human acute myeloid leukemia. *Sci. Transl. Med.* 4, 149ra118.
- Jeon, Y., and Lee, J.T. (2011). YY1 tethers Xist RNA to the inactive X nucleation center. *Cell* 146, 119–133.
- Ji, P., Diederichs, S., Wang, W., Böing, S., Metzger, R., Schneider, P.M., Tidow, N., Brandt, B., Buerger, H., Bulk, E., et al. (2003a). MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene* 22, 8031–8041.
- Ji, P., Diederichs, S., Wang, W., Böing, S., Metzger, R., Schneider, P.M., Tidow, N., Brandt, B., Buerger, H., Bulk, E., et al. (2003b). MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene* 22, 8031–8041.
- Jiang, P.S., and Yung, B.Y. (1999). Down-regulation of nucleophosmin/B23 mRNA delays the entry of cells into

mitosis. *Biochem. Biophys. Res. Commun.* 257, 865–870.

Jiao, Y., Liu, C., Cui, F.-M., Xu, J.-Y., Tong, J., Qi, X.-F., Wang, L.-L., and Zhu, W. (2015). Long intergenic non-coding RNA induced by X-ray irradiation regulates DNA damage response signaling in the human bronchial epithelial BEAS-2B cell line. *Oncol. Lett.* 9, 169–176.

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., et al. (2010). Mediator and cohesin connect gene expression and chromatin architecture. *Nature* 467, 430–435.

Kaikkonen, M.U., Lam, M.T.Y., and Glass, C.K. (2011). Non-coding RNAs as regulators of gene expression and epigenetics. *Cardiovasc. Res.* 90, 430–440.

Kallen, A.N., Zhou, X.-B., Xu, J., Qiao, C., Ma, J., Yan, L., Lu, L., Liu, C., Yi, J.-S., Zhang, H., et al. (2013). The Imprinted H19 LncRNA Antagonizes Let-7 MicroRNAs. *Mol. Cell* 52, 101–112.

Kam, Y., Rubinstein, A., Naik, S., Djavsarov, I., Halle, D., Ariel, I., Gure, A.O., Stojadinovic, A., Pan, H., Tsivin, V., et al. (2013). Detection of a long non-coding RNA (CCAT1) in living cells and human adenocarcinoma of colon tissues using FIT-PNA molecular beacons. *Cancer Lett.* 1–7.

Kanduri, C. (2011). *Kcnq1ot1*: a chromatin regulatory RNA. *Semin. Cell Dev. Biol.* 22, 343–350.

Kanduri, C. (2015). Long noncoding RNAs: Lessons from genomic imprinting. *Biochim. Biophys. Acta.*

Kanduri, C., Thakur, N., and Pandey, R.R. (2006). The length of the transcript encoded from the *Kcnq1ot1* antisense promoter determines the degree of silencing. *EMBO J.* 25, 2096–2106.

Kapoor, A., Goldberg, M.S., Cumberland, L.K., Ratnakumar, K., Segura, M.F., Emanuel, P.O., Menendez, S., Vardabasso, C., LeRoy, G., Vidal, C.I., et al. (2010). The histone variant macroH2A suppresses melanoma progression through regulation of CDK8. *Nature* 468, 1105–1109.

Kapranov, P., Drenkow, J., Cheng, J., Long, J., Helt, G., Dike, S., and Gingeras, T.R. (2005). Examples of the complex architecture of the human transcriptome revealed by RACE and high-density tiling arrays. *Genome Res.* 15, 987–997.

Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yap, C.C., Suzuki, M., Kawai, J., et al. (2005). Antisense transcription in the mammalian transcriptome. *Science* 309, 1564–1566.

Kelley, D., and Rinn, J. (2012). Transposable elements reveal a stem cell-specific class of long noncoding RNAs. *Genome Biol.* 13, R107.

Kelly, L.M., Liu, Q., Kutok, J.L., Williams, I.R., Boulton, C.L., and Gilliland, D.G. (2002). FLT3 internal tandem duplication mutations associated with human acute myeloid leukemias induce myeloproliferative disease in a murine bone marrow transplant model. *Blood* 99, 310–318.

Kertesz, M., Wan, Y., Mazor, E., Rinn, J.L., Nutter, R.C., Chang, H.Y., and Segal, E. (2010). Genome-wide measurement of RNA secondary structure in yeast. *Nature* 467, 103–107.

Khaitan, D., Dinger, M.E., Mazar, J., Crawford, J., Smith, M.A., Mattick, J.S., and Perera, R.J. (2011). The melanoma-upregulated long noncoding RNA *SPRY4-IT1* modulates apoptosis and invasion. *Cancer Res.* 71, 3852–3862.

Khalil, A.M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., Thomas, K., Presser, A., Bernstein, B.E., van Oudenaarden, A., et al. (2009). Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc. Natl. Acad. Sci. U. S. A.* 106, 11667–11672.

Kim, T.-K., Hemberg, M., and Gray, J.M. (2015). Enhancer RNAs: a class of long noncoding RNAs synthesized

at enhancers. *Cold Spring Harb. Perspect. Biol.* 7, a018622.

Kino, T., Hurt, D.E., Ichijo, T., Nader, N., and Chrousos, G.P. (2010). Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. *Sci. Signal.* 3, ra8.

Kiyoi, H., Naoe, T., Nakano, Y., Yokota, S., Minami, S., Miyawaki, S., Asou, N., Kuriyama, K., Jinnai, I., Shimazaki, C., et al. (1999). Prognostic implication of FLT3 and N-RAS gene mutations in acute myeloid leukemia. *Blood* 93, 3074–3080.

Kogo, R., Shimamura, T., Mimori, K., Kawahara, K., Imoto, S., Sudo, T., Tanaka, F., Shibata, K., Suzuki, A., Komune, S., et al. (2011). Long noncoding RNA HOTAIR regulates polycomb-dependent chromatin modification and is associated with poor prognosis in colorectal cancers. *Cancer Res.* 71, 6320–6326.

Koike, A., Nishikawa, H., Wu, W., Okada, Y., Venkitaraman, A.R., and Ohta, T. (2010). Recruitment of phosphorylated NPM1 to sites of DNA damage through RNF8-dependent ubiquitin conjugates. *Cancer Res.* 70, 6746–6756.

de Kok, J.B., Verhaegh, G.W., Roelofs, R.W., Hessels, D., Kiemeny, L.A., Aalders, T.W., Swinkels, D.W., and Schalken, J.A. (2002). DD3(PCA3), a very sensitive and specific marker to detect prostate tumors. *Cancer Res.* 62, 2695–2698.

Kondo, T., Minamino, N., Nagamura-Inoue, T., Matsumoto, M., Taniguchi, T., and Tanaka, N. (1997). Identification and characterization of nucleophosmin/B23/numatrin which binds the anti-oncogenic transcription factor IRF-1 and manifests oncogenic activity. *Oncogene* 15, 1275–1281.

De Koning, L., Corpet, A., Haber, J.E., and Almouzni, G. (2007). Histone chaperones: an escort network regulating histone traffic. *Nat. Struct. Mol. Biol.* 14, 997–1007.

Koreth, J., Schlenk, R., Kopecky, K.J., Honda, S., Sierra, J., Djulbegovic, B.J., Wadleigh, M., DeAngelo, D.J., Stone, R.M., Sakamaki, H., et al. (2009). Allogeneic stem cell transplantation for acute myeloid leukemia in first complete remission: systematic review and meta-analysis of prospective clinical trials. *JAMA* 301, 2349–2361.

Kotake, Y., Nakagawa, T., Kitagawa, K., Suzuki, S., Liu, N., Kitagawa, M., and Xiong, Y. (2011). Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene. *Oncogene* 30, 1956–1962.

Krause, A., and Hoffmann, I. (2010). Polo-like kinase 2-dependent phosphorylation of NPM/B23 on serine 4 triggers centriole duplication. *PLoS One* 5, e9849.

Kretz, M., Webster, D.E., Flockhart, R.J., Lee, C.S., Zehnder, A., Lopez-Pajares, V., Qu, K., Zheng, G.X.Y., Chow, J., Kim, G.E., et al. (2012). Suppression of progenitor differentiation requires the long noncoding RNA ANCR. *Genes Dev.* 26, 338–343.

Kumar, C.C. (2011). Genetic abnormalities and challenges in the treatment of acute myeloid leukemia. *Genes Cancer* 2, 95–107.

Kuo, M.-L., den Besten, W., Bertwistle, D., Roussel, M.F., and Sherr, C.J. (2004). N-terminal polyubiquitination and degradation of the Arf tumor suppressor. *Genes Dev.* 18, 1862–1874.

Lamond, A.I., and Spector, D.L. (2003). Nuclear speckles: a model for nuclear organelles. *Nat. Rev. Mol. Cell Biol.* 4, 605–612.

Lee, J.T., Davidow, L.S., and Warshawsky, D. (1999). Tsix, a gene antisense to Xist at the X-inactivation centre. *Nat. Genet.* 21, 400–404.

Lee, S.Y., Park, J.-H., Kim, S., Park, E.-J., Yun, Y., and Kwon, J. (2005). A proteomics approach for the identification of nucleophosmin and heterogeneous nuclear ribonucleoprotein C1/C2 as chromatin-binding proteins in response to DNA double-strand breaks. *Biochem. J.* 388, 7–15.

- Leong, S.M., Tan, B.X., Bte Ahmad, B., Yan, T., Chee, L.Y., Ang, S.T., Tay, K.G., Koh, L.P., Yeoh, A.E.J., Koay, E.S.-C., et al. (2010). Mutant nucleophosmin deregulates cell death and myeloid differentiation through excessive caspase-6 and -8 inhibition. *Blood* 116, 3286–3296.
- Lerner, M., Harada, M., Lovén, J., Castro, J., Davis, Z., Oscier, D., Henriksson, M., Sangfelt, O., Grandér, D., and Corcoran, M.M. (2009). DLEU2, frequently deleted in malignancy, functions as a critical host gene of the cell cycle inhibitory microRNAs miR-15a and miR-16-1. *Exp. Cell Res.* 315, 2941–2952.
- Leukemia, A.M. (2013). Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N. Engl. J. Med.* 368, 2059–2074.
- Ley, T.J., Ding, L., Walter, M.J., McLellan, M.D., Lamprecht, T., Larson, D.E., Kandoth, C., Payton, J.E., Baty, J., Welch, J., et al. (2010). DNMT3A mutations in acute myeloid leukemia. *N. Engl. J. Med.* 363, 2424–2433.
- Leymarie, V., Galois, A.-C., Falkenrodt, A., and Lessard, M. [Diagnosis of myeloid hematologic malignancies: contributions of the 2001 World Health Organization (WHO) classification]. *Ann. Biol. Clin. (Paris)*. 62, 513–520.
- Li, D., Liu, X., Lin, L., Hou, J., Li, N., Wang, C., Wang, P., Zhang, Q., Zhang, P., Zhou, W., et al. (2011). MicroRNA-99a inhibits hepatocellular carcinoma growth and correlates with prognosis of patients with hepatocellular carcinoma. *J. Biol. Chem.* 286, 36677–36685.
- Li, J., Zhang, X., Sejas, D.P., Bagby, G.C., and Pang, Q. (2004). Hypoxia-induced nucleophosmin protects cell death through inhibition of p53. *J. Biol. Chem.* 279, 41275–41279.
- Li, L., Feng, T., Lian, Y., Zhang, G., Garen, A., and Song, X. (2009). Role of human noncoding RNAs in the control of tumorigenesis. *Proc. Natl. Acad. Sci. U. S. A.* 106, 12956–12961.
- Li, L., Liu, B., Wapinski, O.L., Tsai, M.-C., Qu, K., Zhang, J., Carlson, J.C., Lin, M., Fang, F., Gupta, R.A., et al. (2013a). Targeted disruption of *Hotair* leads to homeotic transformation and gene derepression. *Cell Rep.* 5, 3–12.
- Li, W., Notani, D., Ma, Q., Tanasa, B., Nunez, E., Chen, A.Y., Merkurjev, D., Zhang, J., Ohgi, K., Song, X., et al. (2013b). Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature* 498, 516–520.
- Li, Z., Boone, D., and Hann, S.R. (2008). Nucleophosmin interacts directly with c-Myc and controls c-Myc-induced hyperproliferation and transformation. *Proc. Natl. Acad. Sci. U. S. A.* 105, 18794–18799.
- Lin, L., Gu, Z.-T., Chen, W.-H., and Cao, K.-J. (2015). Increased expression of the long non-coding RNA ANRIL promotes lung cancer cell metastasis and correlates with poor prognosis. *Diagn. Pathol.* 10, 14.
- Lin, M.F., Carlson, J.W., Crosby, M.A., Matthews, B.B., Yu, C., Park, S., Wan, K.H., Schroeder, A.J., Gramates, L.S., St Pierre, S.E., et al. (2007). Revisiting the protein-coding gene catalog of *Drosophila melanogaster* using 12 fly genomes. *Genome Res.* 17, 1823–1836.
- Lindstedt, I., Lindgren, M.A., Andersson, E., and Engström, W. The *WT1* gene--its role in tumourigenesis and prospects for immunotherapeutic advances. *In Vivo* 28, 675–681.
- Lindström, M.S., and Zhang, Y. (2008). Ribosomal protein S9 is a novel B23/NPM-binding protein required for normal cell proliferation. *J. Biol. Chem.* 283, 15568–15576.
- Liu, H., Tan, B.C.-M., Tseng, K.H., Chuang, C.P., Yeh, C.-W., Chen, K.-D., Lee, S.-C., and Yung, B.Y.-M. (2007). Nucleophosmin acts as a novel AP2alpha-binding transcriptional corepressor during cell differentiation. *EMBO Rep.* 8, 394–400.
- Liu, T., Huang, Y., Chen, J., Chi, H., Yu, Z., Wang, J., and Chen, C. (2014). Attenuated ability of BACE1 to cleave the amyloid precursor protein via silencing long noncoding RNA BACE1-AS expression. *Mol. Med. Rep.*

10, 1275–1281.

Low, J.T., and Weeks, K.M. (2010). SHAPE-directed RNA secondary structure prediction. *Methods* 52, 150–158.

Luo, H., Sun, Y., Wei, G., Luo, J., Yang, X., Liu, W., Guo, M., and Chen, R. (2015a). Functional Characterization of Long Noncoding RNA *Lnc_bc060912* in Human Lung Carcinoma Cells. *Biochemistry* 54, 2895–2902.

Luo, M., Li, Z., Wang, W., Zeng, Y., Liu, Z., and Qiu, J. (2013). Upregulated *H19* contributes to bladder cancer cell proliferation by regulating *ID2* expression. *FEBS J.* 280, 1709–1716.

Luo, M., Jeong, M., Sun, D., Park, H.J., Rodriguez, B.A.T., Xia, Z., Yang, L., Zhang, X., Sheng, K., Darlington, G.J., et al. (2015b). Long Non-Coding RNAs Control Hematopoietic Stem Cell Function. *Cell Stem Cell* 16, 426–438.

Maenner, S., Blaud, M., Fouillen, L., Savoye, A., Marchand, V., Dubois, A., Sanglier-Cianfèrani, S., Van Dorselaer, A., Clerc, P., Avner, P., et al. (2010). 2-D structure of the A region of *Xist* RNA and its implication for PRC2 association. *PLoS Biol.* 8, e1000276.

Maggi, L.B., Kuchenruether, M., Dadey, D.Y.A., Schwoppe, R.M., Grisendi, S., Townsend, R.R., Pandolfi, P.P., and Weber, J.D. (2008). Nucleophosmin serves as a rate-limiting nuclear export chaperone for the Mammalian ribosome. *Mol. Cell. Biol.* 28, 7050–7065.

Marahrens, Y., Panning, B., Dausman, J., Strauss, W., and Jaenisch, R. (1997). *Xist*-deficient mice are defective in dosage compensation but not spermatogenesis. *Genes Dev.* 11, 156–166.

Marcucci, G., Haferlach, T., and Döhner, H. (2011). Molecular genetics of adult acute myeloid leukemia: prognostic and therapeutic implications. *J. Clin. Oncol.* 29, 475–486.

Mardis, E.R., Ding, L., Dooling, D.J., Larson, D.E., McLellan, M.D., Chen, K., Koboldt, D.C., Fulton, R.S., Delehaunty, K.D., McGrath, S.D., et al. (2009). Recurring mutations found by sequencing an acute myeloid leukemia genome. *N. Engl. J. Med.* 361, 1058–1066.

Marín-Béjar, O., Marchese, F.P., Athie, A., Sánchez, Y., González, J., Segura, V., Huang, L., Moreno, I., Navarro, A., Monzó, M., et al. (2013). *Pint* lincRNA connects the *p53* pathway with epigenetic silencing by the Polycomb repressive complex 2. *Genome Biol.* 14, R104.

Marques, A.C., and Ponting, C.P. (2009). Catalogues of mammalian long noncoding RNAs: modest conservation and incompleteness. *Genome Biol.* 10, R124.

Martens, J.A., Laprade, L., and Winston, F. (2004). Intergenic transcription is required to repress the *Saccharomyces cerevisiae* *SER3* gene. *Nature* 429, 571–574.

Martianov, I., Ramadass, A., Serra Barros, A., Chow, N., and Akoulitchev, A. (2007). Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. *Nature* 445, 666–670.

Masui, O., and Heard, E. (2006). RNA and protein actors in X-chromosome inactivation. *Cold Spring Harb. Symp. Quant. Biol.* 71, 419–428.

Matouk, I.J., Abbasi, I., Hochberg, A., Galun, E., Dweik, H., and Akkawi, M. (2009). Highly upregulated in liver cancer noncoding RNA is overexpressed in hepatic colorectal metastasis. *Eur. J. Gastroenterol. Hepatol.* 21, 688–692.

Mattick, J.S. (1994). Introns: evolution and function. *Curr. Opin. Genet. Dev.* 4, 823–831.

McHale, C.M., Zhang, L., and Smith, M.T. (2012). Current understanding of the mechanism of benzene-induced leukemia in humans: implications for risk assessment. *Carcinogenesis* 33, 240–252.

Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., Maier, L., Mackowiak, S.D., Gregersen, L.H., Munschauer, M., et al. (2013). Circular RNAs are a large class of animal RNAs with regulatory potency.

Nature 495, 333–338.

Mohammad, F., Mondal, T., and Kanduri, C. (2009). Epigenetics of imprinted long noncoding RNAs. *Epigenetics* 4, 277–286.

Mohammad, F., Mondal, T., Guseva, N., Pandey, G.K., and Kanduri, C. (2010). *Kcnq1ot1* noncoding RNA mediates transcriptional gene silencing by interacting with *Dnmt1*. *Development* 137, 2493–2499.

Mondal, T., Rasmussen, M., Pandey, G.K., Isaksson, A., and Kanduri, C. (2010). Characterization of the RNA content of chromatin. *Genome Res.* 20, 899–907.

Monnier, P., Martinet, C., Pontis, J., Stancheva, I., Ait-Si-Ali, S., and Dandolo, L. (2013). *H19* lncRNA controls gene expression of the Imprinted Gene Network by recruiting *MBD1*. *Proc. Natl. Acad. Sci. U. S. A.* 110, 20693–20698.

Morán, I., Akerman, I., van de Bunt, M., Xie, R., Benazra, M., Nammo, T., Arnes, L., Nakić, N., García-Hurtado, J., Rodríguez-Seguí, S., et al. (2012). Human β cell transcriptome analysis uncovers lncRNAs that are tissue-specific, dynamically regulated, and abnormally expressed in type 2 diabetes. *Cell Metab.* 16, 435–448.

Moshaver, B., van Rhenen, A., Kelder, A., van der Pol, M., Terwijn, M., Bachas, C., Westra, A.H., Ossenkoppele, G.J., Zweegman, S., and Schuurhuis, G.J. (2008). Identification of a small subpopulation of candidate leukemia-initiating cells in the side population of patients with acute myeloid leukemia. *Stem Cells* 26, 3059–3067.

Mourtada-Maarabouni, M., and Williams, G.T. (2014). Role of *GAS5* noncoding RNA in mediating the effects of rapamycin and its analogues on mantle cell lymphoma cells. *Clin. Lymphoma. Myeloma Leuk.* 14, 468–473.

Mourtada-Maarabouni, M., Pickard, M.R., Hedge, V.L., Farzaneh, F., and Williams, G.T. (2009). *GAS5*, a non-protein-coding RNA, controls apoptosis and is downregulated in breast cancer. *Oncogene* 28, 195–208.

Mullighan, C.G., Kennedy, A., Zhou, X., Radtke, I., Phillips, L.A., Shurtleff, S.A., and Downing, J.R. (2007). Pediatric acute myeloid leukemia with *NPM1* mutations is characterized by a gene expression profile with dysregulated *HOX* gene expression distinct from *MLL*-rearranged leukemias. *Leukemia* 21, 2000–2009.

Murano, K., Okuwaki, M., Hisaoka, M., and Nagata, K. (2008). Transcription regulation of the rRNA gene by a multifunctional nucleolar protein, *B23/nucleophosmin*, through its histone chaperone activity. *Mol. Cell. Biol.* 28, 3114–3126.

Nagano, T., Mitchell, J.A., Sanz, L.A., Pauler, F.M., Ferguson-Smith, A.C., Feil, R., and Fraser, P. (2008). The *Air* noncoding RNA epigenetically silences transcription by targeting *G9a* to chromatin. *Science* 322, 1717–1720.

Naganuma, T., and Hirose, T. (2013). Paraspeckle formation during the biogenesis of long non-coding RNAs. *RNA Biol.* 10, 456–461.

Nakamura, Y., Takahashi, N., Kakegawa, E., Yoshida, K., Ito, Y., Kayano, H., Niitsu, N., Jinnai, I., and Bessho, M. (2008). The *GAS5* (growth arrest-specific transcript 5) gene fuses to *BCL6* as a result of *t(1;3)(q25;q27)* in a patient with B-cell lymphoma. *Cancer Genet. Cytogenet.* 182, 144–149.

Namboodiri, V.M.H., Akey, I. V., Schmidt-Zachmann, M.S., Head, J.F., and Akey, C.W. (2004). The structure and function of *Xenopus NO38-core*, a histone chaperone in the nucleolus. *Structure* 12, 2149–2160.

Naoe, T., Suzuki, T., Kiyoi, H., and Urano, T. (2006). Nucleophosmin: a versatile molecule associated with hematological malignancies. *Cancer Sci.* 97, 963–969.

Neame, P.B., Soamboonsrup, P., Browman, G.P., Meyer, R.M., Bengner, A., Wilson, W.E., Walker, I.R., Saeed, N., and McBride, J.A. (1986). Classifying acute leukemia by immunophenotyping: a combined FAB-immunologic classification of AML. *Blood* 68, 1355–1362.

Negi, S.S., and Olson, M.O.J. (2006). Effects of interphase and mitotic phosphorylation on the mobility and

location of nucleolar protein B23. *J. Cell Sci.* 119, 3676–3685.

Ng, L.J., Copley, J.E., Pickett, H.A., Reddel, R.R., and Suter, C.M. (2009). Telomerase activity is associated with an increase in DNA methylation at the proximal subtelomere and a reduction in telomeric transcription. *Nucleic Acids Res.* 37, 1152–1159.

Ng, S.-Y., Johnson, R., and Stanton, L.W. (2012). Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors. *EMBO J.* 31, 522–533.

Niazi, F., and Valadkhan, S. (2012). Computational analysis of functional long noncoding RNAs reveals lack of peptide-coding capacity and parallels with 3' UTRs. *RNA* 18, 825–843.

Noonan, E.J., Place, R.F., Basak, S., Pookot, D., and Li, L.-C. (2010). miR-449a causes Rb-dependent cell cycle arrest and senescence in prostate cancer cells. *Oncotarget* 1, 349–358.

Novikova, I. V, Hennelly, S.P., and Sanbonmatsu, K.Y. (2012). Structural architecture of the human long non-coding RNA, steroid receptor RNA activator. *Nucleic Acids Res.* 40, 5034–5051.

Nowak, T., Januszkiewicz, D., Zawada, M., Pernak, M., Lewandowski, K., Rembowska, J., Nowicka, K., Mankowski, P., and Nowak, J. (2006). Amplification of hTERT and hTERC genes in leukemic cells with high expression and activity of telomerase. *Oncol. Rep.* 16, 301–305.

Ntini, E., Järvelin, A.I., Bornholdt, J., Chen, Y., Boyd, M., Jørgensen, M., Andersson, R., Hoof, I., Schein, A., Andersen, P.R., et al. (2013). Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality. *Nat. Struct. Mol. Biol.* 20, 923–928.

Okuda, M., Horn, H.F., Tarapore, P., Tokuyama, Y., Smulian, A.G., Chan, P.K., Knudsen, E.S., Hofmann, I.A., Snyder, J.D., Bove, K.E., et al. (2000). Nucleophosmin/B23 is a target of CDK2/cyclin E in centrosome duplication. *Cell* 103, 127–140.

Okuwaki, M., Matsumoto, K., Tsujimoto, M., and Nagata, K. (2001a). Function of nucleophosmin/B23, a nucleolar acidic protein, as a histone chaperone. *FEBS Lett.* 506, 272–276.

Okuwaki, M., Iwamatsu, A., Tsujimoto, M., and Nagata, K. (2001b). Identification of nucleophosmin/B23, an acidic nucleolar protein, as a stimulatory factor for in vitro replication of adenovirus DNA complexed with viral basic core proteins. *J. Mol. Biol.* 311, 41–55.

Okuwaki, M., Matsumoto, K., Tsujimoto, M., and Nagata, K. (2001c). Function of nucleophosmin/B23, a nucleolar acidic protein, as a histone chaperone. *FEBS Lett.* 506, 272–276.

Ørom, U.A., Derrien, T., Beringer, M., Gumireddy, K., Gardini, A., Bussotti, G., Lai, F., Zytnicki, M., Notredame, C., Huang, Q., et al. (2010). Long noncoding RNAs with enhancer-like function in human cells. *Cell* 143, 46–58.

Owen, C., Fitzgibbon, J., and Paschka, P. (2010). The clinical relevance of Wilms Tumour 1 (WT1) gene mutations in acute leukaemia. *Hematol. Oncol.* 28, 13–19.

Pagano, A., Castelnuovo, M., Tortelli, F., Ferrari, R., Dieci, G., and Cancedda, R. (2007). New Small Nuclear RNA Gene-Like Transcriptional Units as Sources of Regulatory Transcripts. *PLoS Genet.* 3, e1.

Pandey, R.R., Mondal, T., Mohammad, F., Enroth, S., Redrup, L., Komorowski, J., Nagano, T., Mancini-Dinardo, D., and Kanduri, C. (2008). Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol. Cell* 32, 232–246.

Panzitt, K., Tschernatsch, M.M.O., Guelly, C., Moustafa, T., Stradner, M., Strohmaier, H.M., Buck, C.R., Denk, H., Schroeder, R., Trauner, M., et al. (2007). Characterization of HULC, a novel gene with striking up-regulation in hepatocellular carcinoma, as noncoding RNA. *Gastroenterology* 132, 330–342.

- Paralkar, V.R., Mishra, T., Luan, J., Yao, Y., Kossenkov, A. V, Anderson, S.M., Dunagin, M., Pimkin, M., Gore, M., Sun, D., et al. (2014). Lineage and species-specific long noncoding RNAs during erythro-megakaryocytic development. *Blood* 123, 1927–1937.
- Paraskevoudoulou, M.D., Georgakilas, G., Kostoulas, N., Reczko, M., Maragkakis, M., Dalamagas, T.M., and Hatzigeorgiou, A.G. (2013). DIANA-LncBase: experimentally verified and computationally predicted microRNA targets on long non-coding RNAs. *Nucleic Acids Res.* 41, D239–D245.
- Park, E., and Maquat, L.E. Staufen-mediated mRNA decay. *Wiley Interdiscip. Rev. RNA* 4, 423–435.
- Paschka, P., Schlenk, R.F., Gaidzik, V.I., Habdank, M., Krönke, J., Bullinger, L., Späth, D., Kayser, S., Zucknick, M., Götze, K., et al. (2010). IDH1 and IDH2 mutations are frequent genetic alterations in acute myeloid leukemia and confer adverse prognosis in cytogenetically normal acute myeloid leukemia with NPM1 mutation without FLT3 internal tandem duplication. *J. Clin. Oncol.* 28, 3636–3643.
- Pasic, I., Shlien, A., Durbin, A.D., Stavropoulos, D.J., Baskin, B., Ray, P.N., Novokmet, A., and Malkin, D. (2010). Recurrent focal copy-number changes and loss of heterozygosity implicate two noncoding RNAs and one tumor suppressor gene at chromosome 3q13.31 in osteosarcoma. *Cancer Res.* 70, 160–171.
- Pasmant, E., Sabbagh, A., Vidaud, M., and Bièche, I. (2011). ANRIL, a long, noncoding RNA, is an unexpected major hotspot in GWAS. *FASEB J.* 25, 444–448.
- Pastore, F., Greif, P.A., Schneider, S., Ksienzyk, B., Mellert, G., Zellmeier, E., Braess, J., Sauerland, C.M., Heinecke, A., Krug, U., et al. (2014). The NPM1 mutation type has no impact on survival in cytogenetically normal AML. *PLoS One* 9, e109759.
- Patel, J.P., Gönen, M., Figueroa, M.E., Fernandez, H., Sun, Z., Racevskis, J., Van Vlierberghe, P., Dolgalev, I., Thomas, S., Aminova, O., et al. (2012). Prognostic relevance of integrated genetic profiling in acute myeloid leukemia. *N. Engl. J. Med.* 366, 1079–1089.
- Paul, I.J., and Duerksen, J.D. (1975). Chromatin-associated RNA content of heterochromatin and euchromatin. *Mol. Cell. Biochem.* 9, 9–16.
- Peng, W., Fan, H., Wu, G., Wu, J., and Feng, J. (2015). Upregulation of long noncoding RNA PEG10 associates with poor prognosis in diffuse large B cell lymphoma with facilitating tumorigenicity. *Clin. Exp. Med.* 1–6.
- Peng, X., Gralinski, L., Armour, C.D., Ferris, M.T., Thomas, M.J., Proll, S., Bradel-Tretheway, B.G., Korth, M.J., Castle, J.C., Biery, M.C., et al. (2010). Unique signatures of long noncoding RNA expression in response to virus infection and altered innate immune signaling. *MBio* 1.
- Peterlin, B.M., Brogie, J.E., and Price, D.H. 7SK snRNA: a noncoding RNA that plays a major role in regulating eukaryotic transcription. *Wiley Interdiscip. Rev. RNA* 3, 92–103.
- Petrovics, G., Zhang, W., Makarem, M., Street, J.P., Connelly, R., Sun, L., Sesterhenn, I.A., Srikantan, V., Moul, J.W., and Srivastava, S. (2004). Elevated expression of PCGEM1, a prostate-specific gene with cell growth-promoting function, is associated with high-risk prostate cancer patients. *Oncogene* 23, 605–611.
- Pierce, D.A., Shimizu, Y., Preston, D.L., Vaeth, M., and Mabuchi, K. (2012). Studies of the mortality of atomic bomb survivors. report 12, part I. *Cancer: 1950-1990. 1996. Radiat. Res.* 178, AV61–AV87.
- Poliseno, L., Salmena, L., Zhang, J., Carver, B., Haveman, W.J., and Pandolfi, P.P. (2010). A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* 465, 1033–1038.
- Pollard, K.S., Salama, S.R., Lambert, N., Lambot, M.-A., Coppens, S., Pedersen, J.S., Katzman, S., King, B., Onodera, C., Siepel, A., et al. (2006). An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* 443, 167–172.
- Popescu, N.C., and Zimonjic, D.B. Chromosome-mediated alterations of the MYC gene in human cancer. *J. Cell.*

Mol. Med. 6, 151–159.

Prensner, J.R., Iyer, M.K., Balbin, O.A., Dhanasekaran, S.M., Cao, Q., Brenner, J.C., Laxman, B., Asangani, I. a, Grasso, C.S., Kominsky, H.D., et al. (2011). Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. *Nat. Biotechnol.* 29, 742–749.

Prensner, J.R., Chen, W., Iyer, M.K., Cao, Q., Ma, T., Han, S., Sahu, A., Malik, R., Wilder-Romans, K., Navone, N., et al. (2014). PCAT-1, a long noncoding RNA, regulates BRCA2 and controls homologous recombination in cancer. *Cancer Res.* 74, 1651–1660.

Qiu, J.-J., Lin, Y.-Y., Ding, J.-X., Feng, W.-W., Jin, H.-Y., and Hua, K.-Q. (2015). Long non-coding RNA ANRIL predicts poor prognosis and promotes invasion/metastasis in serous ovarian cancer. *Int. J. Oncol.* 46, 2497–2505.

Qu, S., Yang, X., Li, X., Wang, J., Gao, Y., Shang, R., Sun, W., Dou, K., and Li, H. (2015). Circular RNA: a new star of noncoding RNAs. *Cancer Lett.* 365, 141–148.

Quénet, D., and Dalal, Y. (2014). A long non-coding RNA is required for targeting centromeric protein A to the human centromere. *Elife* 3, e03254.

Rai, R., Varma, S.P.M. V, Shinde, N., Ghosh, S., Kumaran, S.P., Skariah, G., and Laloraya, S. (2011). Small ubiquitin-related modifier ligase activity of Mms21 is required for maintenance of chromosome integrity during the unperturbed mitotic cell division cycle in *Saccharomyces cerevisiae*. *J. Biol. Chem.* 286, 14516–14530.

Ranzani, V., Rossetti, G., Panzeri, I., Arrigoni, A., Bonnal, R.J.P., Curti, S., Gruarin, P., Provasi, E., Sugliano, E., Marconi, M., et al. (2015). The long intergenic noncoding RNA landscape of human lymphocytes highlights the regulation of T cell differentiation by linc-MAF-4. *Nat. Immunol.* 16, 318–325.

Rao, A. V, Valk, P.J.M., Metzeler, K.H., Acharya, C.R., Tuchman, S.A., Stevenson, M.M., Rizzieri, D.A., Delwel, R., Buske, C., Bohlander, S.K., et al. (2009). Age-specific differences in oncogenic pathway dysregulation and anthracycline sensitivity in patients with acute myeloid leukemia. *J. Clin. Oncol.* 27, 5580–5586.

Redon, S., Reichenbach, P., and Lingner, J. (2010). The non-coding RNA TERRA is a natural ligand and direct inhibitor of human telomerase. *Nucleic Acids Res.* 38, 5797–5806.

Ribeiro, A.F.T., Pratcorona, M., Erpelinck-Verschueren, C., Rockova, V., Sanders, M., Abbas, S., Figueroa, M.E., Zeilemaker, A., Melnick, A., Löwenberg, B., et al. (2012). Mutant DNMT3A: a marker of poor prognosis in acute myeloid leukemia. *Blood* 119, 5824–5831.

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S. a, Goodnough, L.H., Helms, J. a, Farnham, P.J., Segal, E., et al. (2007a). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* 129, 1311–1323.

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E., et al. (2007b). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* 129, 1311–1323.

Ripoche, M.A., Kress, C., Poirier, F., and Dandolo, L. (1997). Deletion of the H19 transcription unit reveals the existence of a putative imprinting control element. *Genes Dev.* 11, 1596–1604.

Rossignol, F., Vaché, C., and Clottes, E. (2002). Natural antisense transcripts of hypoxia-inducible factor 1alpha are detected in different normal and tumour human tissues. *Gene* 299, 135–140.

Saito, Y., Uchida, N., Tanaka, S., Suzuki, N., Tomizawa-Murasawa, M., Sone, A., Najima, Y., Takagi, S., Aoki, Y., Wake, A., et al. (2010). Induction of cell cycle entry eliminates human leukemia stem cells in a mouse model of AML. *Nat. Biotechnol.* 28, 275–280.

Salameh, A., Lee, A.K., Cardó-Vila, M., Nunes, D.N., Efstathiou, E., Staquicini, F.I., Dobroff, A.S., Marchiò, S.,

- Navone, N.M., Hosoya, H., et al. (2015). *PRUNE2 is a human prostate cancer suppressor regulated by the intronic long noncoding RNA PCA3*. *Proc. Natl. Acad. Sci. U. S. A.* 112, 8403–8408.
- Salditt-Georgieff, M., and Darnell, J.E. (1982). *Further evidence that the majority of primary nuclear RNA transcripts in mammalian cells do not contribute to mRNA*. *Mol. Cell. Biol.* 2, 701–707.
- De Santa, F., Barozzi, I., Mietton, F., Ghisletti, S., Polletti, S., Tusi, B.K., Muller, H., Ragoussis, J., Wei, C.-L., and Natoli, G. (2010). *A large fraction of extragenic RNA pol II transcription sites overlap enhancers*. *PLoS Biol.* 8, e1000384.
- Sati, S., Ghosh, S., Jain, V., Scaria, V., and Sengupta, S. (2012). *Genome-wide analysis reveals distinct patterns of epigenetic features in long non-coding RNA loci*. *Nucleic Acids Res.* 40, 10018–10031.
- Savkur, R.S., and Olson, M.O. (1998). *Preferential cleavage in pre-ribosomal RNA by protein B23 endoribonuclease*. *Nucleic Acids Res.* 26, 4508–4515.
- Schaukowitch, K., Joo, J.-Y., Liu, X., Watts, J.K., Martinez, C., and Kim, T.-K. (2014). *Enhancer RNA Facilitates NELF Release from Immediate Early Genes*. *Mol. Cell* 56, 29–42.
- Schlenk, R.F., Döhner, K., Krauter, J., Fröhling, S., Corbacioglu, A., Bullinger, L., Habdank, M., Späth, D., Morgan, M., Benner, A., et al. (2008). *Mutations and treatment outcome in cytogenetically normal acute myeloid leukemia*. *N. Engl. J. Med.* 358, 1909–1918.
- Schmitz, K.-M., Mayer, C., Postepska, A., and Grummt, I. (2010). *Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes*. *Genes Dev.* 24, 2264–2269.
- Schneider, C., King, R.M., and Philipson, L. (1988). *Genes specifically expressed at growth arrest of mammalian cells*. *Cell* 54, 787–793.
- Schnittger, S., Schoch, C., Kern, W., Mecucci, C., Tschulik, C., Martelli, M.F., Haferlach, T., Hiddemann, W., and Falini, B. (2005). *Nucleophosmin gene mutations are predictors of favorable prognosis in acute myelogenous leukemia with a normal karyotype*. *Blood* 106, 3733–3739.
- Schnittger, S., Kern, W., Tschulik, C., Weiss, T., Dicker, F., Falini, B., Haferlach, C., and Haferlach, T. (2009). *Minimal residual disease levels assessed by NPM1 mutation-specific RQ-PCR provide important prognostic information in AML*. *Blood* 114, 2220–2231.
- Schoeftner, S., and Blasco, M.A. (2008). *Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II*. *Nat. Cell Biol.* 10, 228–236.
- Scholl, S., Fricke, H.-J., Sayer, H.G., and Höffken, K. (2009). *Clinical implications of molecular genetic aberrations in acute myeloid leukemia*. *J. Cancer Res. Clin. Oncol.* 135, 491–505.
- Schuler, G.D., Boguski, M.S., Stewart, E.A., Stein, L.D., Gyapay, G., Rice, K., White, R.E., Rodriguez-Tom, P., Aggarwal, A., Bajorek, E., et al. (1996). *A Gene Map of the Human Genome*. *Science* (80-). 274, 540–546.
- Sehgal, L., Mathur, R., Braun, F.K., Wise, J.F., Berkova, Z., Neelapu, S., Kwak, L.W., and Samaniego, F. (2014). *FAS-antisense 1 lncRNA and production of soluble versus membrane Fas in B-cell lymphoma*. *Leukemia* 28, 2376–2387.
- Sharma, S., Findlay, G.M., Bandukwala, H.S., Oberdoerffer, S., Baust, B., Li, Z., Schmidt, V., Hogan, P.G., Sacks, D.B., and Rao, A. (2011). *Dephosphorylation of the nuclear factor of activated T cells (NFAT) transcription factor is regulated by an RNA-protein scaffold complex*. *Proc. Natl. Acad. Sci. U. S. A.* 108, 11381–11386.
- Shia, W.-J., Li, B., and Workman, J.L. (2006). *SAS-mediated acetylation of histone H4 Lys 16 is required for H2A.Z incorporation at subtelomeric regions in Saccharomyces cerevisiae*. *Genes Dev.* 20, 2507–2512.
- Shlush, L.I., Zandi, S., Mitchell, A., Chen, W.C., Brandwein, J.M., Gupta, V., Kennedy, J.A., Schimmer, A.D.,

- Schuh, A.C., Yee, K.W., et al. (2014). Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. *Nature* 506, 328–333.
- Sidi, A.A., Ohana, P., Benjamin, S., Shalev, M., Ransom, J.H., Lamm, D., Hochberg, A., and Leibovitch, I. (2008). Phase I/II marker lesion study of intravesical BC-819 DNA plasmid in H19 over expressing superficial bladder cancer refractory to bacillus Calmette-Guerin. *J. Urol.* 180, 2379–2383.
- Sleutels, F., Zwart, R., and Barlow, D.P. (2002). The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature* 415, 810–813.
- Smith, C.M., and Steitz, J.A. (1998). Classification of gas5 as a multi-small-nucleolar-RNA (snoRNA) host gene and a member of the 5'-terminal oligopyrimidine gene family reveals common features of snoRNA host genes. *Mol. Cell. Biol.* 18, 6897–6909.
- Somarowthu, S., Legiewicz, M., Chillón, I., Marcia, M., Liu, F., and Pyle, A.M. (2015). HOTAIR forms an intricate and modular secondary structure. *Mol. Cell* 58, 353–361.
- Song, X., Sui, A., and Garen, A. (2004). Binding of mouse VL30 retrotransposon RNA to PSF protein induces genes repressed by PSF: effects on steroidogenesis and oncogenesis. *Proc. Natl. Acad. Sci. U. S. A.* 101, 621–626.
- Spector, D.L., Ochs, R.L., and Busch, H. (1984). Silver staining, immunofluorescence, and immunoelectron microscopic localization of nucleolar phosphoproteins B23 and C23. *Chromosoma* 90, 139–148.
- Spiekermann, K., Dirschinger, R.J., Schwab, R., Bagrintseva, K., Faber, F., Buske, C., Schnittger, S., Kelly, L.M., Gilliland, D.G., and Hiddemann, W. (2003). The protein tyrosine kinase inhibitor SU5614 inhibits FLT3 and induces growth arrest and apoptosis in AML-derived cell lines expressing a constitutively activated FLT3. *Blood* 101, 1494–1504.
- Srikantan, V., Zou, Z., Petrovics, G., Xu, L., Augustus, M., Davis, L., Livezey, J.R., Connell, T., Sesterhenn, I.A., Yoshino, K., et al. (2000). PCGEM1, a prostate-specific gene, is overexpressed in prostate cancer. *Proc. Natl. Acad. Sci. U. S. A.* 97, 12216–12221.
- Steck, E., Boeuf, S., Gabler, J., Werth, N., Schnatzer, P., Diederichs, S., and Richter, W. (2012). Regulation of H19 and its encoded microRNA-675 in osteoarthritis and under anabolic and catabolic in vitro conditions. *J. Mol. Med. (Berl)*. 90, 1185–1195.
- Struhl, K. (2007). Transcriptional noise and the fidelity of initiation by RNA polymerase II. *Nat. Struct. Mol. Biol.* 14, 103–105.
- Sun, J., Li, W., Sun, Y., Yu, D., Wen, X., Wang, H., Cui, J., Wang, G., Hoffman, A.R., and Hu, J.-F. (2014a). A novel antisense long noncoding RNA within the IGF1R gene locus is imprinted in hematopoietic malignancies. *Nucleic Acids Res.* 42, 9588–9601.
- Sun, M., Jin, F., Xia, R., Kong, R., Li, J., Xu, T., Liu, Y., Zhang, E., Liu, X., and De, W. (2014b). Decreased expression of long noncoding RNA GAS5 indicates a poor prognosis and promotes cell proliferation in gastric cancer. *BMC Cancer* 14, 319.
- Swaminathan, V., Kishore, A.H., Febitha, K.K., and Kundu, T.K. (2005). Human histone chaperone nucleophosmin enhances acetylation-dependent chromatin transcription. *Mol. Cell. Biol.* 25, 7534–7545.
- Takahashi, Y., Sawada, G., Kurashige, J., Uchi, R., Matsumura, T., Ueo, H., Takano, Y., Eguchi, H., Sudo, T., Sugimachi, K., et al. (2014). Amplification of PVT-1 is involved in poor prognosis via apoptosis inhibition in colorectal cancers. *Br. J. Cancer* 110, 164–171.
- Tallman, M.S. (2005). New strategies for the treatment of acute myeloid leukemia including antibodies and other novel agents. *Hematology Am. Soc. Hematol. Educ. Program* 143–150.

- Tano, K., Mizuno, R., Okada, T., Rakwal, R., Shibato, J., Masuo, Y., Ijiri, K., and Akimitsu, N. (2010). MALAT-1 enhances cell motility of lung adenocarcinoma cells by influencing the expression of motility-related genes. *FEBS Lett.* 584, 4575–4580.
- Tell, G., Wilson, D.M., and Lee, C.H. (2010). Intrusion of a DNA repair protein in the RNome world: is this the beginning of a new era? *Mol. Cell. Biol.* 30, 366–371.
- Tessema, M., Länger, F., Bock, O., Seltsam, A., Metzsig, K., Hasemeier, B., Kreipe, H., and Lehmann, U. (2005). Down-regulation of the IGF-2/H19 locus during normal and malignant hematopoiesis is independent of the imprinting pattern. *Int. J. Oncol.* 26, 499–507.
- Thiede, C., Steudel, C., Mohr, B., Schaich, M., Schäkel, U., Platzbecker, U., Wermke, M., Bornhäuser, M., Ritter, M., Neubauer, A., et al. (2002). Analysis of FLT3-activating mutations in 979 patients with acute myelogenous leukemia: association with FAB subtypes and identification of subgroups with poor prognosis. *Blood* 99, 4326–4335.
- Thomson, P.A., MacIntyre, D.J., Hamilton, G., Dominiczak, A., Smith, B.H., Morris, A., Evans, K.L., and Porteous, D.J. (2013). Association of DISC1 variants with age of onset in a population-based sample of recurrent major depression. *Mol. Psychiatry* 18, 745–747.
- Thrash-Bingham, C.A., and Tartof, K.D. (1999). aHIF: a natural antisense transcript overexpressed in human renal cancer and during hypoxia. *J. Natl. Cancer Inst.* 91, 143–151.
- Tiberi, G., Pekowska, A., Oudin, C., Ivey, A., Autret, A., Prebet, T., Koubi, M., Lembo, F., Mozziconacci, M.-J., Bidaut, G., et al. (2015). PcG methylation of the HIST1 cluster defines an epigenetic marker of acute myeloid leukemia. *Leukemia* 29, 1202–1206.
- Tominaga-Yamanaka, K., Abdelmohsen, K., Martindale, J.L., Yang, X., Taub, D.D., and Gorospe, M. (2012). NF90 coordinately represses the senescence-associated secretory phenotype. *Aging (Albany, NY)*. 4, 695–708.
- Tomonaga, T., Matsushita, K., Yamaguchi, S., Oohashi, T., Shimada, H., Ochiai, T., Yoda, K., and Nomura, F. (2003). Overexpression and mistargeting of centromere protein-A in human primary colorectal cancer. *Cancer Res.* 63, 3511–3516.
- Trimarchi, T., Bilal, E., Ntziachristos, P., Fabbri, G., Dalla-Favera, R., Tsigirgos, A., and Aifantis, I. (2014). Genome-wide mapping and characterization of Notch-regulated long noncoding RNAs in acute leukemia. *Cell* 158, 593–606.
- Tripathi, V., Ellis, J.D., Shen, Z., Song, D.Y., Pan, Q., Watt, A.T., Freier, S.M., Bennett, C.F., Sharma, A., Bubulya, P.A., et al. (2010). The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol. Cell* 39, 925–938.
- Tsai, M.-C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E., and Chang, H.Y. (2010a). Long noncoding RNA as modular scaffold of histone modification complexes. *Science* 329, 689–693.
- Tsai, M.-C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E., and Chang, H.Y. (2010b). Long noncoding RNA as modular scaffold of histone modification complexes. *Science* 329, 689–693.
- Tycowski, K.T., Shu, M.D., and Steitz, J.A. (1996). A mammalian gene with introns instead of exons generating stable RNA products. *Nature* 379, 464–466.
- Tyner, J.W., Erickson, H., Deininger, M.W.N., Willis, S.G., Eide, C.A., Levine, R.L., Heinrich, M.C., Gattermann, N., Gilliland, D.G., Druker, B.J., et al. (2009). High-throughput sequencing screen reveals novel, transforming RAS mutations in myeloid leukemia patients. *Blood* 113, 1749–1755.
- Uchida, T., Rossignol, F., Matthay, M.A., Mounier, R., Couette, S., Clottes, E., and Clerici, C. (2004). Prolonged hypoxia differentially regulates hypoxia-inducible factor (HIF)-1alpha and HIF-2alpha expression in lung epithelial cells: implication of natural antisense HIF-1alpha. *J. Biol. Chem.* 279, 14871–14878.

- Ulveling, D., Francastel, C., and Hubé, F. (2011). Identification of potentially new bifunctional RNA based on genome-wide data-mining of alternative splicing events. *Biochimie* 93, 2024–2027.
- Valdez, B.C., Perlaky, L., Henning, D., Saijo, Y., Chan, P.K., and Busch, H. (1994). Identification of the nuclear and nucleolar localization signals of the protein p120. Interaction with translocation protein B23. *J. Biol. Chem.* 269, 23776–23783.
- Vallot, C., Huret, C., Lesecque, Y., Resch, A., Oudrhiri, N., Bennaceur-Griscelli, A., Duret, L., and Rougeulle, C. (2013). XACT, a long noncoding transcript coating the active X chromosome in human pluripotent cells. *Nat. Genet.* 45, 239–241.
- Vardiman, J.W., Thiele, J., Arber, D.A., Brunning, R.D., Borowitz, M.J., Porwit, A., Harris, N.L., Le Beau, M.M., Hellström-Lindberg, E., Tefferi, A., et al. (2009). The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood* 114, 937–951.
- Vassiliou, G.S., Cooper, J.L., Rad, R., Li, J., Rice, S., Uren, A., Rad, L., Ellis, P., Andrews, R., Banerjee, R., et al. (2011). Mutant nucleophosmin and cooperating pathways drive leukemia initiation and progression in mice. *Nat. Genet.* 43, 470–475.
- Venkatraman, A., He, X.C., Thorvaldsen, J.L., Sugimura, R., Perry, J.M., Tao, F., Zhao, M., Christenson, M.K., Sanchez, R., Yu, J.Y., et al. (2013). Maternal imprinting at the H19-Igf2 locus maintains adult haematopoietic stem cell quiescence. *Nature* 500, 345–349.
- Verhaak, R.G.W., Goudswaard, C.S., van Putten, W., Bijl, M.A., Sanders, M.A., Hagens, W., Uitterlinden, A.G., Erpelinck, C.A.J., Delwel, R., Löwenberg, B., et al. (2005). Mutations in nucleophosmin (NPM1) in acute myeloid leukemia (AML): association with other gene abnormalities and previously established gene expression signatures and their favorable prognostic significance. *Blood* 106, 3747–3754.
- Villegas, V.E., and Zaphiropoulos, P.G. (2015). Neighboring gene regulation by antisense long non-coding RNAs. *Int. J. Mol. Sci.* 16, 3251–3266.
- Wagner, L.A., Christensen, C.J., Dunn, D.M., Spangrude, G.J., Georgelas, A., Kelley, L., Esplin, M.S., Weiss, R.B., and Gleich, G.J. (2007). EGO, a novel, noncoding RNA gene, regulates eosinophil granule protein transcript expression. *Blood* 109, 5191–5198.
- Wan, Y., Kertesz, M., Spitale, R.C., Segal, E., and Chang, H.Y. (2011). Understanding the transcriptome through RNA structure. *Nat. Rev. Genet.* 12, 641–655.
- Wang, D., Baumann, A., Szebeni, A., and Olson, M.O. (1994). The nucleic acid binding activity of nucleolar protein B23.1 resides in its carboxyl-terminal end. *J. Biol. Chem.* 269, 30994–30998.
- Wang, H., Li, W., Guo, R., Sun, J., Cui, J., Wang, G., Hoffman, A.R., and Hu, J.-F. (2014a). An intragenic long noncoding RNA interacts epigenetically with the RUNX1 promoter and enhancer chromatin DNA in hematopoietic malignancies. *Int. J. Cancer* 135, 2783–2794.
- Wang, H.-F., Takenaka, K., Nakanishi, A., and Miki, Y. (2011a). BRCA2 and nucleophosmin coregulate centrosome amplification and form a complex with the Rho effector kinase ROCK2. *Cancer Res.* 71, 68–77.
- Wang, K.C., Yang, Y.W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., Lajoie, B.R., Protacio, A., Flynn, R.A., Gupta, R.A., et al. (2011b). A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* 472, 120–124.
- Wang, P., Xue, Y., Han, Y., Lin, L., Wu, C., Xu, S., Jiang, Z., Xu, J., Liu, Q., and Cao, X. (2014b). The STAT3-binding long noncoding RNA *lnc-DC* controls human dendritic cell differentiation. *Science* 344, 310–313.
- Wang, Y., Xu, Z., Jiang, J., Xu, C., Kang, J., Xiao, L., Wu, M., Xiong, J., Guo, X., and Liu, H. (2013). Endogenous miRNA sponge *lincRNA-RoR* regulates *Oct4*, *Nanog*, and *Sox2* in human embryonic stem cell self-renewal. *Dev. Cell* 25, 69–80.

- Wang, Y., Xiao, M., Chen, X., Chen, L., Xu, Y., Lv, L., Wang, P., Yang, H., Ma, S., Lin, H., et al. (2015a). *WT1 recruits TET2 to regulate its target gene expression and suppress leukemia cell proliferation*. *Mol. Cell* 57, 662–673.
- Wang, Y., Zhong, H., Xie, X., Chen, C.Y., Huang, D., Shen, L., Zhang, H., Chen, Z.W., and Zeng, G. (2015b). *Long noncoding RNA derived from CD244 signaling epigenetically controls CD8+ T-cell immune responses in tuberculosis infection*. *Proc. Natl. Acad. Sci. U. S. A.*
- Wang, Y.P., Yi, S., Wen, L., Zhang, B.P., Zhao, Z., Hu, J.Y., Zhao, F., He, J., Fang, J., Zhang, C., et al. (2014c). *Nontoxic-dose of deguelin induce NPMc+ AML cell differentiation by selectively targeting Mt NPM1/SIRT1 instead of HDAC1/3*. *Curr. Cancer Drug Targets* 14, 685–699.
- Wanzel, M., Russ, A.C., Kleine-Kohlbrecher, D., Colombo, E., Pelicci, P.-G., and Eilers, M. (2008). *A ribosomal protein L23-nucleophosmin circuit coordinates Miz1 function with cell growth*. *Nat. Cell Biol.* 10, 1051–1061.
- Weinberg, R.A., and Penman, S. (1968). *Small molecular weight monodisperse nuclear RNA*. *J. Mol. Biol.* 38, 289–304.
- Willingham, A.T., Orth, A.P., Batalov, S., Peters, E.C., Wen, B.G., Aza-Blanc, P., Hogenesch, J.B., and Schultz, P.G. (2005). *A strategy for probing the function of noncoding RNAs finds a repressor of NFAT*. *Science* 309, 1570–1573.
- Wilusz, J.E., Freier, S.M., and Spector, D.L. (2008). *3' end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA*. *Cell* 135, 919–932.
- Wu, G., Broniscer, A., McEachron, T.A., Lu, C., Paugh, B.S., Becksfors, J., Qu, C., Ding, L., Huether, R., Parker, M., et al. (2012a). *Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas*. *Nat. Genet.* 44, 251–253.
- Wu, M.H., Chang, J.H., and Yung, B.Y.M. (2002). *Resistance to UV-induced cell-killing in nucleophosmin/B23 over-expressed NIH 3T3 fibroblasts: enhancement of DNA repair and up-regulation of PCNA in association with nucleophosmin/B23 over-expression*. *Carcinogenesis* 23, 93–100.
- Wu, Q., Qian, Y.-M., Zhao, X.-L., Wang, S.-M., Feng, X.-J., Chen, X.-F., and Zhang, S.-H. (2012b). *Expression and prognostic significance of centromere protein A in human lung adenocarcinoma*. *Lung Cancer* 77, 407–414.
- Xie, C., Yuan, J., Li, H., Li, M., Zhao, G., Bu, D., Zhu, W., Wu, W., Chen, R., and Zhao, Y. (2014). *NONCODEv4: exploring the world of long non-coding RNA genes*. *Nucleic Acids Res.* 42, D98–D103.
- Yakovchuk, P., Goodrich, J.A., and Kugel, J.F. (2014). *B2 RNA represses TFIIH phosphorylation of RNA polymerase II*. *Transcription* 2, 45–49.
- Yamada, K., Kano, J., Tsunoda, H., Yoshikawa, H., Okubo, C., Ishiyama, T., and Noguchi, M. (2006). *Phenotypic characterization of endometrial stromal sarcoma of the uterus*. *Cancer Sci.* 97, 106–112.
- Yan, J., Friedrich, S., and Kurgan, L. (2015). *A comprehensive comparative review of sequence-based predictors of DNA- and RNA-binding residues*. *Brief. Bioinform.*
- Yang, F., Huo, X., Yuan, S., Zhang, L., Zhou, W., Wang, F., and Sun, S. (2013). *Repression of the long noncoding RNA-LET by histone deacetylase 3 contributes to hypoxia-mediated metastasis*. *Mol. Cell* 49, 1083–1096.
- Yang, Z., Zhou, L., Wu, L.-M., Lai, M.-C., Xie, H.-Y., Zhang, F., and Zheng, S.-S. (2011). *Overexpression of long non-coding RNA HOTAIR predicts tumor recurrence in hepatocellular carcinoma patients following liver transplantation*. *Ann. Surg. Oncol.* 18, 1243–1250.
- Yap, K.L., Li, S., Muñoz-Cabello, A.M., Raguz, S., Zeng, L., Mujtaba, S., Gil, J., Walsh, M.J., and Zhou, M.-M. (2010). *Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb*

CBX7 in transcriptional silencing of *INK4a*. *Mol. Cell* 38, 662–674.

Yarden, Y., Kuang, W.J., Yang-Feng, T., Coussens, L., Munemitsu, S., Dull, T.J., Chen, E., Schlessinger, J., Francke, U., and Ullrich, A. (1987). Human proto-oncogene *c-kit*: a new cell surface receptor tyrosine kinase for an unidentified ligand. *EMBO J.* 6, 3341–3351.

Yi, S., Wen, L., He, J., Wang, Y., Zhao, F., Zhao, J., Zhao, Z., Cui, G., and Chen, Y. (2015a). Deguelin, a selective silencer of the *NPM1* mutant, potentiates apoptosis and induces differentiation in AML cells carrying the *NPM1* mutation. *Ann. Hematol.* 94, 201–210.

Yi, S., Wen, L., He, J., Wang, Y., Zhao, F., Zhao, J., Zhao, Z., Cui, G., and Chen, Y. (2015b). Deguelin, a selective silencer of the *NPM1* mutant, potentiates apoptosis and induces differentiation in AML cells carrying the *NPM1* mutation. *Ann. Hematol.* 94, 201–210.

Yildirim, E., Kirby, J.E., Brown, D.E., Mercier, F.E., Sadreyev, R.I., Scadden, D.T., and Lee, J.T. (2013). *Xist* RNA is a potent suppressor of hematologic cancer in mice. *Cell* 152, 727–742.

Yin, Q.-F., Yang, L., Zhang, Y., Xiang, J.-F., Wu, Y.-W., Carmichael, G.G., and Chen, L.-L. (2012). Long noncoding RNAs with *snoRNA* ends. *Mol. Cell* 48, 219–230.

Yoon, J.-H., Abdelmohsen, K., Kim, J., Yang, X., Martindale, J.L., Tominaga-Yamanaka, K., White, E.J., Orjalo, A. V., Rinn, J.L., Kreft, S.G., et al. (2013). Scaffold function of long non-coding RNA *HOTAIR* in protein ubiquitination. *Nat. Commun.* 4, 2939.

Yotova, I.Y., Vlatkovic, I.M., Pauler, F.M., Warczok, K.E., Ambros, P.F., Oshimura, M., Theussl, H.-C., Gessler, M., Wagner, E.F., and Barlow, D.P. (2008). Identification of the human homolog of the imprinted mouse *Air* non-coding RNA. *Genomics* 92, 464–473.

You, B.J., Huang, I.J., Liu, W.H., Hung, Y.B., Chang, J.H., and Yung, B.Y. (1999). Decrease in nucleophosmin/B23 mRNA and telomerase activity during indomethacin-induced apoptosis of gastric KATO-III cancer cells. *Naunyn. Schmiedebergs. Arch. Pharmacol.* 360, 683–690.

Yu, Y., Maggi, L.B., Brady, S.N., Apicelli, A.J., Dai, M.-S., Lu, H., and Weber, J.D. (2006). Nucleophosmin is essential for ribosomal protein L5 nuclear export. *Mol. Cell. Biol.* 26, 3798–3809.

Zatsepina, O. V, Rousselet, A., Chan, P.K., Olson, M.O., Jordan, E.G., and Bornens, M. (1999). The nucleolar phosphoprotein B23 redistributes in part to the spindle poles during mitosis. *J. Cell Sci.* 112 (Pt 4, 455–466.

Zhang, D.E., Zhang, P., Wang, N.D., Hetherington, C.J., Darlington, G.J., and Tenen, D.G. (1997). Absence of granulocyte colony-stimulating factor signaling and neutrophil development in CCAAT enhancer binding protein alpha-deficient mice. *Proc. Natl. Acad. Sci. U. S. A.* 94, 569–574.

Zhang, E., Kong, R., Yin, D., You, L., Sun, M., Han, L., Xu, T., Xia, R., Yang, J., De, W., et al. (2014a). Long noncoding RNA *ANRIL* indicates a poor prognosis of gastric cancer and promotes tumor growth by epigenetically silencing of *miR-99a/miR-449a*. *Oncotarget* 5, 2276–2292.

Zhang, H., Nestor, C.E., Zhao, S., Lentini, A., Bohle, B., Benson, M., and Wang, H. (2013). Profiling of human CD4+ T-cell subsets identifies the TH2-specific noncoding RNA *GATA3-AS1*. *J. Allergy Clin. Immunol.* 132, 1005–1008.

Zhang, H., Zeitz, M.J., Wang, H., Niu, B., Ge, S., Li, W., Cui, J., Wang, G., Qian, G., Higgins, M.J., et al. (2014b). Long noncoding RNA-mediated intrachromosomal interactions promote imprinting at the *Kcnq1* locus. *J. Cell Biol.* 204, 61–75.

Zhang, L., Xu, H.-G., and Lu, C. A novel long non-coding RNA *T-ALL-R-LncR1* knockdown and *Par-4* cooperate to induce cellular apoptosis in T-cell acute lymphoblastic leukemia cells. <http://dx.doi.org.gate2.inist.fr/10.3109/10428194.2013.829574>.

Zhang, X., Lian, Z., Padden, C., Gerstein, M.B., Rozowsky, J., Snyder, M., Gingeras, T.R., Kapranov, P., Weissman, S.M., and Newburger, P.E. (2009). A myelopoiesis-associated regulatory intergenic noncoding RNA transcript within the human HOXA cluster. *Blood* 113, 2526–2534.

Zhang, X., Weissman, S.M., and Newburger, P.E. (2014c). Long intergenic non-coding RNA HOTAIRM1 regulates cell cycle progression during myeloid maturation in NB4 human promyelocytic leukemia cells. *RNA Biol.* 11, 777–787.

Zhao, H., Zhang, X., Frazão, J.B., Condino-Neto, A., and Newburger, P.E. (2013). HOX antisense lincRNA HOXA-AS2 is an apoptosis repressor in all trans retinoic acid treated NB4 promyelocytic leukemia cells. *J. Cell. Biochem.* 114, 2375–2383.

Zhao, J., Sun, B.K., Erwin, J.A., Song, J.-J., and Lee, J.T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* 322, 750–756.

Zirwes, R.F., Kouzmenko, A.P., Peters, J.M., Franke, W.W., and Schmidt-Zachmann, M.S. (1997). Topogenesis of a nucleolar protein: determination of molecular segments directing nucleolar association. *Mol. Biol. Cell* 8, 231–248.

Ziv, O., Zeisel, A., Mirlas-Neisberg, N., Swain, U., Nevo, R., Ben-Chetrit, N., Martelli, M.P., Rossi, R., Schiesser, S., Canman, C.E., et al. (2014). Identification of novel DNA-damage tolerance genes reveals regulation of translesion DNA synthesis by nucleophosmin. *Nat. Commun.* 5, 5437.

Estimation nationale de l'incidence des cancers en France entre 1980 et 2012.

AUTEUR : Etienne DE CLARA

TITRE : Etude des longs ARNs non codants dans la leucémie aiguë myéloblastique à caryotype normal

DIRECTEURS DE THESE : Professeur Pierre BROUSSET et Dr. Marina BOUSQUET

LIEU ET DATE DE SOUTENANCE : Le 26 novembre 2015, Toulouse

RESUME:

Les longs ARN non codants (lncRNAs) sont définis comme des transcrits de plus de 200nt et n'ayant pas de potentiel codant. Des études récentes ont démontré que les lncRNAs pouvaient être impliqués dans la régulation de la transcription, de la traduction, de la différenciation cellulaire, de l'expression génique, du cycle cellulaire et des modifications de la chromatine. De plus, il a été montré un impact fonctionnel de certains lncRNAs dans le processus de cancérogenèse mais nos connaissances actuelles sur ces molécules dans le cancer, et plus particulièrement dans la leucémie, restent extrêmement limitées. Au cours de cette étude, nous avons analysé l'expression des lncRNAs par RNA-sequencing sur 40 patients atteints de leucémie aiguë myéloblastique (LAM) à caryotype normal. Parmi les 11065 lncRNAs exprimés dans nos échantillons, nous avons identifié une signature de lncRNAs associée à la mutation de NPM1. Afin de mettre en évidence les fonctions putatives des lncRNAs sélectionnés, nous avons utilisé un algorithme de prédiction d'interaction protéine/ARN. De manière intéressante, plus de la moitié des lncRNAs présentent des sites d'interactions potentiels à SUZ12, une sous unité du complexe PRC2 (Polycomb repressive complex 2), connu pour être recruté par des lncRNAs pour la régulation épigénétique de gènes cibles. Par RNA immunoprécipitation (RIP) de SUZ12, nous avons pu démontrer que le lncRNA XLOC_087120 interagissait avec SUZ12. De plus, son expression est anti-corrélée avec celle des gènes voisins codants des histones, suggérant un rôle dans la régulation négative des histones par ce lncRNA. L'impact de la dérégulation de XLOC_087120 sur les histones a été confirmé par des expériences de surexpression et d'inhibition de ce lncRNA dans des lignées de LAM. De plus, même si la mutation NPM1 ne semble pas affecter directement l'expression de ce lncRNA, des expériences d'infection de la forme mutée de NPM1 dans une lignée LAM ont montré que NPM1 pourrait réguler la localisation nucléaire/cytoplasmique de XLOC_087120 et moduler sa fonction de répresseur. En conclusion, ces données suggèrent que les lncRNAs sont des facteurs clés dans la pathogenèse des LAMs.

MOTS-CLES : Leucémie aiguë myéloblastique ; long ARN non codant ; RNA-sequencing ; Régulation Epigénétique.

TITLE: Study of long non coding RNAs in acute myeloid leukemia with normal karyotype

ABSTRACT:

Long noncoding RNAs (lncRNAs) are defined as RNA transcripts that are larger than 200 nt but do not appear to have protein-coding potential. Recent studies have demonstrated that lncRNAs regulate many processes such as transcription, translation, cellular differentiation, gene expression regulation, cell cycle regulation, and chromatin modification. Cumulative evidence points towards an important role of lncRNAs in cancer initiation, development, and progression. However, our overall knowledge of lncRNAs in cancer, including leukemia, remains extremely limited. In this study, we investigated lncRNA expression by RNA-sequencing in 40 acute myeloid leukemia (AML) patients with normal karyotype. Among 11065 lncRNA expressed in our samples, we identified specific lncRNA signature associated with the presence of NPM1 mutation. To go further into the putative function of these lncRNAs, we used catRAPID Omics algorithm to predict potential protein partners. Interestingly, the majority of the selected lncRNAs contains putative SUZ12 binding sites, a PRC2 (Polycomb Repressive Complex 2) component known to be linked to lncRNAs and to epigenetically regulates target genes. By using SUZ12 RNA Immunoprecipitation, we identify one lncRNA named XLOC_087120 linked to SUZ12. XLOC_087120 is located in a region enriched in histone genes. Pearson correlation showed a significative anti-correlation between XLOC_087120 and histone neighboring coding gene expression suggesting a role of this lncRNA in the regulation of histone genes. The impact on histone genes expression was confirmed by overexpression and inhibition of XLOC_087120 in AML cell lines. Overexpression of NPM1 mutant in an AML cell line showed that NPM1 modulates the nuclear/cytoplasmic localization of XLOC_087120 and consequently its repressive function. Altogether, these data suggest that lncRNAs should be considered as key players in the pathogenesis of acute myeloid leukemias.

KEY WORDS: Acute Myeloid Leukemia; long noncoding RNA; RNA-sequencing; Epigenetic regulation.

DISCIPLINE ADMINISTRATIVE: Cancérologie

LABORATOIRE: Centre de Recherche en Cancérologie de Toulouse (CRCT), Inserm 1037, 2 avenue Hubert Curien, Oncopôle entrée C, CS 53717, 31037 Toulouse CEDEX 1