



HAL
open science

Evaluation du rôle des longs ARN non codants dans les carcinomes mammaires infiltrants

Didier Meseure

► **To cite this version:**

Didier Meseure. Evaluation du rôle des longs ARN non codants dans les carcinomes mammaires infiltrants. Cancer. Université Paris Saclay (COMUE), 2016. Français. NNT : 2016SACLS471 . tel-01942165

HAL Id: tel-01942165

<https://theses.hal.science/tel-01942165>

Submitted on 3 Dec 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.

NNT : 2016SACLS471

THESE DE DOCTORAT
DE
L'UNIVERSITE PARIS-SACLAY
PREPAREE A
L'UNIVERSITE PARIS-SUD

ECOLE DOCTORALE N° 582
CBMS Cancérologie : biologie - médecine - santé

Thèse de Doctorat de Sciences
Aspects moléculaires et cellulaires de la biologie

Par

Didier MESEURE

**Evaluation du rôle des longs ARNs non codants dans les
carcinomes mammaires infiltrants**

Thèse présentée et soutenue à l'Institut Curie le 2 décembre 2016

Composition du Jury :

Dr Patricia de CREMOUX, Hopital St Louis	Rapporteur
Dr Alexandra HENRION-CAUDE, Hopital Necker	Rapporteur
Dr Antonin MORILLON, Institut Curie	Examineur
Pr Jean-Yves PIERGA, Institut Curie	Président
Pr Yves ALLORY, Hopital Henri Mondor	Examineur
Pr Jean-Christophe SABOURIN, Hopital Charles Nicolle Rouen	Examineur
Dr Ivan BIECHE, Institut Curie	Directeur de thèse

REMERCIEMENTS

Je tiens tout d'abord à remercier mon directeur de thèse, le Dr Ivan BIECHE, pour m'avoir chaleureusement accueilli et dirigé durant ce travail. J'ai pu apprécier pendant ces trois années sa grande rigueur scientifique et sa disponibilité permanente. Merci d'avoir courageusement tenté de reprogrammer un morphologiste convaincu en un biopathologiste passionné par l'épigénétique et le microenvironnement. J'adresse également mes remerciements à l'équipe du laboratoire de Pharmacogénomique, et plus particulièrement à Rosette LIDEREAU de l'Hopital René Huguenin, et à Sophie VACHER, pour sa disponibilité lors de nos nombreux échanges technique et scientifiques.

Je remercie Patricia de CREMOUX, Alexandra HENRION-CAUDE, Antonin MORILLON, Jean-Yves PIERGA, Yves ALLORY et Jean-Christophe SABOURIN, de m'avoir fait l'honneur de constituer le jury de thèse et d'examiner mon travail.

Je souhaite exprimer mes remerciements au Comité Médico-Scientifique de l'Institut Curie et au SIRIC, qui ont autorisé le soutien institutionnel de mon temps médical protégé, et plus particulièrement à Olivier DELATTRE, Geneviève ALMOUZNI, Mounira AMOR-GUERET et Fatima MECHTA-GRIGORIOU.

Je tiens également à remercier mes collègues pathologistes et les équipes des deux sites de Saint-Cloud et de Paris de l'Institut Curie, et plus particulièrement Jean-Marc GUINEBRETTIERE, Xavier SASTRE et Anne VINCENT-SALOMON, qui ont su m'encourager, me soutenir et m'aider à participer au développement de la Pathologie Expérimentale à l'Institut Curie.

J'adresse tout particulièrement mes remerciements à l'équipe de la Plateforme de Pathologie Expérimentale et à tous les chercheurs qui nous accordent leur confiance. A André NICOLAS, pour sa personnalité solaire, son expérience professionnelle irremplaçable et sa créativité et à Laetitia LESAGE, Renaud LECLERE et Gabriel CHAMPENOIS, qui se sont rapidement impliqués avec passion dans l'accompagnement des projets de recherche. A Sergio ROMAN-ROMAN, Vassili SOUMELIS et à Michel HUERRE, pour le soutien permanent qu'ils apportent au développement de la Plateforme.

Merci à ma famille et à nos amis, pour leur soutien et leur patience. A mon épouse Catherine, à ma fille Lara-Dévi et à ma mère. A Theth, Hour Chantha, To Riya, Daniel, Kinan, Jean-François, Muriel, Jean-Luc, Geneviève, Pierre, Anaïs, Laetitia, Nicole et Régine.

A mon père et à mon maître, le Pr Jean KERMAREC, je leur dédie cette thèse.

Si nous cherchons à contempler le miroir en soi, nous ne découvrons finalement rien d'autre que les choses qui s'y reflètent. Si nous voulons saisir les choses, nous n'atteignons finalement rien d'autre que le miroir. Telle est l'histoire universelle de la connaissance.

Friedrich Nietzsche, *Aurore*.

SOMMAIRE

REMERCIEMENTS.....	2
SOMMAIRE.....	3
ABREVIATIONS.....	11
PRINCIPAUX LONGS ARNs NON CODANTS.....	12
FIGURES.....	15
TABLEAUX.....	17
RESUME.....	18
ABSTRACT.....	19
AVANT-PROPOS.....	20
CHAPITRE I : INTRODUCTION.....	21
1 GENERALITES SUR LES CANCERS DU SEIN.....	23
1.1 Fréquence.....	23
1.2 Facteurs de risque.....	25
1.2.1 Facteurs hormonaux.....	25
1.2.2 Facteurs liés à la reproduction.....	25
1.2.3 Facteurs génétiques.....	26
1.2.4 Autres facteurs de risque.....	27
1.3 Anatomie, histologie et physiologie de la glande mammaire.....	28
1.3.1 Anatomie.....	28
1.3.2 Histologie.....	29
1.3.3 Physiologie.....	30
1.4 Classifications clinique, morphologique et immunophénotypique.....	31
1.4.1 Classification clinique.....	31
1.4.2 Classification morphologique.....	32
1.4.2.1 Types histopathologiques.....	32
1.4.2.2 Classification PTNM.....	34
1.4.2.3 Grade et index histopronostiques.....	34
1.4.2.4 Autres facteurs histopronostiques.....	35

1.4.3 Classification immunophénotypique.....	35
1.5 Classifications moléculaires.....	38
1.5.1 Biologie moléculaire des cancers du sein.....	38
1.5.1.1 Concept actuel de la cancérogénèse.....	38
1.5.1.2 Caractéristiques des cellules cancéreuses.....	41
1.5.1.3 Mécanismes moléculaires de la cancérogénèse mammaire.....	41
1.5.1.3.1 Oncogènes et gènes onco-suppresseurs.....	41
1.5.1.3.2 Instabilité génomique et anomalies génétiques/épigénétiques.....	43
1.5.1.3.2.1 Instabilité génomique.....	43
1.5.1.3.2.2 Anomalies génétiques/épigénétiques.....	44
1.5.1.3.2.3 Chromotripsis.....	46
1.5.1.3.3 Anomalies des voies de signalisation.....	47
1.5.2 Classifications moléculaires des cancers du sein.....	49
1.5.2.1 Classifications génomiques et transcriptomiques.....	49
1.5.2.1.1 Classification transcriptomique intrinsèque.....	49
1.5.2.1.2 Classification génomique et transcriptomique.....	52
1.5.2.2 Classification moléculaire intégrée des cancers du sein.....	53
1.5.2.3 Classification moléculaire intégrée des cancers du sein triple-négatifs.....	57
2 LES LONGS ARNs NON CODANTS	59
2.1 Nouvelle organisation fonctionnelle du génome	59
2.2 ARNs non codants.....	61
2.2.1 Petits ARNs non codants.....	62
2.2.2 Longs ARNs non codants.....	64
2.3 Longs ARNs non codants.....	64
2.3.1 Définition des longs ARNs non codants.....	64
2.3.2 Méthodes d'investigation et d'annotation des lncRNAs.....	66
2.3.3 Caractéristiques des lncRNAs	68
2.3.3.1 Structures biochimiques.....	68
2.3.3.2 Organisation structurale et fonctionnelle.....	69
2.3.3.3 Organisation topographique.....	73

2.3.3.4 Taille.....	79
2.3.3.5 Epissage alternatif	80
2.3.3.6 Niveaux d'expression et spécificité.....	81
2.3.3.7 Stabilité.....	82
2.3.3.8 Localisation tissulaire et cellulaire.....	82
2.3.3.9 Régulation.....	84
2.3.3.10 Conservation.....	84
2.3.3.11 Evolution.....	85
2.3.3.12 Origine.....	86
2.3.3.13 Classification fonctionnelle.....	87
2.3.4 Mécanismes d'action des lncRNAs.....	89
2.3.4.1 Régulation de l'expression des gènes.....	89
2.3.4.2 Reconnaissance de sites et de séquences fonctionnelles génomiques.....	93
2.3.4.3 Edification de domaines intranucléaires.....	93
2.3.5 Fonctions moléculaires des lncRNAs.....	93
2.3.5.1 Généralités.....	93
2.3.5.2 Fonctions régulatrices des lncRNAs.....	94
2.3.5.2.1 Régulation au niveau épigénétique.....	97
2.3.5.2.1.1 Expression allélique.....	97
2.3.5.2.1.1.1 Inactivation du chromosome X.....	97
2.3.5.2.1.1.2 Empreinte génomique parentale.....	97
2.3.5.2.1.2 Positionnement des nucléosomes.....	98
2.3.5.2.1.3 Méthylation de l'ADN.....	99
2.3.5.2.1.4 Modification du code histone.....	99
2.3.5.2.2 Régulation au niveau de l'ADN.....	100
2.3.5.2.3 Régulation au niveau de l'ARN.....	101
2.3.5.2.4 Régulation au niveau des protéines.....	103

2.3.5.2.5 Régulation complexe à de multiples niveaux.....	104
2.3.5.2.6 Régulation inductible.....	105
2.3.5.3 Fonctions structurales des lncRNAs.....	105
2.3.5.3.1 Stabilité génomique et chromosomique.....	106
2.3.5.3.1.1 Domaines régulateurs intrachromosomiques.....	106
2.3.5.3.1.2 Domaines régulateurs interchromosomiques.....	106
2.3.5.3.1.3 Boucles d'interaction <i>enhancer</i> -promoteur.....	106
2.3.5.3.2 Edification de corps intra-nucléaires.....	107
2.3.5.4 Biogénèse de petits ARNs non codants.....	108
2.3.5.5 lncRNAs circulants.....	109
2.3.6 Fonctions cellulaires des lncRNAs.....	109
2.3.6.1 Développement embryonnaire, organogénèse et équilibre homéostatique.....	109
2.3.6.2 Rôles des lncRNAs dans le développement et la physiologie mammaire.....	114
2.3.6.2.1 lncRNAs régulant le développement de la glande mammaire.....	114
2.3.6.2.2 lncRNAs contrôlant la différenciation et l'homéostasie mammaire.....	115
3 LES LONGS ARNs NON CODANTS ET LE CANCER.....	116
3.1 Généralités.....	116
3.2 lncRNAs et mécanismes de la cancérogénèse.....	117
3.2.1 lncRNAs et génétique du cancer.....	117
3.2.2 lncRNAs et épigénétique du cancer.....	118
3.2.2.1 Généralités sur l'épigénétique du cancer.....	118
3.2.2.2 Les lncRNAs comme régulateurs épigénétiques.....	119
3.2.3 lncRNAs et <i>hallmarks</i> du cancer.....	120
3.2.4 lncRNAs et voies de signalisation du cancer.....	121
3.2.5 lncRNAs, progression tumorale et évolution métastatique.....	122

3.2.6 LncRNAs et interactions avec les différentes classes d'ARNs.....	123
3.2.7 LncRNAs, sensibilité et résistance aux thérapies anti-cancéreuses.....	123
4 LES LONGS ARNs NON CODANTS DANS LE CANCER DU SEIN.....	127
4.1 Dérégulation des lncRNAs dans les cancers du sein.....	127
4.1.1 Dérégulation globale des lncRNAs dans les cancers du sein.....	127
4.1.2 LncRNAs oncogéniques et oncosuppresseurs.....	130
4.1.2.1 LncRNAs oncogéniques.....	130
4.1.2.1.1 Prolifération cellulaire et apoptose.....	130
<i>UCA1</i>	130
<i>LSINCT5</i>	130
<i>LncRNA-Smad7</i>	131
<i>LOC554202</i>	131
<i>lncRNA-JADE</i>	131
<i>PVT1</i>	132
4.1.2.1.2 Invasion et évolution métastatique.....	133
<i>treRNA</i>	133
<i>HOTAIR</i>	133
<i>BCYRN1/BC200</i>	134
<i>EFNA3</i>	135
<i>NEAT</i>	135
<i>CCAT2</i>	136
4.1.2.1.3 Acquisition de caractères de cellules souches.....	137
<i>SOX2OT</i>	137
<i>FAL1</i>	137

4.1.2.2 LncRNAs oncosuppresseurs.....	137
4.1.2.2.1 Prolifération et apoptose.....	138
<i>GAS5</i>	138
<i>PTENP1</i>	138
<i>Zfas1</i>	139
4.1.2.2.2 Invasion et évolution métastatique.....	139
<i>Xist</i>	139
<i>NKILA</i>	140
<i>LIMT</i>	141
4.1.2.3 LncRNAs controversés dans la cancérogénèse mammaire.....	141
<i>H19</i>	141
<i>SRA</i>	143
4.2 Rôles des lncRNAs au cours des étapes de la progression tumorale.....	144
4.3 Classifications moléculaires des cancers du sein intégrant des lncRNAs.....	147
5 LES LONGS ARN NON CODANTS COMME BIOMARQUEURS ET CIBLES THERAPEUTIQUES POTENTIELS.....	150
5.1 Les lncRNAs comme biomarqueurs.....	150
5.1.1 Généralités.....	150
5.1.2 Les lncRNAs comme biomarqueurs dans différents autres types de cancers.....	150
5.2 Les lncRNAs comme cibles thérapeutiques potentielles.....	157
5.2.1 Généralités.....	157
5.2.2 Techniques de ciblage thérapeutique des lncRNAs.....	157
6 PERSPECTIVES.....	161

CHAPITRE II : PROBLEMATIQUE DE LA RECHERCHE	163
CHAPITRE III : RESULTATS.....	183
ARTICLE 1 : Analyse de l'expression des gènes appartenant au complexe ANRIL - Polycomb - CDKN2B / CDKN2A / ARF dans les tumeurs du sein : identification d'une signature pronostique indépendante EZH2 / CBX7.....	185
7.1 Objectif.....	185
7.2 Résumé.....	186
7.3 Résultats.....	186
7.4 Conclusion.....	187
7.5 Article 1.....	188
ARTICLE 2 : Valeur pronostique d'un variant d'épissage alternatif du long ARN non codant MALATI dans les carcinomes mammaires infiltrants.....	223
8.1 Objectif.....	223
8.2 Résumé.....	224
8.3 Résultats.....	224
8.4 Conclusion.....	225
8.5 Article 2.....	226
RESULTATS COMPLEMENTAIRES : Etude préliminaire de l'épitranscriptome non codant des carcinomes mammaires infiltrants.....	249
9.1 Objectif.....	249
9.2 Résumé.....	249
9.3 Résultats.....	249
9.4 Conclusion.....	250
9.5 Etude préliminaire du transcriptome des carcinomes mammaires	251
9.5.1 Matériel.....	251
9.5.2 Protocole d'analyse bioinformatique.....	252

9.5.3 Signatures transcriptomiques.....	252
9.5.3.1 Signature diagnostique globale.....	252
9.5.3.2 Signature diagnostique antisens	258
9.5.3.3 Signature pronostique préliminaire.....	260
9.5.3.4 Validation du <i>set</i> d'intérêt.....	261
9.5.3.5 Conclusion.....	261
CHAPITRE IV : CONCLUSION ET PERSPECTIVES.....	263
BIBLIOGRAPHIE.....	275
ANNEXES.....	339
10.1 Annexe 1: Article collaboratif “Characterization of human antisense long non-coding RNA reveals markers for high risk prostate cancer”.....	341
10.2 Annexe 2: Revue “Long Noncoding RNAs as New Architects in Cancer Epigenetics, Prognostic Biomarkers, and Potential Therapeutic Targets”.....	367
10.3 Annexe 3: Articles soumis, en révision et acceptés pour publication (2013-2016).....	383

ABBREVIATIONS

ASO: *antisense oligonucleotide* **CCI:**
carcinome canalaire invasif **ChIP:**
chromatin immunoprecipitation **ChIP-**
seq: *ChIP sequencing*

CSTN: carcinome du sein triple négatif

EMT: *epithelial-to-mesenchymal transition*

ENCODE: *encyclopedia of DNA elements*

GRO-seq: *global nuclear run on sequencing*

GWAS: *genome-wide association study*

IHC: *immunohistochemistry*

kb: kilobase

lncRNA: *long non-coding RNA*

mRNA: *messenger RNA*

miRNA: *micro RNA*

MREs: *miRNAs response elements*

NAT: *natural antisens transcript*

ncRNA: *non-coding RNA*

nt: nucléotide

PCR: *polymerase chain reaction*

RNA-FISH: *RNA fluorescence in situ hybridization*

RNA-seq: *RNA sequencing*

RPKM: *read per kilobase per million*

RT: *reverse transcript*

siRNA: *small interfering RNA*

PRINCIPAUX LONGS ARNs NON CODANTS

ANRIL/ CDKN2B-AS : *CDKN2B antisense RNA 1* (9p21.3)
AIR: *antisense of IGF2R non-protein coding RNA* (6q25.3)
Anti-NOS2 : *nitric oxide synthase 2 antisense* (17q11.2)
ASIDHRS4: *DHRS4 antisense RNA 1* (14q11.2)
BANCR: *BRAF-activated non-protein coding RNA* (9q21.11)
BCYRN1/BC200: *brain cytoplasmic RNA 1* (2p16)
BRAVEHEART: *braveheart long non-coding RNA* (18; 18 E1)
CTBP1-AS: *CTBP1 antisense RNA* (4p16.3)
CCAT1: *colon cancer associated transcript 1* (8q24.21) **CCAT2**:
colon cancer associated transcript 2 (8q24.21) **CRNDE**:
colorectal neoplasia differentially expressed (16q12.2)
DALI: *DNMT1-associated long intergenic non-coding RNA* (2q12)
DLX1-AS: *distal-less homeobox 1 antisense* (2q32)
EFNA3: *ephrin-A3* (1q21.22)
FAL1/FALEC: *focally amplified long non-coding RNA in epithelial cancer* (chromosome: 1)
FENDRR: *FOXF1 adjacent non-coding developmental regulatory RNA* (16q24.1)
FIRRE: *firre intergenic repeating RNA element* (Xq26.2)
GACAT1: *gastric cancer associated transcript 1* (2q12.3)
GAS5: *growth arrest specific 5* (1q25.1)
GOMAFU/MIAT: *myocardial infarction associated transcript* (22q12.1)
HEIH: *hepatocellular carcinoma up-regulated EZH2-associated* (5q35.3)
HIF1A-AS1: *HIF1A antisense RNA 1* (14q23.2)
HOTAIR: *HOX transcript antisense RNA* (12q13.13)
HOTAIRM1: *HOXA transcript antisense RNA, myeloid-specific 1* (7p15.2)
HOTTIP: *HOXA distal transcript antisense RNA* (7p15.2)
HULC: *hepatocellular carcinoma up-regulated long non-coding RNA* (6p24.3)

H19: H19 imprinted maternally expressed transcript (11p15.5)

KCNQ1OT1: *KCNQ1* opposite strand/antisense transcript 1 (11p15.5)

LINC00461: long intergenic non-protein coding RNA 461 (5q14.3)

LINC-MD: long intergenic non-protein coding RNA (6p12.2)

LincRNA-p21: tumor protein p53 pathway corepressor 1 (6p21.2)

LincRNA-VLDLR: VLDLR antisense RNA 1 (9p24.2)

Linc-ROR: long intergenic non-protein coding RNA, regulator of reprogramming (18q21.31)

Lnc-DC/WFDC21P: WAP four-disulfide core domain 21, pseudogene (17q23.1)

LncRNA-JADE: JADE1 adjacent regulatory RNA (4q28.2)

LncRNA-ES1/ LINC01108: long intergenic non-protein coding RNA 1108 (6p23)

LncRNA-Smad7: SMAD family member 7 (18q21.1)

LSINCT5: long stress-induced non-coding transcript 5 (5p15.33)

MALAT1: metastasis associated lung adenocarcinoma transcript 1 (11q13.1)

MEG3: maternally expressed 3 (14q32.2)

MEG9: maternally expressed 9 (chromosome 14)

MIR31AG/ LOC554202: *MIR31* host gene (9p21.3)

MPINC: pregnancy induced noncoding RNA (1; 1 C3)

NEAT1: nuclear paraspeckle assembly transcript 1 (11q13.1)

NEST: *IFNG* antisense RNA 1 (12q15)

NKILA: NF-kappaB interacting long non-coding RNA (20q13.31)

NRON: non-protein coding RNA, repressor of NFAT (9q33.3)

PACERR: *PTGS2* antisense *NFKB1* complex-mediated expression regulator RNA (1q31.1)

PANDA: promoter of *CDKN1A* antisense DNA damage activated RNA (6p21.2)

PAUPAR: *PAX6* upstream antisense RNA (11p13)

PCAT1: prostate cancer associated transcript 1 (8q24.21)

PCA3: prostate cancer associated 3 (9q21.2)

PCGEM1: *PCGEM1*, prostate-specific transcript (2q32.3) **PRNCRI**: prostate cancer associated non-coding RNA 1(8q24.21) **PTCSC3**: papillary thyroid carcinoma susceptibility candidate 3 (14q13.3)

PTENP1-AS: *PTENP1 antisense RNA* (9p13.3)

PVT1: *plasmocytoma variant translocation 1* (8q24)

RIAN/MEG8: *maternally expressed 8* (14q32.31)

SAMMSON: *survival associated mitochondrial melanoma specific oncogenic ncRNA*
(chromosome 3)

SOX2-OT: *SOX2 overlapping transcript* (3q26.33)

SPRY4-ITI: *SPRY4 intronic transcript Iprovided* (5q31.3)

SRA1: *steroid receptor RNA activator 1* (5q31.3)

TERRA: *telomeric repeat-containing RNAs*

THRIL: *TNF and HNRNPL related immunoregulatory long non-coding RNA* (12q24.31)

TINCR: *tissue differentiation-inducing non-protein coding RNA* (19p13.3)

TreRNA: *translation regulatory long non-coding RNA* (20q13.13)

Tsix: *Tsix transcript, Xist antisense RNA* (Xq13.2)

TUG1: *taurine up-regulated 1* (22q12.2)

TUNA: *TCL1 upstream neural differentiation-associated RNA* (14q32.2)

UCA1: *urothelial cancer associated 1* (19p13.12)

Xist: *X inactive specific transcript* (Xq13.2)

ZEB2-AS1: *ZEB2 antisense RNA 1* (2q22.3)

Zfas1: *ZNF1 antisense RNA 1* (20q13.13)

FIGURES

Figure 1 : Incidence du cancer du sein dans le monde en 2012.....	23
Figure 2 : Incidence des cancers solides chez la femme en France en 2012.....	24
Figure 3 : Principaux gènes de prédisposition au cancer du sein.....	27
Figure 4 : Schémas anatomique (4A) et physiologique (4B) de la glande mamma.....	28
Figure 5 : Schéma histologique de l'unité ductulo-alvéolaire mammaire.....	29
Figure 6 : Histologie mammaire: 6A (HES), 6B (Immunofluorescence).....	30
Figure 7 : Etapes classiques de la cancérogénèse mammaire.....	33
Figure 8 : Principaux types histopathologiques des carcinomes mammaires infiltrants.....	33
Figure 9 : Classification immunophénotypique des carcinomes mammaires infiltrants.....	38
Figure 10 : Progression tumorale des carcinomes mammaires.....	39
Figure 11 : Cadre actuel de la cancérogénèse.....	40
Figure 12 : Caractéristiques principales du cancer et leur ciblage thérapeutique.....	41
Figure 13 : Fréquence des mutations dans différents types de carcinomes.....	45
Figure 14 : Chromotripsis	46
Figure 15 : Voie de signalisation PI3K/Akt/mTOR.....	48
Figure 16 : Classification moléculaire intrinsèque des carcinomes mammaires infiltrants.....	50
Figure 17 : Pronostic des sous-types moléculaires des cancers du sein.....	52
Figure 18 : Classification génomique et transcriptomique.....	53
Figure 19 : Classification moléculaire intégrée des cancers du sein.....	54
Figure 20 : Analyse comparée des altérations moléculaires des carcinomes mammaire de	

type triple négatif et ovariens séreux de haut grade.....	55
Figure 21 : Biogénèse des ARNs codant pour les protéines et des ARNs non codants.....	62
Figure 22 : Structure second-tertiaire et interactions des lncRNAs.....	72
Figure 23 : Classification des lncRNAs selon leur topographie génomique.....	76
Figure 24 : Mécanismes d'action en <i>cis</i> et en <i>trans</i> des lncRNAs au niveau de leurs cibles.....	90
Figure 25 : Mécanismes d'action moléculaires des lncRNAs.....	92
Figure 26 : Coordination de l'expression des gènes par les lncRNAs.....	96
Figure 27a : Principaux lncRNAs impliqués dans la cancérogénèse mammaire.....	145
Figure 27b : Principaux lncRNAs et leurs cible dans la cancérogénèse mammaire.....	146
Figure 28 : Classifications moléculaires intégrant les lncRNAs.....	149
Figure 29 : Interactions du lncRNA ANRIL et des complexes polycomb PRC2/PRC1 au niveau du locus p16-CDKN2A/p15-CDKN2B/p14-ARF.....	166
Figure 30 : Mécanismes d'action du lncRNA MALAT1 au niveau des speckles intra-nucléaires.....	178

TABLEAUX

Tableau 1 : Caractères génomiques et protéomiques des sous-types moléculaires.....	56
Tableau 2 : Types et fonctions des principaux petits et longs ARNs non codants.....	63
Tableau 3 : Fonctions moléculaires et cellulaires des lncRNAs.....	95
Tableau 4 : Fonctions et rôles biologiques des lncRNAs.....	111
Tableau 5 : Mécanismes d'action des lncRNAs dans l'évolution métastatique.....	124
Tableau 6 : Principaux lncRNAs impliqués dans la cancérogénèse mammaire.....	128
Tableau 7 : LncRNAs représentant des biomarqueurs et des cibles thérapeutiques potentiels.....	154

RESUME

Le cancer du sein représente à l'échelle mondiale le deuxième cancer le plus fréquent et la première des tumeurs malignes de la femme. Actuellement, seuls certains biomarqueurs (RO, RP, récepteur HER2, index Ki67) et la signature transcriptomique PAM50 sont pris en compte dans la classification morphologique et l'orientation thérapeutique. Les analyses transcriptomiques à haut débit ont révélé que plus de 80% du génome humain est transcrit en ARN. Parmi les ARNs non codants, les transcrits dont la longueur est supérieure à 200 nt sont arbitrairement qualifiés de longs ARNs non codants (lncRNAs). Les lncRNAs jouent un rôle crucial dans le maintien de l'homéostasie cellulaire et présentent des profils d'expression anormaux dans diverses pathologies, dont le cancer. L'objectif principal de mon projet de thèse a consisté à analyser l'expression des lncRNAs, leur fonctionnalité et leur rôle dans l'oncogénèse mammaire. La première partie s'est focalisée sur l'étude des gènes *ANRIL* (ainsi que 10 gènes de la même voie de signalisation) et *MALATI*, deux lncRNAs dont les mécanismes d'action et la signification clinique au cours de la cancérogénèse mammaire sont encore controversés. *ANRIL* et *MALATI* sont respectivement surexprimés dans 20% et 14% des tumeurs de notre série, confirmant leurs rôles pro-oncogénique dans la cancérogénèse mammaire. La surexpression de *MALATI* se traduit en RNA-FISH par la présence de volumineux *speckles* intranucléaires. La complexité de leur dérégulation est liée à la présence d'isoformes et de réseaux d'interactions avec les mRNAs et les miRNAs. Concernant les sous-unités appartenant aux complexes Polycomb PRC2 et PRC1 qui interagissent avec *ANRIL*, *EZH2* (PRC2) est normalement ciblé par 3 miRNAs onco-suppresseurs (miR-26A1, miR-125B et miR-214) qui sont sous-exprimés dans notre série de CCI. Les 2 oncomiRs miR-181B1 and miR-181A2 qui ciblent et inactivent *CBX7* (PRC1), apparaissent surexprimés dans notre série, en relation avec l'activation de l'oncogène *HMGAI*. Concernant *MALATI*, le complexe de biogénèse des miRNAs Drosha-DGCR8-Microprocesseur régule l'expression d'un variant d'épissage Δ -*MALATI* et ce dernier est impliqué dans l'activation de la voie PI3K/Akt. Des corrélations significatives sont observées entre *MALATI* et des gènes impliqués dans l'épissage alternatif, le cycle cellulaire, l'apoptose, la réparation de l'ADN et la migration cellulaire. Les profils transcriptomiques aberrants de ces 2 lncRNAs semblent caractéristiques des carcinomes mammaires. Ainsi, *ANRIL* (i) présente une association positive inattendue avec le *cluster p16-CDKN2A/p15-CDKN2B/p14-ARF* dans notre série, alors que cette association apparait négative dans les carcinomes de prostate et (ii) inactive épigénétiquement les miRNAs onco-suppresseurs miR-99a/miR-449a dans les carcinomes gastriques et non dans notre série de CCI. D'un point de vue clinique, deux signatures pronostiques indépendantes ont pu être identifiées, l'une intégrant les 2 partenaires protéiques d'*ANRIL* appartenant aux complexes Polycomb (surexpression d'*EZH2* / sous-expression de *CBX7*), et l'autre représentée par la sous-expression de Δ -*MALATI*, observée dans 20% des tumeurs de notre série. La présence de variants d'épissage alternatifs, de réseaux d'interactions multiples et d'une spécificité d'organe devra être prise en compte lors de l'évaluation des thérapies épigénétiques ciblant *ANRIL* (inhibiteurs des bromodomaines et des oncoMIRs) ou *MALATI* (ASOs) dans les cancers du sein. La deuxième partie du projet de thèse a consisté en l'analyse du transcriptome non codant des carcinomes mammaires par une stratégie pangénomique, afin d'identifier de nouveaux types de lncRNAs, tels les nouveaux lncRNAs antisens, circulaires et associés à des séquences ultra-conservées ou induisant des résistances médicamenteuses. L'analyse préliminaire réalisée sur une petite cohorte de carcinomes mammaires (n=8) permettra la mise en place de l'étude principale (n=40) pour renforcer la robustesse des signatures identifiées sur la cohorte initiale.

ABSTRACT

Breast cancer is the second most common cancer and the first malignancy of women. Currently, only few biomarkers (ER, PR, receptor HER2, index Ki67) and transcriptomic signature PAM50 are included in the morphological classification and therapeutic orientation. Transcriptome genome-wide analyses unexpectedly revealed that over 80% of the DNA is transcribed into RNA. Among these noncoding RNAs, transcripts longer than 200 nt are arbitrarily qualified as long noncoding RNAs (lncRNAs). LncRNAs play a crucial role in maintenance of cellular homeostasis and present abnormal expression patterns in various diseases, including cancer. The main objective of my project was to analyze expression of lncRNAs, their functionality and their roles in breast oncogenesis. The first part focused on the study of *ANRIL* and *MALAT1* genes, two lncRNAs whose mechanisms of action and clinical significance in breast carcinogenesis are still controversial. *ANRIL* and *MALAT1* respectively overexpressed in 20% and 14% of tumors in our series, confirming their pro-oncogenic roles in mammary carcinogenesis. *MALAT1* overexpression results in RNA-FISH by presence of huge intranuclear speckles. Complexity of their deregulation is associated with presence of various isoforms and interaction networks with miRNAs, mRNAs and other lncRNAs. Concerning PRC2/PRC1 polycomb sub-units interacting with *ANRIL*, *EZH2* (PRC2) is normally targeted by 3 onco-suppressor miRNAs (miR-26A1, miR-125B and miR-214) that are under-expressed in our series of CCIs. The 2 oncomiRs miR-181B1 and miR-181A2 that normally target and inactivate *CBX7* (PRC1) appear overexpressed in our series of CCIs, resulting from activation of the oncogene *HMGAI1*. Concerning *MALAT1*, the miRNAs biogenesis complex Drosha-DGCR8-Microprocessor regulates expression levels of the splicing variant Δ -*MALAT1* and the latter is involved in activation of PI3K/Akt pathway. Significant correlations were observed between *MALAT1* and genes involved in alternative splicing, cell cycle, apoptosis, DNA repair and migration. Aberrant transcriptomic profiles of these two lncRNAs seem characteristics of mammary carcinomas. Thus, *ANRIL* (i) presents an unexpected positive association with the *p16-CDKN2A/p15-CDKN2B/p14-ARF* cluster in our series of CCIs, whereas this association appears negative in prostate carcinomas and (ii) epigenetically inactivates onco-suppressor miRNAs miR99a/miR-449a in gastric carcinomas, but not in our series. From a clinic point of view, two independent prognostic signatures were identified, one incorporating two protein partners of *ANRIL* belonging to the polycomb complexes (*EZH2* overexpression / *CBX7* under-expression) and the other represented by under-expression of the variant Δ -*MALAT1* observed in 20% of tumors in our series. The presence of alternative splice variants, multiple interactions with mRNAs and miRNAs and organ specificity should be considered when evaluating epigenetic antitumoral drugs designed to target *ANRIL* (bromodomains and oncoMIRs inhibitors) and *MALAT1* (ASOs) in breast cancers. The second part of the project involved analysis of non-coding transcriptome of mammary carcinomas to identify new types of lncRNAs, including new antisens lncRNAs, circular lncRNAs, induced lncRNAs, noncoding ultraconserved transcripts and lncRNAs associated with resistance to systemic treatments. The preliminary analysis performed on a small cohort of breast cancers (n=8) will allow the implementation of the main (n=40) which will enhance robustness of identified signatures.

AVANT-PROPOS

Ce projet de thèse a été réalisé au sein de l'Unité de Pharmacogénomique dirigée par le Dr Ivan Bièche du Service de Génétique de l'Ensemble Hospitalier de l'Institut Curie. L'analyse par RNA-seq et l'étude bioinformatique ont été effectuées au sein de l'unité CNRS UMR 3244 « Epigénétique, ARN non codant et fluidité du génome » de l'Institut Curie dirigée par le Dr Antonin Morillon, en collaboration avec la plateforme de séquençage ICGEX de l'Institut Curie et de la plateforme Imagif du CNRS de Gif-Sur-Yvette.

Ce travail a été dirigé par le Dr Ivan BIECHE et a donné lieu à la rédaction de deux articles scientifiques.

Article 1

Expression of *ANRIL* – Polycomb complexes - *CDKN2B/CDKN2A/ARF* genes in breast tumors: identification of a two-gene (*EZH2/CBX7*) expression signature of independent prognostic value.

Didier Meseure, Sophie Vacher, Kinan Drak Alsibai, Andre Nicolas, Walid Chemlali, Martial Caly, Rosette Lidereau, Eric Pasmant, Celine Callens and Ivan Bieche.

Article publié dans la revue *Molecular Cancer Research* 2016;14:623-33.

Article 2

Prognostic value of a new identified *MALATI* alternatively spliced transcript in breast cancer.

Didier Meseure, Sophie Vacher, François Lallemand, Kinan Drak Alsibai, Rana Hatem, Walid Chemlali, Andre Nicolas, Leanne De Koning, Eric Pasmant, Celine Callens, Rosette Lidereau, Antonin Morillon and Ivan Bieche.

Article publié dans la revue *British Journal of Cancer* 2016;114:1395-404.

Durant la préparation de cette thèse (2013-2016), j'ai également participé, en tant que premier (ou dernier) auteur ou co-auteur, à d'autres projets scientifiques qui ont donné lieu à 16 articles ou revues publiés, en révision ou soumis dans des revues internationales que je joins en annexes.

INTRODUCTION

1 GENERALITES SUR LES CANCERS DU SEIN

1.1 Fréquence

Le cancer du sein représente un problème majeur de santé publique. Il est en effet le deuxième cancer le plus fréquent à l'échelle mondiale (12 %) et la première des tumeurs malignes de la femme avec plus de 1.300.000 nouveaux cas et 450.000 décès chaque année dans le monde. La prévalence de ce cancer est très variable selon les régions géographiques et le taux d'incidence du cancer du sein aux Etats-Unis, au Canada et en Europe est très élevé par rapport à l'Asie et à l'Afrique (**Fig. 1**). Ces différences sont en grande partie dues à la pratique d'un dépistage dans les pays développés, mais suggèrent également des influences génétiques et/ou environnementales variables d'une région à l'autre. Cependant, l'évolution actuelle tend vers une stabilisation relative dans les pays occidentaux et une augmentation de l'incidence en Asie et dans de nombreux pays en voie de développement.

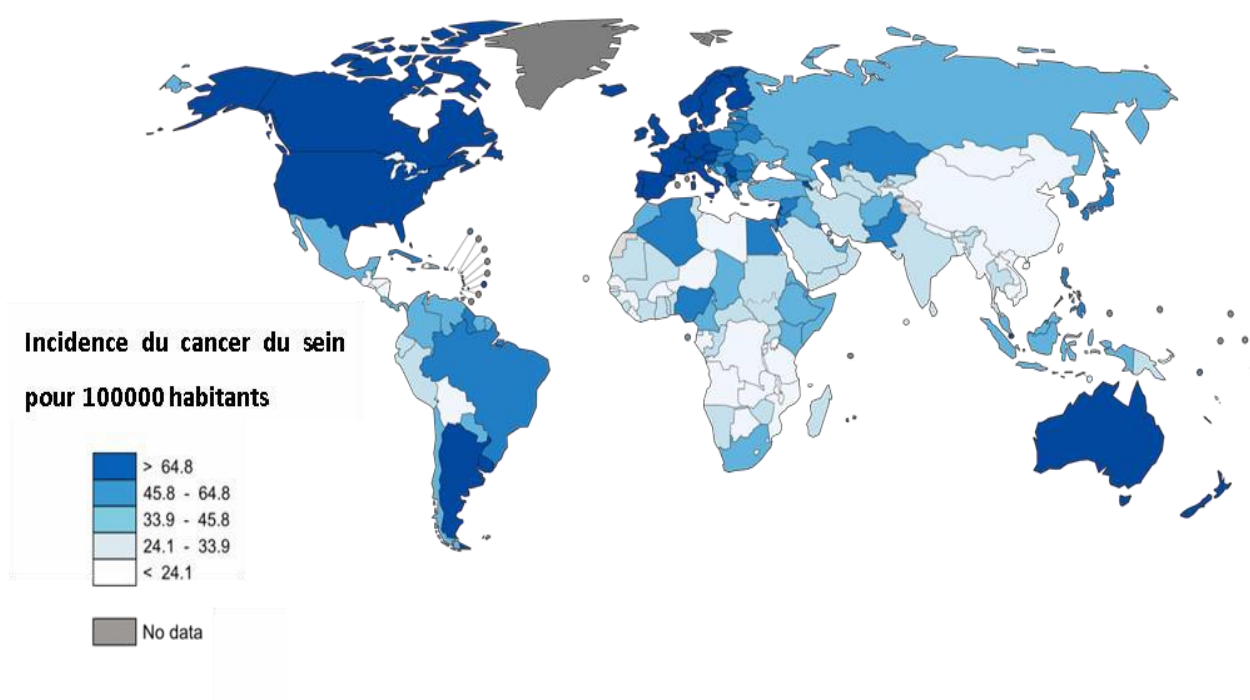


Figure 1 : Incidence du cancer du sein dans le monde en 2012
Source: World Health Organization, IARC (International Agency for Research on Cancer), GLOBOCAN 2012

En France, le cancer du sein constitue la tumeur maligne la plus fréquente chez la femme et représente 33.4 % de l'ensemble des cancers et 37.9 % des cancers solides chez la femme en

2012 (www.e-cancer.fr: La situation du cancer en France en 2012) (**Fig. 2**). On estime actuellement qu'une femme sur 8 sera atteinte d'un cancer du sein au cours de sa vie. Ce cancer peut aussi survenir très rarement chez les hommes (Ferlay et al., 2012).

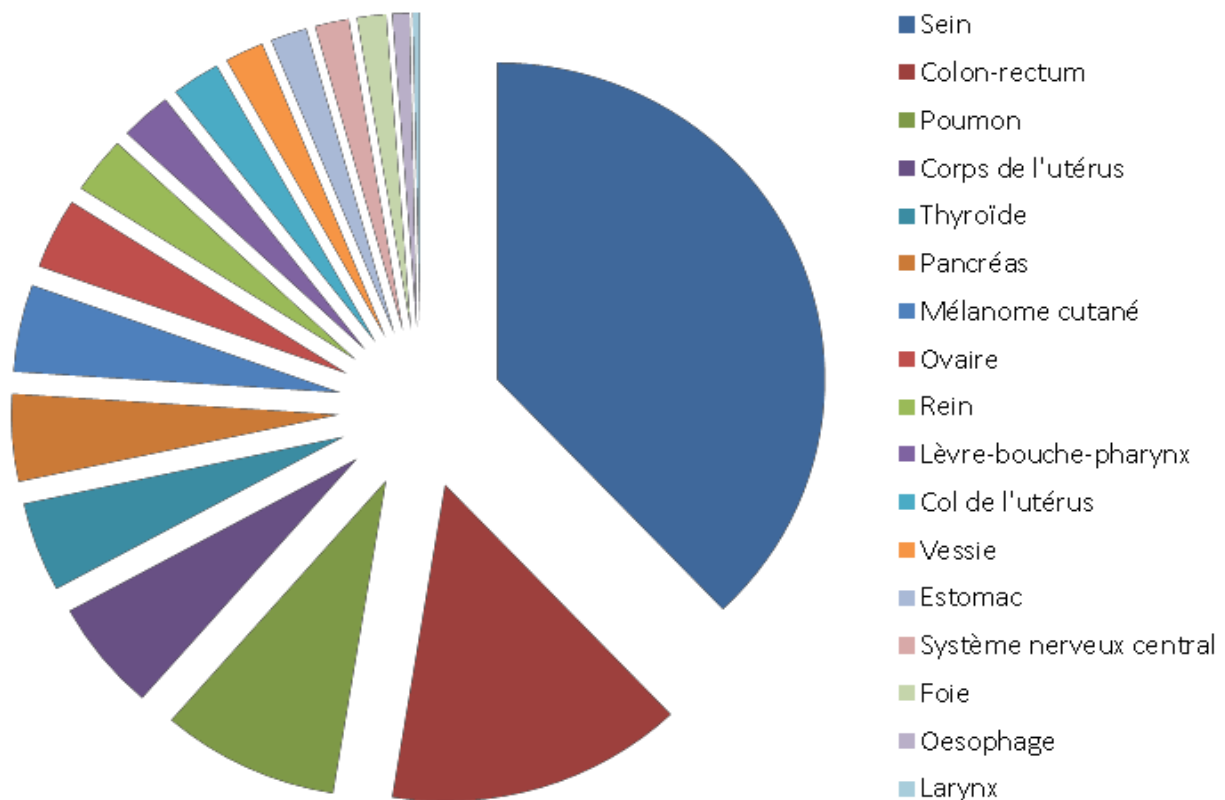


Figure 2 : Incidence des cancers solides chez la femme en France en 2012

D'après les données de l'INCa, le cancer du sein constitue le premier cancer féminin en France (www.e-cancer.fr « la situation du cancer en 2012 »).

1.2 Facteurs de risque

Plusieurs études épidémiologiques et expérimentales menées à travers le monde ont mis en évidence des facteurs génétiques et environnementaux pouvant augmenter le risque de survenue du cancer du sein.

1.2.1 Facteurs hormonaux

Ils sont essentiellement liés à l'allongement du temps d'exposition aux oestrogènes (Nkondjock and Ghadirian, 2005) et comprennent :

- **Les facteurs hormonaux endogènes** : L'apparition du cycle menstruel à un jeune âge (avant 12 ans) et son arrêt tardif (ménopause) s'accompagnent d'un risque légèrement plus élevé de cancer du sein (Brinton et al., 1988).
- **Les facteurs hormonaux exogènes** : Le traitement hormonal substitutif de la ménopause est un facteur de risque modéré de survenue d'un cancer du sein (Rossouw et al., 2002), de même que la prise d'une contraception orale œstroprogestative. Néanmoins, ce sur-risque potentiel lié aux contraceptifs disparaîtrait dès l'interruption du traitement de sorte qu'après l'arrêt de son utilisation, aucune augmentation de risque n'est observée (*Collaborative group on hormonal factors in breast cancer*, 1996).

1.2.2 Facteurs liés à la reproduction

Le développement du cancer du sein étant lié à la vitesse de prolifération des cellules épithéliales mammaires et inversement corrélé au degré de différenciation, **la nulliparité** ou **une première grossesse tardive** sont des facteurs augmentant le risque de survenue du cancer du sein.

1.2.3 Facteurs génétiques

- **Histoire familiale de cancers du sein** : Des antécédents familiaux au premier degré de cancer du sein doublent le risque. Ce risque est moindre s'il s'agit d'antécédents au deuxième degré.
- **Prédisposition génétique** : 5 à 10% des cancers du sein sont la conséquence d'anomalies génétiques. Les principales mutations germinales inactivatrices observées sont celles de *BRCA1* et *BRCA2*, deux gènes codant pour des protéines de réparation des cassures double-brin de l'ADN. Ces mutations sont transmises selon un mode autosomal dominant avec une forte pénétrance.

60 à 65 % des patientes porteuses de la mutation *BRCA1* et 45 à 55 % de celles porteuses de la mutation *BRCA2* sont susceptibles de développer un cancer du sein. Dans la population générale, les femmes porteuses des mutations *BRCA1* et *BRCA2* sont rares (0.11% et 0.12%, respectivement). 6 % des femmes âgées de moins de 50 ans et présentant un cancer du sein sont porteuses d'une mutation germinale de l'un de ces deux gènes, alors que seulement 1 à 1.5% des femmes âgées de plus de 50 ans en sont affectées.

Les autres gènes dont les mutations sont transmises avec une forte pénétrance sont les gènes *TP53* (syndrome de Li-Fraumeni) et *PTEN* (syndrome tumoral hamartomateux *PTEN*).

D'autres gènes de susceptibilité plus rares et conférant aux porteuses de ces mutations un risque faible à modéré de survenue d'un cancer du sein sont *CHEK2*, *ATM*, *PALB2*, *BRIP1*, *CDH1*, *RAD50*, *RAD51C* et *RAD51D*, tous impliqués dans les mécanismes de réparation de l'ADN (**Fig. 3**).

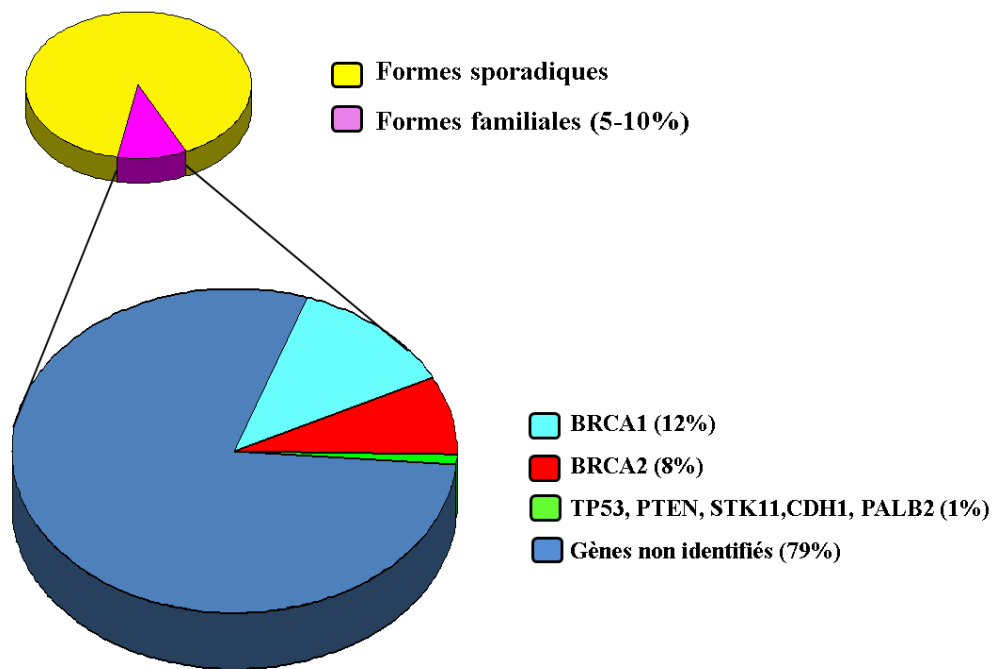


Figure 3 : Principaux gènes de prédisposition au cancer du sein

1.2.4 Autres facteurs de risque

- **Age** : L'âge est le facteur de risque le plus important. La maladie est rare chez les femmes de moins de 30 ans et le risque augmente significativement entre 50 et 75 ans (près des deux tiers des cancers du sein).
- **La consommation d'alcool et le tabagisme** : L'alcool est le seul facteur nutritionnel établi de risque de cancer du sein. Le tabagisme accroît légèrement le risque de survenue de carcinomes mammaires.
- **La densité mammographique et les lésions bénignes du sein** : Le risque augmente avec le niveau de densité du tissu mammaire (risque multiplié par deux à six). Les lésions bénignes sont divisées en deux groupes, les lésions non proliférantes et les lésions prolifératives avec ou sans atypies. Les lésions non proliférantes ne sont généralement pas associées à un risque accru de cancer du sein. Les lésions

prolifératives sans atypies multiplient le risque par deux, alors que les lésions prolifératives avec atypies augmentent ce risque par quatre (Key et al., 2001).

- **Les radiations ionisantes** : Le sein est l'un des organes les plus sensibles aux effets des radiations (Key et al., 2001). Le risque est similaire pour une exposition unique ou pour des expositions multiples à intensité totale égale (Little et al., 1999).

1.3 Anatomie, histologie et physiologie de la glande mammaire

1.3.1 Anatomie

Le sein est subdivisé en 10 à 20 lobes entourés et séparés par un tissu adipeux. Chaque lobe est constitué de lobules drainés par des canaux galactophoriques. Chaque lobule est organisé en acini regroupés autour de canalicules intralobulaires. Ces derniers se réunissent pour former des canaux lobulaires, puis des canaux galactophoriques qui convergent et constituent des sinus lactifères sous mamelonnaires (**Fig. 4B**).

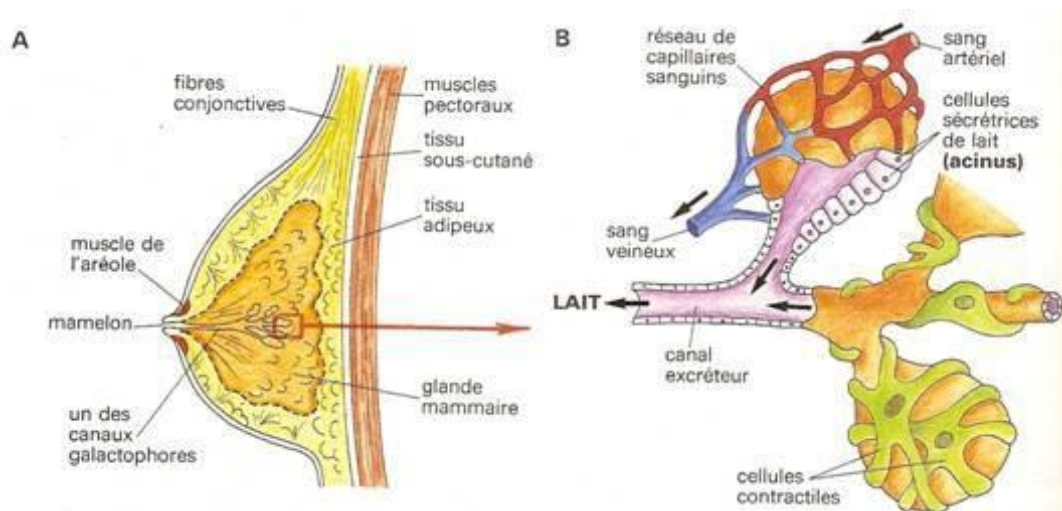


Figure 4 : Schémas anatomique (4A) et physiologique (4B) de la glande mammaire (D'après Russo et al., 2001).

1.3.2 Histologie

La glande mammaire est constituée de deux composantes tissulaires, un compartiment épithélial/myoépithélial édifiant les canaux et les unités ductulo-alvéolaires terminales et un tissu conjonctif paucicellulaire, le stroma palléal. Les cellules luminales sécrétrices sont situées au contact des lumières et sont entourées d'une couche de cellules myoépithéliales, évacuant le lait en se contractant, et contribuant à l'homéostasie des structures épithéliales en élaborant une membrane basale (**Fig 5 et 6**). Le stroma palléal comprend une matrice extracellulaire et divers types de cellules mésenchymateuses, dont des fibroblastes, des myofibroblastes, des cellules endothéliales/péricytaire ainsi que de rares leucocytes. Ce stroma palléal assure le soutien et la nutrition de l'organe et régule les fonctions de l'épithélium mammaire en favorisant des interactions réciproques de nature physique, paracrine et endocrine (Polyak and Kalluri, 2010). Le tissu mammaire est en effet influencé par l'action des hormones, œstrogènes et progestérone, produites chez la femme en quantité variable tout au long de sa vie (puberté, grossesse, allaitement).

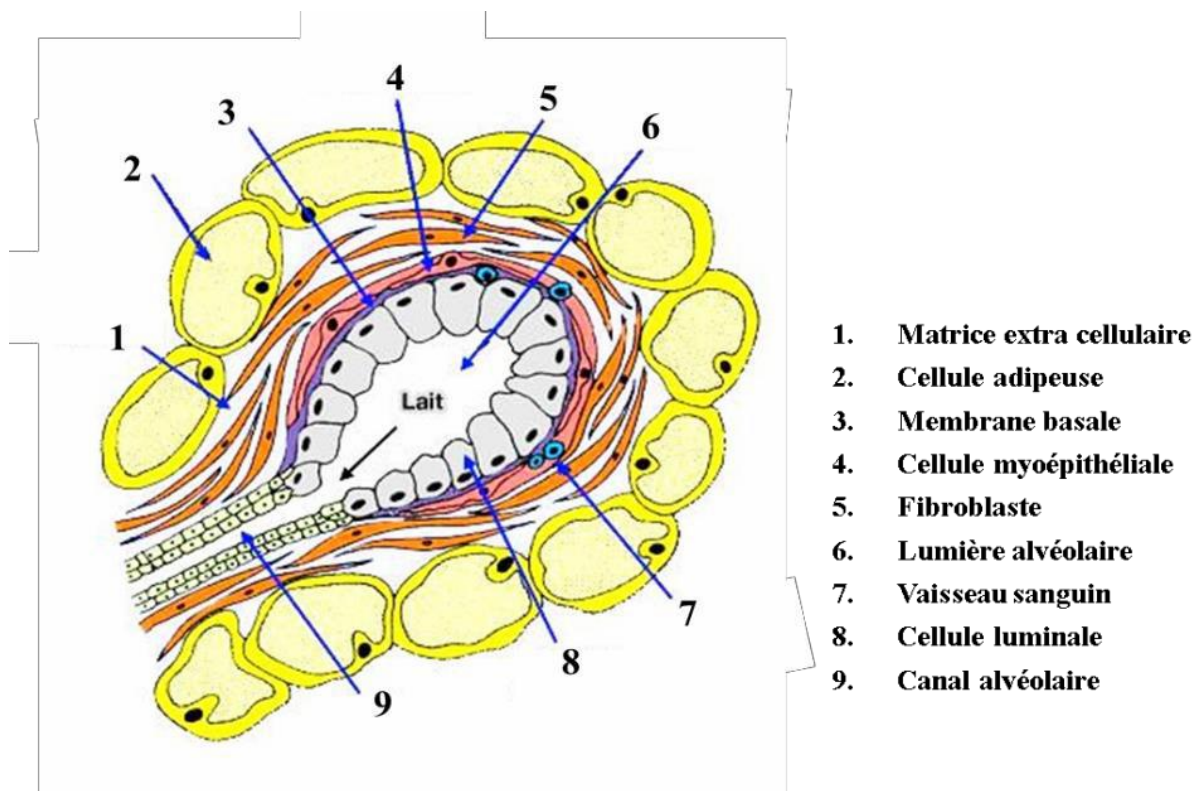


Figure 5 : Schéma histologique de l'unité ductulo-alvéolaire mammaire (D'après Delouis, 1978)

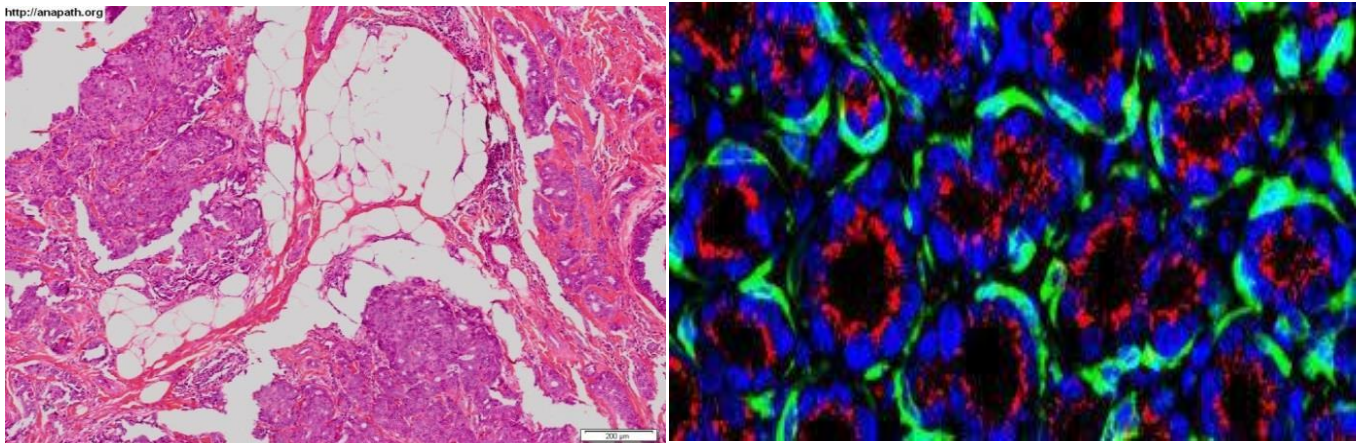


Figure 6 : Histologie mammaire. 6A : Lobules et canaux galactophoriques (HES x 50), **6B** : Polarisation architecturale des alvéoles et canalicules (cellules épithéliales bleues, cellules myoépithéliales vertes, Immunofluorescence x 200)(anapath.org).

1.3.3 Physiologie

Le sein est une glande exocrine dont la fonction est de produire et d'excréter le lait (**Fig. 4B**). La glande mammaire est un organe unique car la majeure partie de sa croissance et de sa différenciation surviennent après la naissance (Cowin and Wysolmerski, 2010). À la naissance, le coussinet adipeux mammaire contient une structure épithéliale canalaire simple localisée sous le mamelon qui demeure quiescente jusqu'à la survenue de la puberté (**Fig. 4A**). A la puberté, les hormones ovariennes et hypophysaires, dont l'oestrogène, la progestérone et l'hormone de croissance, stimulent la prolifération des cellules épithéliales mammaires, afin de générer des bourgeons situés à l'extrémité terminale de la structure canalaire (*terminal end buds*, TEB) (Brisken and O'Malley, 2010). Les TEBs sont des entités dynamiques constituées de cellules épithéliales prolifératives et apoptotiques, nécessaires à l'allongement des structures canalaire et à la formation des lumières. Ils favorisent l'extension et la bifurcation des canaux vers les extrémités du coussinet adipeux pour édifier une structure finale ramifiée et arborisée emplissant le coussinet. Chez l'adulte, les canaux épithéliaux sont constitués d'une couche interne de cellules épithéliales luminales et d'une couche externe de cellules myoépithéliales/basales (**Fig. 4B**) supposées provenir d'une cellule souche bipotente commune présente lors du développement embryonnaire (van Keymeulen et al., 2011). Les cellules luminales peuvent être subdivisées en progéniteurs luminaux et en cellules luminales différenciées et matures pourvues de récepteurs hormonaux. La couche externe est constituée d'une population de cellules basales souches/progénitrices et de cellules myoépithéliales

présentant à la fois des fonctions protectrices et oncosuppressives et des fonctions contractiles permettant l'expulsion du lait hors de l'alvéole vers le mamelon pendant l'allaitement. Les cellules progénitrices luminales et basales donnent naissance à des progéniteurs alvéolaires qui, pendant la grossesse, se différencient en cellules alvéolaires matures productrices de lait (van Amerongen et al., 2012). Au cours des grossesses, la glande mammaire adulte subit des cycles multiples caractérisés par l'expansion et la différenciation terminale des cellules alvéolaires nécessaires à la lactation, suivies de leur élimination par apoptose durant l'involution (Anderson et al., 2007). En début de grossesse, les cellules épithéliales mammaires prolifèrent pour générer des structures ductulo-alvéolaires au sein du coussinet adipeux mammaire. A la fin de la grossesse, les taux de prolifération diminuent à mesure que les progéniteurs alvéolaires commencent à se différencier et à produire de faibles niveaux de protéines du lait. En post-partum, les progéniteurs alvéolaires subissent une différenciation sécrétoire terminale permettant de produire et de sécréter un grand volume de lait nécessaire à l'allaitement. Lors de l'involution, la majorité des cellules alvéolaires différenciées subissent une apoptose, la glande mammaire recouvrant finalement une organisation structurale et fonctionnelle ayant existé avant la grossesse. Des progéniteurs alvéolaires persistent dans les lobules régressifs de la glande involutive et pourront intervenir lors des grossesses ultérieures.

1.4 Classifications clinique, morphologique et phénotypique des cancers du sein

1.4.1 Classification clinique

La classification TNM est un système international de classement des cancers reposant sur leur extension anatomique. Ce système a été initialement proposé par le chirurgien français Pierre Denoix de l'Institut Gustave-Roussy. Plusieurs révisions de cette classification ont été publiées par l'UICC (Union internationale Contre le Cancer), la dernière étant la 7ème édition publiée en 2010.

La très grande majorité des tumeurs malignes du sein peuvent être classées selon 3 critères principaux :

- La taille tumorale sur le site de la tumeur primitive (T)
- L'extension tumorale ganglionnaire locorégionale (N)
- L'extension métastatique à distance (M).

Le stade TNM défini par l'UICC permet d'établir l'opérabilité de la tumeur et d'obtenir une première évaluation pronostique.

Cependant, des types particuliers de carcinomes mammaires peuvent être classés selon leur mode de présentation clinique. Le carcinome inflammatoire du sein est un cancer rare et agressif de pronostic péjoratif défini cliniquement par un sein érythémateux, oedémateux et hyperthermique avec souvent un aspect caractéristique « en peau d'orange ».

De plus, la survenue d'un cancer du sein au cours de la grossesse et en post-partum est considérée comme un facteur de pronostic défavorable.

1.4.2 Classification morphologique

Les critères morphologiques importants sur les plans pronostique et thérapeutique sont le type histopathologique, le stade pTNM, le grade SBR, le statut des récepteurs hormonaux et du récepteur HER2 et l'index Ki-67.

1.4.2.1 Type histopathologique

Les tumeurs malignes du sein sont des carcinomes dans 90% des cas. Les carcinomes mammaires sont de type canalaire (issus des canaux galactophoriques) dans 70% des cas et de type lobulaire (issus des acini) dans 10% des cas. Les carcinomes mammaires sont dans 15% des cas des carcinomes non infiltrants (carcinomes canaux et lobulaires *in situ*) (**Fig. 7**) et dans 75% des cas des carcinomes infiltrants (carcinomes canaux infiltrants et carcinomes lobulaires infiltrants) (Mallon et al., 2000).

Dans 10% des cas, les carcinomes mammaires infiltrants sont d'un type morphologique particulier, incluant les carcinomes tubuleux, mucineux, cribriformes, papillaires, médullaires, micropapillaires et neuroendocrines (**Fig. 8**). Les tumeurs malignes infiltrantes du sein peuvent être subdivisées en tumeurs d'excellent pronostic (carcinome tubuleux, carcinome

cribriforme, carcinome mucineux), de bon pronostic (carcinome canalaire bien différencié, carcinome lobulaire, carcinome papillaire), de pronostic moyen (carcinome canalaire moyennement différencié, carcinome HER2+) et de mauvais pronostic (carcinome canalaire peu différencié, carcinome triple-négatif, carcinome inflammatoire, carcinome métaplasique).

Membrane basale
Cellules myoépithéliales
Epithélium canalaire

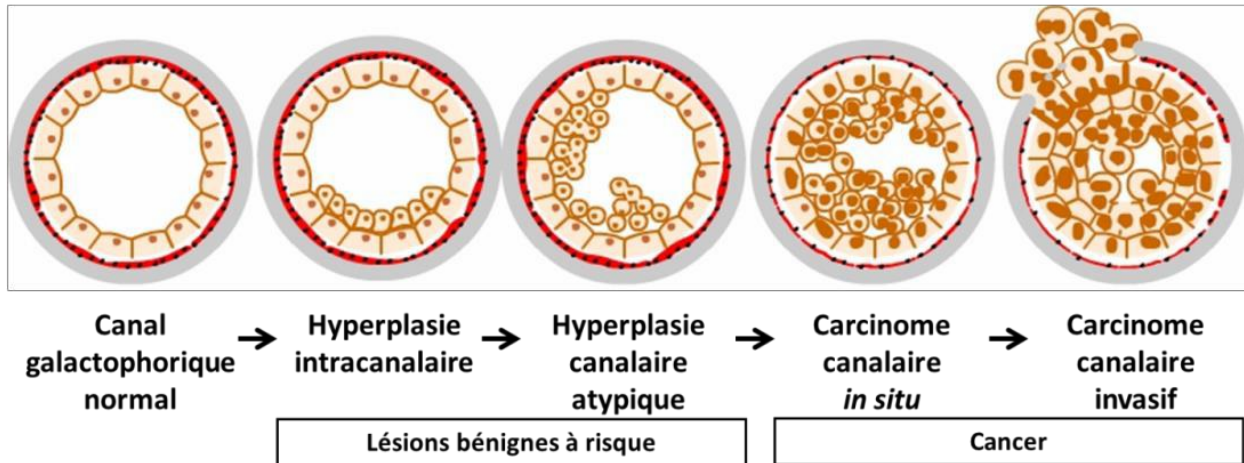


Figure 7 : Etapes classiques de la cancérogénèse mammaire

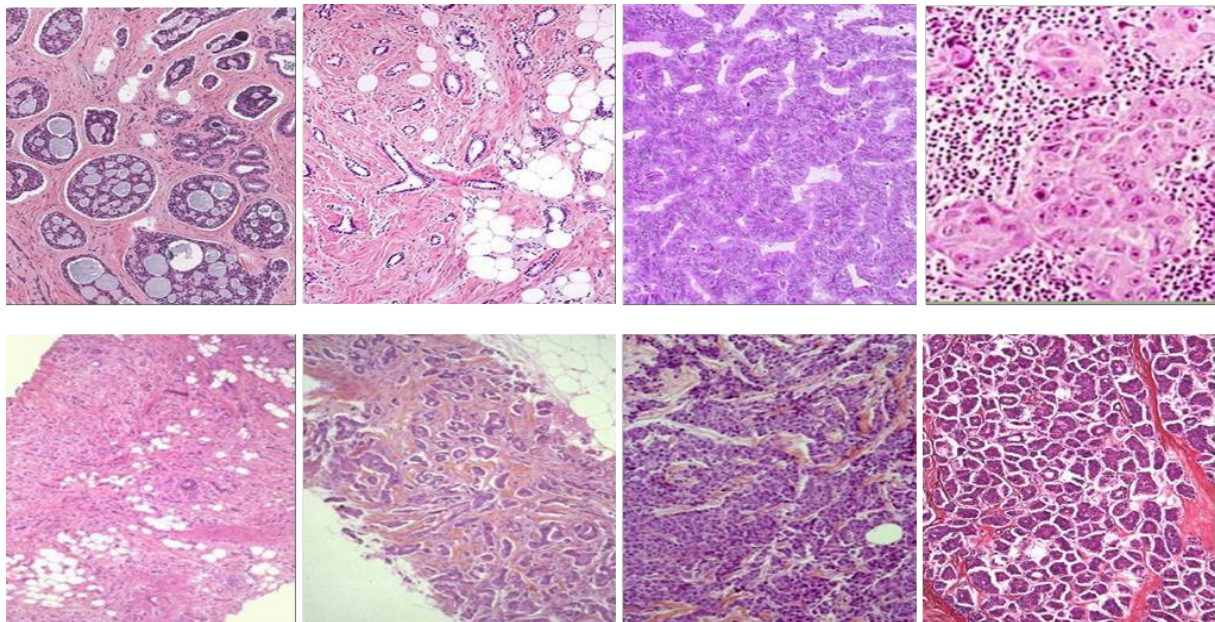


Figure 8 : Principaux types histopathologiques des carcinomes mammaires infiltrants. De haut en bas et de gauche à droite: carcinome cribriforme, carcinome tubuleux, carcinome papillaire, carcinome HER2+, carcinome lobulaire, carcinome bien différencié, carcinome triple-négatif, carcinome micropapillaire.

1.4.2.2 Classification pTNM

Elle est établie en post-chirurgical par les pathologistes.

- **La taille tumorale**

La taille tumorale mesurée lors de l'examen macroscopique et microscopique constitue un facteur pronostique important. Elle est corrélée à l'envahissement des ganglions axillaires. Ceux-ci sont envahis dans 20% des tumeurs de moins de 10mm, dans 30% des tumeurs de moins de 20mm et dans 70% des tumeurs de plus de 50mm. La survie globale à 5 ans est de 91% pour les tumeurs de moins de 10mm, de 80% pour les tumeurs de 20 à 50mm et de 63% pour les tumeurs de plus de 50mm.

- **L'envahissement des ganglions axillaires**

Il constitue actuellement le facteur pronostique le plus important pour prédire les rechutes métastatiques et la survie des patientes. Un nombre minimal de 10 ganglions doivent être examinés. La survenue d'un envahissement ganglionnaire par des cellules tumorales est un facteur pronostique défavorable (75 % de survie globale à 10 ans chez les patientes N- contre 50% chez les patientes N+). Le nombre de ganglions axillaires envahis est également un critère pronostique important (survie à 10 ans : 75% chez les patientes N-, 58% chez les patientes présentant 1 à 3 ganglions envahis, 29% chez les patientes présentant plus de 4 ganglions envahis).

1.4.2.3 Grade et index histopronostiques

L'examen morphologique microscopique, effectué à partir de prélèvements biopsiques ou de pièces opératoires, permet de classer les carcinomes mammaires selon le grade histopronostique de Scarff-Bloom-Richardson ou grade « SBR ».

Trois paramètres sont pris en compte : l'architecture tumorale, l'anisocaryose et le nombre de mitoses par 10 champs au grossissement 400. A chacun de ces paramètres est attribué un score de 1 à 3, dont la somme permet de définir 3 grades : (i) Grade 1 : cancer de croissance lente et de risque évolutif faible, (ii) Grade 2 : cancer de croissance moyenne et de risque évolutif médian et (iii) Grade 3 : cancer de croissance rapide et de risque évolutif élevé. Les

tumeurs mammaires infiltrantes dont le grade est le plus élevé sont considérées comme les plus agressives et au pronostic le plus défavorable.

L'index pronostique de Nottingham peut également être utilisé pour prédire l'évolution des carcinomes mammaires en considérant plusieurs critères, incluant la taille tumorale, le nombre de ganglions lymphatiques métastatiques et le grade histologique de la tumeur.

1.4.2.4 Autres facteurs histopronostiques

Les autres facteurs pronostiques qui doivent être pris en compte lors de l'analyse morphologique sont les emboles vasculaires péri-tumoraux, les engainements péri-nerveux, une composante intracanaulaire associée, des territoires de nécrose tumorale, l'envahissement des marges d'excision, les cellules tumorales isolées (pN0i+) et les micrométastases ganglionnaires (pN1mi).

1.4.3 Classification immunophénotypique

Quatre des plus importants biomarqueurs des cancers du sein sont le récepteur des œstrogènes (RE), le récepteur de la progestérone (RP), le récepteur humain de croissance épidermique de type 2 (HER2) et éventuellement l'index de prolifération cellulaire Ki-67. En pratique médicale courante, leurs niveaux d'expression sont mis en évidence par immunohistochimie.

Les récepteurs nucléaires RE et RP : Le seuil de positivité des récepteurs RE et RP en immunohistochimie (IHC) varie selon les pays. En Europe et particulièrement en France, une intensité d'expression protéique faible identifiée dans les noyaux d'au moins 10 % des cellules tumorales (score IHC 1+) constitue le seuil de positivité habituellement retenu.

Le récepteur membranaire à activité tyrosine kinase HER2

- **La surexpression du récepteur HER2** est définie par un score IHC et/ou FISH (Wolff et al., 2013).

Score IHC 3+ : Marquage membranaire continu d'intensité forte de la quasi-totalité des cellules tumorales.

Score IHC 2+ : Marquage continu ou discontinu d'intensité modérée d'une majorité de cellules tumorales associé à une amplification du gène *HER2* ($HER2 > 6$ copies ou $HER2/CEN17 > 2.2$) mise en évidence par la technique de FISH (*Fluorescence In Situ Hybridization*).

- **L'absence de surexpression du récepteur *HER2*** est définie par un score IHC/FISH :
Score 0 : Absence de marquage.
Score 1+ : Marquage discontinu d'intensité faible d'une minorité de cellules.
Score 2+ et absence d'amplification du gène *HER2* : $HER2 < 4$ copies ou $HER2/CEN17 < 2.2$) (recommandations conjointes de l'ASCO et du Collège des pathologistes américains ; Wolff et al., 2007).

- **Résultat ambigu**
 $4 < HER2 < 6$ copies ou $1.8 < HER2/CEN17 < 2.2$.

L'index de prolifération cellulaire Ki-67

L'index de prolifération Ki-67 présente une importante valeur diagnostique, pronostique et prédictive dans les carcinomes mammaires. Il est un marqueur de prolifération exprimé dans le noyau des cellules lors des phases G1, S, G2 et M du cycle cellulaire. Sur le plan diagnostique, il permet de séparer les carcinomes de type luminal A (index Ki-67 < 20%) des carcinomes de type luminal B (index Ki-67 > 20%). Sur le plan pronostique, il est un facteur indépendant mais de moindre importance que le grade SBR ou l'index mitotique. Sur le plan prédictif, l'index Ki-67 évalué sur micro-biopsies de tumeurs mammaires avant chimiothérapie néoadjuvante apparaît être un important facteur de réponse tumorale, les tumeurs présentant un index élevé régressant mieux que les tumeurs présentant un index bas.

Test d'expression génique PAM50

Dans le cadre de la "médecine personnalisée", des tests innovants de biologie moléculaire permettent une approche concrète de la prise en charge théranostique des patients. Ces tests sont appelés tests "d'expression génique" ou de "signature génique" ou "GEP" (*Gene Expression Profiling*). Ils sont basés sur l'identification et la quantification de mRNAs ou de cDNAs de gènes impliqués essentiellement dans la prolifération tumorale. Le principal intérêt

de ces tests est de permettre pour chaque patiente, une analyse plus fine que les techniques immunohistochimiques de stratification du risque, et de fournir au clinicien une information supplémentaire quant à l'indication éventuelle d'une chimiothérapie ajoutée au traitement standard. Ces GEPs sont réalisés sur des prélèvements de tissu tumoral (biopsie ou pièce opératoire). Parmi les tests proposés, le test Prosigna® (PAM50) a été récemment intégré dans l'arbre décisionnel théranostique des cancers du sein à l'Institut Curie. Les critères d'inclusion pour ce test moléculaire associent: (i) des patientes post-ménopausées, avec un diagnostic de cancer du sein au stade précoce, ayant subi une mastectomie ou un traitement chirurgical conservateur du sein, avec atteintes ganglionnaires (N1 à N3) ou sans (N0), (ii) des tumeurs avec récepteurs hormonaux positifs (RH+), (iii) des tumeurs avec absence de surexpression (ou d'amplification) de HER2 (HER2-), (iv) des patientes éligibles à un traitement par chimiothérapie adjuvante. Le test PAM50 fournit une évaluation du risque rapide, fiable et individualisée avec en particulier la notion d'indicateur pronostique de survie sans récurrence à 10 ans. Il est nécessaire pour réaliser le test PAM50 de disposer de deux variables anatomo-cliniques : la taille de la tumeur (\leq à 2 cm ou $>$ à 2cm) et l'atteinte ganglionnaire (0 à 3). L'infiltration tumorale est vérifiée par un médecin anatomopathologiste sur le bloc tumoral technique. Le dosage Prosigna® est effectué sur l'ARN isolé à partir d'échantillons tumoraux mammaires fixés au formol et inclus en paraffine. Le test PAM50 mesure simultanément par RT-PCR les niveaux d'expression de 50 gènes utilisés par l'algorithme de classification des sous types intrinsèques. Le test est effectué sur le système d'analyse nCounter® DX de Nanostring. L'algorithme renvoie un score RDR (Risque De Récurrence) qui est corrélé avec la probabilité de récurrence à distance à 10 ans. Le test PAM50 a été validé au plan analytique par plusieurs études. Les résultats de nombreuses études, dont TransATAC *clinical validation study* et ABCSG-8 *validation study*, ont permis de valider le modèle qui lie le score RDR du test PAM50 à la probabilité de récurrence à distance dans la population testée, avec un intervalle de confiance de 95 % (Gnant et al., 2014 ; Filipits et al., 2015 ; Sestak et al., 2015).

Les tumeurs de type « triple négatif » sont caractérisées par l'absence d'expression des récepteurs RE et RP et l'absence de surexpression du récepteur HER2.

En pratique médicale quotidienne, ces biomarqueurs sont à la fois utilisés comme facteurs diagnostiques, pronostiques et prédictifs et comme des cibles thérapeutiques. Environ 80%

des tumeurs malignes du sein surexpriment RE (**Fig. 9A**) et 60% surexpriment RP (**Fig. 9B**). Le récepteur HER2 est surexprimé/amplifié dans environ 15% des carcinomes mammaires (**Fig. 9C et 9D**).

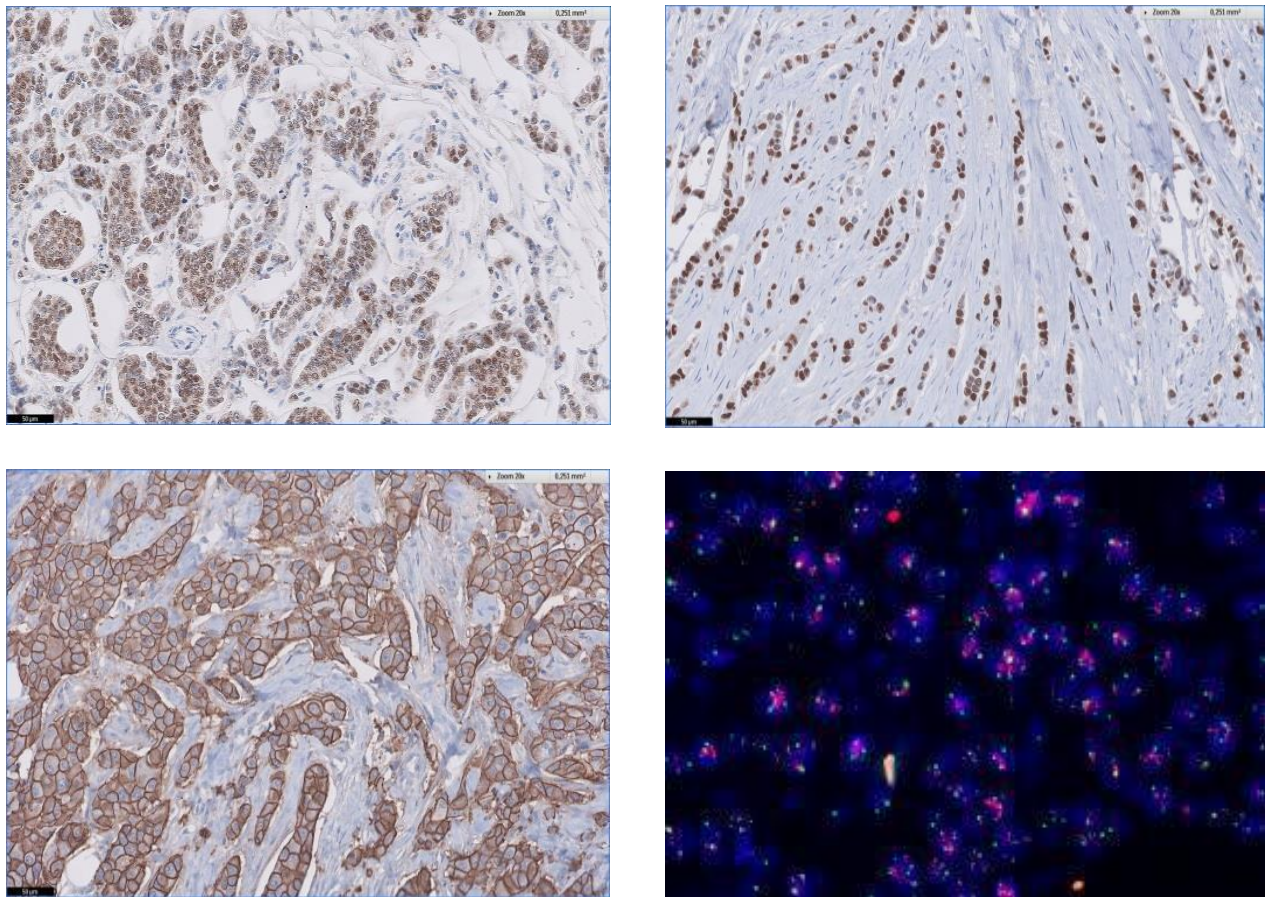


Figure 9 : Classification immunophénotypique habituelle des carcinomes mammaires infiltrants.
9A : RE score 3+, 95% ; **9B** : RP score 3+, 95% ; **9C** : HER2 score 3+ ; **9D** : Amplification de *HER2* (FISH).

1.5 Classifications moléculaires des carcinomes du sein

1.5.1 Biologie moléculaire du cancer du sein

1.5.1.1 Concept général actuel de la cancérogénèse

Le cadre paradigmatique du cancer a récemment évolué et la cancérogénèse est actuellement considérée comme un processus complexe comportant de multiples étapes initiées par des altérations génétiques et épigénétiques des cellules épithéliales et des modifications du microenvironnement normal (Elshamy and Duhé, 2013 ; Hanahan and Coussens, 2012 ; Byun and Gardner, 2013). Les interactions réciproques développées entre les différentes populations épithéliales atypiques et stromales conduisent à une reprogrammation des cellules stromales, au remodelage de la matrice extra-cellulaire (MEC) et à l'acquisition par les cellules cancéreuses de multiples stratégies adaptatives et d'un phénotype plus agressif (**Fig. 10**).

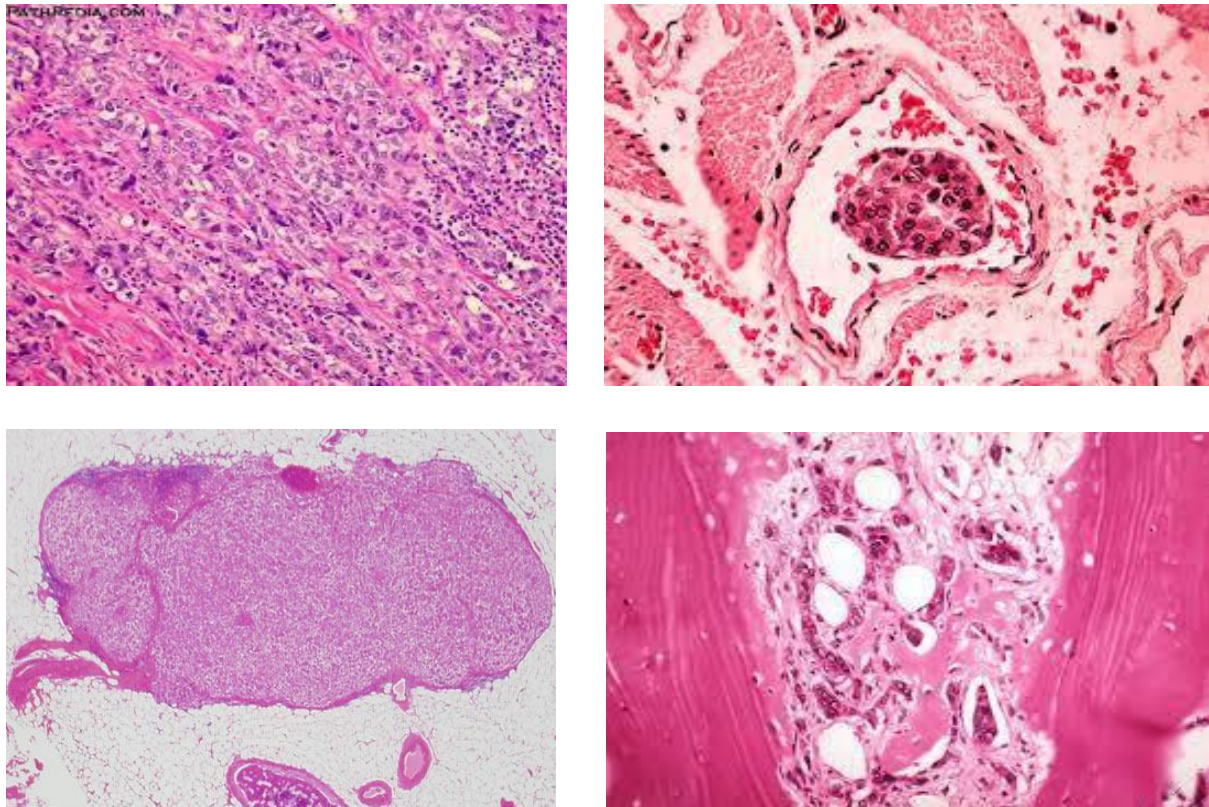


Figure 10 : Progression tumorale des carcinomes mammaires. Infiltration locale (a), invasion vasculaire (b), métastase ganglionnaire (c) et métastase osseuse (d) (anapath.org).

La transformation d'un tissu conjonctif normal en un microenvironnement tumoral est caractérisée par l'acquisition d'une hypoxie chronique, de propriétés oxydantes, d'une reprogrammation métabolique, d'une acidose, d'une néoangiogénèse, d'une inflammation et d'une immunosuppression. Les différentes populations cellulaires tumorales sélectionnées

dans ce microenvironnement modifié, présentent des phénotypes plus agressifs caractérisés à des degrés variables par l'acquisition d'une plasticité épithélio-mésenchymateuse et vasculaire et de propriétés de cellules souches, une reprogrammation métabolique, une réponse au stress du réticulum endoplasmique (RE), une autophagie, la sécrétion d'exosomes et un échappement au système immunitaire. Ces nouvelles propriétés développées au sein de la tumeur primitive par les cellules cancéreuses permettent l'apparition d'une hétérogénéité intratumorale responsable en partie de l'acquisition d'une mobilité cellulaire et d'une résistance à l'anoïkis, à l'origine de l'invasion, de l'intravasation, de l'évolution métastatique et de la résistance aux thérapies systémiques (**Fig. 11**).

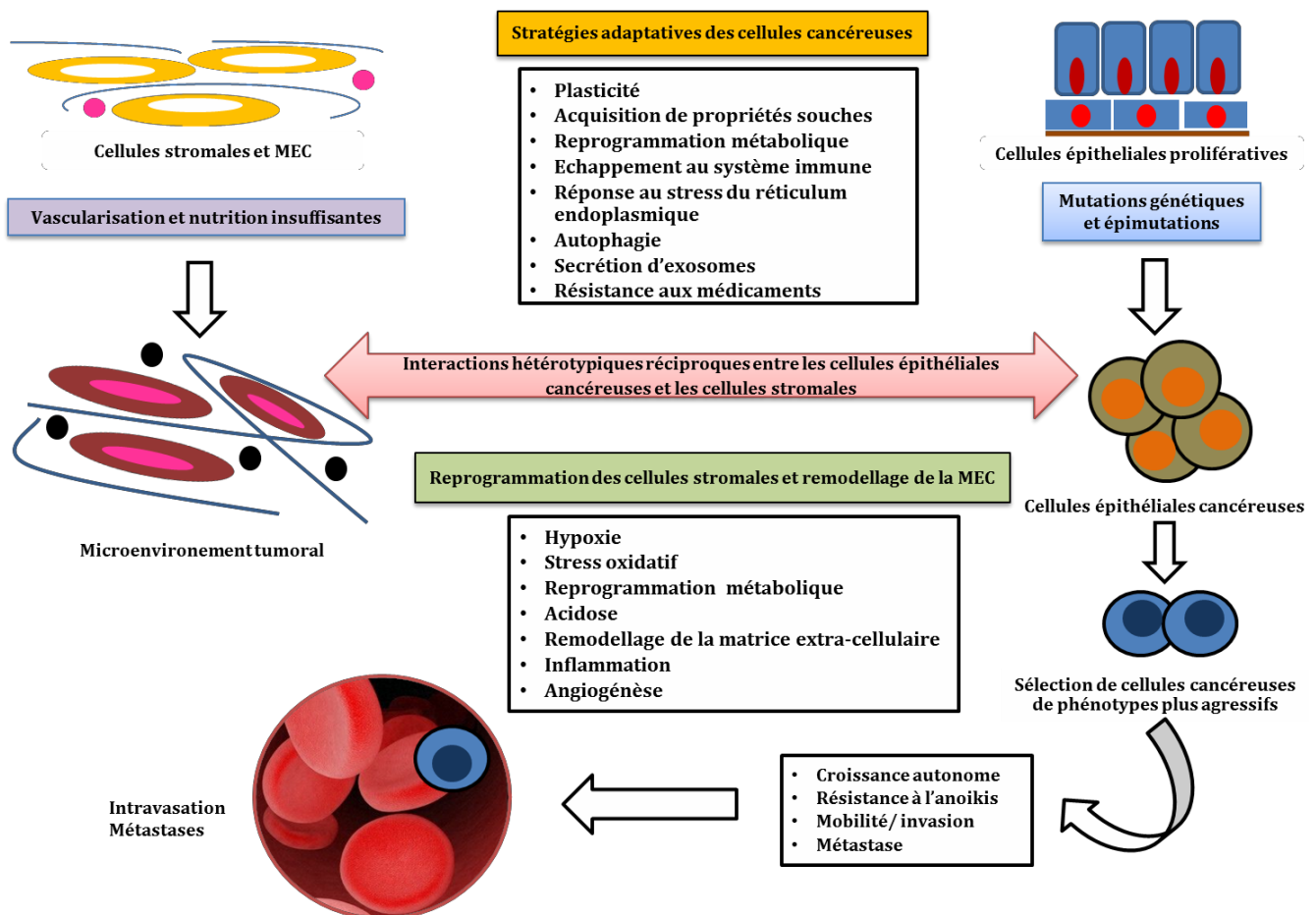


Figure 11 : Cadre paradigmatique actuel de la cancérogénèse. Altérations génétiques/épigénétiques des cellules cancéreuses et interactions réciproques entre les cellules cancéreuses et le microenvironnement (RE: réticulum endoplasmique, MEC : matrice extra-cellulaire).

1.5.1.2 Caractéristiques des cellules cancéreuses

La transformation d'une cellule normale en une cellule cancéreuse se traduit par l'acquisition de nouvelles caractéristiques (Hanahan and Weinberg, 2000). Initialement, selon l'équipe de R. Weinberg, ces dernières étaient au nombre de 6 et comprenaient l'auto-suffisance en signaux de croissance, l'insensibilité vis-à-vis des suppresseurs de croissance, l'inhibition de l'apoptose, une replication illimitée, une néoangiogénèse continue et un potentiel infiltrant et métastatique. Plus récemment, des caractéristiques supplémentaires ont été ajoutées par l'équipe de R. Weinberg, incluant l'immunosuppression, l'inflammation chronique, la reprogrammation métabolique et épigénétique, et l'acquisition de mutations/épimutations et d'une instabilité génomique (Hanahan and Weinberg, 2011) (**Fig. 12**).

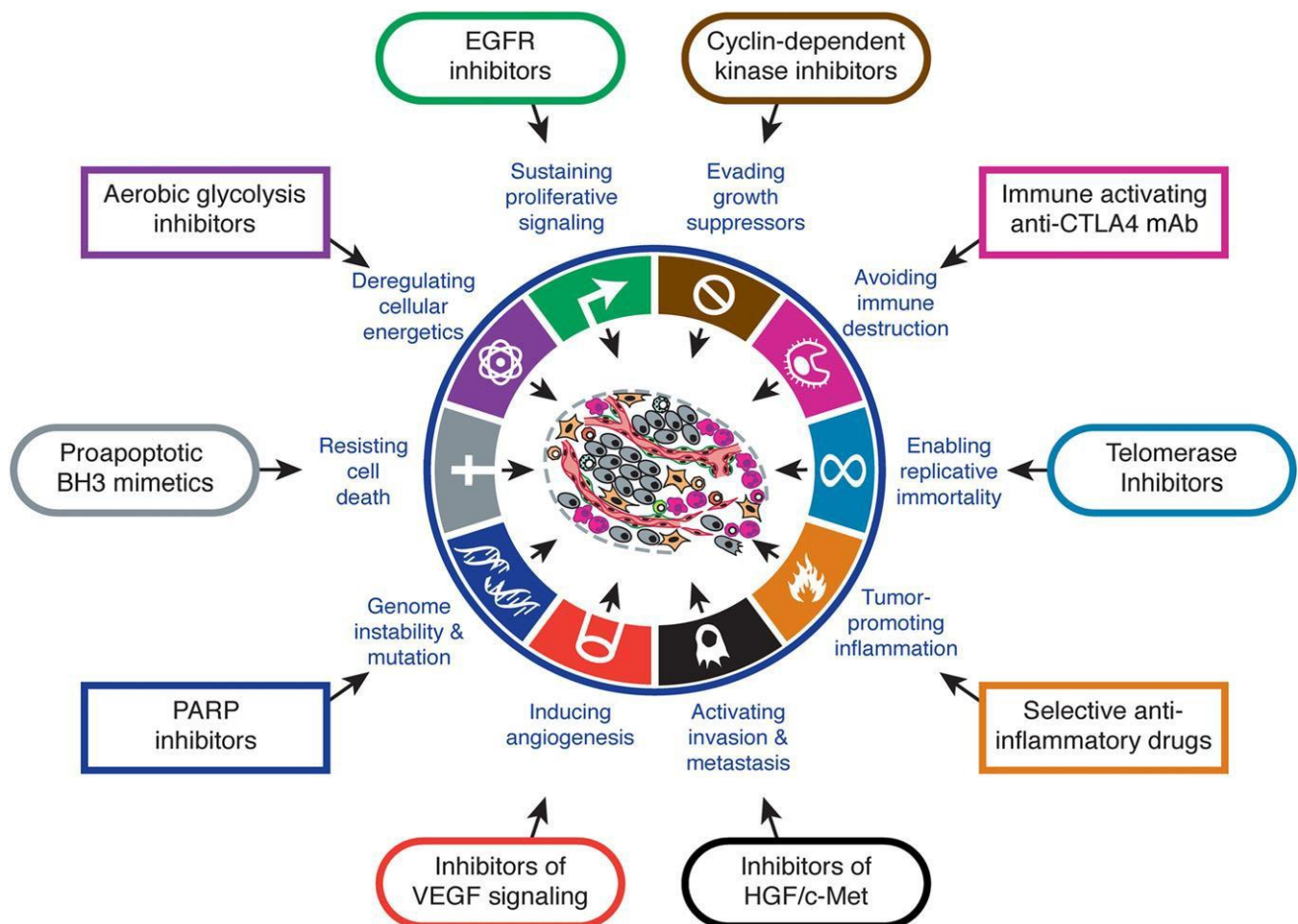


Figure 12 : Caractéristiques principales du cancer et leur ciblage thérapeutique
(Hanahan and Weinberg, 2011)

1.5.1.3 Mécanismes moléculaires de la cancérogénèse mammaire

1.5.1.3.1 Oncogènes et gènes suppresseurs de tumeur

Au cours des décennies précédentes, les progrès réalisés dans la compréhension de la biologie des tumeurs malignes ont permis la découverte de l'implication des oncogènes et des gènes suppresseurs de cancer dans la cancérogénèse mammaire.

Les oncogènes, découverts en 1976, sont des gènes qui acquièrent un pouvoir transformant suite à des modifications qualitatives ou quantitatives, pour devenir des oncogènes (Stehelin et al., 1976). Leurs mécanismes d'action sont dominants et conduisent à un gain de fonction puisqu'une seule altération sur l'un des deux allèles est suffisante pour assurer leur oncogénicité. Les oncogènes peuvent être activés par des mécanismes variés tels que l'intégration virale, la mutation ponctuelle, la microdélétion avec conservation du cadre de lecture, le réarrangement structural, l'amplification génique et la dérégulation de mRNAs codant pour des oncoprotéines.

Les gènes suppresseurs de cancer regroupent deux familles principales : les gènes de contrôle du cycle cellulaire (*Gatekeepers*) et les gènes de réparation de l'ADN (*Caretakers*). Les gènes de contrôle du cycle cellulaire sont capables de réguler négativement ce dernier et induisent l'apoptose, inhibant ainsi la croissance cellulaire. Les gènes de réparation de l'ADN sont responsables du maintien de l'intégrité du génome suite à un endommagement de l'ADN. Il s'agit principalement du système de réparation des mésappariements MMR (*Mismatch Repair*), du système de réparation des nucléotides NER (*Nucleotide Excision Repair*) et du système de réparation des cassures double-brin HR (*Homolog Recombinaison*). Ces gènes suppresseurs de tumeur peuvent être altérés et intervenir alors indirectement dans l'oncogénèse. Les mécanismes qui aboutissent à la perte de fonction de ces gènes sont récessifs car ils s'accompagnent de l'altération simultanée de leurs deux allèles. Les oncosuppresseurs peuvent être inactivés par des mécanismes variés tels les mutations ponctuelles (mutations non sens ou d'épissage, microdélétions ou microinsertions avec modification du cadre de lecture) ainsi que par des délétions, des insertions, des surexpressions de miRNAs et l'hyperméthylation des îlots CpG des promoteurs inhibant la transcription (Volgenstein and Kinzler, 2004).

1.5.1.3.2 Instabilité génomiques et anomalies génétiques/épigénétiques

Au cours des cinq dernières années, les techniques de séquençage massif parallèle de l'ADN et de l'ARN effectuées à partir de milliers de génomes de tumeurs malignes, ont permis d'approfondir nos connaissances sur les altérations génétiques dans l'oncogénèse et leur fréquence dans différents types de cancers, dont les carcinomes mammaires. La majorité des caractéristiques des cellules cancéreuses mammaires résultent de leur instabilité génomique et de l'acquisition progressive d'anomalies génétiques/épigénétiques.

1.5.1.3.2.1 Instabilité génomique

Les mécanismes moléculaires responsables de l'instabilité génomique et de la sélection de génomes complexes dans les cancers du sein sont encore largement inconnus, mais résultent en partie d'altérations de la réparation de l'ADN et des *checkpoints* et du maintien de la longueur des télomères *via* la réactivation de la télomérase. Cette dernière peut résulter de divers mécanismes tels l'existence (i) de mutations du promoteur de *hTERT* (< 5% dans les cancers du sein) et des gènes *ATRX* et *DAXX* de la voie alternative, (ii) de la prolifération de cellules souches ou progénitrices télomérase+, (iii) de dérégulations épigénétiques (acétylation, hyperméthylation) et (iv) d'altérations de facteurs de régulation (surexpression de c-Myc, sous-expression de Mad-1 et WT1). En plus du maintien de la longueur des télomères, la réactivation de la télomérase est impliquée dans l'activation des voies NF- κ B et Wnt/ β -caténine, l'adhésion et la prolifération cellulaire, l'EMT et la migration. Cette réactivation permet surtout aux cellules cancéreuses de contourner les mécanismes physiologiques de la sénescence et de continuer à se diviser jusqu'à ce que l'apparition de multiples petits télomères ayant atteint une réduction critique de leur taille, déclenche la « crise télomérique ». Celle-ci est caractérisée par une période de sénescence replicative complète, des fusions inter-chromosomiques localisées à leurs extrémités et une apoptose extensive (Capper et al., 2007 ; Jafri et al., 2016). Ces courts télomères dysfonctionnels ont été détectés dans de nombreux types de tumeurs malignes, dont les cancers du sein (Chin et al., 2004 ; Roger et al., 2013). L'association apparemment paradoxale de courts télomères et d'une activité télomérasique accrue et non corrélée aux niveaux d'expression de hTERT, a été récemment observée dans les sous-types les plus agressifs des carcinomes mammaires. Ainsi, les CSTNs sont caractérisés par l'activité télomérasique la plus élevée, les plus courts télomères, une distribution nucléaire diffuse de hTERT, un statut muté de TP53 et une surexpression simultanée des métalloprotéinases MMP2 et MMP7 (Ceja-Rangel et al., 2016).

1.5.1.3.2.2 Anomalies génétiques et épigénétiques

Ces altérations comprennent principalement les translocations chromosomiques, les amplifications et les délétions géniques, les mutations ponctuelles et les modifications épigénétiques. Trois études pangénomiques (Kandoth et al., 2013 ; Zack et al., 2013; Stransky et al., 2014) ont analysé les profils des translocations chromosomiques, les altérations du nombre de copies géniques et les mutations somatiques dans une douzaine de types tumoraux, incluant plus de 1000 tumeurs du sein (**Fig. 13**).

- **Translocations** : L'analyse du transcriptome de 7000 tumeurs appartenant à 20 types différents a révélé que les transcrits de fusion impliquant les différentes kinases sont rares dans les cancers du sein (< 2%). Les principales translocations chromosomiques observées dans les cancers du sein concernent *ALK*, *RET*, *PIK3CA*, *PDGFRA*, *NTRK3* et *FGFR1/2* (Stransky et al., 2014).
- **Altérations du nombre de copies d'ADN** : Les trois études pangénomiques précédemment citées ont révélé pour l'ensemble des tumeurs étudiées que les amplifications/délétions somatiques ont tendance à se situer dans les mêmes régions chromosomiques. Seules un tiers des régions identifiées comportent des oncogènes et des oncosuppresseurs préalablement décrits. Les gènes les plus fréquemment amplifiés dans les carcinomes mammaires sont *MYC*, *CCND1*, *EGFR*, *MDM2*, *CCNE1*, *ERBB2*, *ZNF703* et *ZNF217*. La délétion homozygote la plus fréquemment observée dans l'ensemble des cancers étudiés, dont le cancer du sein, est localisée en 9p21 et a pour cible le gène suppresseur de tumeur *CDKN2A* (*Cyclin-Dependent Kinase Inhibitor 2A*). Les autres gènes délétés dans les cancers du sein sont *PTEN*, *INPP4B*, *PPP2R2A*, *MLL3* et *MAP2K4*.
- **Mutations ponctuelles somatiques** : Les études pangénomiques récentes ont identifié 127 gènes significativement mutés dans les différents types de cancers. Ces gènes appartiennent pour la plupart aux voies de signalisation bien connues dans l'oncogénèse, dont les voies MAPK, PI3K et Wnt/ β -caténine, mais aussi à des voies émergentes impliquant la modification des histones, le métabolisme et la protéolyse. Le gène suppresseur de tumeur le plus fréquemment muté dans les tumeurs humaines est le gène *TP53* (42% pour l'ensemble des tumeurs et 33% pour les tumeurs du sein).

avec une prédominance pour le sous-type triple-négatif). L'oncogène le plus fréquemment muté est *PIK3CA* (18 % pour l'ensemble des cancers et 34 % pour le cancer du sein). Ces études ont de plus révélé que le cancer du sein est l'un des cancers présentant le plus faible taux de mutations somatiques, après la leucémie myéloïde chronique (Kandoth et al., 2013) (**Fig. 13**). Ainsi, seuls 7 gènes sont mutés dans les cancers du sein dont la fréquence est supérieure à 5% : *TP53* (37%), *PIK3CA* (36%), *GATA3* (11%), *MAP3K1* (8%), *MLL3* (7%) et *CDH1* (7%). 17 gènes supplémentaires sont mutés avec une fréquence variant de 1% à 5%. La fréquence des mutations est la plus élevée dans les carcinomes mammaires de type luminal B et la plus basse dans les carcinomes de type triple-négatif (Curtis et al., 2012).

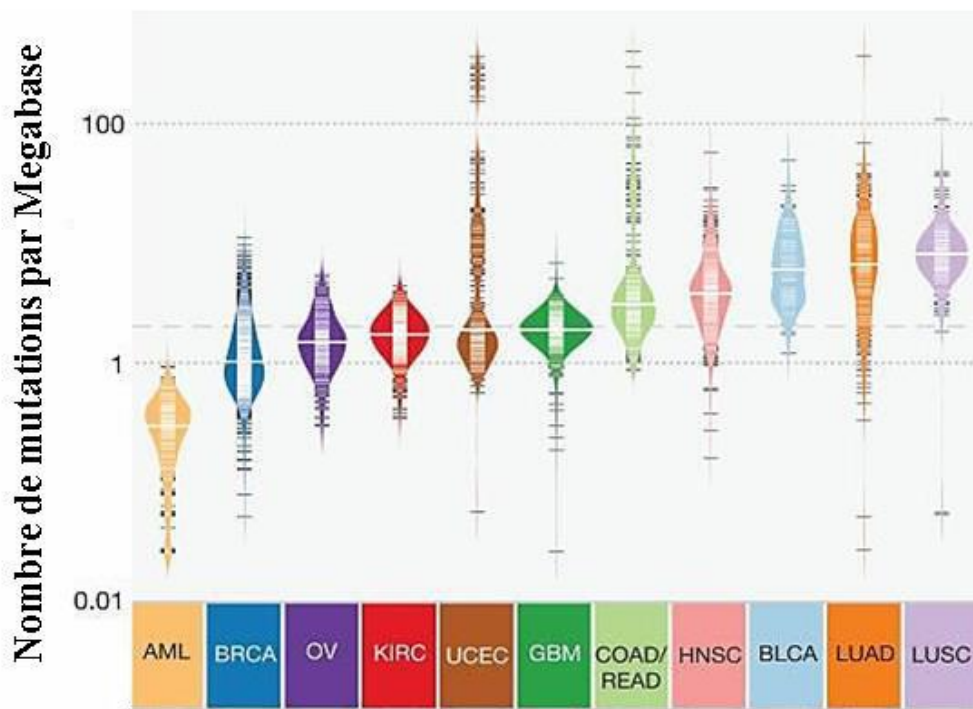


Figure 13 : Fréquence des mutations dans 12 types différents de tumeurs malignes

AML : acute myeloid leukaemia ; BRCA : breast carcinoma ; OV : ovarian serous carcinoma; KIRC : kidney renal clear cell carcinoma; UCEC : uterine corpus endometrial carcinoma; GBM : glioblastoma multiforme; COAD, READ : colon and rectal carcinoma; HNSC : head and neck squamous cell carcinoma; BLCA : bladder urothelial carcinoma; LUAD : lung adenocarcinoma; LUSC : lung squamous cell carcinoma (D'après Kandoth, McLellan et al., 2013).

1.5.1.3.2.3 Chromotripsis

Le modèle classique de la progression tumorale multi-étapes postule que les mutations s'accumulent progressivement selon un processus de sélection darwinienne (Nowell, 1976). Toutefois, une minorité de tumeurs mammaires agressives sont caractérisées par l'apparition brutale de nombreux réarrangements chromosomiques réalisant une *chromotripsis* (Stephens et al., 2011 ; Goncalves et al., 2014). Ainsi, de nombreuses mutations peuvent survenir durant de très courtes périodes pendant lesquelles des centaines de réarrangements chromosomiques localisés se produisent lors d'une seule division cellulaire catastrophique (Cai et al., 2014). Ces cassures génèrent des remaniements majeurs du génome pouvant favoriser la transformation cancéreuse par le biais de nombreuses amplifications d'oncogènes et de délétions de gènes oncosuppresseurs survenant simultanément au voisinage des translocations (Fig. 14). La prévalence de la *chromotripsis* a été évaluée à 41% dans les cancers du sein à haut risque (Przybytkowski et al., 2014) et pourrait expliquer le développement rapide de certaines tumeurs pendant les intervalles séparant deux consultations de dépistage (Nik-Zainal et al., 2012).

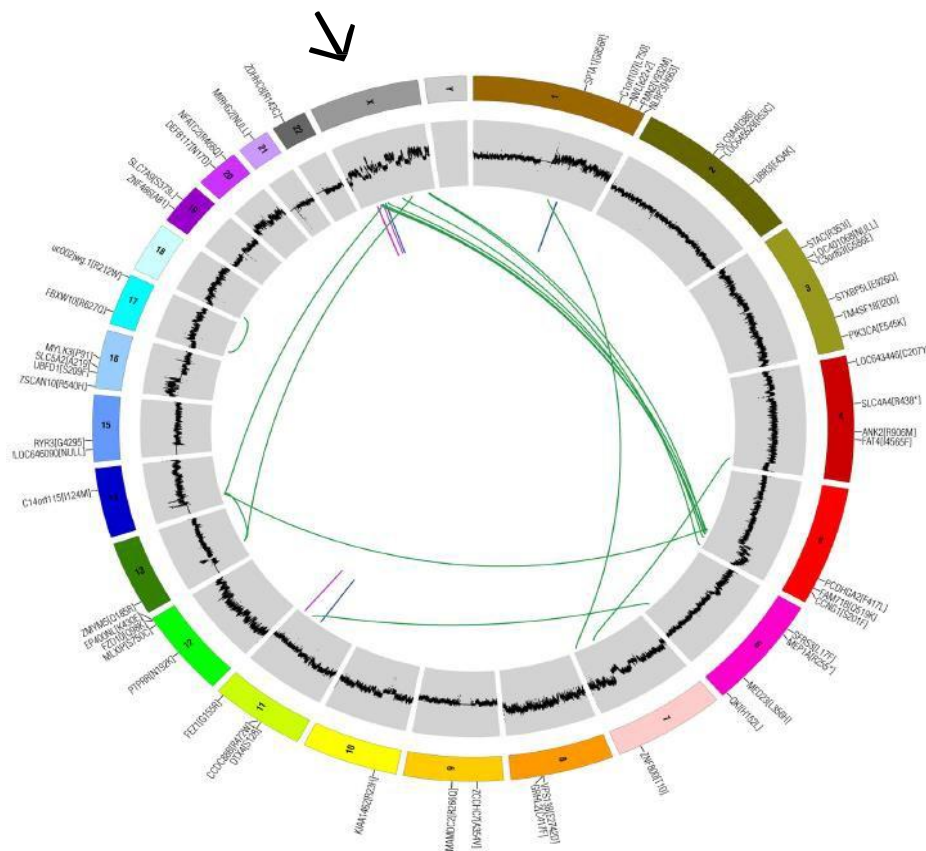


Figure 14 : Chromotripsis (→): Amplifications d'oncogènes et délétions d'oncosuppresseurs présentes simultanément à proximité de multiples translocations (Goncalves et al., 2014). La flèche → indique la région de la chromotripsis sur le chromosome 2.

Ces techniques globales ont donc identifié un degré de complexité supplémentaire dans le génome des cancers du sein et ce niveau croissant de complexité est corrélé à un pronostic plus défavorable. Ces analyses ont de plus révélé que la progression multi-étapes avec acquisition progressive de mutations génétiques et épimutations peut se combiner avec des dysfonctionnements télomériques, des défauts de réparation de l'ADN et une *chromotrypsis*, pour produire des niveaux élevés d'hétérogénéité intra-tumorale et inter-tumorale. L'identification des altérations génétiques clés dans les cancers du sein a servi de support à la recherche et à la sélection de cibles pertinentes en oncologie à l'origine du développement de nouvelles thérapies moléculaires (**Fig. 12**).

1.5.1.3.3 Anomalies des voies de signalisation

Les analyses pangénomiques réalisées dans les différents types de cancers, ont permis d'identifier les principales voies de signalisations altérées. Il s'agit des voies p53/apoptose/cycle cellulaire, Wnt, Notch, récepteurs à activité tyrosine kinase et ses voies d'aval MAPK et PI3K/Akt/mTOR. Pour les carcinomes mammaires, ces études réalisées sur plus de 1000 tumeurs du sein (Kandoth et al., 2013; Zack et al., 2013; Stransky et al., 2014) ont identifié un panel de 127 gènes mutés appartenant aux voies régulant le cycle cellulaire (*TP53*, *p27/CDKN1B*, *RB1*), la voie EGF (*PIK3CA*, *PTEN*, *PIK3R1*, *AKT1*, *NF1*), la voie Stress-JNK (*MAP3K1*, *MAP2K4*), la transcription des récepteurs oestrogéniques (*GATA3*, *NCOR1*, *RUNX1/CBFB*, *CTCF*, *FOXA1*), l'épissage des pré-RNAs (*SF3B1*), la méthylation des histones (*MLL3*) et le développement de la glande mammaire (*TBX3*).

Le TCGA (The Cancer Genome Atlas), publié en 2012, a permis d'analyser les carcinomes mammaires primitifs de 825 patientes en combinant plusieurs techniques innovantes à haut débit telles que SNP array, méthylation de l'ADN, séquençage des exomes, cDNA microarrays, séquençage des microRNAs et RPPA.

- ✓ Ces analyses ont identifié les récepteurs à activité tyrosine kinase les plus fréquemment altérés dans le cancer du sein, dont ERBB2/HER2 (21%), IGF1R (7%) et EGFR (10%).
- ✓ Elles ont également révélé que la voie PI3K/Akt/mTOR est la voie la plus fréquemment altérée dans le cancer du sein (**Fig. 15**) où elle apparaît fortement activée dans tous les sous-types de carcinomes mammaires. Les anomalies génétiques de la voie PI3K

comportent (i) des altérations des gènes *PIK3CA* (36%) et *AKT1* (2%) avec des mutations présentes dans les sous-types luminaux et HER2+ et des amplifications dans les CSTNs, (ii) des inactivations de *PTEN* résultant de délétions dans les CSTNs et de rares mutations dans les différents autres sous-types, (iii) des délétions du gène *INPP4B* principalement dans les CSTNs et (iv) de rares mutations de *PIK3R1* et d'*AKT1*. Les analyses RPPA ont confirmé la perte des deux gènes suppresseurs *PTEN* et *INPP4B* de la voie PI3K/Akt/mTOR, perte qui est plus fréquente dans les CSTNs que dans les autres sous-types tumoraux de carcinomes mammaires.

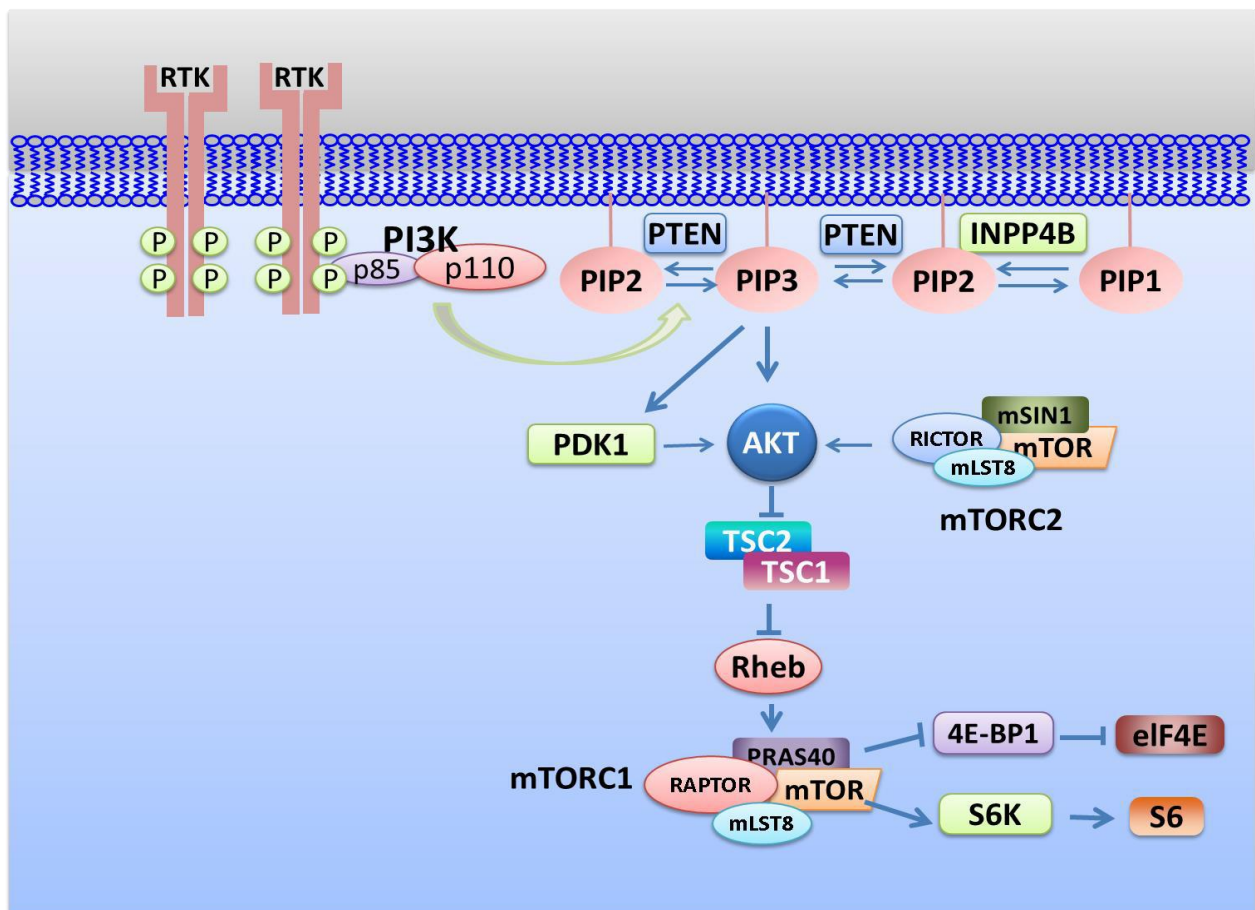


Figure 15 : Voie de signalisation PI3K/Akt/mTOR (D'après Fasolo and Sessa, 2012)

1.5.2 Classifications moléculaires des cancers du sein

Les différentes classifications cliniques, morphologiques et immunophénotypiques ne permettent cependant pas d'explorer la totalité de l'hétérogénéité intratumorale propre à chaque patiente, notamment au niveau génomique. Ainsi, l'analyse des mécanismes moléculaires des cancers du sein a fait l'objet de recherches intensives au cours des dernières décennies. La publication par une équipe de chercheurs de l'université de Stanford (Perou et al., 2000) d'un article identifiant les profils moléculaires des tumeurs du sein a révolutionné l'approche de la cancérogénèse mammaire. Depuis, confirmant la pertinence des classifications morphologiques, le cancer du sein est apparu comme une maladie très hétérogène caractérisée par des sous-types morphologiques et moléculaires pourvus de profils génomiques, épigénomiques et transcriptomiques distincts. Grâce au développement des techniques d'analyses à haut débit, la classification des cancers du sein a été considérablement enrichie par la prise en compte de marqueurs moléculaires spécifiques exprimés par les cellules tumorales mammaires.

1.5.2.1 Classifications génomiques et transcriptomiques

1.5.2.1.1 Classification transcriptomique « intrinsèque »

Depuis l'an 2000, il est devenu possible d'appréhender le transcriptome d'un tissu donné dans sa globalité. Ainsi, l'utilisation des puces à ADN (*cDNA microarrays*) a permis de mesurer simultanément l'expression de milliers de gènes au niveau ARN et de dresser un profil moléculaire transcriptomique de chaque échantillon. En établissant de tels profils d'expression pour l'ensemble des tumeurs mammaires, une première étude transcriptomique a permis de les classer en 5 catégories (**Fig. 16**) sur la base d'une signature « intrinsèque » correspondant à l'expression de 534 gènes (Perou, Sorlie et al., 2000 ; Sorlie, Perou et al., 2001). Ces 5 catégories comprennent les sous-types Luminal A, Luminal B, ERBB2+, *Basal-like* et *Normal-like*.

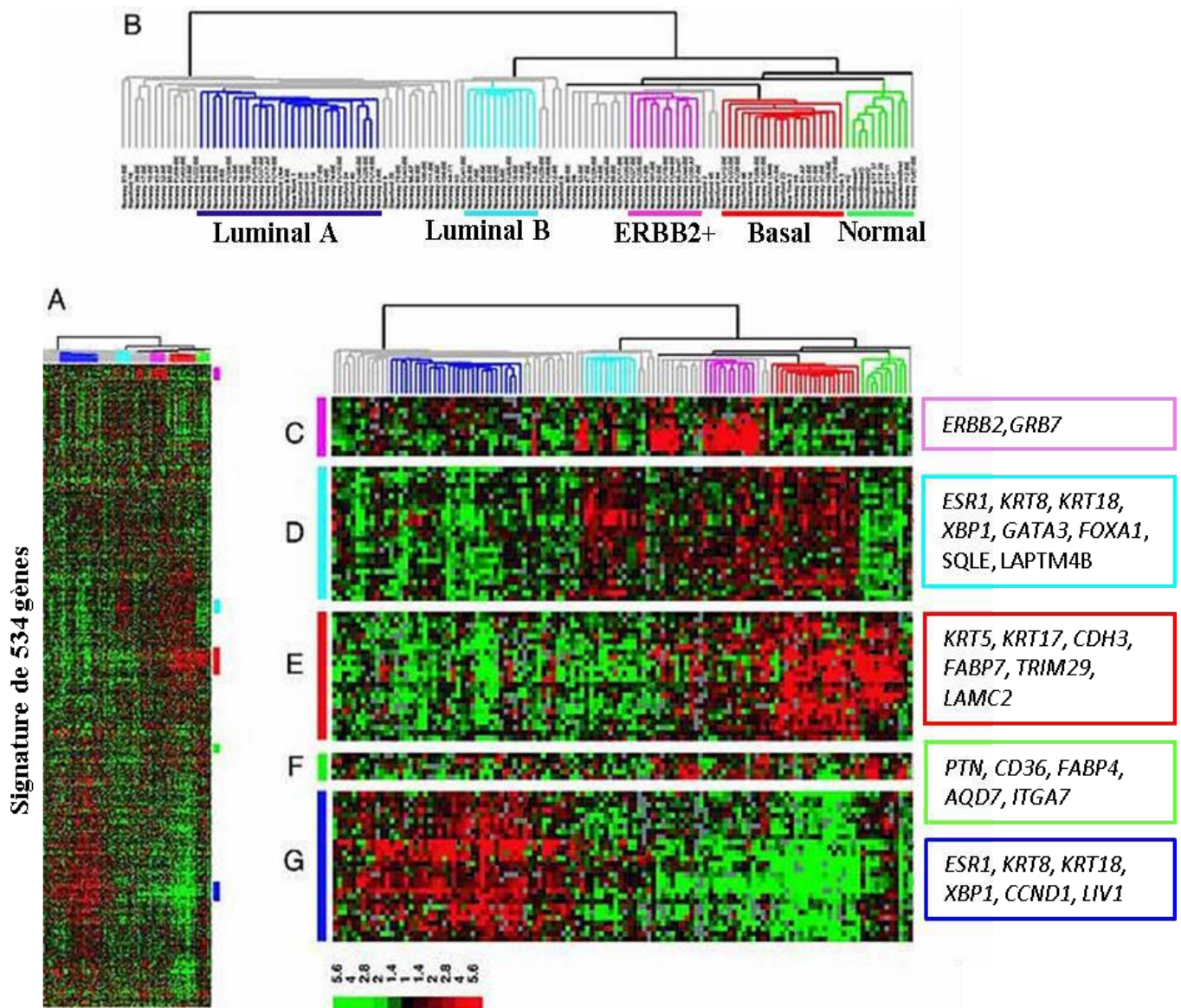


Figure 16 : Classification moléculaire intrinsèque des carcinomes mammaires infiltrants

A. Classification hiérarchique par le biais de la signature intrinsèque d’une série de 115 tumeurs du sein et 7 tissus mammaires sains.

B. Dendrogramme caractérisant les cinq sous-groupes de tumeurs.

C-G. Principaux gènes différentiellement exprimés dans les 5 sous-types tumoraux : ERBB2 (C), Luminal B (D), *Basal-like* (E), *Normal-like* (F) et Luminal A (G).

La barre d’échelle représente le niveau d’expression de chaque gène relativement au niveau médian de son expression dans tous les échantillons (D’après Sorlie et al., 2003).

Cette classification « intrinsèque » suppose que les différentes catégories sont issues de types cellulaires normaux différents. Les 5 sous-types peuvent être regroupés en 2 catégories : les tumeurs RE positives (RE+) exprimant le gène *ESR1* (80%) et les tumeurs RE négatives (RE-) n'exprimant pas le gène *ESR1* (20%).

Les tumeurs RE+ se subdivisent en tumeurs de type luminal A (60%) et luminal B (20%), qui présentent un profil d'expression proche de celui des cellules épithéliales lumorales mammaires. Les tumeurs de type luminal A expriment très fortement le récepteur aux œstrogènes et faiblement les gènes liés à la prolifération. Elles sont généralement de faible grade et de bon pronostic, tandis que les tumeurs de type luminal B expriment plus faiblement le récepteur aux œstrogènes, comportent une expression élevée des gènes liés à la prolifération et sont de moins bon pronostic (**Fig 17**).

Les tumeurs RE- comportent trois sous-types : ERBB2+, *Basal-like* et *Normal-like*.

- **Le sous-type ERBB2+** est caractérisé par une forte surexpression du gène *HER2/ERBB2* et des gènes associés à l'amplicon localisé sur le chromosome 17q12 ainsi que par la surexpression de gènes liés à la prolifération.
- **Le sous-type Basal-like** est caractérisé par l'expression de gènes de cellules myoépithéliales/basales normales de sein ainsi qu'une forte expression de gènes liés à la prolifération. Ce sous-type semble lui-même extrêmement hétérogène et représente la grande majorité des tumeurs de la catégorie triple-négative, caractérisées par étude immunohistochimique (Nielsen et al., 2004 ; Livasy et al., 2006). Le sous-type *basal-like* est fréquemment observé chez les femmes porteuses d'une mutation constitutionnelle du gène *BRCA1*. Ces tumeurs héréditaires partagent des caractéristiques génomiques transcriptomiques et morphologiques avec les tumeurs sporadiques de type *basal-like* (Foulkes et al., 2003), en particulier les mutations du gène *TP53*, l'expression des cytokératines et le profil triple-négatif des récepteurs. Les tumeurs *basal-like* sporadiques peuvent également présenter une inactivation bi-allélique somatique de *BRCA1* par méthylation ou inhibition du promoteur du gène *BRCA1* (Turner and Reis-Filho, 2006 ; Turner et al., 2007).
- **Le sous-type Normal-like** est caractérisé par l'expression des gènes de cellules épithéliales basales et de tissu adipeux. L'existence de ce sous-type est actuellement

contestée car il pourrait être issu d'un artéfact lié à une contamination par le tissu mammaire normal environnant (Parker et al., 2009).

Une classification « intrinsèque » simplifiée, nommée PAM50 et basée non plus sur 534 gènes mais sur 50 gènes, semble être la signature d'expression génique la plus utilisée aujourd'hui en pratique clinique.

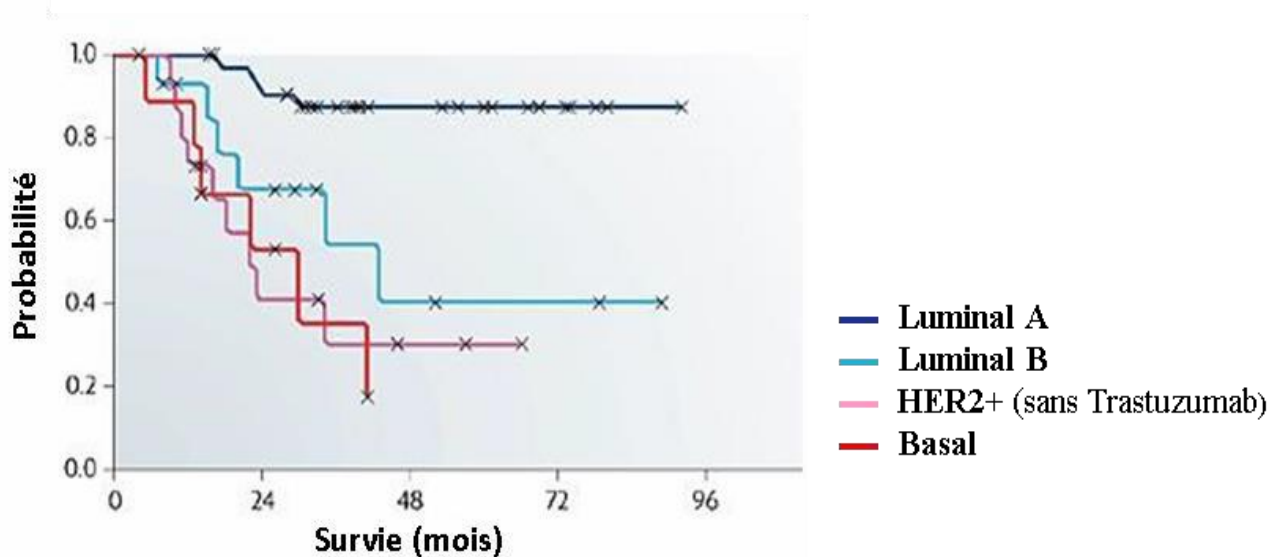


Figure 17 : Pronostic des différents sous-types moléculaires de cancers du sein représenté par la survie globale. NB : Depuis 2006, le traitement anti-HER2 Trastuzumab (Herceptine®) a radicalement amélioré le pronostic du sous-type HER2+ (D'après Sorlie et al., 2003).

1.5.2.1.2 Classification génomique et transcriptomique

Plus récemment, d'autres sous-types moléculaires ont été identifiés (Curtis et al., 2012) en utilisant une approche associant des données génomiques et transcriptomiques associées à des altérations somatiques acquises du nombre de copies (CNAs). L'analyse de la classification hiérarchique non supervisée des profils appariés d'ADN-ARN a ainsi identifié 10 sous-groupes présentant des évolutions cliniques distinctes (**Fig. 18**).

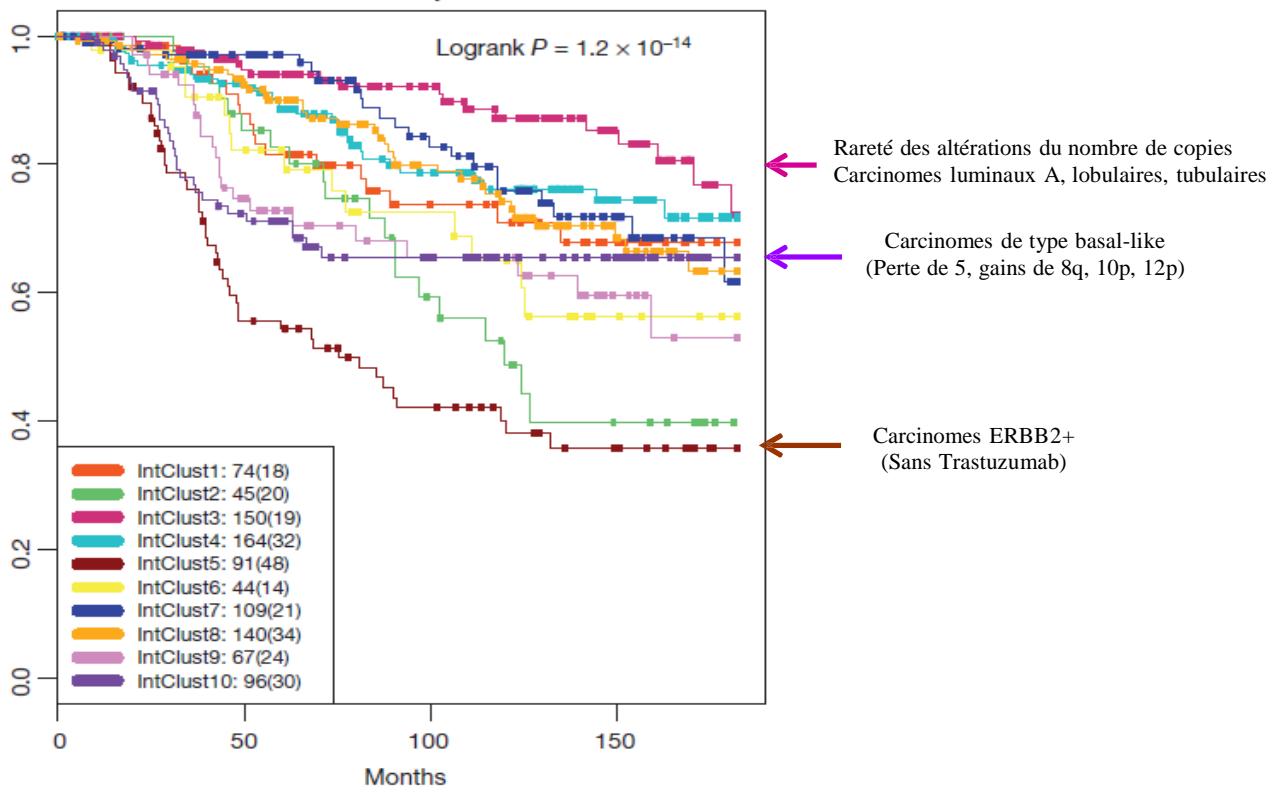


Figure 18: Classification génomique et transcriptomique intégrant les anomalies du nombre de copies. Classification hiérarchique non supervisée révélant 10 sous-groupes tumoraux d'évolution clinique distincte (D'après Curtis et al., 2012).

1.5.2.2 Classification moléculaire intégrée des cancers du sein

Une nouvelle classification moléculaire des carcinomes mammaires a été récemment proposée en analysant et en intégrant les données issues des techniques de CGH array, SNP array, séquençage d'exome, méthylation de l'ADN, cDNA array, miRNAseq et RPPA (TCGA, 2012) (**Fig. 19**). Quatre sous-types moléculaires ont été individualisés dont les anomalies génétiques et épigénétiques pourraient en grande partie rendre compte de l'hétérogénéité clinique et morphologique des tumeurs analysées (**Tableau 1**). De plus, la comparaison des tumeurs mammaires de type basal avec des tumeurs ovariennes séreuses de haut grade et les tumeurs endométriales, a révélé de nombreuses anomalies moléculaires communes telles (i) l'inactivation de *BRCA1*, (ii) la perte de *RBI* et l'amplification de *cycline E1*, (iii) une surexpression de *AKT3*, (iv) l'amplification et la surexpression de *MYC* et (v) une fréquence élevée de mutations de *TP53*. Ces anomalies partagées suggèrent la possibilité d'approches thérapeutiques communes, telles que les dérivés du platine (**Figure 20a et b**).

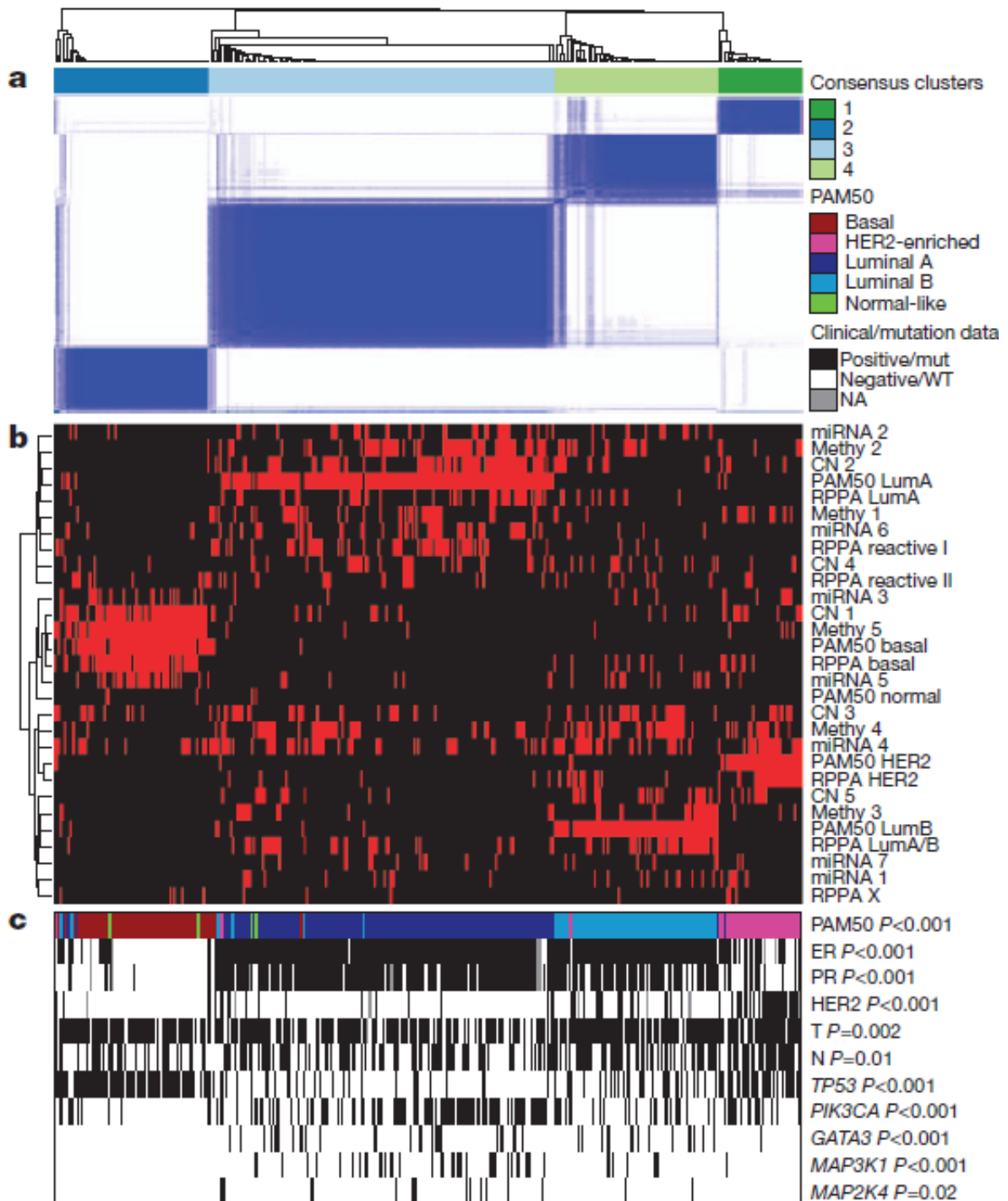


Figure 19 : Analyse coordonnée des différents sous-types moléculaires de cancers du sein (TCGA, 2012)

- (a) L'analyse des *clusters consensus* permet l'identification de 4 groupes principaux.
- (b) *Heat-map* des différents sous-types définis par l'analyse des miRNAs, de la méthylation de l'ADN, le nombre de copies, l'expression des mRNAs (PAM50 et RPPA).
- (c) Corrélations avec les anomalies moléculaires.

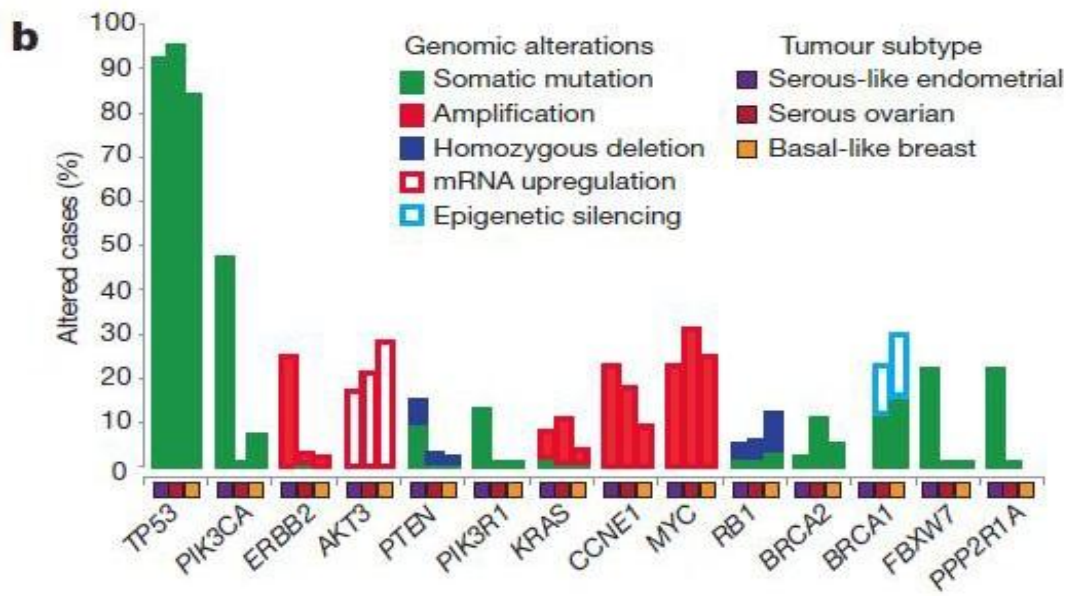
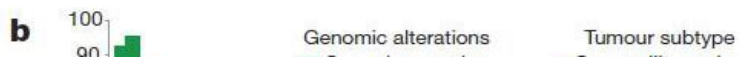
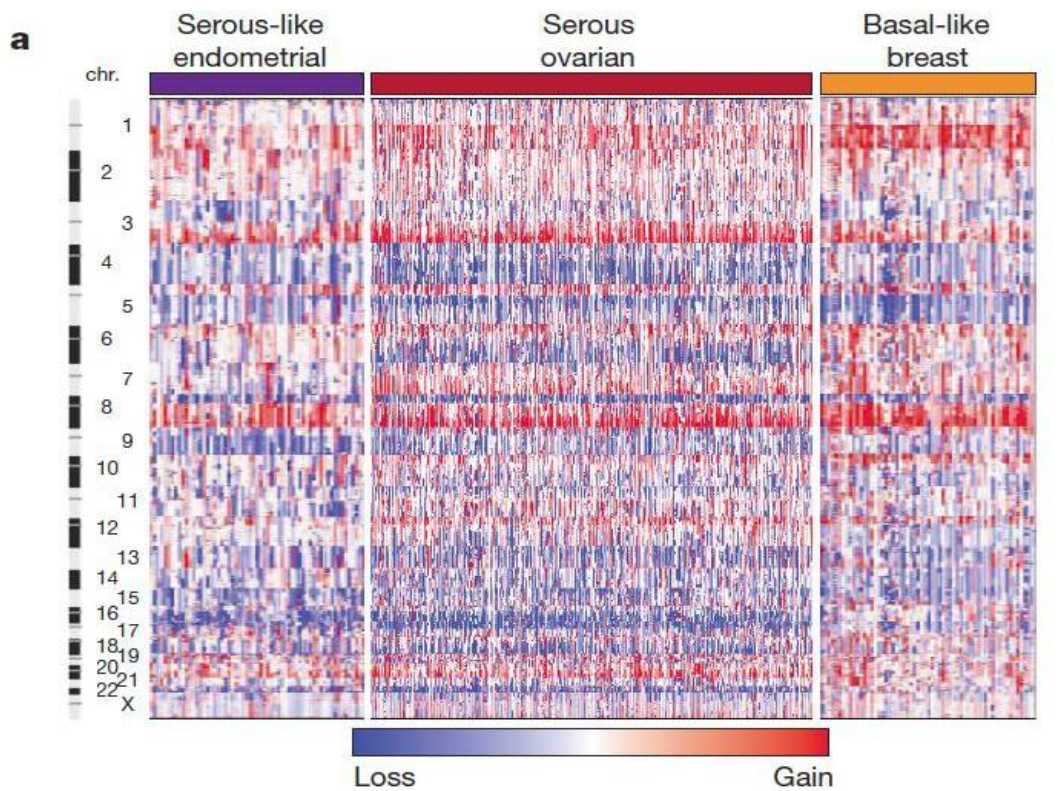


Figure 20a et 20b : Etude comparée des altérations moléculaires des cancers mammaires de type triple-négatif, des carcinomes séreux ovariens de haut grade et des carcinomes endométriaux (TCGA, 2012).

Sous-types	Luminal A	Luminal B	Triple-négatif	HER2+
Voie TP53	mutation de <i>TP53</i> (12%) gain de <i>MDM2</i> (14%)	mutation de <i>TP53</i> (32%) gain de <i>MDM2</i> (31%)	mutation de <i>TP53</i> (84%) gain de <i>MDM2</i> (14%)	mutation de <i>TP53</i> (75%) gain de <i>MDM2</i> (30%)
Voie PIK3CA/PTEN	mutation de <i>PIK3CA</i> (49%) mutation/perte de <i>PTEN</i> (13%) perte d' <i>INPP4B</i> (9%)	mutation de <i>PIK3CA</i> (32%) mutation/perte de <i>PTEN</i> (24%) perte d' <i>INPP4B</i> (16%)	mutation de <i>PIK3CA</i> (7%) mutation/perte de <i>PTEN</i> (35%) perte d' <i>INPP4B</i> (30%)	mutation de <i>PIK3CA</i> (42%) mutation/perte de <i>PTEN</i> (19%) perte d' <i>INPP4B</i> (30%)
Voie RB1	amplification de <i>Cycline D1</i> (29%) gain de <i>CDK4</i> (14%) sous-expression de <i>CDKN2C</i> surexpression de <i>RB1</i>	amplification de <i>Cycline D1</i> (58%) gain de <i>CDK4</i> (25%)	mutation/perte de <i>RB1</i> (20%) gain de <i>CDK4</i> (14%) surexpression de <i>CDKN2A</i> sous-expression de <i>RB1</i>	amplification de <i>Cycline D1</i> (38%) gain de <i>CDK4</i> (24%)
Expression des mRNAs	cluster RE+++ prolifération +	cluster RE +/- prolifération +/-	signature basale prolifération +++	signature amplicon HER2 Prolifération +++
Nombre de copies	Diploidie ; gain de 1q, 8q, 8p11 perte de 8p, 16q amplification de 11q13.3 (24%)	aneuploidie, amplifications focales gain de 1q, 8q, 8p11; Perte de 8p, 16q ; amplifications de 11q13.3 (51%) et 8p11.23 (28%)	aneuploidie, instabilité génomique; gain de 1q, 10p; perte de 8p, 5q; gain focal de MYC (40%)	instabilité génomique ; gain de 1q, 8q; perte de 8p; amplification focale de 17q12 ERBB2 (71%)
Mutations de l'ADN	<i>PIK3CA</i> (49%); <i>TP53</i> (12%); <i>GATA3</i> (14%); <i>MAP3K1</i> (14%)	<i>TP53</i> (32%); <i>PIK3CA</i> (32%); <i>MAP3K1</i> (5%)	<i>TP53</i> (84%); <i>PIK3CA</i> (7%)	<i>TP53</i> (75%); <i>PIK3CA</i> (42%); <i>PIK3R1</i> (8%)
Méthylation de l'ADN		hyperméthylation	hypométhylation	
Expression protéique	signalisation oestrogénique +++ expression élevée de MYB	signalisation oestrogénique + expression élevée de FOXM1, MYC	expression élevée des protéines de réparation de l'ADN Perte d'expression de PTEN et INPP4B	surexpression et phosphorylation de EGFR and HER2

Tableau 1 : Caractères génomiques, protéomiques et cliniques des sous-types tumoraux de la classification moléculaire intégrée

1.5.2.3 Classification moléculaire intégrée des cancers du sein de type triple-négatif

La première classification intrinsèque de T. Sorlie et al. a révélé l'hétérogénéité moléculaire intra-tumorale des cancers du sein et identifié en particulier le sous-type *basal-like* (BL) par l'absence d'expression des récepteurs RE, RP et HER2 et la surexpression des cytokératines de type basal. Des études ultérieures ont révélé que: les CSTNs et les carcinomes de type BL ne sont pas des entités homogènes: (i) les carcinomes de type BL sont caractérisés par des signatures de perte d'hétérozygotie spécifiques contenant de multiples allèles (Wang et al., 2004), (ii) les carcinomes de type BL sont le plus souvent triple-négatifs en IHC (RE-, RP-, HER2-) et dans plus de 50% des cas positifs pour EGFR/HER1 (Nielsen et al., 2004), (iii) les CSTNs ne sont pas tous des carcinomes BL et n'ont pas tous un pronostic défavorable. Les CSTNs de bon pronostic présentent un marquage positif pour le récepteur aux androgènes (RA) et une absence de marquage pour les marqueurs basaux (cytokératines basales, EGFR, cadhérine P, p53), tandis que les CSTNs de mauvais pronostic sont négatifs pour le RA et positifs pour les marqueurs basaux (Rakha et al., 2007), (iv) les carcinomes de type BL peuvent être HER2+ en IHC (Matos et al., 2005), (v) des types histologiques spécifiques de type BL peuvent être de pronostic très favorable, tels le carcinome adénoïde kystique (Azoulay et al., 2005), (vi) la majorité des cancers du sein présentant des mutations du gène *BRCA1* sont de type BL (Livasy et al., 2006), (vii) les carcinomes de type BL comportent une importante réponse stromale lymphocytaire (Livasy et al., 2006).

Depuis 2005, une meilleure identification des profils génomiques, transcriptomiques et phénotypiques des CSTNs a pu être réalisée grâce au développement intensif des techniques d'analyse à haut débit et à une meilleure caractérisation immunophénotypique des CSTNs. Plusieurs sous-catégories de CSTNs ont été ainsi définies, reposant sur l'association de leur profil morphologique et immunohistochimique, de leurs anomalies moléculaires et des particularités de leur microenvironnement tumoral. Ces nouvelles sous-classifications moléculaires des CSTNs ont l'intérêt de présenter une implication thérapeutique potentielle pour l'éligibilité des patients dans des essais cliniques.

10% des CSTNs appartiennent à des sous-types spécifiques: (i) carcinomes adénoïdes kystiques avec t(6;9)(q22-23;p23-24) et fusion *MYB-NFIB* associées à des mutations de *BRAF*, *MYB*, *FBXW7*, *SMARCA5*, *SF3B1*, *FGFR2* (Martelotto et al., 2015), (ii) carcinomes du sein sécrétoires avec t(12;15) et fusion *ETV6-NTRK3* (del Castillo et al., 2016), (iii)

cancers du sein de type médullaire comportant un architecture épithéliale syncytiale, un infiltrat stromal lymphocytaire diffus intra et péri-tumoral et des mutations du gène *BRCA1* (Dieci et al., 2014b), (iv) carcinomes apocrines exprimant le RA et (v) carcinomes métaplasiques agressifs comportant une amplification de *EGFR/HER1*, des mutations de *PIK3CA*, des altérations de la voie Wnt/ β -caténine, une faible expression de la claudine et un pourcentage important de cellules souches cancéreuses et en EMT (Cadoo et al., 2012).

Pour les 90% des CSTNs appartenant au sous-type « non spécifique », les analyses à haut débit ont révélé que les CSTNs représentent un groupe de tumeurs très hétérogènes différant par leurs caractères morphologiques, leurs profils d'expression génique, leur évolution clinique, leur pronostic et leur sensibilité aux thérapies systémiques. Ces CSTNs constituent une entité clinique composée de tous les sous-types moléculaires intrinsèques. Les tumeurs de type *basal-like* (*claudin-low* et *basal-like*) représentent 70% des CSTNs, alors que les CSTNs non *basal-like* (Luminal A, luminal B, enrichi en HER2) représentent 30% des CSTNs. Les CSTNs et les tumeurs de type *basal-like* comportent des discordances dans 25% des cas et ne doivent pas être considérés comme des entités synonymes (Prat et al., 2013). L'hétérogénéité des CSTNs reflètent non seulement l'hétérogénéité des cellules tumorales mais aussi celle du microenvironnement.

Plusieurs classifications récentes ont identifié des sous-groupes de CSTNs liés aux caractéristiques de la population de cellules cancéreuses, du stroma tumoral et du système immunitaire: (i) 6 sous-types : *basal-like* 1 et 2 (BL1, BL2), *mesenchymal-like* (M), *mesenchymal stem-like* (MSL), *immunomodulatory* (IM) et *luminal androgen receptor* (LAR) (Lehmann et al., 2011), (ii) 4 sous-groupes : *luminal androgen receptor* (LAR), *mesenchymal* (MES), *basal-like immunosuppressed* (BLIS) et *basal-like immune-activated* (BLIA) (Burstein et al., 2015) et (iii) 4 *clusters* : *immunity cluster*, *proliferation/DNA damage cluster*, *androgen receptor (AR)-like cluster* et *matrix/invasion cluster* (Bonsang-Kitzis et al., 2015).

Malgré ces résultats décisifs, des études concernant de nouveaux biomarqueurs, dont les longs ARNs non codants, apparaissent très souhaitables, afin d'améliorer la compréhension de l'hétérogénéité de ces tumeurs et de stratifier les tumeurs mammaires et en particulier les CSTNs, dans des groupes biologiques encore plus homogènes.

2 LES LONGS ARNs NON CODANTS

2.1 Nouvelle organisation fonctionnelle du génome

Au cours des deux dernières décennies, le dogme central de la biologie concernant les flux d'informations génétiques et leurs conversions en fonctions biologiques a été très largement réévalué. Selon ce dogme « protéino-centrique », l'information génétique contenue dans l'ADN est transcrite en ARNs et traduite en protéines, l'ARN jouant uniquement un rôle de médiateur de la synthèse protéique, tandis que les protéines représentent les effecteurs structuraux et régulateurs responsables des grandes fonctions biologiques participant à l'homéostasie cellulaire et tissulaire.

Le concept de transcrits fonctionnels non traduits avait été initialement proposé par Francis Crick dès la fin des années 1950, lorsqu'il supposa l'existence « d'ARNs métaboliques » (F.H. Crick, 1958). Cette hypothèse fut confirmée quelques années plus tard avec l'identification des premiers ARNs non codants, tels les ARNs de transfert, ARNs ribosomaux et petits ARNs non codants nucléaires (Holley et al., 1965 ; Busch et al., 1982). Cependant, malgré la découverte ultérieure de nombreux autres ARNs non codants, ces derniers ont été longtemps considérés comme des exceptions acceptées par le dogme central (ARNt, ARNr), un « bruit transcriptionnel » ou des produits marginaux issus de la transcription (Comings, 1972). Toutefois, à la fin des années 1990, plusieurs études ont été consacrées au potentiel biologique d'ARNs non codants, tels *Xist* et *H19*. Au cours des années 2000, le développement des technologies génomiques à grande échelle combinées aux analyses bioinformatiques, a constitué une révolution dans la compréhension du flux d'information dans les cellules eucaryotes, en révélant une complexité, une flexibilité et une dynamique insoupçonnées du transcriptome humain. Les techniques de séquençage à haut débit de l'ADN et l'ARN ont permis avant tout de séquencer de plus en plus rapidement et avec un coût de plus en plus accessible, des génomes et des transcriptomes entiers. Elles ont également contribué à la découverte de nouvelles entités biologiques clés, dont les longs ARNs non codants (lncRNAs), les microDNAs (petits ADN circulaires extra-chromosomiques) et les petits fragments dérivés des tRNAs (Reon and Dutta, 2016).

Ainsi, les études effectuées grâce aux analyses de *cDNA microarray* de haute résolution et de séquençage massif parallèle (NGS), ont révélé des niveaux significatifs d'activité transcriptionnelle au sein des régions génomiques annotées et non annotées :

- La grande majorité du génome est transcrite en ARNs (*pervasive transcription*), la plupart n'ayant pas de potentiel codant pour les protéines (Berretta and Morillon, 2009 ; Clark et al., 2011b). 85% du génome humain est transcrit de façon bidirectionnelle et est essentiellement constitué de régions non codantes, le plus souvent non annotées sur le plan fonctionnel (Okazaki et al., 2002 ; *The Encode Project Consortium*, 2012 ; Djebali et al., 2012), alors que seulement 1.2% des séquences génomiques codent pour des protéines (Lander et al., 2001 ; *Human Genome sequencing Consortium International*, 2004 ; Ponting, 2009).
- 80% des séquences du génome semblent potentiellement fonctionnelles car elles présentent des marqueurs biochimiques de transcription active, incluant des sites de fixation de facteurs de transcription (FT) ou de DNase I et des domaines de modification de la chromatine (*Encode Project Consortium*, 2012). Bien que la transcription ne soit pas toujours synonyme de fonctionnalité, l'analyse de ces signatures reproductibles a ainsi permis d'identifier dans le génome humain 399.124 régions comportant des sites *enhancers*, 70.292 régions présentant des sites promoteurs, et 3 à 8% des bases nucléotidiques sous sélection négative (Projet ENCODE). En outre, les études pan-génomiques par RNA-seq ont révélé que les transcrits provenant des régions non codantes prédominent dans la population cellulaire des ARNs non ribosomiaux et non mitochondriaux.
- Les transcrits isoformes provenant de sites d'épissage et de promoteurs alternatifs représentent la majorité des lncRNAs et cette complexité transcriptomique inattendue présente un faible niveau d'expression.

Les techniques de séquençage à haut débit ont aussi permis de réévaluer le concept traditionnel de l'organisation fonctionnelle génomique, défini par la présence d'îlots de gènes codant pour des protéines qui sont dispersés dans un « océan » de séquences répétées et non-transcrites (Kapranov et al., 2002 ; Okazaki et al., 2002). Ce changement paradigmatique

majeur a été rendu possible grâce à l'identification de nouveaux ARNs non codant pour des protéines (ncRNAs), dont le nombre s'est révélé supérieur à celui des ARNs codants (Carninci et al., 2005 ; Trapnell et al., 2010). Les ARNs sont actuellement considérés comme des molécules très versatiles impliquées dans la transmission de l'information *via* des interactions multiples avec les protéines, l'ARN et l'ADN et une organisation en structures secondaires et quaternaires. Grâce à l'utilisation croissante des techniques NGS, le rôle biologique des ARNs, initialement limité à la synthèse protéique, s'est ainsi étendu à un large spectre structural, régulateur et catalytique, incluant l'organisation de l'architecture nucléaire et chromosomique et la coordination de l'expression des gènes codant pour les protéines (Reis and Verjovski-Almeida, 2012 ; Wahlestedt, 2013 ; St Laurent et al., 2014).

2.2 Les ARNs non-codants

Les ARNs non codants (ncRNAs) représentent la grande majorité du transcriptome humain. Ils sont exprimés de façon ubiquitaire ou peuvent être spécifiques d'un tissu donné. Ils comportent diverses fonctions architecturales, catalytiques et régulatrices, particulièrement critiques dans de nombreuses fonctions biologiques, telles la pluripotence, le développement embryonnaire, l'organogénèse, la réponse au stress et la sénescence. Les ncRNAs peuvent être classés en deux catégories en fonction de leur taille et de leurs propriétés structurales et régulatrices. Cette longueur qui est arbitrairement fixée à 200 nucléotides (nt), correspond au seuil de sensibilité des méthodes d'extraction de l'ARN et permet de différencier les petits ARNs non codants des longs ARNs non codants. Bien que la majorité des ncRNAs soient encore peu caractérisés, un nombre croissant d'études semblent les impliquer dans les fonctions biologiques fondamentales et différents états pathologiques, dont les maladies neurodégénératives, infectieuses, cardiovasculaires, auto-immunes et le cancer (Wapinski and Chang, 2011). Il est hautement probable qu'au cours des prochaines années, de nouvelles classifications intégrant les ncRNAs enrichiront et affineront la pratique médicale cancérologique, que des ncRNAs seront identifiés comme des biomarqueurs diagnostiques, pronostiques et prédictifs et que certains d'entre eux constitueront des cibles thérapeutiques potentielles.

2.2.1 Petits ARNs non codants

Les petits ARNs ont une taille variant de 18 nt à 200 nt. Ils peuvent être divisés en deux sous-familles, les petits ARNs interférents et les autres petits ARNs non codants (Farazi et al., 2008 ; Spicuglia et al., 2013 ; Weick and Miska, 2014). Ces ARN non codants sont impliqués dans la régulation de l'expression des gènes, contrôlent la réponse antivirale, assurent la méthylation et la pseudo-uridilation d'autres RNAs et favorisent l'orientation de l'hétérochromatine dans la formation des centromères (Morris and Mattick, 2014). Ils sont également impliqués dans le fonctionnement des complexes d'épissage et peuvent intervenir dans la régulation de l'activité des transposons et de l'état de la chromatine (Mattick and Makunin, 2006 ; Liu and Paroo, 2010 ; Tisseur, Kwapisz and Morillon, 2011 ; Dogini et al., 2014 ; Ha et al., 2014b). En outre, les petits ARNs non codants pourraient jouer un rôle important dans l'activité des promoteurs et des terminateurs (Tisseur et al., 2011). Les ARNs interférents comprennent les siRNAs endogènes (*short interfering RNAs*), les piRNAs (*Piwi-interacting RNAs*) et les miRNAs (*microRNAs*). Leur taille varie de 19 nt à 31 nt (Fig. 21). Les autres petits ARNs non codants, dont la taille varie de 32 à 200 nt, comprennent les snoRNAs impliqués dans les modifications des rRNAs, les PASRs participant à la régulation de la transcription des gènes codant pour les protéines, les TSSa-RNAs assurant le maintien de la transcription, les PROMPTs activant la transcription et les crasiRNAs recrutant des protéines hétérochromatiniennes et centromériques (Tableau 2).

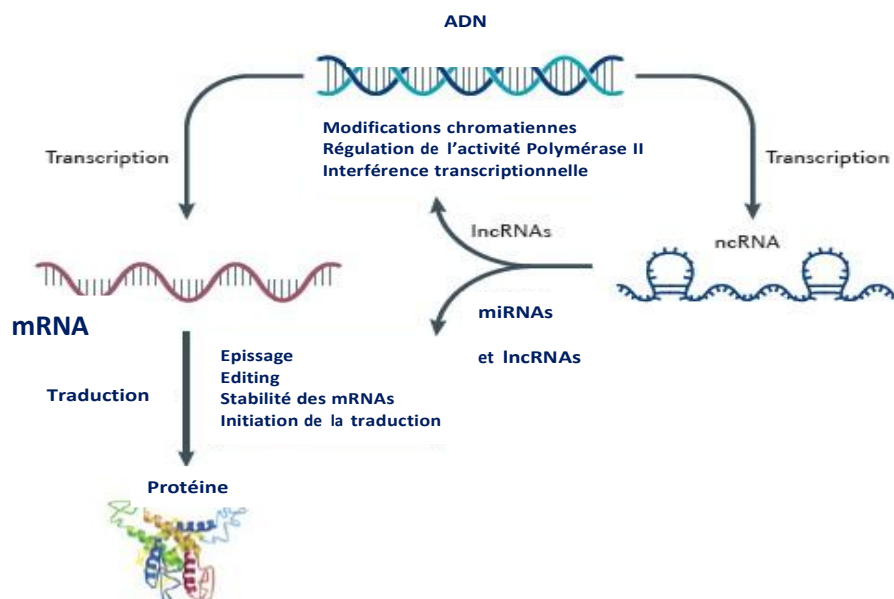


Figure 21 : Biogénèse des ARNs codant pour des protéines et des ARNs non codants

Type	Taille (nt)	Fonctions
Petits ncRNAs	< 200	
Courts ncRNAs	19-31	
miRNAs		ciblage des mRNAs et régulation de la prolifération, de la différenciation et de l'apoptose
siRNAs		<i>silencing</i> génique post-transcriptionnel et défense contre les acides nucléiques pathogènes
tiRNAs		régulation de la transcription par ciblage de complexes épigénétiques inhibiteurs
piRNAs		répression des transposons et <i>silencing</i> des retrotransposons, méthylation de l'ADN et renouvellement des cellules souches
tel-sRNAs		régulation épigénétique
Moyens ncRNAs	32-200	
snoRNAs		modifications des rRNAs
PASRs		régulation de la transcription des gènes codants
TSSa-RNAs		maintien de la transcription
PROMPTs		activation de la transcription
crasiRNAs		recrutement des protéines constituant l'hétérochromatine et les centromères
Longs ncRNAs	>200	
lncRNAs intergéniques		régulation épigénétique des gènes codants, pluripotence
lncRNAs sens		régulation de l'expression des gènes codant pour des protéines, parfois double fonction codante et non codante
lncRNAs antisens		régulation de l'expression des miRNAs et mRNAs et inhibition anti-sens des gènes codants et des autres ncRNAs
lncRNAs introniques		régulation de l'expression des gènes codants et des gènes parentaux
lncRNAs bifonctionnels		modulation de l'expression génique.

Tableau 2 : Principaux petits et longs ARNs non codants (D'après Kim and Reitmair, 2013)

2.2.2 Longs ARNs non codants

Les longs ARNs non codants (lncRNAs) ont été découverts et analysés plus récemment que les petits ARNs non codants à partir des études de séquençage du cDNA de génomes murins. Les lncRNAs représentent actuellement le sous-groupe majoritaire des ncRNAs (Wapinski and Cheng, 2011 ; Ma et al., 2013). Il s'agit d'une famille de molécules beaucoup plus hétérogènes que les petits ncRNAs et dotées de structures et de fonctions extrêmement variées. En plus de leur longueur, les lncRNAs présentent des différences significatives avec les petits ncRNAs ainsi que des similarités inattendues avec les mRNAs. Alors que les petits ncRNAs régulent l'expression génique *via* leur liaison avec des séquences spécifiques (*microRNA response elements*, MREs), les lncRNAs coordonnent l'expression des gènes par de multiples mécanismes encore largement méconnus et présentent une spécificité tissulaire et développementale caractéristiques (Cawley et al., 2004 ; Ravasi et al., 2006 ; Amaral and Mattick, 2008 ; Mercer et al., 2009 ; Whitehead et al., 2009). De plus, des études convergentes ont révélé que les lncRNAs sont particulièrement impliqués dans la régulation de l'homéostasie cellulaire (Guttman et al., 2011 ; Pauli et al., 2011).

2.3 Longs ARNs non codants

2.3.1 Définition des lncRNAs

Les lncRNAs ont été initialement définis par une longueur supérieure à 200 nt et une absence de traduction en protéines (Frith et al., 2006 ; Pang et al., 2006 ; Kapranov et al., 2007 ; Wang and Cheung, 2011a ; Prabakaran et al., 2014).

Cette définition non standardisée a été depuis lors remise en cause :

- La longueur > 200 nt n'est pas un critère absolu puisque des ncRNAs de moins de 200 nt (*BCI*, *snaR*) sont inclus dans les lncRNAs.
- L'absence de cadre de lecture bien défini ne peut être considérée comme un facteur discriminant entre les lncRNAs et les mRNAs :
 - ✓ 50% des lncRNAs intergéniques contiennent des cadres de lecture ouverts plus étendus ainsi qu'un potentiel traductionnel élevé, comparable à celui des

mRNAs, suggérant qu'ils sont activement exportés vers le cytoplasme et traduits en courts polypeptides. Ainsi, les cadres de lecture des lncRNAs *H19* et *TUG1* comportent respectivement 256 et 82 acides aminés potentiels.

- ✓ De rares lncRNAs codent pour de courts polypeptides, nécessitant de nouvelles approches méthodologiques afin de distinguer les lncRNAs des RNAs codant pour des polypeptides. Bien que la majorité des lncRNAs soient associés à des ribosomes, très peu sont traduits en oligopeptides fonctionnels en raison de leur instabilité, mais pourraient jouer un rôle modulateur dans la traduction des protéines (Anderson et al., 2015).
- La machinerie transcriptionnelle, en particulier l'utilisation de l'ARN polymérase II, est commune aux lncRNAs et aux mRNAs et des polypeptides de moins de 100 acides aminés ont été identifiés comme fonctionnels.
- Le caractère non codant n'est pas spécifique des lncRNAs :
 - ✓ Les techniques de séquençage NGS ont révélé l'existence de gènes bifonctionnels. Ces transcrits codent non seulement pour des protéines mais fonctionnent aussi comme des lncRNAs régulateurs. Ainsi, l'intron du gène codant pour la protéine *SPRY4* code également pour le lncRNA *SPRY4* qui joue un rôle important dans la prolifération et la mobilité des mélanocytes.
 - ✓ Certains mRNAs perdent leur capacité à coder pour des protéines (*Xist*) tandis que d'autres acquièrent une fonction codante (*SRA*) (Wapinski et al., 2011).
 - ✓ Des mRNAs peuvent agir comme des molécules régulatrices analogues à des ncRNAs. Ainsi, le domaine mRNA de *TP53* codant pour le site de fixation de MDM2 interagit directement avec ce dernier, neutralise son activité ligase et promeut la traduction du mRNA de *TP53*.
 - ✓ L'épissage alternatif, en incorporant de multiples exons issus de gènes codants et non codants, génère des transcrits ambigus qui échappent à toute classification.

Néanmoins, cette définition est aujourd'hui la meilleure qui puisse être utilisée en attendant une classification fonctionnelle des lncRNAs.

2.3.2 Méthodes d'investigation et d'annotation des lncRNAs

A l'ère pré-génomique, les premiers lncRNAs ont été découverts en utilisant des approches génétiques fonctionnelles reposant sur leurs relations avec des mécanismes cellulaires spécifiques tels que (i) la répression par interférence transcriptionnelle d'*Ubx* induite par le lncRNA *bdx*, (ii) l'inactivation du chromosome X par le lncRNA *Xist* et (iii) l'inactivation de loci par empreinte génomique induite en particulier par les lncRNAs *Airn* et *Kcnq1*. Ces études initiales ont établi le concept de lncRNAs fonctionnels agissant comme des régulateurs transcriptionnels impliqués dans divers processus biologiques fondamentaux (Vallot et al., 2013).

A l'ère génomique, au cours des deux dernières décennies, le séquençage systématique global des banques d'ADNc a abouti à une connaissance approfondie de la composante transcrite non codante du génome. Ces analyses transcriptomiques basées essentiellement sur la technologie *microarray*, ont identifié plusieurs dizaines de milliers de lncRNAs et les techniques d'immunoprécipitation de l'ARN et d'inactivation induite par des siRNAs ont conduit à la validation fonctionnelle d'un nombre encore très limité de lncRNAs (Guttman et al., 2009 ; Ulitsky et al., 2013).

A l'ère post-génomique, l'identification des lncRNAs et leur caractérisation fonctionnelle repose sur une approche expérimentale combinant (i) la découverte de nouveaux transcrits non codants par RNA-seq et ChIP-seq, (ii) l'annotation de l'étendue des régions transcrites et la quantité de produits transcrits par analyse bioinformatique, (iii) la quantification de l'expression des lncRNAs dans différents types et conditions cellulaires et tissulaires, (iv) les analyses de co-expression et co-régulation et (v) les tests fonctionnels de gain et de perte de fonction sur des lignées cellulaires et des xéno greffes utilisant les techniques siRNA, ASO, TALEN et CRISPR/Cas9. Ces différentes techniques systématiques ont permis de révéler l'existence de lncRNAs de manière non ambiguë et d'apprécier la complexité de leurs mécanismes d'action (Derrien et al., 2012 ; Spicuglia et al., 2013). L'avantage de la technique de RNA-seq réside dans sa sensibilité, permettant aux lncRNAs faiblement exprimés d'être

identifiés. Toutefois, en raison de la très faible expression habituelle des lncRNAs et de la fréquente complexité de leurs structures exon/intron, il peut être difficile d'identifier les différents transcrits produits par épissage alternatif à partir d'un gène de lncRNA. De plus, les techniques de *microarray* et de RNA-seq favorisent l'identification de lncRNAs polyA+. Il est alors nécessaire de combiner l'approche classique par RNA-seq avec d'autres techniques telles que les analyses épigénétiques (marqueurs de régions promotrices ou de gène entier), le séquençage ciblé de transcrits d'intérêt après capture sur une puce à ADN (*Capture-RNAseq*) et l'analyse permettant l'identification des extrémités 5' et 3' terminales des gènes (Spicuglia et al., 2013 ; Jiang et al., 2014 ; Park et al., 2014 ; Ye et al., 2014).

Au cours des 15 dernières années, des projets internationaux ont contribué à la caractérisation systématique des lncRNAs des mammifères. Grâce aux approches précédemment décrites, couplées à des analyses bio-informatiques de plus en plus performantes, il a été possible d'identifier un très grand nombre de lncRNAs exprimés dans divers tissus et lignées cellulaires. Le projet **FANTOM** (*Functional Annotation of Mammalian Genome*) est un projet international initié par le Japon en l'an 2000 visant à identifier et annoter l'ensemble des transcrits de souris (*FANTOM Consortium and the RIKEN PMI and CLST*, 2014). Les résultats publiés ont identifié 15.000 lncRNAs en 2002, 23.000 lncRNAs en 2005 et 34.000 lncRNAs en 2006 avec un pourcentage du génome transcrit évalué à 63%. Le projet **ENCODE** (*Encyclopedia Of DNA Elements*), lancé en 2003, a été conçu pour identifier, cartographier et rendre public tous les éléments fonctionnels du génome humain en intégrant différents niveaux d'analyse (transcription, distribution des marques chromatiniennes, facteurs de transcription). Des progrès décisifs concernant la lecture, l'intégration et l'interprétation de diverses bases de données ont rendu possible le déploiement du consortium **GENCODE**. Ainsi, l'annotation Gencode (version 24 de l'annotation officielle du projet ENCODE) maintient à jour la liste la plus exhaustive de lncRNAs exprimés chez l'homme et comporte ainsi 57.820 gènes, dont 16000 gènes de lncRNAs, 26414 lncRNAs transcrits et 20.345 gènes codant pour des protéines (Derrien et al., 2012a ; Harrow et al., 2012). L'annotation GENCODE M8 évalue à 9000 le nombre de lncRNAs murins.

Actuellement, les études consacrées à la biologie des lncRNAs génèrent un nombre considérable de publications et de bases de données telles que lncRNAdb, NONCODE, lncRNAator, lncRNADisease, ncFANs, InCeDB, LNCipedia, psiDR et MyTranscriptome (Derrien et al., 2012b ; Park et al., 2014 ; Das et al., 2014 ; Frankish et al., 2014). Le nombre

de lncRNAs fonctionnels est en constante augmentation. Les bases de données disponibles en 2015 du TCGA, de MyTranscriptome, de NONCODE et de LNCipedia ont identifié respectivement 58.648, 66.758, 92.343 et 107.000 gènes de lncRNAs (Iyer et al., 2015 ; Volders et al., 2015).

Au cours des prochaines années, il est très probable que la recherche concernant les lncRNAs s'intensifiera et s'élargira à tous les domaines de l'homéostasie cellulaire et que de nouveaux outils permettront de déterminer la fonctionnalité et les mécanismes d'action des dizaines de milliers de lncRNAs découverts (Iwakiri, 2016 ; Fang and Fullwood, 2016).

2.3.3 Caractéristiques des lncRNAs

2.3.3.1 Structures biochimiques des lncRNAs

La majorité des lncRNAs subissent un processus de maturation canonique

- **Les gènes des lncRNAs partagent de nombreuses caractéristiques biochimiques avec les gènes codant pour les protéines** : action prédominante de l'ARN Polymérase II, modifications co-transcriptionnelles (*cap* en 5', épissage alternatif des pré-lncRNAs au niveau de sites canoniques et polyadénylation partielle en 3'), épissage des introns, 2 ou plus de 2 exons, modifications épigénétiques (« domaine K4-K36 », H3K4me3 au niveau du promoteur, H3K36me3 dans le corps des gènes, H3K4me1 au niveau des *enhancers*), taille des exons et des introns, processus de biosynthèse et association avec des polysomes. La majorité des lncRNAs est située à moins de 10kb des gènes codant pour les protéines. De nombreux lncRNAs sont antisens de gènes codants ou sont introniques (Kapranov et al., 2007 ; Guttman et al., 2009 ; Mercer et al., 2009 ; Derrien et al., 2012a ; Niazi et al., 2012).
- **Toutefois, les gènes des lncRNAs présentent des différences notables avec les gènes codant pour les protéines** : Les lncRNAs sont moins stables, plus courts (longueur médiane de 592 nt comparée à la médiane des mRNAs qui est de 2453 nt) et possèdent moins d'exons (Derrien et al., 2012a). Ils ne présentent pas un cadre de lecture bien défini, hébergent de nombreuses séquences répétées (LINE, SINE), ont un niveau d'expression habituellement faible et une moindre association avec les ribosomes. Ils ne sont pas soumis à un export cytoplasmique comme les mRNAs et

leurs séquences sont dans l'ensemble peu conservées par rapport à celles des gènes codant pour les protéines (Pang et al., 2006 ; Louro et al., 2009 ; Mercer et al., 2009). Toutefois, il a été plus récemment démontré que des sous-classes de lncRNAs possèdent une grande stabilité (lncRNAs circulaires) et sont conservées entre les espèces (T-UCRs) (Clark and Mattick, 2011 ; Ulitsky and Bartel, 2013). Enfin, les lncRNAs doivent s'associer avec des *sets* spécifiques de protéines et former des complexes ribonucléoprotéiques pour exprimer leurs fonctions.

Certains lncRNAs subissent un processus de maturation non canonique

- Des lncRNAs transcrits par la RNA Polymérase II défont ce dogme tout en maintenant une stabilité assurant leurs fonctions. Ils peuvent manquer de queue poly A+ remplacée par une structure triple-hélicoidale (*MALAT1*, *MEN β*), présenter des terminaisons de type snoRNAs (sno-lncRNAs) ou être dépourvus de structures terminales 5' et 3' (lncRNAs circulaires).
- Deux groupes minoritaires de lncRNAs encore mal caractérisés sont représentés par des lncRNAs non-polyadénylés synthétisés par l'ARN polymérase III et par des lncRNAs obtenus par épissage alternatif.

2.3.3.2 Organisation structurale et fonctionnelle des lncRNAs

Le génome possède une architecture modulaire composée de régions transcriptionnelles complexes, caractérisées par des liens étroits entre les séquences nucléotidiques organisées en sens/antisens et des transcrits codants/non codants. Ainsi, plus de 50% des gènes codant pour des protéines sont associés à des transcrits complémentaires antisens non codants régulant *in cis* la chromatine et l'expression des gènes adjacents. En outre, l'épissage alternatif et l'initiation/terminaison transcriptionnelle utilisent cette architecture modulaire pour assurer une diversification transcriptionnelle aboutissant à un très grand nombre d'isoformes de lncRNAs (Yan et al., 2016).

L'organisation en domaines fonctionnels des lncRNAs semble provenir de 2 mécanismes distincts : (i) l'adoption d'une structure secondotertiaire médiant les interactions avec les partenaires protéiques garantissant la stabilité du génome et (ii) l'hybridation basée

sur la complémentarité des bases médiant les interactions avec les acides nucléiques (ARN et ADN). La plasticité et la flexibilité des lncRNAs permettent d'assurer des activités organisationnelles, catalytiques, régulatrices et structurales. Ces propriétés semblent mieux expliquer la grande diversité de leurs mécanismes d'action que la composition de leur séquence nucléotidique. L'organisation des lncRNAs en structures secondaires et tertiaires contribue à la création de domaines fonctionnels interagissant avec des protéines (petits ligands, complexes multiprotéiques) ou s'hybridant avec des acides nucléiques (mRNA, miRNA, DNA) (Reuter and Mathews, 2010 ; Wan et al., 2011 ; Bida and Maher, 2012). Il semble exister des similitudes entre les lncRNAs et les régions 3' non traduites (3' UTRs) des ARNs codant pour les protéines en ce qui concerne leurs structures et leurs séquences (Niazi and Valadkhan, 2012). Leurs interactions conduisent à leur tour à des changements conformationnels allostériques des lncRNAs leur permettant de se lier à d'autres acteurs moléculaires impliqués dans l'expression des gènes et la traduction des mRNAs (Brown et al., 2012). Les structures tertiaires des lncRNAs qui résultent d'interactions entre diverses structures secondaires jouent d'importants rôles régulateurs et dans la stabilité du génome (Novikova et al., 2012). Ainsi, les lncRNAs initient et orchestrent des réseaux de régulation très complexes aux niveaux épigénétique, transcriptionnel, post-transcriptionnel et traductionnel, afin de transmettre et de coordonner le flux d'information au travers des voies de signalisation nécessaires au fonctionnement des cellules eucaryotes (Gong and Maquat, 2011 ; Carrieri et al., 2012 ; Hung et al., 2013 ; Kung et al., 2013 ; Schmitz et al., 2016) (**Fig. 22**).

Cette organisation modulaire pourrait être liée à l'existence d'éléments transposables (TEs) qui constitueraient des domaines régulateurs dont les combinaisons détermineraient les fonctions des lncRNAs (Guttman and Rinn, 2012 ; Johnson and Guigo, 2014). Les TEs représentent des séquences répétitives et mobiles d'origine et de stade évolutif variés, constituant entre 50% et 66% du génome humain, et jouent un rôle majeur dans l'évolution génétique. Le processus par lequel les TEs concourent à l'apparition de nouvelles fonctions exercées par le génome hôte est l'insertion de séquences transposables préformées, appelée exaptation ou co-option (de Souza et al., 2013). Les TEs peuvent être classés en 2 catégories : les transposons et les rétrotransposons (LTR, LINE, SINE). Ils pourraient constituer des structures fonctionnelles ou des séquences conférant une activité biologique aux lncRNAs. Selon cette hypothèse, appelée « *RIDL hypothesis* » (*Repeat Insertion Domain of LncRNA* ; Johnson and Guigó, 2014), les TEs peuvent être subdivisés en TEs de type I, *via*

leurs structures et leurs interactions avec les protéines et en TEs de type II, *via* leur séquence et leurs interactions avec les acides nucléiques. Ces domaines pourraient être échangés entre lncRNAs et conduire à de nouvelles fonctions biologiques. Ainsi, les TEs représenteraient une source majeure et versatile de domaines fonctionnels favorisant l'apparition de nouveaux lncRNAs. Bien qu'initialement considérés comme non fonctionnels, les TEs semblent jouer un rôle déterminant dans de multiples processus biologiques, tels le développement embryonnaire précoce, le contrôle épigénétique des cellules souches et l'épissage alternatif. Ainsi, l'embryon est le siège d'une importante reprogrammation au cours de laquelle entrent en compétition les TEs libérés par la déméthylation globale et la machinerie épigénétique suppressive représentée par TRIM28, SETDB1, CoREST, PRC1, PRC2, HDACs et NuRD. Durant l'évolution, cette compétition a conduit à la co-option de TEs sous la forme de lncRNAs ou de domaines modulateurs des lncRNAs médiés par des TEs. Les lncRNAs comportant des TEs jouent un rôle majeur dans le maintien de la pluripotence. 127 lncRNAs contenant des éléments *HERVH* sont surexprimés dans les cellules pluripotentes. *LINC-ROR* est presque totalement constitué de TEs, *LINC01108* contient 2 TEs et *lncRNA-p21* comporte 7 TEs. Des TEs fonctionnels ont été identifiés dans certains lncRNAs (SINEB2 dans *Uchl1as*, Alu dans *ANRIL*, ERVB5 dans *Xist*, LINE1 dans *Fendrr*).

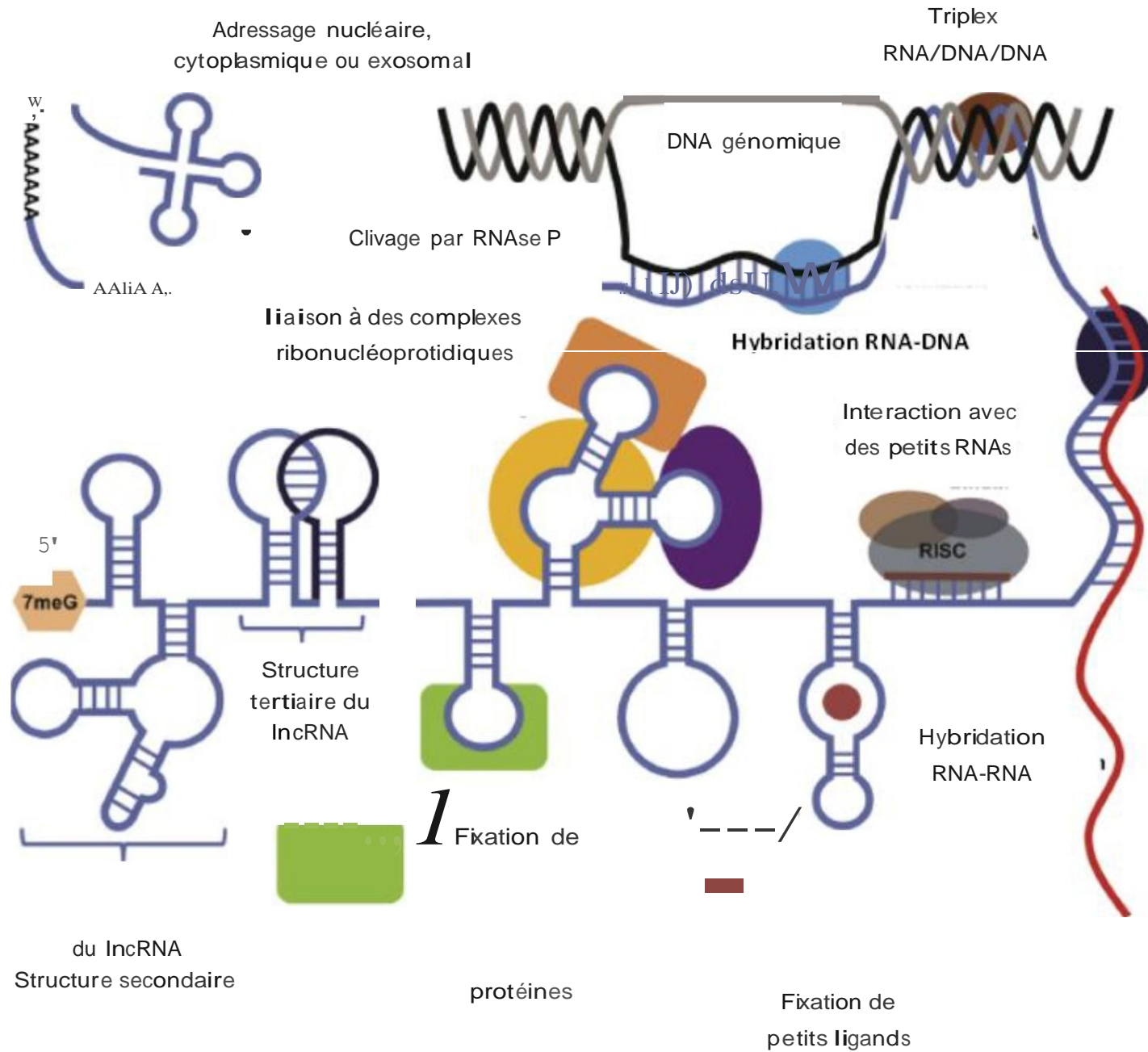


Figure 22 : Structure secondaire et tertiaire des lncRNAs et interactions des lncRNAs avec l'ADN, l'ARN et les protéines.
(Adapté d'après Li and Chen, 2013)

2.3.3.3 Organisation topographique des lncRNAs

Cette classification repose sur la découverte d'une organisation générale du génome humain en « forêts transcriptionnelles » (Consortium FANTOM). Les lncRNAs intragéniques peuvent être subdivisés en fonction de la façon dont ils chevauchent les gènes codant pour les protéines ou de leur orientation par rapport à ceux-ci. La situation des lncRNAs par rapport aux mRNAs des gènes annotés codant pour des protéines les plus proches a servi de base à la classification GENCODE en lncRNAs intergéniques et intragéniques (Tisseur et al., 2011 ; Derrien et al., 2012b ; Harrow et al., 2012).

Groupe I : lncRNAs situés à proximité des gènes codant pour les protéines

Les lncRNAs peuvent ainsi être divisés en 5 groupes topographiques principaux :

- **Les lncRNAs intergéniques (lincRNAs)** sont situés dans des régions génomiques où il n'existe aucun gène annoté sur une distance d'au moins 1kb (Orom et al., 2010). Les lincRNAs représentent la classe la mieux étudiée des lncRNAs (Khalil et al., 2009 ; Ulitsky and Bartel, 2013). Ils mesurent quelques centaines de pb à plusieurs kb. Ce sont des lncRNAs épissés et polyadénylés fonctionnels caractérisés par leur stabilité, un domaine transcriptionnel actif K4H36 et une expression cellulaire et tissulaire spécifiques. Les lincRNAs agissent à distance du lieu de synthèse au niveau de loci chromosomiques différents (mécanisme en *trans*) et sont principalement impliqués dans la régulation épigénétique de l'expression des gènes codant pour les protéines *via* leur interaction avec PRC2 ou d'autres complexes modifiant la chromatine (**Fig. 23, position 1**).
- **Les lncRNAs introniques (linRNAs)** peuvent être transcrits à partir des introns des gènes codant pour des protéines (Louro et al., 2009) (**Fig. 23, position 2**). Les linRNAs constituent la composante majeure du transcriptome des ncRNAs et représentent 70% de tous les ncRNAs nucléaires et 50% des ncRNAs cellulaires (St Laurent et al., 2012). Contrairement aux lincRNAs, des dizaines de milliers de linRNAs ont été identifiés, mais seulement quelques-uns ont été analysés sur le plan fonctionnel. Un sous-groupe de linRNAs a été récemment découvert, les lncRNAs

introniques circulaires (ciRNAs), qui s'accumulent dans les cellules et régulent l'expression de leurs gènes parentaux (Zhang et al., 2013).

- **Les lncRNAs bidirectionnels** ont leur séquence située sur le brin opposé au gène codant pour une protéine. De nombreux gènes transcrits en lncRNAs intergéniques possèdent un site d'initiation de la transcription proche de celui d'un gène codant, la transcription se faisant alors sur le brin opposé (transcription inverse). Ces gènes associés à ces transcrits divergents codent préférentiellement pour des régulateurs transcriptionnels impliqués dans le développement embryonnaire et la différenciation cellulaire (Lepoivre et al., 2013) (**Fig. 23, position 3**).
- **Les lncRNAs exoniques sens (sRNAs)** sont transcrits à partir du brin sens et sont chevauchants avec un ou plusieurs exons de gènes codant pour des protéines (Denoeud et al., 2007 ; Djebali et al., 2008). Les sRNAs peuvent avoir un certain chevauchement avec des gènes codants ou couvrir toute la séquence d'un intron. Cette catégorie inclut des sRNAs non épissés partiellement introniques (Nakaya et al., 2007) et des transcrits épissés qui combinent des exons provenant de régions codantes et non codantes du gène (Makrythanasis et al., 2009). Si la plupart des sRNAs n'ont pas de potentiel codant pour des protéines, plusieurs sRNAs fonctionnent à la fois comme des ncRNAs et comme des gènes codant pour des protéines (*SRA*, *ENOD40*) (**Fig. 23, position 4**).
- **Les lncRNAs exoniques antisens ou *natural antisens transcripts* (NATs)** sont exprimés en antisens des gènes codant pour les protéines (Kunej et al., 2014). Ils sont transcrits à partir du brin d'ADN opposé à celui du gène codant pour des protéines (**Fig. 23, position 5**). Les NATs peuvent aussi présenter un chevauchement avec les régions exoniques ou introniques ou couvrir toute la séquence d'un intron. Ils sont très fréquents dans le génome et représentent 32% des lncRNAs chez l'homme. 20% des gènes humains forment des paires sens-antisens et 72% des gènes ont une activité transcriptionnelle présente sur le brin opposé. Certains NATs sont des gènes codant pour des protéines alors que d'autres sont des lncRNAs. Les NATs présentent une organisation génomique divergente (quand le promoteur génère une transcription bidirectionnelle) ou convergente (quand la transcription débute à partir d'un autre promoteur ou à partir de l'extrémité 3' située sur l'autre brin). Ils sont observés dans

de nombreuses espèces animales, les plantes, les levures, les procaryotes et les virus mais, à la différence de certains lincRNAs qui sont hautement conservés, les NATs n'ont aucune séquence ou structure conservée qui pourrait être évocatrice d'une fonction particulière. Ils exercent des influences variées sur leur transcrits sens, activation ou répression. Les niveaux d'expression des paires sens-antisens sont généralement concordantes mais une expression antagoniste peut exister. Les NATs peuvent agir selon plusieurs mécanismes (plateforme, collision transcriptionnelle). *Wrap53* est un NAT régulant l'expression de *TP53* en ciblant la région 5' non traduite de l'ARNm de *TP53*. Cette interaction inhibitrice prévient l'induction de *TP53* après altération de l'ADN.

- **Certains lincRNAs chevauchent de petits ARNs**, tels des snoRNAs ou des miRNAs, avec lesquels ils entretiennent des liens fonctionnels, comme dans les régions soumises à l'empreinte génomique (Koerner et al., 2009 ; Abel et al., 2014).
- **Des lincRNAs plus récemment identifiés sont constitués seulement de séquences exoniques** pouvant provenir (i) de diverses parties de mRNAs ayant subi un processus de maturation post-transcriptionnelle (*Affymetrix ENCODE transcriptome Project*, 2009), tels les secteurs 3'UTR (Mercer et al., 2011) et 5'cap (Abdelhamid et al., 2014), et (ii) d'autres processus cellulaires, tels l'épissage inverse produisant des lincRNAs exoniques circulaires (Jeck et al., 2013), le trans-épissage conduisant à la synthèse de lincRNAs chimériques (Djebali et al., 2012), la copie de RNAs produisant des lincRNAs antisens « en miroir » (Kapranov et al., 2010) et l'épissage de transcrits pourvus d'une partie 3'UTR de grande taille à l'origine de lincRNAs participant au maintien de la configuration chromatiniennne interphasique (Caudron-Herger et al., 2011).

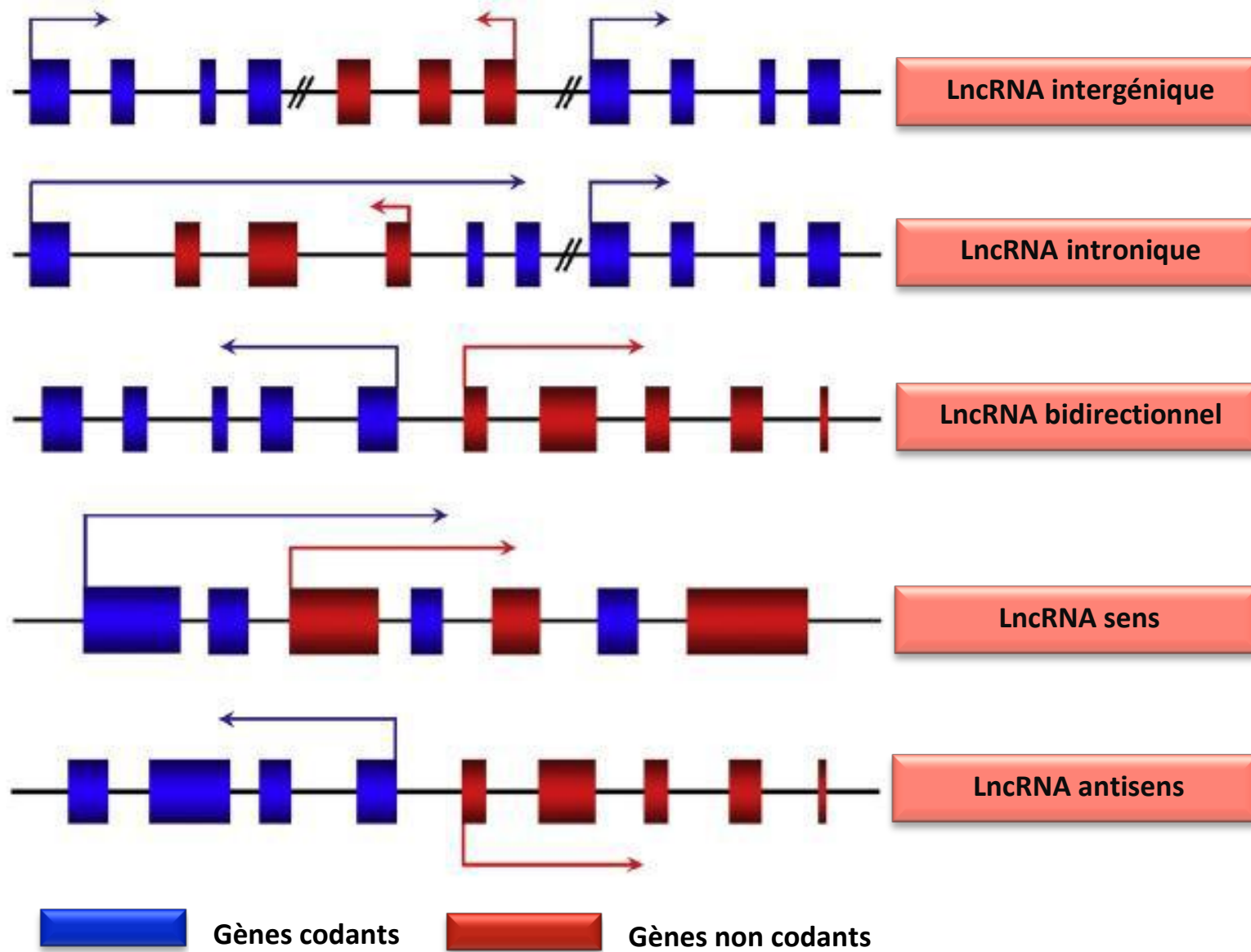


Figure 23 : Classification des lncRNAs selon leur topographie génomique (Adapté d'après Ma et al., 2013)

Groupe II : LncRNAs situés au niveau des régions régulatrices des gènes

- **Les *enhancer ncRNAs* (eRNAs)** sont transcrits à partir d'un *enhancer* (Bulger and Groudine, 2011 ; Xie and Ren, 2013). On distingue les eRNAs qui sont transcrits dans les 2 directions et les elncRNAs transcrits unidirectionnellement. Des analyses convergentes ont révélé que les lncRNAs peuvent réguler directement l'expression des gènes en contrôlant l'activité des *enhancers* des promoteurs et que le niveau d'expression des eRNAs est corrélé positivement avec le niveau d'expression des mRNAs des gènes voisins codant pour des protéines. De plus, des lncRNAs peuvent agir comme des *enhancers* et certains *enhancers* sont le lieu de synthèse de lncRNAs impliqués dans la régulation positive et négative des gènes, *via* la formation de boucles chromosomiques (Kim et al., 2010 ; Orom et al., 2010 ; Lai et al., 2013 ; Mousavi et al., 2013 ; Shibayama et al., 2014). En outre, des *super-enhancers* caractérisés par des groupes d'*enhancers* associés à des gènes impliqués dans le maintien de l'identité cellulaire, ont été récemment découverts (Lovén et al., 2013 ; Whyte et al., 2013). Ainsi, le lncRNA *CCAT-1* est transcrit à partir d'un *super-enhancer* situé en amont du locus *MYC* et agit comme un eRNA régulant la transcription de *MYC* (Xiang et al., 2014).
- **Les *promoter associated lncRNAs* (paRNAs)** sont transcrits à partir des domaines promoteurs situés en amont des gènes codant pour les protéines. La transcription divergente au niveau des régions promotrices de différents gènes donne lieu à la production de lncRNAs qui régulent la transcription des gènes codants voisins (Core et al., 2008 ; Seila et al., 2008 ; Wang et al., 2008 ; Taft et al., 2009). Ainsi, l'expression du lncRNA *PANDA*, issu du promoteur du gène *CDKN1A-p21*, est induite après une altération de l'ADN et favorise la prolifération cellulaire en inhibant le gène pro-apoptotique *CDKN1A-p21* (Wang et al., 2008 ; Hung et al., 2011). De même, le lncRNA *APTR* inhibe la transcription de *CDKN1A-p21* en recrutant et guidant le complexe polycomb PRC2 au niveau de son promoteur (Negishi et al., 2014). En outre, le lncRNA *Linc-p21* agit en inactivant les promoteurs de gènes réprimés par p53 *via* hnRNP-K (Huarte et al., 2010 ; Dimitrova et al., 2014).

- **Les lncRNAs associés à l'extrémité 3'-UTR (uaRNAs)** sont transcrits à partir de la région 3' non traduite des gènes codant pour les protéines.

Groupe III : LncRNAs situés au niveau de régions chromosomiques spécifiques

- **Les lncRNAs associés aux télomères**

Un groupe de lncRNAs, nommé *TERRA*, est transcrit des loci sub-téломériques et est impliqué dans la replication des chromosomes et dans l'homéostasie télomérique *via* la séquestration de la protéine hnRNPA1, la régulation directe de l'activité télomérasique et le maintien de la voie télomérique alternative (Azzalin et al., 2007; Redon et al., 2010). *TERRA* est considéré comme un régulateur négatif de la télomérase et semble impliqué dans la cancérogénèse lorsqu'il est sous-exprimé (Ng et al., 2009 ; Caslini, 2010 ; Cusanelli et al., 2013 ; Yu et al., 2014).

- **Les lncRNAs associés à des séquences répétitives**

50% du génome humain est constitué de séquences répétitives de diverses catégories, dont les séquences LINE, les séquences rétrovirales endogènes transcrites, les répétitions satellitaires, les promoteurs situés au sein de séquences répétitives et les pseudogènes transcrits (Belancio et al., 2009 ; Faulkner et al., 2009). De nombreuses régions génomiques transcrites en ncRNAs présentent des chevauchements avec ces séquences répétitives (Kelley et al., 2012 ; Kapusta et al., 2013). En particulier, les lncRNAs *PTENP1* et *KRASPI* sont transcrits à partir de pseudogènes. Les lncRNAs transcrits à partir de séquences répétitives sont préférentiellement impliqués dans le maintien de la pluripotence cellulaire (Johnsson et al., 2013).

- **Les lncRNAs associés à des séquences conservées**

De nombreux lncRNAs sont transcrits à partir de séquences conservées entre les espèces, telles les régions ultraconservées transcrites (T-UCR), les transcrits ultraconservés non codants induits par l'hypoxie (HINCUT), les transcrits non codants induits par le stress (LSINCT) et les *very long intergenic non-coding RNAs* (vlincRNAs). Les transcrits des régions ultraconservées ont une taille comprise entre 200 nt et 779 nt et comportent une expression tissulaire spécifique. De nombreux T-

UCRs présentent une altération de leur expression dans divers cancers (Scaruffi et al., 2009 ; Silva et al., 2010 ; Braconi et al., 2011 ; Ferdin et al., 2013).

- **Les lncRNAs associés à des « régions humaines accélérées »**

Plus de 200 « régions génomiques humaines accélérées », caractérisées par un taux accru de substitutions nucléotidiques entre le singe et l'homme, ont été récemment identifiées, dont 96% sont situées dans des segments génomiques non codants (Pollard et al., 2006). Parmi ces régions présentant un très faible degré de conservation entre les séquences des primates et celles des autres espèces, le lncRNA *HARI* est impliqué dans le développement du néocortex. Son expression spatio-temporelle associée à sa structure secondaire sont fortement suggestives d'une fonctionnalité spécifique (Beniaminov et al., 2008).

2.3.3.4 Taille des lncRNAs

Les lncRNAs possèdent une taille très variable, comprise pour la plupart d'entre eux entre 340 pb (*7SK*) et 91.5kb (*KCNQ1OT1*) (Wang et al., 2011a). La taille moyenne des lncRNAs est de 2.4 kb mais une fraction des lncRNAs est constituée de molécules de grande taille servant principalement de plateformes interactives, dont les lincRNAs, les macroRNAs et les vlincRNAs. Les analyses par RNA-seq de la fraction polyA+ ont révélé que les lincRNAs sont surtout de topographie intergénique et sont impliqués dans le contrôle transcriptionnel *via* leurs interactions préférentielles avec PRC2 et d'autres complexes modifiant la chromatine, tels *ANRIL*, *H19*, *HOTAIR*, *HOTTIP*, *lincRNA-p21*, *Xist* et *Paupar*.

Toutefois, des milliers de très longs lncRNAs, les macroRNAs et les vlincRNAs, dont la taille s'étend de 50 kb à plusieurs mégabases (Mb), ont été identifiés grâce à l'utilisation de la technologie RNA-seq à l'ensemble des RNAs (St Laurent et al., 2013 ; Hackermuller et al., 2014 ; Lazorthes et al., 2015).

- Les macroRNAs sont non épissés, de topographie nucléaire et leur longueur s'étend sur plusieurs dizaines de kb (*KCNQ1OT1* : 91.5kb). Ils sont surtout impliqués dans la régulation de l'empreinte génomique.

- Les vlincRNAs ont une longueur pouvant atteindre l'ordre de la Mb (*HELLP*, *vlinc_21*, *vlinc_185*, *vlinc_377*, *vlinc_500*). Ils constituent des plateformes géantes connectant différentes parties du génome. Le vlincRNA *Firre*, d'une longueur de 5 Mb et situé sur le chromosome X, se lie *via* des domaines répétés de 156 bp, aux ribonucléoprotéines hnRNP-U. *Firre* et hnRNP-U coordonnent l'organisation topographique de multiples chromosomes *via* une plateforme agissant en *trans* et établissent des interactions multichromosomiques.

Les vlincRNAs et les macroRNAs sont plus particulièrement impliqués dans le maintien de la pluripotence et l'empreinte génomique, l'apoptose (St Laurent et al., 2013), le cycle cellulaire (van Dijk et al., 2012), la sénescence (Lazorthes et al., 2015) et la cancérogénèse (Hackermuller et al., 2014).

2.3.3.5 Epissage alternatif

Les lncRNAs se présentent comme des isoformes linéaires et circulaires. Les isoformes les mieux caractérisés sont les isoformes linéaires qui apparaissent dans l'ensemble moins bien conservés que les mRNAs, globalement peu stables, faiblement exprimés et pourvus d'extrémités polarisées 5' et 3'. Les isoformes circulaires des lncRNAs (circRNAs) ont été redécouverts très récemment et font depuis lors l'objet de nombreuses études en raison de leurs particularités biologiques et de leurs rôles dans la cancérogénèse. Les circRNAs avaient été initialement observés dès la fin des années 70 dans les virus à ARN (Sanger et al., 1976), mais considérés par la suite comme des produits résultant d'erreurs d'épissage. Avec le développement des techniques de séquençage et de la bioinformatique, des centaines de nouveaux circRNAs ont été identifiés depuis 2012. Les principales bases de données concernant les circRNAs comprennent find_circ, Mapslice, CIRCexplorer et circRNAFinder. Les circRNAs sont des lncRNAs dont les extrémités sont dépourvues de polarité 5'3' ou de queue polyadénylée, mais jointes par des liaisons covalentes. Ils sont essentiellement dérivés d'exons de gènes codant pour les protéines *via* un mécanisme de *backsplicing*, bien qu'ils puissent aussi provenir d'introns, de régions intergéniques, de régions UTR et de brins antisens. De plus, de nombreux circRNAs peuvent être générés par le même locus (circularisation alternative). Il existe trois types de circRNAs : les circRNAs exoniques (EciRNAs), les circRNAs introniques (IciRNAs) et les circRNAs exoniques-introniques (EIciRNAs). Les circRNAs présentent des caractéristiques globalement opposées à celles de

leurs isoformes linéaires (Chen et al., 2016a). Ils sont abondants et présentent une très grande diversité avec pour certains lncRNAs, 10 fois plus de circRNAs que d'isoformes linéaires correspondants. Grace à leur structure fermée, les circRNAs sont plus stables que les lncRNAs linéaires en raison de l'absence de dégradation par les RNA exonucléases et la RNase R. Leur expression présente une spécificité cellulaire et tissulaire et est liée au stade de développement. Les circRNAs sont conservés entre les espèces. La majorité des circRNAs est d'origine endogène et des circRNAs artificiels comportant des IRES peuvent être traduits en oligopeptides. Ils jouent d'importants rôles dans la régulation transcriptionnelle et post-transcriptionnelle et fonctionnent comme des éponges à microRNAs, régulent l'épissage alternatif et modulent l'expression génique (Ebbesen et al., 2016). Les circRNAs composés d'exons sont localisés dans le cytoplasme et agissent comme des compétiteurs endogènes (ciRS-7/CDR1as et Sry). Les circRNAs composés d'introns ou d'exons et d'introns sont surtout situés dans le noyau et régulent l'expression de leurs gènes parentaux. Certains circRNAs régulent la transcription d'ARN linéaire en se liant à des protéines fixant l'ARN et en formant des complexes ribonucléoprotéiques qui interagissent avec les transcrits linéaires. Les circRNAs sont impliqués dans l'athérosclérose, les maladies neurodégénératives, les maladies à prions et le cancer (Salzman, 2016).

2.3.3.6 Niveaux d'expression et spécificité

L'expression des lncRNAs est faible, présente une spécificité spatio-temporelle et est dynamique (Gloss and Dinger, 2016). Leur faible niveau global d'expression peut néanmoins représenter une simplification abusive et traduire en fait des situations différentes, telles une expression à un moment particulier du développement embryonnaire, pendant un moment très court du cycle cellulaire ou induite par une sous population cellulaire spécifique. Cette expression est difficile à détecter avec les approches courantes de séquençage, justifiant le choix d'analyses plus performantes, telles les techniques de Capture-seq, HITS-CLIP et RIP-seq. Les faibles niveaux d'expression des lncRNAs semblent donc partiellement liés à leur spécificité et à la nature hautement dynamique de leur mode d'action. Ainsi, les analyses à partir des données du TCGA ont révélé dans une série de 5000 tumeurs comportant approximativement le même nombre de gènes codants et de lncRNAs dérégulés, que 60% des lncRNAs présentent une spécificité pour seulement un seul type de cancer (Schmitz et al., 2016). Leur spécificité temporo-spatiale (cellulaire, tissulaire, développementale) constitue la

caractéristique qui sépare actuellement le mieux les lncRNAs des mRNAs et le meilleur argument en faveur de leur fonctionnalité. Cette expression est spécifique mais aussi dynamique, dépendant du contexte spatio-temporel (*HOTTIP*). Elle rend compte de l'intérêt des lncRNAs comme biomarqueurs et comme cibles thérapeutiques.

2.3.3.7 Stabilité

Les lncRNAs sont considérés comme des molécules souvent instables en raison d'une faible expression globale et de l'existence de transcrits à demi-vie courte. Cependant, une étude récente a montré que seule une minorité de transcrits est instable (29%) avec une demi-vie inférieure à 2 heures, alors que 6% des lncRNAs analysés sont très stables avec une demi vie supérieure à 12 heures (Wilusz, 2016).

2.3.3.8 Localisation tissulaire et cellulaire

Les lncRNAs sont localisés dans de nombreux tissus mais une majorité de lncRNAs est située dans le cerveau, où leur diversité est la plus importante. Au niveau de la cellule, le consortium ENCODE a analysé le profil transcriptionnel en lncRNAs des différents compartiments sub-cellulaires et des divers sous-compartiments intra-nucléaires, tels que la chromatine, le nucléole et la matrice nucléaire (Djebali et al., 2012). La localisation cellulaire des lncRNAs représente une importante clé dans la détermination de leurs fonctions biologiques et de l'efficacité de nouvelles thérapies ciblées basées sur les RNAs (Kapranov et al., 2007).

Les fonctions régulatrices et structurales des lncRNAs sont associées à leurs localisations sub-cellulaires. Les lncRNAs matures peuvent s'accumuler au niveau de leur site de transcription, dans le nucléoplasme ou être exportés dans le cytoplasme pour exécuter leurs fonctions. Une large majorité de lncRNAs sont localisés dans le noyau, suggérant leur rôle primordial dans la régulation spatio-temporelle de l'architecture nucléaire et de l'expression des gènes codant pour des protéines (Mondal et al., 2010 ; Derrien et al., 2012a). Divers types de lncRNAs intranucléaires peuvent être observés, tels les lncRNAs associés à la chromatine et interconnectant la chromatine, les lncRNAs associés aux corpuscules intranucléaires, les lncRNAs interagissant avec les complexes multiprotéiques modifiant la chromatine et les lncRNAs associés à la matrice nucléaire (Bergmann and Spector, 2014). Ainsi, les lncRNAs

favoriseraient la transmission du flux d'informations génétiques en contrôlant et en liant la dynamique de l'architecture nucléaire, la plasticité de la signalisation chromatinienne et la régulation de l'expression des gènes. Des lncRNAs sont situés dans le cytosol et sont associés aux mitochondries (*SLC7A2-IT1A/B*) et aux ribosomes (Rackham et al., 2011 ; Demongeot et al., 2013 ; van Heesch et al., 2014). Une minorité de lncRNAs sont impliqués dans le trafic nucléocytoplasmique (*NRON*), la stabilité des mRNAs et la régulation transcriptionnelle tandis que d'autres lncRNAs sont transportés dans les milieux extracellulaires *via* des exosomes.

Une classification liant la localisation cellulaire des lncRNAs à leurs fonctions a été récemment proposée (Chen, 2016b).

- Le premier groupe de lncRNAs intranucléaires est représenté par des lncRNAs qui s'accumulent au niveau de leur site de transcription et qui peuvent agir en *cis* ou en *trans*. Les lncRNAs qui agissent en *cis* modulent localement l'expression des gènes en recrutant des facteurs de transcription (YY1), des organisateurs (*CCAT1* et CTCF) ou des complexes de remodelage chromatinien (*Air* et G9a, *KCNQ1OT1* et G9a/PRC2) et en formant des triplex RNA-DNA (*Khps1* et le promoteur de *SPHK1*). Les lncRNAs qui agissent en *trans* peuvent influencer l'expression génique en établissant des contacts trans-chromosomiques (*Firre*) ou en séquestrant des facteurs d'épissage, tels Fox2 (sno-lncRNAs).
- Le deuxième groupe de lncRNAs intranucléaires est constitué de lncRNAs relocalisés à distance de leur site de transcription à l'ensemble du nucléoplasme ou dans des corps intranucléaires. Les lncRNAs nucléoplasmiques agissent globalement (*HOTAIR*, *BCAR4* et *PVT1*) ou au sein de corpuscules spécifiques (*MIAT/Gomafu*, *NEAT1*, *MALAT1*).
- Le troisième groupe est représenté par les lncRNAs exportés dans le cytoplasme qui régulent l'expression des gènes en interagissant avec des modifications post-traductionnelles (*NKILA* et NF-kB), des miRNAs (*lincMD1* et miR-133), des protéines (*NORAD* et les protéines PUMILIO) et la machinerie traductionnelles (*lincRNA-p21*).

De multiples facteurs, tels des domaines fonctionnels (les éléments ribonucléiques de rétention), des protéines nucléaires, des signaux de rétention lors de la maturation des

lncRNAs et des conformations ribonucléiques (structures secondaires et tertiaires) et chromosomiques, assurent la spécificité topographique cellulaire des lncRNAs.

2.3.3.9 Régulation des lncRNAs

La régulation des lncRNAs est actuellement encore largement méconnue. Elle semble toutefois similaire à celle contrôlant les gènes codant pour les protéines et les miRNAs. La régulation épigénétique semble jouer un rôle de moindre importance dans l'expression de nombreux lncRNAs, car les marques de méthylation de l'ADN et des histones diffèrent considérablement entre les mRNAs et les lncRNAs. Cependant, la régulation épigénétique joue un rôle important dans l'inactivation des T-UCRs. Des facteurs de transcription, tels p53, B-MYB et c-Myc régulent de nombreux lncRNAs au niveau transcriptionnel. Les miRNAs qui diminuent la stabilité des lncRNAs (*HOTTIP*, *linc-p21*, *MALAT1*) jouent un rôle déterminant dans la régulation post-transcriptionnelle des lncRNAs.

2.3.3.10 Conservation des lncRNAs

Une large majorité de lncRNAs présentent un degré de conservation plus faible en terme de séquences que les gènes codant pour les protéines, mais plus élevé que les régions évolutives neutres du génome et que les introns des gènes codant pour les protéines, suggérant un processus d'évolution rapide soumis à des contraintes modérées (Wang and Leung, 2004 ; Pang et al., 2006). Bien que les lncRNAs présentent une faible conservation globale de leur séquence avec une prévalence forte chez les primates, 3% des lncRNAs sont apparus il y a 300 millions d'années et sont présents chez le xénope, le poulet et l'homme. De plus, des sous-groupes de lncRNAs très conservés ont été plus récemment identifiés, tels les T-UCRs, les lncRNAs circulaires et les lincRNAs. Cette moindre conservation globale peut s'expliquer par les fonctions spécifiques des lncRNAs reposant davantage sur leurs structures tridimensionnelles que sur leurs séquences, à l'inverse des RNAs codant pour les protéines (Carninci et al., 2005 ; Ponjavic et al., 2007). Les lncRNAs ont été fréquemment identifiés dans différentes espèces au sein de régions génomiques synténiques (Schmitz et al., 2016) et cette conservation est liée à l'édification d'une structure tridimensionnelle. La conservation de leur structure secondaires et tertiaires, de leurs régions régulatrices et de leurs sites d'épissage semble jouer ainsi un rôle déterminant dans le maintien de leurs fonctions biologiques (*Xist*,

TUNA/MEGAMIND). La conservation de leurs structures au cours de l'évolution pourrait de plus être liée à une sur-représentation de séquences transposables (TEs) dans les lncRNAs par rapport aux ARNs codants. Ces transposons maintiennent en effet la stabilité et la complexité des structures secondaires des lncRNAs. Le faible degré de conservation des lncRNAs pourrait être ainsi le reflêt d'une apparition récente et d'une évolution extrêmement rapide.

2.3.3.11 Evolution des lncRNAs

Le séquençage de l'ensemble du génome de différents organismes a révélé que le nombre de leurs gènes codant pour des protéines n'est pas corrélé à leur degré de complexité évolutionnelle. Cette dernière peut être attribuée à l'épissage alternatif qui majore la complexité protéique, mais surtout à la présence des ncRNAs qui représentent plus de 95% du génome. Les lncRNAs sont observés chez de nombreux eucaryotes et la proportion des lncRNAs par rapport aux RNAs codants chez l'homme est la plus élevée (Taft et al., 2007). A l'inverse des séquences codant pour des protéines, les lncRNAs ont rapidement évolué, expliquant que les orthologues des lncRNAs chez les mammifères ne sont observés que chez les vertébrés (Khaitovitch et al., 2006 ; Ulitsky et al., 2011). Même chez ces derniers, la conservation des séquences d'ADN est faible et suggère que la plupart des lncRNAs ne sont pas fonctionnels en raison d'une pression de sélection insuffisante. Néanmoins, d'autres critères que la séquence, pourraient être pris en compte dans l'évaluation de leur conservation, tels la localisation génomique, le profil transcriptionnel et la structure secondotertiaire. La complexité actuelle de la physiologie humaine ne peut en effet être expliquée par l'expression de moins de 20.000 gènes codant pour des protéines, nombre comparable à celui observé chez *Drosophila melanogaster* et *Caenorhabditis elegans*, mais bien plutôt par le développement parallèle d'un génome non codant (Hezroni et al., 2015).

Contrairement au protéome, la quantité de ncRNAs produite a fortement progressé au cours de l'évolution, et il semble y avoir une corrélation significative entre les niveaux d'expression des lncRNAs et le niveau de complexité biologique, expliquant que la plupart des primates expriment la très grande majorité des lncRNAs. En effet, le nombre de gènes codant pour des protéines ne peut pas expliquer le fonctionnement d'organes très finement régulés tels que le cerveau. Ce niveau supplémentaire de complexité observé chez les vertébrés serait en partie lié à l'expression de lncRNAs pourvus d'une haute spécificité spatio-temporelle et à leurs interactions avec l'ADN, l'ARN messager et les protéines. La transcription des lncRNAs

semble en effet plus particulièrement associée à des phénotypes cellulaires et tissulaires complexes (cerveau, testicule) et plastiques (système immunitaire, pluripotence, développement embryonnaire). Les lncRNAs pourraient alors être considérés comme des senseurs moléculaires des changements environnementaux conférant une plasticité évolutionnelle qui contribue au développement de la complexité du vivant. En réaction à ces modifications environnementales, des lncRNAs situés dans les séquences intergéniques (lincRNAs) serviraient de supports pour de nouvelles fonctions permettant à l'organisme de s'adapter à de nouvelles contraintes (Diederichs, 2014).

Les lncRNAs apparaissent ainsi comme des régulateurs plus performants que les protéines au cours des changements évolutifs majeurs survenus chez les mammifères. Les lncRNAs sont immédiatement fonctionnels lors de la transcription et n'ont pas à être traduits et transportés dans le cytoplasme. Ils agissent directement au niveau de leur site de transcription (*cis*) ou à distance (*trans*). Ils sont intrinsèquement versatiles dans leurs interactions moléculaires avec les protéines, l'ARN et l'ADN *via* leurs séquences et leurs structures. Les lncRNAs sont des molécules très plastiques et malléables sur le plan évolutif, car leur séquence peut tolérer des insertions et des délétions, à la différence des protéines dont les mutations entraînent généralement une perte de fonction.

2.3.3.12 Origine des lncRNAs

L'émergence récente de nombreux lncRNAs suggère qu'ils continuent à être activement synthétisés. Bien que l'origine des lncRNAs soit encore largement méconnue, plusieurs mécanismes aboutissant à la formation de lncRNAs ont été proposés :

- Des gènes codant pour des protéines pourraient générer des lncRNAs et à l'inverse des protéines seraient synthétisées à partir de lncRNAs. Des données récentes ont en effet identifié 24 gènes codant pour des protéines humaines dont les gènes homologues dans d'autres espèces sont des gènes non codants (Nam and Bartel, 2012).
- La duplication des gènes et des chromosomes est un important mécanisme d'évolution biologique. Elle peut être la résultante d'un *crossing over* inégal, d'une retroposition ou d'une duplication d'un chromosome ou d'un génome entier (Magadum et al., 2013). Les gènes dupliqués peuvent alors devenir inactifs (pseudogènes), coder pour

des protéines pourvues d'autres fonctions biologiques ou être transcrits en lncRNAs (*Xist*) (Duret et al., 2006).

- L'insertion d'éléments génétiques mobiles contribue largement à l'évolution des génomes eucaryotes (Hutchins and Pei, 2015). Parmi ces éléments, les rétrotransposons et les transposons constituent plus de 50% du génome des mammifères et peuvent fournir des exons, des sites d'épissage, des *enhancers*, des promoteurs et des domaines de lncRNAs se liant à l'ADN, à l'ARN et aux protéines (Sytnikova et al., 2014). Parmi les lncRNAs, les éléments transposables (rétrotransposons et transposons) représentent plus de 40% des nucléotides des lincRNAs (Kelley and Rinn, 2012 ; Johnson and Guigo, 2014). Ainsi, le nombre d'exons du gène d'*ANRIL* s'est accru par insertion de transposons et d'exons *de novo* (He et al., 2013). Des séquences promotrices issues de *HERVH* ont été insérées dans des régions génomiques inactives (Kelley and Rinn, 2012).
- Des régions particulières du génome pourvues d'un degré de conservation très variable peuvent générer des lncRNAs, telles les séquences répétitives, les séquences conservées et les régions humaines accélérées.
- La synthèse de lncRNAs surviendrait enfin *de novo* à partir de l'ADN intergénomique.

2.3.3.13 Classification fonctionnelle

Les analyses réalisées *in vitro* et *in vivo* sur un nombre jusqu'à présent très réduit de lncRNAs ont néanmoins permis d'identifier certaines de leurs caractéristiques générales, telles une plus grande spécificité spatio-temporelle, une action rapide et une demi-vie courte, de nombreux isoformes par épissage alternatif et un rôle fondamental dans l'organisation nucléaire, la régulation de l'expression des gènes et l'homéostasie. Les lncRNAs sont préférentiellement exprimés par des organisations biologiques complexes et plastiques. Si certains lncRNAs exercent des fonctions cellulaires de base ou ont un phénotype visible et sont indispensables à la vie, la plupart ne semblent jouer qu'un rôle modulateur dans certaines conditions spécifiques, expliquant ainsi en partie leur faible niveau global d'expression. Cette expression spécifique et dynamique dépend du contexte tissulaire

et environnemental. Il existe une faible corrélation entre le niveau d'expression des lncRNAs et leur importance fonctionnelle. Les lncRNAs jouent un rôle tantôt régulateur tantôt modulateur de certaines sous populations cellulaires, justifiant la poursuite de leurs investigations par des techniques *one-cell*. La pénétrance des gènes transcrits en lncRNAs est souvent incomplète suggérant que les fonctions de certains lncRNAs ne sont pas normalement requises mais deviennent essentielles dans des conditions critiques.

Le débat sur la fonctionnalité des lncRNAs demeure toujours d'actualité. Les lncRNAs sont-ils la conséquence marginale de processus transcriptionnels régulateurs ou possèdent-ils des fonctions spécifiques ?

Les arguments en faveur de leur fonctionnalité sont de plus en plus nombreux: (i) leur profil d'expression est souvent lié à certains stades de développement et survient dans des sous-populations cellulaires spécifiques, (ii) leur topographie génomique est préférentiellement située à proximité des gènes codant pour les protéines, (iii) leur localisation cellulaire est observée dans des compartiments spécifiques, (iv) les lncRNAs sont impliqués dans les grandes fonctions biologiques et l'homéostasie et (v) la dérégulation de l'expression des lncRNAs est observée dans diverses pathologies.

Cependant, la majorité des travaux de recherche ont été conduits *in vitro* et la plupart de leurs fonctions sont inconnues puisque seulement 0.1% les lncRNAs ont été analysés sur le plan fonctionnel.

De plus, certains secteurs génomiques non codants et actuellement inactifs pourraient servir de « réservoirs évolutifs » et acquérir une fonctionnalité ultérieure par co-option d'éléments transposables.

Toutefois, une classification des lncRNAs basée sur leurs fonctions moléculaires peut être actuellement ébauchée et pourrait comprendre les lncRNAs architecturaux, les lncRNAs régulateurs épigénétiques, les lncRNAs antisens naturels, les lncRNAs activateurs, les lncRNAs compétiteurs endogènes linéaires et circulaires, les lncRNAs inductibles, les lncRNAs liés à la résistance thérapeutique et les lncRNAs précurseurs de petits ARNs.

La future classification fonctionnelle des lncRNAs devra intégrer l'ensemble des données issues de leur caractérisation qui est actuellement très parcellaire : *RNA identity elements*, organisation génomique, niveau d'expression, localisation subcellulaire, *processing*, isoformes, structure biochimique, réseaux d'interactions, mécanismes d'action, fonctions moléculaires et cellulaires, mécanismes de dérégulation dans les pathologies.

2.3.4 Mécanismes d'action des lncRNAs

Des études convergentes impliquent les lncRNAs dans la coordination de nombreux réseaux génétiques et la régulation des diverses voies de signalisation durant le développement embryonnaire et le maintien de l'homéostasie à l'âge adulte (Nagano et al., 2011 ; Wang et al., 2011b ; Geisler et al., 2013).

2.3.4.1 Régulation de l'expression des gènes

Afin de moduler l'expression des gènes, les lncRNAs peuvent interagir localement (en *cis*) sur les gènes voisins (**Fig. 24A**) ou à distance (en *trans*), lorsque leurs mécanismes d'action s'effectuent indépendamment de la localisation des gènes cibles (**Fig. 24B**).

Les lncRNAs interagissent de façon très variable avec des complexes modifiant la chromatine, divers sous types de RNAs et des protéines (**Fig. 25**).

Les *cis*-lncRNAs agissent quand le lncRNA régulateur et le gène cible sont tous deux transcrits à partir du même locus. Cependant, certains *cis*-lncRNAs peuvent agir à distance *via* des boucles d'ADN (*HOTTIP* avec WDR5/MLL, *PANDA* avec PRC2/NFYA, *ANRIL* avec PRC2/PRC1).

Les *trans*-lncRNAs interagissent à distance avec des régions spécifiques de la chromatine sur des chromosomes différents et modulent l'expression génique en ciblant des protéines et des ncRNAs *via* des triplex RNA/DNA.

L'architecture intrinsèque des lncRNAs est le facteur clé déterminant leurs interactions spécifiques (*PRNCR1/PCGEM1*, *HOTAIR*).

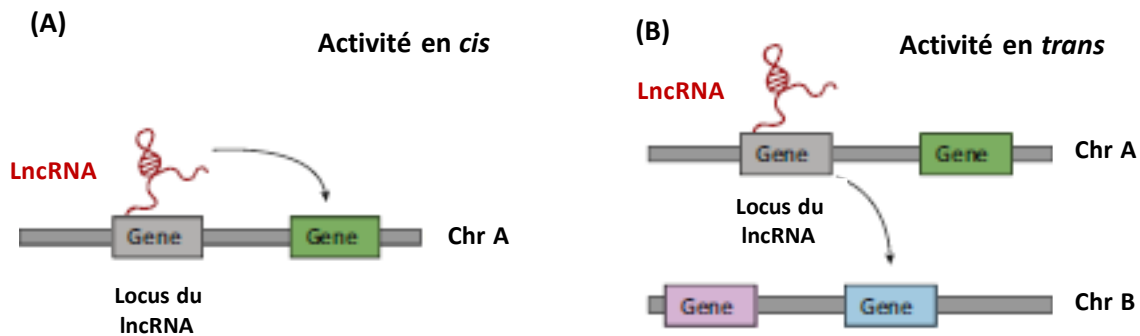


Figure 24 : Mécanismes d'action en *cis* (A) et en *trans* (B) des lncRNAs

A la différence des petits ncRNAs, leurs mécanismes d'action sont multiples (Fig. 25):

- **LncRNA « Guide »** : Ils recrutent les molécules des complexes modifiant la chromatine et les acheminent au niveau de loci spécifiques (*HOTTIP*, *Xist*) (Lee et al., 2009).
- **LncRNA « Leurre »** : Ils interagissent avec des protéines ou des ARNs (Kino et al., 2010). Dans le noyau, ils peuvent se lier à des facteurs de transcription ou à l'ADN méthyltransférase1 (*PANDA*, *GAS5*, *MALATI*), alors que dans le cytoplasme, ils peuvent fonctionner comme des « éponges moléculaires » pour attirer des complexes protéiques et des miRNAs (*PTENP1*, *HULC*) (Bak et al., 2014).
- **LncRNA « Signal »** : Ils régulent l'activité transcriptionnelle des gènes codant pour les protéines (*lincRNA-p21*) (Tang et al., 2015).
- **LncRNA « Plateforme ribonucléoprotidique »** : Ils constituent des adaptateurs structuraux qui se lient à des complexes moléculaires et régulent l'expression des gènes (*HOTAIR*, *ANRIL*, *TERC/TERT*) (Tsai et al., 2010 ; Kotake et al., 2011).

D'autres mécanismes d'action des lncRNAs ont été plus récemment identifiés :

- **Les lncRNAs précurseurs de petits ncRNAs**, tels les t-RNA-like, les snoRNAs, les piRNAs et les miRNAs (Wilusz et al., 2008 ; Valen et al., 2011 ; Pircher et al., 2014).
- **Les lncRNAs activateurs (ncRNA-a)** se distinguent des eRNAs par la régulation positive de gènes situés à leur proximité immédiate. Ainsi, le lncRNA *ncRNA-a7* active le facteur de transcription SNAI1, un des régulateurs majeurs de l'EMT (Orom et al., 2010).
- **Les lncRNAs compétiteurs endogènes (ceRNA)** partagent des séquences similaires avec des transcrits non codants ou codant pour des protéines (Tay et al., 2014a). Ils assurent la régulation de ces transcrits par un mécanisme de compétition. Les pseudogènes représentent une importante fraction des ceRNAs (*PTENP1* et *KRASPI*) qui constitue une matrice régulatrice intracytoplasmique complexe agissant au travers d'un gradient d'affinité entre différentes classes d'ARNs (Poliseno et al., 2010).
- **Les LncRNAs traduits** : Des analyses protéomiques ont récemment identifié des sous-types de lncRNAs pouvant être traduits en petites protéines (Cohen, 2014) ou en micropeptides fonctionnels (Anderson et al., 2015 ; Laouressergues et al., 2015 ; Nelson et al., 2016).
- **Les lncRNAs induits** : L'expression de lncRNAs peut être induite par les facteurs de croissance (Sas-Chen et al., 2016) ou par le stress cellulaire (de Andres-Pablo et al., 2016).

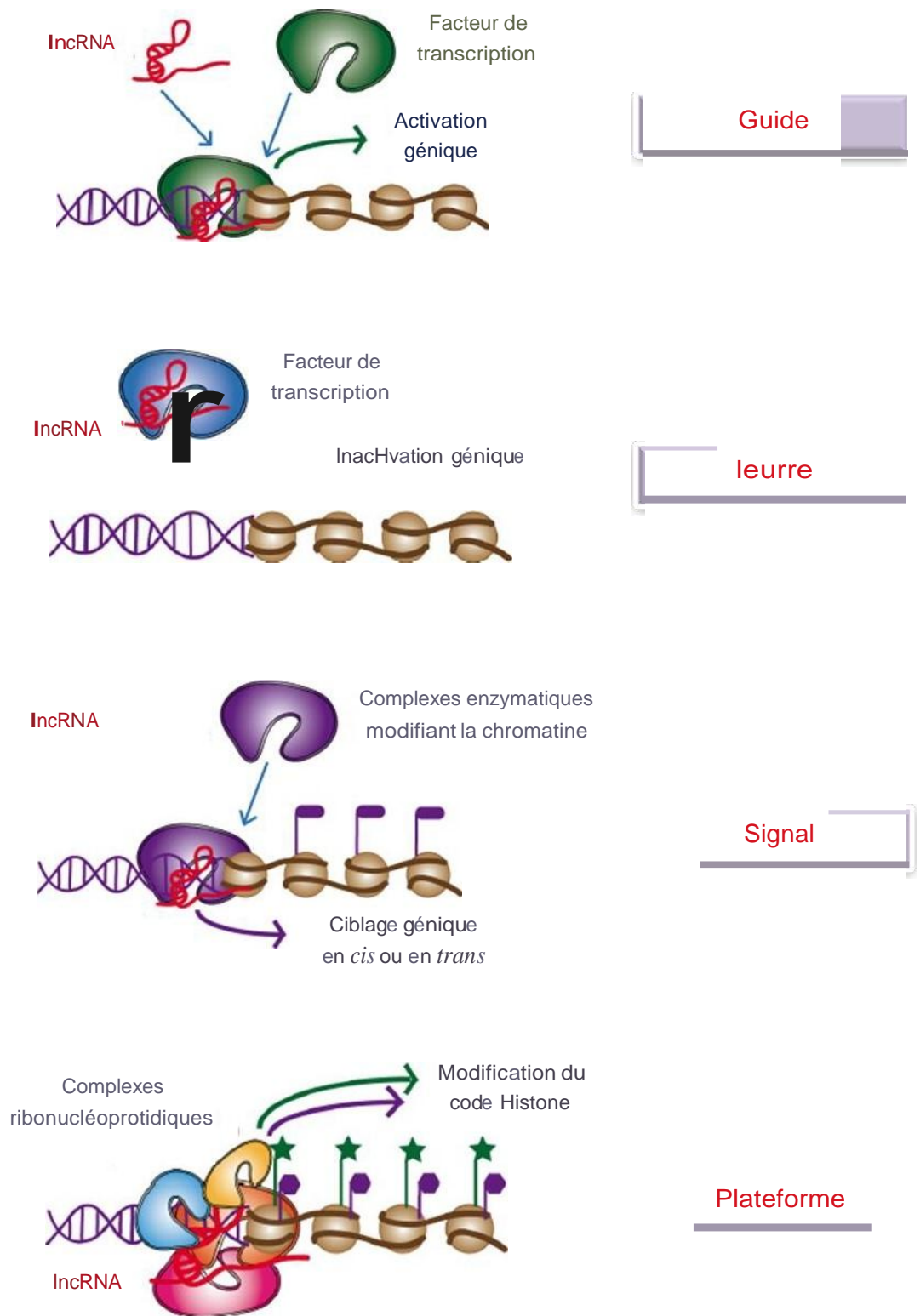


Figure 25 : Mécanismes d'action moléculaires des lncRNAs (guide, leurre, signal, plateforme) (D'après da Sacco et al., 2012)

2.3.4.2 Reconnaissance de sites et de séquences fonctionnelles génomiques

- L'ARN polymérase II peut lier un lncRNA au niveau de son site de transcription à partir duquel le lncRNA pourra agir sur les gènes de voisinage (*NEATI*).
- Les lncRNAs peuvent interagir avec l'ADN par hybridation directe (*TERRA*) ou en créant des triplex ARN-ADN (lncRNAs associés au promoteur de l'ADN ribosomal).
- Les lncRNAs peuvent interagir physiquement avec les protéines se liant à l'ADN (*Xist*, *Firre*).

2.3.4.3 Edification de domaines intranucléaires

Les lncRNAs peuvent créer des compartiments dynamiques intra-nucléaires rapprochant ou repositionnant des gènes (Bergmann and Spector, 2014 ; Quinodoz and Guttman, 2014). Cette stratégie, qui favorise la proximité spatiale, pourrait en partie expliquer le faible niveau d'expression des lncRNAs (*NEATI*, *Xist*) par rapport aux gènes codant pour les protéines.

2.3.5 Fonctions moléculaires des lncRNAs

2.3.5.1 Généralités

Contrairement aux protéines qui possèdent souvent des domaines fonctionnels bien définis, il est actuellement impossible de prédire les fonctions des lncRNAs à partir de leur seule séquence nucléotidique (Yang et al., 2014b).

Les fonctions moléculaires et cellulaires des lncRNAs demeurent donc encore largement méconnues pour la plupart d'entre eux. Seul un nombre très limité de lncRNAs ont été bien caractérisés chez l'homme, tels *Xist*, *KCNQ10T1*, *AIR*, *HOTAIR*, *ANRIL*, *HOTTIP*, *MALAT1*, *TERRA* et *HULC*.

Des études de gain et de perte de fonction ont révélé que les lncRNAs présentent la plus grande diversité fonctionnelle parmi les ncRNAs et que leur fréquente spécificité tissulaire suggère leur implication dans des processus biologiques majeurs (Gibb et al., 2011 ; Clark et al., 2013 ; Kung et al., 2013 ; Clark and Blackshaw, 2014) (**Tableau 3**).

Les lncRNAs sont surtout des régulateurs transcriptionnels et post-transcriptionnels de l'expression des gènes. L'utilisation de modèles d'inactivation (*KO*) a préférentiellement révélé des anomalies légères du développement, suggérant une modulation fine dans des processus biologiques complexes, dynamiques et plastiques (di Gesualdo et al., 2014).

Les fonctions structurales des lncRNAs comportent l'édification de corps intranucléaires et le maintien de la stabilité et de l'intégrité du génome et des chromosomes.

Les différents réseaux d'effecteurs moléculaires, tels les lncRNAs, les miRNAs, les facteurs de transcription, les mRNAs et les protéines modifiant la chromatine, sont indispensables à la régulation temporo-spatiale de l'expression des gènes et sont des composants structuraux nucléaires clés en formant des complexes avec les différents acides nucléiques sous la forme de duplex RNA/RNA et de triplex RNA/DNA double brin.

2.3.5.2 Fonctions régulatrices des lncRNAs

Les lncRNAs sont actuellement considérés comme des coordonnateurs et des régulateurs transcriptionnels et post-transcriptionnels contrôlant l'expression des gènes codant pour les protéines.

Ils agissent au niveau épigénétique, de l'ADN, de l'ARN et des protéines, par des mécanismes variés dépendant de leur localisation cellulaire et conduisant à des modifications de la chromatine, de la transcription et de la traduction des protéines (**Fig. 26**).

Fonctions moléculaires des lncRNAs

Réarrangement de la chromatine
Modification de la méthylation de l'ADN et du code histone
Epissage alternatif des mRNAs
Boucles chromatinienne
Co-régulation transcriptionnelle
Activation ou répression transcriptionnelle
Trafficking de facteurs de transcription
Protection des miRNAs Compétition endogène
Précurseurs de ncRNAs (miRNAs, piRNAs, snoRNAs)
Traduction

Fonctions cellulaires des lncRNAs

Expression des gènes codants
Pluripotence
Cycle cellulaire
Sénescence et apoptose
Réparation de l'ADN
Métabolisme
Sécrétion des protéines
Dégradation protéosomale
Différenciation cellulaire
Angiogénèse
Sécrétion d'exosomes
Edification de structures nucléaires
Régulation du système immunitaire

Tableau 3 : Fonctions moléculaires et cellulaires des lncRNAs

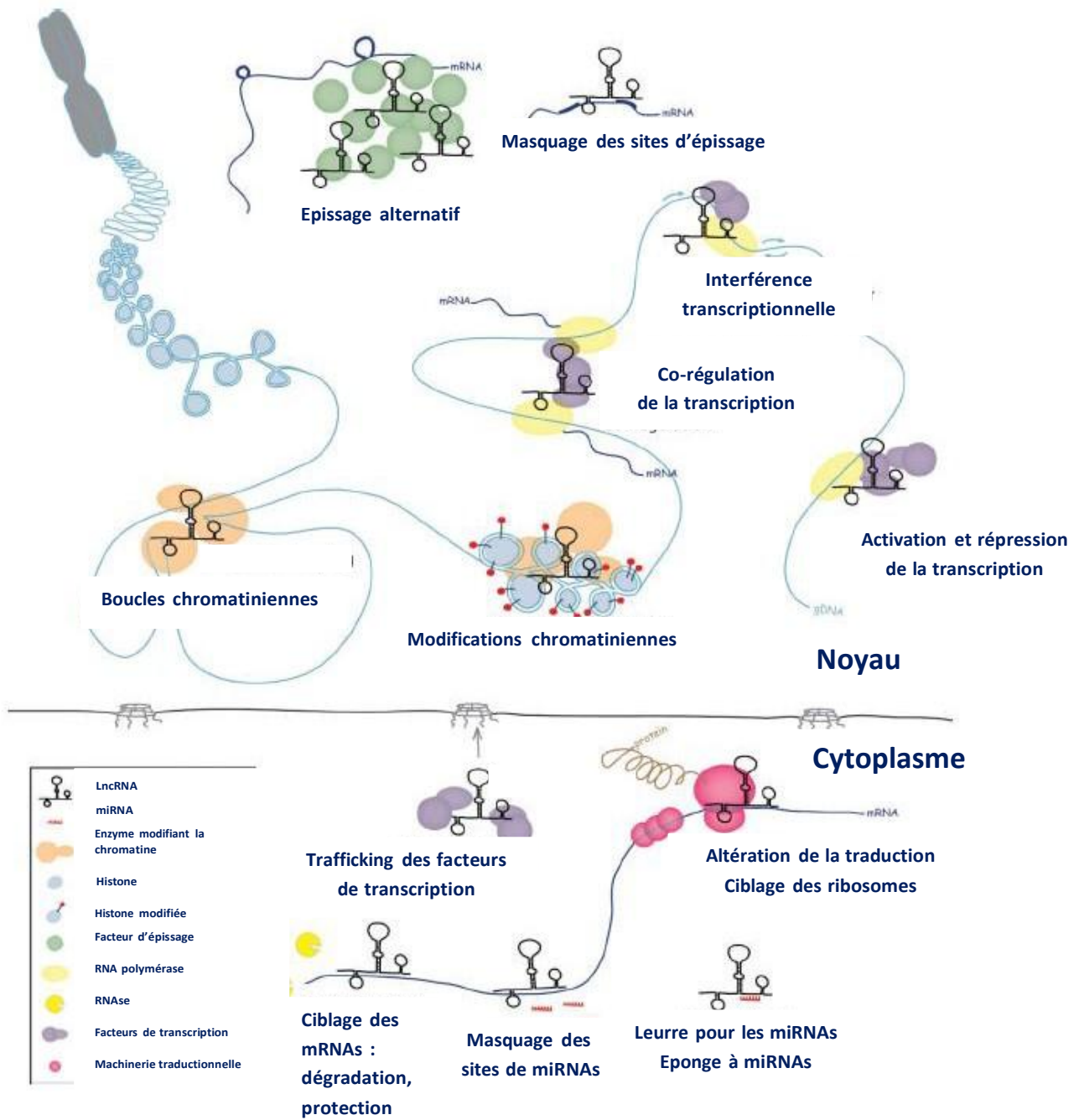


Figure 26 : Coordination de l'expression génique par les lncRNAs (d'après Tye et al., 2015)

2.3.5.2.1 Régulation au niveau épigénétique

Les lncRNAs contrôlent l'expression des gènes *via* la régulation de la chromatine. La plupart des lncRNAs sont des répresseurs transcriptionnels (*HOTAIR*, *ANRIL*, *Xist*, *H19*, *AIR*) mais certains sont des activateurs de la transcription (*HOTTIP*).

2.3.5.2.1.1 Expression allélique

2.3.5.2.1.1.1 Inactivation du chromosome X

L'inactivation d'un des chromosomes X chez la femme *via* le centre d'inactivation du X (*Xic*), conduit à une expression génique équilibrée du chromosome X entre les 2 sexes (Brown et al., 1992). *Xist* est transcrit à partir du *Xic* et peut couvrir l'ensemble du chromosome X inactif (*Xi*). Il inactive l'expression des gènes sur le *Xi* en recrutant les complexes de remodelage de la chromatine dont PRC2 et en induisant la triméthylation H3K27 des histones et la méthylation des ilots CpG (Payer and Lee, 2008 ; Kanduri et al., 2009 ; Leeb et al., 2009). **D'autres lncRNAs sont produits à partir du *Xic* : *Tsix*** est un antisens de *Xist* qui l'antagonise par compétition pour PRC2 avec *RepA*. ***RepA*** recrute PRC2 et permet la transcription de *Xist* (Marahrens et al., 1997 ; Lee et al., 1999). ***Xite*** agit comme un *enhancer* de *Tsix*. ***Jpx*** est un activateur de *Xist* (Tian et al., 2010). ***XACT*** est produit par les cellules pluripotentes humaines et peut couvrir le chromosome X actif en l'absence d'expression de *Xist*.

2.3.5.2.1.1.2 Empreinte génomique parentale

L'empreinte parentale est le processus par lequel la copie d'un gène héritée de l'un des parents est épigénétiquement réprimée (Wood and Oakey, 2006). Ainsi, certains gènes présentent une activation transcriptionnelle de l'allèle paternel et une inactivation de l'allèle maternel, tandis que d'autres possèdent un profil d'expression opposé. L'ICR (*Imprinting Control Region*) coordonne l'empreinte génomique des locus concernés (Bartolomei et al., 2009). Les régions soumises à l'empreinte parentale contiennent une large proportion de gènes d'ARNs non codants qui favorisent l'inactivation génique *via* divers mécanismes, dont le remodelage chromatinien et la compétition avec les *enhancers* (Wan and Bartolomei, 2008). Le dépôt des marques épigénétiques d'inactivation est guidé par des lncRNAs exprimés depuis les brins antisens aux loci codants soumis à l'empreinte génomique. De

nombreux lncRNAs sont impliqués dans le processus de l’empreinte parentale par divers mécanismes incluant en particulier le remodelage de la chromatine (PRC1 et PRC2) et l’activation des promoteurs. *H19* participe au contrôle des gènes sous empreinte parentale impliqués dans la croissance chez l’embryon en recrutant la protéine MBD1 (*Methyl-CpG-Binding Domain protein 1*) et en formant un complexe qui réprime 5 gènes du réseau de l’empreinte parentale, dont *IGF2* (Bartolomei et al., 1991 ; Ripoche et al., 1997). Pour assurer la régulation du gène *IGF2*, les loci *H19* et *IGF2* partagent le même *enhancer*, mais *IGF2* est exprimé uniquement à partir de l’allèle paternel et *H19* à partir de l’allèle maternel. Sur l’allèle maternel, le facteur de transcription CTCF (*CCCTC-binding factor*) se fixe sur l’ICR non méthylé, empêchant l’*enhancer* d’interagir avec *IGF2* et permettant à *H19* d’accéder à l’*enhancer*. Sur l’allèle paternel, la méthylation de l’ICR s’étend au promoteur de *H19*, qui inactive l’expression de *H19* et prévient la fixation du CTCF sur l’ICR. De plus, *H19* est le précurseur du miRNA miR-675 qui assure la régulation de la croissance placentaire (Monnier et al., 2013). *KCNQ1OT1* est exprimé au sein du locus du gène *KCNQ1* sur le chromosome paternel et est réprimé sur le chromosome maternel via la méthylation de l’ICR et de son promoteur (Mohammad et al., 2008). L’expression de *KCNQ1OT1* par le chromosome paternel conduit à la répression d’une dizaine de gènes dans un locus *KCNQ1* de 1 Mb via son interaction avec EZH2 et SUZ12 appartenant au complexe PRC2 et avec l’histone méthyl-transférase G9a (Nakano et al., 2006 ; Pandey et al., 2008 ; Wagschal et al., 2008 ; Kanduri, 2011). *AIR* est impliqué dans l’empreinte génomique parentale du domaine *IGF2R*. Dans ce dernier, *IGF2R* et le locus produisant le lncRNA *AIR* sont transcrits selon des directions opposées. L’ICR est situé dans le 2^o intron d’*IGFR2*, qui contient le promoteur d’*AIR*. L’ICR est non méthylé sur l’allèle paternel, conduisant à l’expression du locus *AIR* par évacuation de la RNA Polymérase II du promoteur (Braidotti et al., 2004). Le transcrit *AIR* couvre le promoteur de *IGFR2*, inhibant ainsi son expression. Sur l’allèle maternel la méthylation de l’ICR inhibe la transcription d’*AIR* et permet l’expression d’*IGF2R* (Sleutels et al., 2002 ; Nagano et al., 2008).

2.3.5.2.1.2 Positionnement des nucléosomes

Les nucléosomes peuvent être déstabilisés et restructurés par des complexes de remodelage appartenant à 4 familles (SWI/SNF, ISWI, CHD et INO80) (Yen et al., 2012). Le positionnement des nucléosomes et leur remodelage régulent l’expression génique en

perturbant l'accessibilité des séquences régulatrices d'ADN par des facteurs de transcription et de la machinerie de transcription. Le lncRNA *SChLAPI* interagit avec la sous-unité SNF5 du complexe SWI/SNF et inhibe sa fixation à la chromatine, entraînant une large dérégulation de l'activité génique. De même, *Myheart* se lie à la sous-unité BRG1 du complexe SWI/SNF et prévient le remodelage chromatinien. A l'opposé, *Evf2* favorise la fixation du complexe SWI/SNF au niveau des *enhancers* des gènes *Dlx5* et *Dlx6*.

2.3.5.2.1.3 Méthylation de l'ADN

Durant le développement embryonnaire, l'établissement et le maintien de la méthylation de l'ADN au niveau des îlots CpG des promoteurs conduit à la modification de l'expression des gènes. *Khps1a* se lie à des sites CC (A/T) GG, réduit la méthylation de ses îlots CpG et conduit à l'expression du gène *Sphk1* (sphingosine kinase 1) (Mattick, 2009). *EcCEBP* se lie à l'ADN méthyltransférase 1 et inhibe la méthylation du locus *CEBPA* (di Ruscio et al., 2013). Les lncRNAs peuvent modifier la méthylation de l'ADN en interagissant directement avec divers membres de la famille des DNA méthyltransférases (DNMTs) : (i) DNMT1 en *cis* (extra-coding *CEBPA*, *ecCEBP*) ou en *trans* (*Dali*) et (ii) DNMT1, DNMT3a et DNMT3b (*developmental pluripotency-associated 2 Dppa2*, *Dum*). Des lncRNAs (*lincRNA-p21*) peuvent également interagir indirectement sur la méthylation de l'ADN via des interactions avec des protéines liant l'ARN (hnRNPk) qui servent de plateformes à divers méthyltransférases (SETDB1 et DNMT1) (Zhao et al., 2016).

2.3.5.2.1.4 Modification du code histone

Les lncRNAs constituent un réseau de modulateurs épigénétiques permettant un remodelage chromatinien. Ils agissent comme des guides, signaux et plateformes pouvant recruter des protéines régulant la chromatine et les guider vers les gènes cibles et constituer des plateformes multimoléculaires pour des complexes protéiques ne possédant pas habituellement d'interactions protéine-protéine. Leurs sous-unités catalytiques interagissent en modifiant le code histone et le profil de méthylation avec comme conséquence une activation ou plus fréquemment une inactivation de l'expression des gènes cibles (Khalil et al., 2009). *ANRIL* recrute les complexes répresseurs polycomb PRC1 et PRC2 et induit l'inactivation des gènes du locus *INK4B-ARF-INK4A* et plus particulièrement de *p15* via

SUZ12 de PRC2 (Mattick and Gagen, 2001 ; Mattick, 2009 ; Margueron and Reinberg, 2011 ; Wan et al., 2013a). **HOTAIR** recrute et sert de plateforme d'échafaudage pour les 2 complexes PRC2 et LSD1, entraînant simultanément le dépôt de la marque répressive H3K27me3 et le retrait de la marque activatrice au niveau du locus des gènes *HOXD* (Shi et al., 2005 ; Rinn et al., 2007 ; Tsai et al., 2010). **MISTRAL** et **HOTTIP** faciliteraient l'expression des gènes *HOXA* en recrutant le complexe épigénétique activateur WDR5/MLL, qui induit le dépôt de la marque épigénétique d'activation H3K4me3 au niveau des gènes cibles (Schuettengruber et al., 2007 ; Bertani et al., 2011 ; Wang et al., 2011b). De même, **KCNQ10T1** se lie à G9a et PRC2, **TUG1** et **Fendrr** à PRC2.

2.3.5.2.2 Régulation au niveau de l'ADN

Les lncRNAs peuvent modifier l'expression génique en se liant à des domaines spécifiques des gènes cibles adjacents en *cis*, en interagissant avec l'ARN Polymérase II, en coopérant avec des protéines et en fonctionnant comme des co-activateurs.

Interactions avec les *enhancers* et promoteurs : Les lncRNAs associés aux promoteurs (PaRNAs) et les lncRNAs de type *enhancer* (eRNAs) peuvent directement réguler la transcription de gènes cibles adjacents par les mécanismes de l'activation et de la suppression transcriptionnelles. Dans le cadre de la suppression transcriptionnelle, les lncRNAs agissent par interférence transcriptionnelle dans la région promotrice des gènes codants. **SRGI** inhibe l'initiation de la transcription du gène *SER3* (Martens et al., 2004). **Up-DHFR** forme une triple hélice ARN-ADN avec le promoteur du gène *DHFR*, interagit directement avec le facteur de transcription TFIIB, dissocie le complexe de pré-initiation et réprime *DHFR* (Martianov et al., 2007).

Interactions avec l'ARN Polymérase II : Les lncRNAs peuvent interagir avec l'ARN polymérase II et bloquer la transcription en *trans*. **B2 et Alu** sont des lncRNAs produits par des séquences SINE. Leur induction par un choc thermique s'accompagne d'une diminution de la stabilité des complexes de pré-initiation et d'une inhibition des gènes cibles en interagissant avec le facteur de transcription TFIIF (Allen et al., 2004 ; Mariner et al., 2008).

Interactions avec les protéines : Les lncRNAs peuvent indirectement affecter la transcription en *cis* en se liant à des protéines spécifiques, en activant des facteurs de transcription accessoires ou en oligomérisant des protéines activatrices. *Up-CCND1*, induit par des altérations de l'ADN, coopère avec la protéine TLS (*translocated in liposarcoma*) au niveau du promoteur de *CCND1*. TLS se lie à CBP et p300 et est activée allostériquement. L'activation de TLS s'accompagne de l'inhibition de l'activité acétyltransférase de CBP et p300, supprimant ainsi la transcription des gènes cibles (Wang et al., 2008). *HSR1* favorise la trimérisation de la protéine HSF1 qui induit l'expression des HSPs (*heat shock proteins*) (Shamovsky et al., 2006). *NRON* régule le trafic intracellulaire en réponse à un stress et permet de faire circuler le facteur de transcription NFAT depuis le cytoplasme de manière à promouvoir l'expression des gènes cibles (Sharma et al., 2011). *THRIL*, impliqué dans la réponse immune, se lie à la protéine hnRNP-1 et induit la transcription du gène *TNF α* en se fixant au niveau de son promoteur (Li et al., 2014a).

Co-activation transcriptionnelle : Certains lncRNAs agissent comme des co-activateurs permettant de contrôler la transcription des gènes cibles. *LincRNA-p21* interagit en tant que co-activateur avec hnRNP-K pour induire la transcription en *cis* du gène adjacent *p21* et l'expression des gènes polycomb (Dimitrova et al., 2014).

2.3.5.2.3 Régulation au niveau de l'ARN

Les lncRNAs sont largement impliqués dans les processus post-transcriptionnels. Ils participent aux mécanismes de la compétition endogène avec divers types de RNAs et à la biogénèse de siRNAs endogènes.

Interactions avec les mRNAs : La capacité des lncRNAs à reconnaître des séquences complémentaires leur permet d'interagir spécifiquement avec des mRNAs cibles. Ils peuvent ainsi réguler plusieurs processus tels que la biogénèse et la maturation des mRNAs (*capping*, polyadénylation et épissage alternatif des pré-mRNAs), la stabilité et l'activité transcriptionnelle. Ils agissent également sur le trafic, leur localisation cellulaire et leur dégradation (Tripathi et al., 2010 ; Gong et al., 2011 ; Yoon et al., 2012). Concernant la maturation des mRNAs, les lncRNAs peuvent agir par 2 mécanismes : (i) les transcrits antisens naturels (NATs), en formant des duplex RNA-RNA avec leurs transcrits sens correspondants, altèrent directement la reconnaissance du site d'épissage et le recrutement du

spliceosome, (ii) les lncRNAs peuvent agir indirectement en régulant la phosphorylation des facteurs d'épissage et en interagissant avec des composants du *spliceosome* (petits RNAs nucléaires U1). *L'anti-sens de ZEB2* peut masquer les sites d'épissage en agissant en *cis* et ainsi provoquer un épissage de l'ARNm de *ZEB2* qui est alternativement épissé quand son lncRNA antisens est épissé, provoquant la rétention de la séquence 5'-UTR et une augmentation du niveau de traduction de *ZEB2* (Beltran et al., 2008). *MALATI* régule l'activité des protéines riches en sérine/arginine (SR) qui sont impliquées dans l'épissage alternatif de multiples protéines (Hutchinson et al., 2007 ; Tripathi et al., 2010 ; Bardou et al., 2014). Les lncRNAs peuvent stabiliser ou déstabiliser leur transcrits cibles. Ainsi, *BACE1-AS* augmente la stabilité du mRNA *BACE1* en s'appariant parfaitement avec sa séquence (Faghihi et al., 2008). *I/2sbsRNA* présente un appariement imparfait avec les mRNAs contenant des sites de fixation pour Staufen-1 et des séquences Alu. Ce processus favorise la combinaison de Staufen-1 et des séquences Alu qui entraîne une dégradation du mRNA (Gong and Maquat, 2011).

Interactions avec les miRNAs : Les miRNAs régulent la traduction des mRNAs en assurant leur dégradation *via* le complexe RISC ou leur séquestration dans les corps GW/P ou les granules de stress. Les lncRNAs peuvent soit interagir directement avec des miRNAs de manière à réguler leur biogénèse et leur activité et ainsi contrôler la traduction et la dégradation de leurs mRNAs cibles, soit se fixer sur des séquences de mRNAs se liant aux miRNAs (Bak et al., 2014). Les RNAs compétiteurs endogènes (ceRNAs) ont été initialement découverts chez les plantes. Ils comportent des pseudogènes, des lncRNAs, des mRNAs et des lncRNAs circulaires. Ils représentent un nouveau mécanisme de régulation post-transcriptionnelle (Shi et al., 2015a). Les ceRNAs possèdent des miRNAs *response elements* (MREs) et inhibent l'expression de leurs miRNAs cibles *via* 2 mécanismes, la compétition et « l'éponge moléculaire » (Yang et al., 2016a). Pandolfi et al. ont proposé l'hypothèse du réseau de compétiteurs endogènes (Salmena et al., 2011). Selon cette hypothèse, les MREs sont les lettres d'un langage constitué d'ARN et les transcrits possédant des MREs spécifiques peuvent communiquer avec d'autres transcrits *via* la médiation d'un miRNA. Les MREs sont situés dans les régions non codantes 3'UTR ou dans les régions codantes. Tout transcrit possédant des MREs peut ainsi fonctionner comme un compétiteur endogène et déréprimer l'activité d'autres RNAs pourvus de MREs similaires. Des protagonistes sont nécessaires afin de constituer un réseau de compétiteurs endogènes : les miRNAs comme « messagers » et les transcrits comme compétiteurs endogènes, à l'origine de deux types de

réseaux d'interactions complexes intracellulaires : (i) mRNA/lncRNA/miRNA et (ii) mRNA/lncRNA/facteurs de transcription/miRNA. Les petits ARNs non codants ne contiennent pas suffisamment de MREs mais interfèrent avec la régulation du réseau de compétiteurs endogènes *via* leur abondance et leur disponibilité aux niveaux transcriptionnel et post-transcriptionnel (Yang et al., 2016a). Les ceRNAs peuvent ainsi neutraliser les miRNAs (Deng et al., 2016b ; Liz and Esteller, 2016). *Uc.283+A* inhibe la maturation du pri-miR-195 en miRNA mature en prévenant son clivage par Drosha (Liz et al., 2014). *Linc-MDI* supprime l'action de miR-133 and miR-135 et contrôle la traduction de *MEF2C* et *MAMALI* (Cesana et al., 2011 ; von Roretz et al., 2011). *HULC* et *CDRI-as/ciRS-7* séquestrent les miRNAs et les empêchent de se fixer et d'inactiver leurs mRNAs cibles (Memczak et al., 2013). *PTENPI* module le niveau d'expression endogène de *PTEN* en séquestrant les miRNAs le ciblant (Deng and Sui, 2013). *Linc ROR* inhibe miR-145 (Takahashi et al., 2014a). L'hypothèse du réseau compétiteur endogène a été validée par de nombreuses études et permet d'expliquer des résultats expérimentaux inattendus: (i) la découverte d'un mRNA de *ZEB2* suppresseur de mélanome suggère qu'un mRNA codant pour une protéine peut fonctionner non seulement comme un lncRNA, mais aussi que le rôle de la protéine peut être différent voire opposé à celui du ncRNA (*ZEB2*), (ii) certains transcrits peuvent avoir différents rôles dans différents types tumoraux, dépendant du *pool* de miRNAs et d'autres transcrits du réseau de compétiteurs endogènes (Versican 3'UTR), (iii) des lncRNAs agissent comme compétiteurs endogènes dans le tissu normal mais non dans le cancer ou le réseau est aboli (*PTENPI*), (iv) des lncRNAs compétiteurs endogènes, bien que surexprimés dans le cancer, ne sont plus fonctionnels en raison de la transcription préférentielle d'isoformes dépourvus de MREs dans le cancer du sein (*PVTI*) et (v) les mRNAs pourraient constituer de nouvelles cibles thérapeutiques (mRNA de *HMGA2* et *BRAF*).

Interactions avec les siRNAs endogènes : La formation de duplex RNA/RNA constitués de séquences codantes et de lncRNAs peut générer la production de siRNAs endogènes (*CENDI*, *FERM*) (Watanabe et al., 2008).

2.3.5.2.4 Régulation au niveau des protéines

Comme tous les RNAs, les lncRNAs ont la particularité biochimique de reconnaître et de se lier à la plupart des biomolécules, dont les protéines. Ils peuvent de ce fait contrôler également l'homéostasie tissulaire en se liant et en modifiant l'activité des protéines. De plus,

certaines lncRNAs régulent l'expression des gènes *via* leurs interactions avec des récepteurs hormonaux. *Evi2* édifie des complexes stables avec les membres de la famille de facteurs de transcription Dlx/Dll impliqués dans le développement embryonnaire. *SRA* agit comme un co-activateur de différents récepteurs hormonaux stéroïdiens, tels les récepteurs aux oestrogènes, à la progestérone, aux androgènes et aux glucocorticoïdes (Lanz et al., 1999). *GAS5* est également impliqué dans la régulation génique médiée par les hormones (Kino et al., 2010). *PRNCRI* et *PCGEMI* se lient à la partie COOH terminale du récepteur androgénique (AR) et sont impliqués dans son activation indépendante du ligand. Leur surexpression est habituellement observée dans les carcinomes prostatiques réfractaires aux androgènes. Toutefois, leur rôle dans la carcinogénèse prostatique a été récemment remis en cause (Prensner et al., 2014b).

2.3.5.2.5 Régulation complexe à de multiples niveaux

En réalité, plusieurs niveaux de régulation peuvent s'associer et être utilisés par les lncRNAs dans le contrôle de l'expression génique. L'oncosuppresseur *PTEN* est ainsi l'objet d'une régulation complexe par son pseudogène *PTENP1*. Lorsqu'il est transcrit en sens, le locus *PTENP1* produit le lncRNA *PTENP1*. Transcrit en antisens, le locus *PTENP1* crée 2 isoformes d'un autre lncRNA, α et β . L'isoforme α se lie au promoteur de *PTEN*, recrute la DNA méthyltransférase DNMT3A et les histones méthyltransférases EZH2 et G9a et induit l'inactivation épigénétique de *PTEN* *via* la méthylation de H3K27 et H3K9. L'isoforme β constitue un duplex RNA/RNA avec le lncRNA sens *PTENP1*, renforce la stabilité de ce dernier et favorise sa fonction « d'éponge moléculaire » pour les miRNAs, résultant en une augmentation de l'activité de *PTEN* (Deng and Sui, 2013). *HOTAIR* en plus de ses fonctions régulatrices au niveau épigénétique, agit également au niveau post-transcriptionnel en servant de plateforme pour l'ubiquitinylation des protéines. *PANDA* inhibe l'expression des gènes de l'apoptose en séquestrant le facteur de transcription NF- κ B. *LncRNA-ATB* agit comme un compétiteur endogène en séquestrant miR-200 de sa cible ZEB1/2 et promeut ainsi les fonctions pro-métastatiques de TGF- β (Shi et al., 2015b).

2.3.5.2.6 Régulation inductible

Certains lncRNAs semblent fonctionner comme des molécules régulatrices dans certaines circonstances et interagir avec la machinerie traductionnelle pour produire des peptides fonctionnels sous d'autres conditions. Ainsi, certains lncRNAs cytoplasmiques pourvus de petits cadres de lectures ouverts (smORFs), observés chez la levure (*Xrn1-sensitive unstable transcripts*, XUTs) mais aussi chez la drosophile, le poisson-zèbre et la souris, peuvent être traduits en peptides impliqués en particulier dans les mouvements embryonnaires et l'activité cardiaque *in vivo* (Magny et al., 2013 ; Pueyo et al., 2016). Le système NMD (*Nonsense-Mediated Decay*) est un système de surveillance lié à la machinerie traductionnelle, qui assure l'élimination des mRNAs pourvus d'une terminaison aberrante et des lncRNAs cryptiques. Le système NMD contribue également à la régulation de l'expression génique médiée par les lncRNAs en réduisant physiologiquement le niveau d'expression des lncRNAs régulateurs antisens qui inhibent normalement la transcription de leurs transcrits appariés sens codant pour des protéines. L'inhibition du système NMD et de la traduction par le stress conduit à la stabilisation des lncRNAs et à la réduction de la transcription des transcrits sens appariés (de Andres-Pablo et al., 2016). D'autres mécanismes liés au stress hypoxique assurent la stabilisation de HIF1 (*lncRNA LET*) ou sont médiés par HIF1 et induisent la transcription d'isoformes non codants et codants à partir d'un même locus (*EFNA3*). Les isoformes non codants surexprimés neutralisent alors par compétition endogène les miRNAs qui inhibent les les transcrits codants, aboutissant à leur surexpression (Gomez-Maldonado et al., 2015).

2.3.5.3 Fonctions structurales des lncRNAs

Les lncRNAs peuvent également interagir avec leurs cibles en exploitant la conformation tridimensionnelle du noyau. Ils façonnent ainsi l'architecture nucléaire selon plusieurs niveaux d'organisation et apparaissent essentiels à l'édification et au maintien de différents domaines intranucléaires (Clemson et al., 2009 ; Bergmann and Spector, 2014 ; Hacısuleyman et al., 2014 ; Quinodoz and Guttman, 2014). Ces domaines sont caractérisés par une concentration élevée des lncRNAs à proximité de leur site transcriptionnel et des complexes protéiques interagissant avec les lncRNAs. Les lncRNAs pourraient ainsi constituer des « centres organisationnels » édifiant de multiples domaines fonctionnels et/ou structuraux intranucléaires qui régulent temporairement ou de façon permanente l'expression des gènes.

Ce modèle spatial de régulation épigénétique pourrait expliquer le très faible niveau d'expression habituel des lncRNAs (Chujo et al., 2016).

2.3.5.3.1 Stabilité génomique et chromosomique

2.3.5.3.1.1 Domaines régulateurs intrachromosomiques

La restructuration tridimensionnelle du chromosome X en un corpuscule de Barr est assurée par *Xist* (Plath et al., 2002 ; Splinter et al., 2011). Ce lncRNA inactive progressivement l'ensemble du chromosome X en interagissant à proximité de son locus transcriptionnel avec les gènes non encore inactivés et les repositionne au sein de ce compartiment intrachromosomique périphérique à l'aide de son domaine répétitif A (Chaumeil et al., 2006).

2.3.5.3.1.2 Domaines régulateurs interchromosomiques

Les domaines interchromosomiques sont définis par l'accumulation spatiale de gènes présentant des fonctions communes ou régulés par les mêmes facteurs (Lanctot et al., 2007). *Firre*, impliqué dans l'adipogénèse, édifie un unique domaine nucléaire *trans*-chromosomique hébergeant 5 gènes cibles localisés sur les chromosomes 2, 9, 15 et 17 (Sun et al., 2013). Ces gènes impliqués dans le métabolisme énergétique sont disposés à proximité du locus transcriptionnel de *Firre* situé sur le chromosome X.

2.3.5.3.1.3 Boucles d'interaction *enhancer*-promoteur

- De larges boucles d'ADN rapprochant des couples *enhancer*/promoteur et des promoteurs éloignés sur un même chromosome ou situés sur plusieurs chromosomes ont été découvertes à l'aide des techniques de capture chromosomique conformationnelle, de microscopie haute résolution et de FISH (Lai et al., 2013 ; Li et al., 2013a). Ces larges boucles sont issues d'unités de la taille d'une mégabase dénommées « domaines topologiques » caractérisés par une importante concentration de sites liant le CTCF. Ce dernier participe à l'édification et au maintien de ces larges interactions et est lié à la chromatine par des lncRNAs (Fanucchi et al., 2013 ; Halley et al., 2014 ; Shibayama et al., 2014).

- Des boucles d'interaction plus courtes entre *enhancer* et promoteur peuvent être médiées par des lncRNAs exprimés à partir du site *enhancer* (eRNAs). Ces eRNAs interagissent avec les protéines du complexe Médiateur afin d'initier la flexion de la chromatine. Le complexe Médiateur stabilise ensuite la boucle chromatinienne et assure la transcription. Des lncRNAs, tels *PRNCR1* et *PCGEM1*, dont la synthèse est favorisée par les androgènes édifient des boucles d'ADN entre *enhancers* et promoteurs et favorisent l'activation des gènes induits par les androgènes (Lai et al., 2013 ; Yang et al., 2013a).

2.3.5.3.2 Edification de corps intra-nucléaires

Les corps nucléaires (CN) sont des structures granuleuses dépourvues de membrane (Mao et al., 2011). Ils sont constitués de protéines et d'ARN et sont impliqués dans la biogénèse, la maturation, le stockage et la séquestration de divers ARNs, de protéines et de complexes ribonucléoprotéiques. Certains lncRNAs architecturaux (arcRNAs) servent de plateformes et de matrices à l'édification de différents types de CNs (Chujo et al., 2016). Les arcRNAs sont des lncRNAs localisés spécifiquement avec une concentration élevée dans des CNs. La disparition des arcRNAs entraîne la désintégration des CNs ou altère leur localisation. Ils présentent des caractéristiques communes : (i) les arcRNAs forment des CNs au niveau leur site de transcription, (ii) ils sont transcrits à partir de séquences répétitives ou en comportent, (iii) leur expression est transitoire (stress, étapes du développement, infection virale, cancer), (iv) ils séquestrent diverses protéines régulatrices (protéines fixant l'ADN et l'ARN, ubiquitine ligase) et affectent l'expression des gènes.

Paraspeckles

La formation des *paraspeckles* est corrélée à la transcription active du lncRNA *NEAT1* et des amas de *paraspeckles* se développent autour des loci *NEAT1*. Ces structures ribonucléoprotéiques de 360 nm comportant plus de 60 types de protéines, sont impliquées dans la régulation de l'expression génique en assurant la séquestration de RNAs (Clemson et al., 2009 ; Choudhry et al., 2015 ; Hirose et al., 2014 ; Yamazaki and Hirose, 2015). Ils participent à la formation des corps lutéaux et au développement de la glande mammaire et sont impliqués dans la sclérose latérale amyotrophique et divers cancers, dont les carcinomes du sein et de la prostate.

Corps nucléaires de stress

Le stress induit par un choc thermique conduit à la formation des corps nucléaires de stress en activant la transcription de répétitions en tandem de séquences satellites satIII situées dans les régions péri-centromériques du chromosome 9q12. Dans ces corps de stress, les SatIII lncRNAs semblent affecter l'expression globale en séquestrant des facteurs de transcription et d'épissage, des protéines fixant l'ADN et des complexes de remodelage, dont SWI/SNF (Denegri et al., 2002 ; Rizzi et al., 2004 ; Jolly et al., 2006).

Centres de détention nucléolaires

La formation des centres de détention nucléolaires est liée à un stress cellulaire dû à l'acidose ou à un choc thermique. Elle est dépendante de l'expression de lncRNAs qui capturent et immobilisent les protéines nucléolaires, assurant un arrêt de la biogénèse ribosomale. La disparition des facteurs de stress induit une diminution de leur expression et le retour du nucléole à une conformation active (Audas and Jacob, 2012 ; Jacob et al., 2013).

Les lncRNAs *TUG1* et *MALATI* localisés respectivement dans les corps répresseurs polycomb et les granules activateurs interchromatiniens (*speckles*), ne sont pas *stricto sensu* des arcRNAs, mais contrôlent la localisation de multiples gènes et coordonnent l'inactivation (*TUG1*) ou l'activation (*MALATI*) de leur transcription (Yang et al., 2011b).

Ainsi, le modèle actuel de fonctionnement des lncRNAs suggère que les lncRNAs assurent la coordination et le transfert du flux d'informations entre les structures tridimensionnelles *intra* ou *interchromosomiques*, les modifications épigénétiques de la chromatine et la régulation de la transcription des gènes.

2.3.5.4 Biogénèse de petits ARNs non codants

Les lncRNAs servent de précurseurs pour divers types de RNAs, tels les piRNAs, les snoRNAs et les miRNAs qui assureront à leur tour leurs fonctions régulatrices (*Affimetrix ENCODE Transcriptome Project and Cold Spring Harbor Laboratory ENCODE Transcriptome Project*, 2009 ; Rother and Meister, 2011 ; Ha et al., 2014). En outre, certains lncRNAs antisens transcrits à partir de pseudogènes et des NATs peuvent se lier à des mRNAs spécifiques pour constituer des duplex qui seront ensuite dégradés en siRNAs endogènes régulant l'activité de ces mRNAs (Watanabe et al., 2008).

2.3.5.5 LncRNAs extra-cellulaires circulants

Des lncRNAs ont été identifiés au sein des exosomes présents dans différents fluides biologiques (Liu et al., 2014a). Ainsi, le lncRNA *linc-ROR* apparaît surexprimé dans les vésicules exosomales secrétées par les cellules cancéreuses des hépatocarcinomes (Takahashi et al., 2014a).

2.3.6 Fonctions cellulaires des lncRNAs

2.3.6.1 Développement embryonnaire, organogénèse et équilibre homéostatique

Des études pangénomiques analysant les profils transcriptomiques de cellules souches embryonnaires et adultes ainsi que différents types cellulaires et tissulaires spécifiques, ont identifié de nombreux lncRNAs impliqués dans la régulation du développement embryonnaire, de l'organogénèse et de l'homéostasie cellulaire et tissulaire (Flynn and Chang, 2014 ; Lopez-Pajares, 2016 ; Melissari and Grote, 2016).

Au cours du développement embryonnaire, de nombreux lncRNAs régulateurs sont variablement exprimés selon les stades de différenciation des cellules souches (Rosa and Brivanlou, 2013). Certains lncRNAs sont impliqués dans le maintien de la pluripotence (*HOTAIRM1*, *Rian*, *lncRNA-ES1-3*, *LincRNA-SFMBT2*, *lincRNA-VLDLR*, *MEG9*), alors que d'autres, tels *Mira* favorisent la différenciation des cellules souches (Loewer et al., 2010 ; Cesana et al., 2011 ; Kretz et al., 2012 ; Ng et al., 2012 ; Klattenhoff et al., 2013 ; Berghoff et al., 2013 ; Eades et al., 2014 ; Shore et al., 2014 ; Zhang et al., 2014e). De plus, des lncRNAs soumis à l'empreinte parentale, tels *H19*, sont non seulement exprimés au cours du développement par les cellules souches embryonnaires mais également par les cellules souches adultes, où ils jouent un rôle additionnel dans la régulation de l'équilibre entre auto-renouvellement et différenciation dans divers tissus adultes (Mattick, 2009 ; Jeon et al., 2012). En outre, les lncRNAs présentent souvent chez l'adulte une expression temporo-spatiale précise, témoignant de leur rôle dans l'engagement et la différenciation au cours du développement (Clark and Mattick, 2011a ; Derrien et al., 2012a).

De nombreux lncRNAs sont exprimés au cours de l'organogénèse (**Tableau 4**). Ces lncRNAs participent au développement du système nerveux central (*Six3os*, *Dlx1as*, *TUNA*, *RMST*, *Paupar*), des muscles et du cœur (*Braveheart*, *Fendrr*), du tissu adipeux (*SRA*), du pancréas (*H19*), des vaisseaux (*MALAT1*, *TUG1*, *MEG3*), du poumon (*NANCI*) et du système osseux (*DANCR*, *HOTAIR*, *HOTTIP*) (Boon et al., 2016).

L'homéostasie cellulaire réalise un équilibre dynamique permettant une adaptation aux différents stimuli provenant des cellules voisines et du microenvironnement. La régulation de l'expression génique représente une des réponses cellulaires majeures à ces modifications et son intégrité est assurée par des processus correcteurs situés aux niveaux transcriptionnel et post-transcriptionnel (Shi et al., 2015a). Bien que les protéines demeurent les effecteurs essentiels des mécanismes régulateurs de l'expression des gènes, les ARNs non codants et en particulier les lncRNAs, ont plus récemment émergé comme des coordonnateurs majeurs de l'expression des gènes impliqués dans le maintien de l'équilibre homéostatique (Audas and Lee, 2016 ; Carpenter, 2016). Les lncRNAs sont pourvus d'une capacité intrinsèque à interagir avec les protéines, l'ARN et l'ADN *via* leur structures secondaires-tertiaires et leurs séquences nucléotidiques. A l'échelle de la cellule, les lncRNAs régulent de nombreux processus physiologiques tels que l'adaptation au stress, la pluripotence, la différenciation, la sénescence, l'apoptose, le cycle cellulaire, le métabolisme et l'immunité (Schmitz et al., 2016 ; Kour and Rath, 2016 ; Carpenter, 2016). Leur expression est essentielle à la survie des cellules car des expériences de *knock-out* de certains lncRNAs chez la souris peuvent s'accompagner d'une létalité péri- ou post-natale ainsi que de divers troubles du développement. De plus, la dérégulation de leur expression est observée dans de multiples pathologies, dont le cancer. De nombreux lncRNAs jouent un rôle fondamental dans le maintien de l'homéostasie cellulaire et tissulaire en participant à la régulation du cycle cellulaire et de l'apoptose (*PTENPpg1*, *LincRNA-RoR*, *LincRNA-EPS*, *FMR4*, *ANRIL*) (Wapinski and Chang, 2011 ; Rossi and Antonangeli, 2014), de la réparation de l'ADN (*DDSR1*) (Sharma et al., 2015), de l'angiogénèse (*MALAT1*, *TUG1*, *MEG3*) (Zhou et al., 2010 ; Michalik et al., 2014 ; Tee et al., 2016), des rythmes circadiens et du métabolisme (*116HG*, *asPer2*, *HULC*) (Kornfeld and Brüning, 2014 ; Cui et al., 2015) et du système immunitaire (*lincRNA-Cox2*, *THRIL*, *NEAT1*, *lnc-IL7R*, *PACER*, *lnc-DC*, *HOTAIRMI*, *NeST*) (Wu et al., 2015)(**Tableau 4**).

Fonctions	Localisation	Rôles physiologiques	Phénotype après KO	Mécanismes et partenaires	Références
Apoptose et cycle cellulaire					
<i>PTENPpg1</i>	noyau, cytoplasme	apoptose <i>via</i> la voie PI3K/AKT	prolifération	répression transcriptionnelle éponge à miRNAs	Wang et al. (2016)
<i>LincRNA-RoR</i>	cytoplasme	inhibition de l'apoptose (p53)	apoptose	hnRNP1	Wang et al. (2013)
<i>LincRNA-EPS</i>	noyau	prévention de l'apoptose	apoptose et arrêt de la prolifération	répression de <i>Pycard, Bad, Bax</i>	Atianand et al. (2016)
<i>FMR4</i>		rôle anti-apoptotique	arrêt du cycle cellulaire et apoptose	activation de <i>FMR1</i>	Peschansky et al. (2015)
<i>ANRIL</i>	noyau	rôle anti-apoptotique	prolifération	répression épigénétique du locus <i>INKA-ARF-INKB</i>	Qiu et al. (2016)
Réparation de l'ADN					
<i>DDSR1</i>	noyau	réparation de l'ADN	défaut de réparation de l'ADN	séquestration de <i>BRAC1/RAP80</i> activation de la voie ATM/NF-kB	Lukas et al. (2015)
Angiogénèse					
<i>MALAT1</i>	noyau	rôle pro-angiogénique	prolifération vasculaire	phénotype vasculaire pro-migratoire et antiprolifératif	Tee et al. (2016)
<i>TUG1</i>	noyau	rôle pro-angiogénique	réduction de l'angiogénèse	prolifération et différenciation vasculaire	Michalik et al. (2014)
<i>MEG3</i>	cytoplasme	rôle anti-angiogénique empreinte	prolifération vasculaire	modification des histones méthylation de l'ADN	Zhou et al. (2012)
Rythme circadien					
<i>116HG</i>	noyau	régulation du rythme circadien cérébral	dérégulation de gènes <i>Clock</i> et <i>Per2</i>	activation du FT RBBP5	Powell et al. (2013)
<i>asPer2</i>	noyau	régulation de l'oscillateur circadien <i>Per2</i>	dérégulation métabolique	oscillations dans les modifications épigénétiques histoniques	Vollmers et al. (2012)
Différenciation et organogénèse					
Epiderme					
<i>TINCR</i>	cytoplasme	différenciation épidermique	dédiérenciation en cellules souches	STAU1	Kretz et al. (2013)
<i>ANCR</i>	noyau	maintien des cellules souches	différenciation en cellules épidermiques	FT MAF/MAFB	Kretz et al. (2012)
Système nerveux					
<i>TUNA</i>	noyau	développement neural	malformations du SN et des yeux	PTBP1, hnRNP-K	Lin et al. (2014)

RMST	noyau	développement neural	accroissement glial	hnRNPA2/B1, SOX2	Ng et al. (2012)
Paupar	noyau	développement neural	croissance neuronale	PAX6	Vance et al. (2014)
Tissu squelettique et membres					
HOTAIR	noyau	développement vertébral	malformations homéotiques	PRC2/LSD1	Hansji et al. (2014)
HOTTIP	noyau	développement des membres	raccourcissement des os/membres	WDR5	Sang et al. (2016)
Tissu musculaire et myocarde					
Braveheart	noyau	développement cardiaque	réduction des corps embryotiques contractiles	modification des histones PRC2	Klattenhoff et al. (2013)
Fendrr	noyau	développement cardiaque et pulmonaire	malformations cardio-pulmonaires	modification des histones PRC2/WDR5	Grote et al. (2013)
NANCI	noyau	développement pulmonaire	réduction des saccules aériens	WDR5	Herriges et al. (2014)
Pancréas					
H19	noyau	développement des ilots de Langhérans	méthylation de l'ADN défaut de croissance des ilots de Langhérans	IGF2, miR-675	Dey et al. (2014)
Tissu adipeux					
SRA	noyau	adipogénèse	suppression de l'adipogénèse	Pparg	Novikova et al. (2012)
Tissu osseux et cartilagineux					
DANCR	noyau	chondrogénèse	suppression de la chondrogénèse	Sox4	Zhu et al. (2013)
Système immunitaire					
Immunité innée					
LincRNA-Cox2	noyau	activation (<i>IL-6, TLR1, IL23a</i>) et répression (<i>CCL5, CX3CL1</i>)		hn-RNP-A/B et A2/B1	Hu et al. (2016)
THRIL	noyau	régulation de <i>TNFα</i>		Hn-RNP-L	Li et al. (2014a)
NEAT1	noyau	activation de gènes pro-inflammatoires		séquestration du FT SFPQ	Imamura et al. (2014)
PACER	noyau	régulation du niveau basal de Cox2		sous-unité p50 de NF-κB	Krawczyk et al. (2014)
Immunité adaptative					
Lnc-DC		différenciation des cellules dendritiques activation des lymphocytes T		STAT3	Wang et al. (2014b)

<i>GAS5</i>		inhibition des lymphocytes T	récepteurs glucocorticoïdes	Pickard et al. (2014)
<i>NeST</i>		différenciation des lymphocytes Th1	WDR5	Zhao et al. (2016)
<i>LincR-Ccr2-5'AS</i>		migration des lymphocytes Th2		Hu et al. (2013)
<i>Linc-MAF-4</i>		différenciation des lymphocytes Th1/2	PRC2/LSD1	Ranzani et al. (2015)
Pluripotence				
<i>Gomafu</i>	noyau	contrôle de Oct4	SF1	Barry et al. (2014)
<i>HOTAIRM1</i>	noyau	contrôle du locus <i>HOXA</i>	PRC1, PRC2, CBX1	Zhang et al. (2014e)
<i>Rian</i>	noyau	répression génique	PRC1, JARID1B/1C, CBX3	Das et al. (2015)

Tableau 4 : Fonctions et rôles biologiques, partenaires moléculaires et mécanismes d'action des lncRNAs dans le maintien de la pluripotence, de la différenciation, de l'organogénèse et de l'homéostasie (adapté d'après Grote and Herrmann, 2015 ; Schmitz et al., 2016). Pages 110-111-112.

2.3.6.2 Rôles des lncRNAs dans le développement et la physiologie mammaire

2.3.6.2.1 LncRNAs régulant le développement de la glande mammaire

H19 et *SRA*, qui ont été parmi les premiers lncRNAs identifiés, semblent jouer un rôle dans le développement de la glande mammaire. *PVT1* a été plus récemment impliqué dans la physiologie mammaire.

H19 est induit par les œstrogènes dans la glande mammaire de la souris et les techniques d'hybridation *in situ* ont révélé l'existence d'une majoration de son expression dans les cellules épithéliales mammaires de souris pubères et dans les cellules alvéolaires de souris gravides (Adriaenssens et al., 1999). Bien qu'*H19* ait été initialement considéré comme un inhibiteur de la croissance durant le développement embryonnaire, des données récentes l'ont identifié comme pouvant maintenir à long terme le *pool* de cellules souches hématopoïétiques adultes (Venkatraman et al., 2013). Si *H19* possède un rôle équivalent dans la glande mammaire, il pourrait assurer la survie et le renouvellement des populations de cellules souches et/ou progénitrices au cours du développement de la glande mammaire et en particulier au cours des cycles hautement prolifératifs que sont la puberté et la grossesse.

SRA favorise la prolifération cellulaire, la ramification canalaire et la différenciation alvéolaire précoce lorsqu'il est surexprimé dans des cellules épithéliales mammaires d'un modèle de souris transgénique (Lanz et al., 2003). *SRA* module l'activité des récepteurs hormonaux nucléaires et est impliqué dans la différenciation spécifique de nombreux tissus (Caretta et al., 2006 ; Xu et al., 2010), suggérant un rôle potentiel dans la différenciation épithéliale mammaire.

PVT1 appartient au réseau compétiteur endogène mammaire MMI (*miRNA mediated sponge interactions*) et cible plus particulièrement la famille mir-200. Ce lncRNA semble impliqué dans le développement de la glande mammaire et le maintien de son homéostasie en préservant l'adhésion inter-cellulaire, la migration, la prolifération et l'apoptose (Cui et al., 2016).

Une étude relativement récente (Askarian-Amiri et al., 2011) a identifié près de 100 lncRNAs exprimés de manière différentielle entre les souris gestantes, allaitantes et en involution post-partum.

2.3.6.2.2 LncRNAs contrôlant la différenciation et l'homéostasie mammaire

Deux lncRNAs sont impliqués dans la différenciation des cellules épithéliales mammaires : *mPINC* et *Zfas1*. Ces deux lncRNAs ont été identifiés par des analyses à grande échelle visant à isoler des gènes différenciellement exprimés pendant le développement mammaire post-pubertaire (Askarian-Amiri et al., 2011 ; Shore and Rosen, 2014). Les premières études *in vitro* ont révélé que l'inactivation de *mPINC* ou de *Zfas1* dans les lignées de cellules mammaires gestantes HC11, conduit à une prolifération accrue, suggérant un rôle suppresseur de croissance de ces 2 lncRNAs (Ginger et al., 2006). Toutefois, d'autres études suggèrent qu'ils jouent un rôle plus spécialisé dans la glande mammaire. *MPINC* et *Zfas1* présentent tous deux une expression spatio-temporelle stricte au cours du développement de la glande mammaire post-pubertaire (Askarian-Amiri et al., 2011 ; Shore et al., 2012). Leurs niveaux d'expression augmentent considérablement pendant la grossesse et l'hybridation *in situ* de glandes mammaires de souris à mi-grossesse révèle que *mPINC* est enrichi dans les cellules alvéolaires, alors que *Zfas1* se localise préférentiellement dans les cellules canalaire et alvéolaires. Les analyses *in vitro* conduites à partir de lignées murines gestantes suggèrent que des niveaux élevés de *mPINC* et *Zfas1* répriment l'étape finale d'alvéologénèse et assurent le maintien d'un pool de progéniteurs alvéolaires. A l'inverse, la réduction importante de leurs niveaux d'expression en début de lactation semble nécessaire à la différenciation alvéolaire terminale. Au début de l'involution, les niveaux d'expression de *mPINC* et de *Zfas1* augmentent à nouveau, permettant la préservation de progéniteurs alvéolaires pour les grossesses ultérieures. La répression de la différenciation alvéolaire par *mPINC* semble régulée épigénétiquement. Les données *in vitro* suggèrent que le complexe PRC2 comprenant *mPINC* et la protéine RbAp46 réprime les gènes nécessaires à la différenciation alvéolaire et maintient ainsi une population de cellules progénitrices alvéolaires (Zhang et al., 2006).

3 LES LONGS ARNs NON CODANTS ET LE CANCER

3.1 Généralités

La plupart des maladies, dont le cancer, sont associées à une altération de leur profil transcriptomique. Cette altération n'est pas limitée à la production de niveaux aberrants d'ARNs codant pour les protéines, mais aussi à la dérégulation de l'expression de nombreux ARNs non codants, dont les miRNAs et les lncRNAs.

En raison de l'implication des lncRNAs dans divers processus physiologiques fondamentaux, leur dysfonctionnement devrait avoir des conséquences néfastes pour le maintien de l'homéostasie cellulaire. La dérégulation des lncRNAs est observée dans divers états pathologiques : maladies auto-immunes, cardiovasculaires, infectieuses, neurodégénératives et les cancers. Plusieurs études récentes ont montré que l'expression de nombreux lncRNAs varie considérablement dans des conditions pathologiques, telles que le cancer par rapport aux tissus sains (Kitagawa et al., 2012 ; di Gesualdo et al., 2014 ; Haemmerle and Gutschner, 2015). Les recherches ayant porté sur la transcription du génome non codant ont ainsi révélé une nouvelle dimension de l'architecture moléculaire du cancer (Chen et al., 2014c ; Malek et al., 2014 ; Sultmann and Diederichs, 2014). L'analyse des profils d'expression de différentes tumeurs et de leurs tissus normaux correspondants ainsi que des études fonctionnelles, ont validé l'expression tissulaire plus spécifique des lncRNAs et confirmé que ces derniers sont dérégulés dans un type particulier de tumeur ou dans de multiples types tumoraux. Ainsi, une étude récente globale concernant 1300 tumeurs appartenant à 4 types tumoraux différents a identifié 80 à 300 lncRNAs *drivers* impliqués dans la progression des types tumoraux (Du et al., 2013). En outre, des études pangénomiques consacrées à divers types de cancers, dont les tumeurs mammaires et hépatiques, ont identifié des mutations dans leurs séquences primaires (Ellis et al., 2012 ; Nik-Zainal et al., 2016 ; Fushimoto et al., 2016). Comme la plupart des variants génétiques identifiés par les études d'association pangénomique (GWAS) sont situées en dehors des gènes codants, certains de ces variants pourraient affecter des lncRNAs (Huarte and Rinn, 2010 ; Gibb et al., 2011 ; Prensner and Chinnalyan, 2011 ; Gutschner and Diederichs, 2012 ; Nie et al., 2012 ; Spizzo et al., 2012 ; Qiu et al., 2013).

Le cancer est caractérisé par des altérations génétiques et épigénétiques survenant dans divers processus qui régulent l'homéostasie cellulaire, tels que le cycle cellulaire, la réparation de l'ADN, la survie et la migration. Des lncRNAs dérégulés qui interviennent habituellement dans divers mécanismes homéostatiques tels l'apoptose, la stabilité génomique, le cycle cellulaire, la réparation de l'ADN, l'angiogénèse et la migration, ont été identifiés dans le cancer (Cheetham et al., 2013). Les lncRNAs *PRNCRI* et *PCGEMI* favorisent la prolifération cellulaire, *PANDA* induit la réparation de l'ADN et la survie et *MALATI* promeut la migration cellulaire et l'angiogénèse. Ainsi, les lncRNAs régulent les principales voies signalétiques impliquées dans l'initiation, la progression tumorale et la dissémination métastatique de façon similaire aux oncogènes et aux gènes oncosuppresseurs codant pour des protéines (Qi and Du, 2013). *PINC* et *PGEMI* ont été les premiers lncRNAs oncogènes à être découverts surexprimés respectivement dans les cancers du sein et de la prostate. Depuis, de nombreux autres lncRNAs ont été identifiés, avec des propriétés oncogéniques (*KRASP*, *HULC*, *HOTAIR*, *MALATI*, *HOTTIP*, *ANRIL* et *RICTOR*) ou des propriétés oncosuppressives (*MEG3*, *GAS5*, *lincRNA-p21*, *PTENP1*, *TERRA*, *CCND1/CyclinD1* et *TUG1*). Plusieurs lncRNAs peuvent présenter à la fois des activités oncogéniques et oncosuppressives selon le contexte cellulaire. De plus *Xist* est surexprimé dans les tumeurs de sexe masculin et sous-exprimé dans les tumeurs de sexe féminin.

3.2 LncRNAs et mécanismes de la cancérogénèse

3.2.1 LncRNAs et génétique du cancer

Le cancer est une maladie génétique résultant d'une dérégulation des gènes contrôlant l'identité, la différenciation et la croissance normales des cellules. Alors qu'une faible proportion de cancers est attribuée à des mutations monogéniques héréditaires de gènes codant pour des protéines, tels *BRCA1* dans le cancer du sein familial (Miki et al., 1994) et *RBI* dans le rétinoblastome familial (Du and Pogoriler, 2006), la majorité des cancers surviennent par l'intermédiaire de mutations somatiques. La prédisposition à développer un cancer sporadique est une donnée complexe à appréhender, comprenant divers facteurs génétiques et environnementaux. Sa composante héréditaire dépend du type de cancer, mais apparaît souvent considérable (sein : 56%, thyroïde : 53%, prostate : 42%, ovaire : 40%, testicule : 25%, col : 22%, colon : 13%, pancréas : 10%, poumon : 8%, leucémie : 1%, estomac : 1%) (Cariaso and Lennon, 2012). Les analyses intensives ayant concerné les

mécanismes génétiques de prédisposition au cancer ont révélé qu'une grande partie de la susceptibilité au cancer ne peut être attribuée à la seule variation des séquences codant pour des protéines. Les études GWAS ont confirmé que la grande majorité des *loci* à risque de cancer (80% des 301 SNPs associés au cancer) sont situés à l'extérieur des régions codant pour des protéines. La plupart des SNPs sont en effet situés dans les introns de gènes codant pour des protéines (40%) ou dans les régions intergéniques (44%), alors que seuls 3% modifient la séquence des acides aminés constituant les protéines. Des études plus récentes ont révélé que les *loci* à risque de cancer sont transcrits en lncRNAs et que les SNPs associés au cancer sont impliqués dans la cancérogénèse *via* des mécanismes modifiant en particulier l'expression de lncRNAs suppresseurs de tumeurs. *PTCSC3*, un lncRNA spécifique de la thyroïde, apparaît fortement associé au SNP rs944289 situé dans la région 14q13.3 et est sous-exprimé dans le carcinome papillaire de la thyroïde. *PTCSC3* est un lncRNA onco-suppresseur contrôlant l'expression de gènes impliqués dans la replication et la réparation de l'ADN, l'apoptose et la migration cellulaire. L'allèle à risque modifie la liaison de la protéine C/EBP avec le promoteur du gène *PTCSC3*, réduisant ainsi l'expression du lncRNA (Jendrzewski et al., 2012). *ANRIL* est situé dans un *hotspot* de SNPs associé à de nombreuses maladies complexes (Pasmant et al., 2011a). Il interagit avec les complexes Polycomb PRC1/PRC2 afin d'apposer des marques histoniques suppressives au niveau du locus inhibiteur de la prolifération cellulaire *p15-CDKN2B/p16-CDKN2A/p14-ARF*. Les polymorphismes situés dans les différentes régions de l'ARN d'*ANRIL* peuvent modifier les interactions ARN-ADN ou ARN-protéine nécessaires à l'inactivation du locus. Il est probable que de nombreux autres lncRNAs sont transcrits à partir de loci associés aux cancers. De plus, la combinaison de l'étude GWAS et de l'analyse de l'ARN par capture-Seq appliqué aux « déserts géniques », a confirmé que de nombreux loci non codants à risque de cancers identifiés par l'étude GWAS sont transcrits en lncRNAs présentant d'importantes fonctions régulatrices dans la biologie du cancer. Ainsi, le lncRNA *PRNCRI* a été localisé dans la région désertique comportant le locus 8q24 lié au cancer de la prostate (Chung et al., 2011).

3.2.2 LncRNAs et épigénétique du cancer

3.2.2.1 Généralités sur l'épigénétique du cancer

L'épigénétique se réfère à des événements réversibles qui modulent l'activité du génome sans changer sa séquence. Les différents acteurs de l'épigénétique participent au contrôle de l'expression globale du génome et à l'établissement de profils d'expression géniques spécifiques du type cellulaire et tissulaire. Les principaux mécanismes épigénétiques comprennent des modifications chimiques de l'ADN et des histones (méthylation des cytosines, modifications post-traductionnelles des histones, remodelage de la chromatine et positionnement des nucléosomes). Ces modifications chimiques sont réversibles et contrôlées tant par des complexes enzymatiques directement connectés aux voies métaboliques et de signalisation que par des senseurs appartenant aux microenvironnements intra et extracellulaires. La dérégulation de ces mécanismes épigénétiques a été observée dans des cellules cancéreuses présentant des altérations de la transcription génique ou de l'expression protéique bien que dépourvues de lésions de l'ADN. Les lncRNAs sont impliqués dans l'épigénétique du cancer (Morris, 2009 ; Saxena and Carninci, 2011 ; Lee, 2012 ; Mercer and Mattick, 2013 ; Timp and Feinberg, 2013).

3.2.2.2 Les lncRNAs comme régulateurs épigénétiques

Méthylation de l'ADN: L'épigénome des cellules cancéreuses se caractérise à la fois par une hypométhylation globale conduisant à une instabilité génomique et par une hyperméthylation des îlots CpG provoquant l'inactivation des gènes oncosuppresseurs impliqués dans les voies de signalisation (réparation de l'ADN, apoptose, cycle cellulaire) ou de l'inactivation de facteurs de transcription impliqués dans le contrôle de ces gènes (Feinberg and Volgenstein, 1983; Wang and Leung, 2004; Mohammad et al., 2012). Les lncRNAs se lient à l'ADN méthyltransférase, la guident vers les promoteurs de gènes oncosuppresseurs et assurent leur répression transcriptionnelle (*AS1DHR54*) (Lee et al., 2012).

Remodelage des nucléosomes : Des mutations et une inactivation de sous-unités de ces complexes ont été observés dans divers types de cancers. De plus, des lncRNAs guidant le complexe de remodelage SWI/SNF ont été récemment découverts et pourraient repositionner les nucléosomes en interagissant et en guidant ces complexes vers des régions génomiques spécifiques où la restructuration des nucléosomes assurerait la répression des gènes oncosupresseurs (Zhu et al., 2013).

Modification des histones : Les lncRNAs recrutent des enzymes modifiant les histones qui agissent à proximité du site de transcription des lncRNAs. Les enzymes modifiant les histones ajoutent ou suppriment des liens covalents avec pour conséquences un changement dans l'accessibilité de la chromatine et la fixation d'effecteurs protéiques non histones qui lisent le code ainsi modifié (Kouzarides, 2007). Des altérations de l'expression des enzymes modifiant les histones (mutations, surexpression) ont été observées dans différents types de carcinomes (Ropero et al., 2006). Ces altérations inhibent la transcription des gènes oncosuppresseurs (histone méthyltransférase EZH2 responsable de la marque répressive H3K27me3 ; histone méthyltransférase G9a induisant la marque répressive H3K9me3 ; histone déacétylase assurant la perte de la lysine acétylée H4) ou activent la transcription d'oncogènes (déméthylases et désacétylases) (Kondo et al., 2008).

3.2.3 lncRNAs et *hallmarks* du cancer

Un nombre croissant de lncRNAs sont impliqués dans les mécanismes caractéristiques du cancer. Ils peuvent agir sur la prolifération par co-activation (*SRAI*) et inhibition de l'élongation (*RN7SK*). Les lncRNAs permettent l'échappement aux voies de signalisation oncosuppressives *via* des mécanismes de compétition (*PFS*), d'inactivation (*ANRIL*) et de répression ribonucléique (*GAS5*). Plusieurs lncRNAs sont impliqués dans l'immortalité répllicative en inhibant la télomérase (*TERRA*). Certains lncRNAs favorisent la néoangiogénèse en inhibant HIF1A (*AHIF*, *MALATI*). Ils induisent une résistance à la mort cellulaire en agissant sur p53 et p21 (*PCGEM1*) et en réduisant l'expression de gènes pro-apoptotiques (*PANDA*). Ils peuvent enfin être régulés par des facteurs de transcription (p53, B-MYB, c-Myc, STAT3).

De nombreux lncRNAs présentent des effets pro-oncogéniques et une expression majorée dans un type tumoral spécifique ou dans divers cancers. Ainsi, **parmi les lncRNAs intergéniques**, *MALATI* promeut la prolifération, la migration et l'invasion dans de nombreux cancers (Ji et al., 2003; Xu et al., 2011; Lai et al., 2012; Gutschner et al., 2013b); *HULC* induit la prolifération médiée par HBV et la croissance indépendante de l'ancrage des hépatocarcinomes (Panzitt et al., 2007); *H19* favorise la croissance après survenue d'une hypoxie, inhibe l'apoptose et induit l'EMT dans divers cancers (Raveh et al., 2015); *HOTAIR* supprime l'expression de nombreux oncosuppresseurs *via* l'induction d'une méthylation de certaines histones (Yang et al., 2011c; Kim et al., 2012; Li et al., 2013b; Nie et al., 2013);

PGEMI induit la prolifération cellulaire et la formation de tumeurs *via* la phosphorylation de Rb dépendante de la voie androgénique dans les cancers de prostate (Petrovics et al., 2004). Parmi les lncRNAs antisens (NATs), *ANRIL* favorise la croissance cellulaire en réprimant l'expression des gènes du *cluster p16-CDKN2A/p15-CDKN2B/p14-ARF* dans divers cancers et hémopathies (Yap et al., 2010) ; *ZEB2-NAT* altère l'épissage alternatif et induit la transcription de *ZEB2* (Beltran et al., 2008); *aHIF* induit la dégradation de HIF1 α et réduit la stabilité de p53 durant l'hypoxie dans les cancers du rein et du sein (Span et al., 2011). Parmi les régions ultra-conservées transcrites, *uc.73* est associé à un pronostic défavorable dans les carcinomes colo-rectaux (Sana et al., 2012) ; *TUC 338* favorise la croissance cellulaire des hépatocarcinomes (Bracconi et al., 2011) ; *uc.460* est activé par l'amplification de *NMyc* et *t-uc.300a* induit la répression d'*INPP5D* dans les neuroblastomes (Mestdagh et al., 2010 ; Watters et al., 2013). Parmi les pseudogènes, *ATP8A2- ψ* augmente l'expression de la protéine ATP8A2 dans les cancers de prostate (Kalyana-Sundaram et al., 2012) ; *Oct4-pg* agit comme un leurre pour les miRNAs et promeut l'expression d'OCT4 dans les cancers du sein, les gliomes et les hémopathies (Guo and Tang, 2012) ; *PTENP1* fonctionne également comme un leurre pour les miRNAs ciblant *PTEN* dans les carcinomes colo-rectaux (Poliseno et al., 2010).

De plus rares lncRNAs présentent une activité oncosuppressive dans les cancers et pourraient être potentiellement contrôlés par des thérapies géniques (*GAS5*, *MEG3*, *linc-p21*).

3.2.4 LncRNAs et voies de signalisation du cancer

Les lncRNAs jouent un rôle majeur dans l'activation des voies de signalisation oncogéniques qui édifient le phénotype cancéreux (Schmitt and Chang, 2016). Ils sont impliqués dans la régulation de la voie MYC (*PVT1* dans les cancers gastriques et mammaires et les hémopathies; lncRNA *XLOC_010588* dans les cancers de vessie) et p53 (*lincRNA-p21* dans les carcinomes pulmonaires; *MALATI* dans divers types tumoraux), des voies PI3K/Akt/mTOR et MAPK (*ANRIL* dans les carcinomes gastriques, *lncRNARMEL3* dans les mélanomes), de la voie NF κ B (*BANCR* dans les carcinomes gastriques), des sphingosines kinases (*HULC* dans les hépatocarcinomes), de l'EMT et de la migration cellulaire (*LncRNA-ATB* dans les carcinomes hépatocellulaires, *MALATI* dans les carcinomes de vessie, *MVIH* dans les carcinomes pulmonaires, *HOTAIR* dans les carcinomes coliques), la

signalisation hormonale (*NEATI* et *PRNCRI* dans les carcinomes prostatiques) et la voie de l'angiogénèse (*HOTTIP* dans les carcinomes hépatocellulaires et *TUGI* dans les carcinomes pulmonaires) (Flippot et al., 2016).

3.2.5 LncRNAs, progression tumorale et évolution métastatique

L'évolution métastatique est un processus multi-étapes comportant habituellement: (i) l'invasion locale par migration collective ou EMT et l'intravasation, (ii) la survie des cellules tumorales (CTC) présentant surtout un phénotype mixte épithélial et mésenchymateux dans la circulation sanguine ou lymphatique, (iii) l'arrêt des CTCs dans les petits vaisseaux d'un organe secondaire, leur transition mésenchymato-épithéliale (MET) et leur extravasation, (iv) la survie des cellules tumorales au contact de niches prémétastatiques induites par la production d'exosomes et (v) le développement et la croissance de micrométastases et de macrométastases (Mehlen et al., 2006; Chaffer and Weinberg, 2011; Spano et al., 2012).

Il est actuellement bien établi que les lncRNAs possèdent des fonctions pro- et anti-métastatiques *via* la survie cellulaire et l'inhibition de l'apoptose, l'hypoxie et la plasticité, le contrôle des voies de signalisation Wnt, TGF- β , Notch, PI3K/Akt-mTOR, JNK, Hedgehog, l'adhésion cellulaire, la migration et la dormance transitoire (Tahira et al., 2011 ; Bhartiya et al., 2013 ; Crea et al., 2014a ; Serviss et al., 2014) (**Tableau 5**). Les mécanismes pro-métastatiques utilisés par les lncRNAs altèrent (i) la régulation épigénétique : *HOTAIR* (Gupta et al., 2010), *PTENP1-AS* (Hesson et al., 2012), *ANRIL* (Congrains et al., 2013), (ii) la régulation transcriptionnelle : *H19* (Matouk et al., 2014), *CT13* (Cruickshanks et al., 2013), (iii) la régulation post-transcriptionnelle : *PTENP1-AS* (Wang et al., 2016), *PCAT1* (Ge et al., 2013), *MALAT1* (Gutschner et al., 2013a,b), *ZEB2 NAT* (Beltran et al., 2008), (iv) la compétition endogène intracytoplasmique entre les divers types de RNAs : *HOTAIR*, *H19*, *EFNA3*, *CCAT2* (Ling et al., 2013; Gómez-Maldonado et al., 2015), (v) la dégradation protéosomale : *LET* (Yang et al., 2013d), (vi) la sécrétion des protéiques : *MVIH* (Yuan et al., 2012), (vii) la composition du cytosquelette : *Dreh* (Huang et al., 2013) et la l'angiogénèse (*ANRIL* et *MALAT1*).

3.2.6 LncRNAs et interactions avec les différentes classes d'ARNs

Les lncRNAs interagissent avec de nombreux mRNAs, miRNAs et lncRNAs impliqués dans les mécanismes de la cancérogénèse. Certains lncRNAs régulent des miRNAs, tels *ANRIL* qui inhibe miR99A et miR449A et induit l'expression de mTOR et d'E2F1 (Zhang et al., 2014d). A l'opposé, des miRNAs peuvent réguler l'activité des lncRNAs, tels miRNA 141 qui inhibe l'expression de *HOTAIR* (Flippot et al., 2016). Les lncRNAs peuvent interagir entre eux dans le cadre de processus régulateurs très complexes (*PTENP1* et *PTENP1-AS* dans la régulation de *PTEN*). Les lncRNAs, qui représentent chez l'homme 25% de l'ARN total cellulaire et 70% des transcrits, constituent un réseau d'interactions pouvant jouer un rôle majeur dans l'acquisition du phénotype cancéreux (Iyer et al., 2015).

3.2.7 LncRNAs, sensibilité et résistance aux thérapies anti-cancéreuses

De nombreux lncRNAs participent aux mécanismes de résistance aux thérapies. *NEATI*, *PRNCRI* et *PCGEMI* activent la voie du récepteur androgénique indépendamment du ligand et la prolifération dans les carcinomes prostatiques (Yang et al., 2013a). *UCA-1* accroît la chimiorésistance des carcinomes vésicaux au cisplatine en régulant la voie Wnt (Fan et al., 2014). *HOTAIR* induit une résistance au carboplatine dans les carcinomes ovariens (Teschendorff et al., 2015). D'autres lncRNAs pourraient représenter des biomarqueurs de réponse aux traitements, tels *CCAT-1* dans les carcinomes coliques, dont la transcription est corrélée à une sensibilité accrue aux inhibiteurs des bromodomaines *in vitro* (McClelland et al., 2016).

lncRNA	Description	Niveau d'expression tumoral	Corrélation avec la survenue de métastases	Topographie cellulaire	Fonctions dans les cellules tumorales	Fonctions et mécanismes d'action moléculaires	dbles moléculaires	Localisations tumorales	Références
ANRIL/CDKN2B-AS	Antisense nc.RNA in the INK4 locus	Haut	Positive	Noyau	Prolifération (+) Migration (+) Invasion (+) Apoptose (-)	Modification des histones PRC1/PRC2 Régulation transcriptionnelle	LOXUS CDKN2-B/CDKN2A/ARF	Sein, ovaire, poumon, foie	Aguiló et al. (2016)
BIWCE	BRAF catalytic domain nc.RNA	Haut Bas	Positive	Noyau cytoplasme	Prolifération (+) Migration (+) EMT (+) Métastases (+)		NF-κB miR-9 Voie MAPK	Melanome, colon, estomac, poumon, thyroïde, oesophage, rétinoblastome	Liu et al. (2016a)
SCAR4	Breast cancer c. ti-estrogen resistance 4	Haut	Positive	Noyau	Prolifération (+) Migration (+) Métastases (+)	Modification des histones Régulation transcriptionnelle	SNP1, PNUMS Voie Hedgehog/GU2	Sein	Xing et al. (2014)
BCYRNJ	Brcin cyto. Jasmic RNA 1	Haut	Positive	Cytoplasme	Migration (+) Invasion (+) Métastases (+)	abaissement de c-Myc Régulation transcriptionnelle	MMP9 MMP13	Poumon	Iarangi et al. (2013)
CCAT1	Colon cancer-associated transcript 1	Haut	Positive	Noyau	Prolifération (+) Migration (+) Invasion (+)	Compétition endogène Co-répression de HMGA2 and c-Myc	let-7	Foie, estomac, coloco	Deng et al. (2015)
CCAT2	Colon cancer-associated transcript 2	Haut	Positive	Noyau	Prolifération (+) Migration (+) Invasion (+) Métastases (+)	Enhancer RNA Régulation transcriptionnelle Activation de la voie Wnt (Myc)	TCE712	Sein, oesophage, estomac, poumon,	Ling et al. (2013)
DRAIC	Downregulated RNA in cancer	Bas	Négative	Noyau Cytoplasme	Transformation (-) Migration (-) Invasion (-)	Activation aberrante du RA	FOXA1, NKX3-1	Prostate	Sakurai et al. (2015)
hDreh	Dau4-regulated expression by HBx	Bas	Négative	Cytoplasme	EMT (-) Invasion (-) Métastases (-)	Modification du cytosquelette	Altération de la vimentine	foie	Huang et al. (2013)
EFNA3	Ephrin-A3 (lqZ1-qll)	Haut	Positive	Cytoplasme	Migration cellulaire Angiogénèse Prolifération (+) Invasion (+)	Compétition endogène Traduction accrue d'EFNA3 et de rEphrineA3	miRNAs inhibent la traduction de la protéine EFNA3 GHR EGFR	Sein foie	Gomez-Maldonado et al. (2015)
EGFR-ASJ	EGF receptor-mt sense 1	Haut	Positive	Cytoplasme	Prolifération (+) Invasion (+)	Activation de EGFR		foie	Qi et al. (2016)
EGOT	Eosinophil granulocyte ontogeny transcript 1	Bas	Négative	cytoplasme	Prolifération (+) Invasion (-)	Oblédé HCV	Antagonise la réponse antivirale	Sein	Camero et al. (2016)
fpist	Esophagus epithelial intergenic specific transcript	Bas	Négative		Migration (-) Invasion (-)			Oesophage	Wejtal. (2015)
GAS5	Growth arrest-specific lincRNA	Bas	Négative	Noyau cytoplasme	Prolifération (+) Migration (+) Invasion (+)	Séquestration de FT Riborépression transcriptionnelle via les récepteurs des glucocorticoïdes	Inhibition de la prolifération induite par le stress et l'apoptose CD44 SNA12 PFSJNONO PRC2	Foie, col, cerveau, sein	Sun et al. (2014a)
GAPUNC	Gastric adenocarcinoma predictive lincRNA	Haut	Positive	cytoplasme	Prolifération (+) Migration (+) Invasion (+)	Compétition endogène		Estomac, colon	Wang et al. (2016)
HJ9	HJ9	Haut	Positive	Noyau Cytoplasme	EMT (+) Métastases (+)	Méthylation de rAON Régulation transcriptionnelle Biogénèse de miRNAs Compétition endogène Inactivation traductionnelle	let-7, miR-138, miR200a IGF2l1, Rb1, TP53, TGF6, Smad1/5	Sein, vessi foie cerveau, estomac, colon, poumon	Matouk et al. (2014)
HOTAIR	HOX transcript antisense	Haut	Positive	Noyau cytoplasme	Prolifération (+) Migration (+) Invasion (+) Métastases (+) Apoptose (-)	Remodelage chromatinien Modification des histones PRC2/LSD1 Compétition endogène miR331-3p Contrôle de la stabilité protéique	locus HOXD HER2	Sein, colon, poumon, tissus mous	Gupta et al. (2010)

<i>HOTTIP</i>	<i>HOXA transcript at the distal tip</i>	Haut	Positive	Noyau	Prolifération (+) Migration (+) Invasion (+) Métastases (+) Apoptose (-)	Remodelage chromatinien Modification des histones WDR5/MLL	HOXA13 Voie Wnt-β caténine	Foie, pancréas, langue	Sang et al. (2016)
<i>HULC</i>	<i>Highly upregulated in liver cancer</i>	Haut	Positive	Noyau cytoplasme	Prolifération (+) Migration (+) EMT (+) Invasion (+) Métastases (+) Apoptose (-)	Compétition endogène	Voie miR-200a/3p- ZEB1 TNFα Voie PI3K/Akt/mTOR	Foie, colon, estomac, pancréas, os	Xie et al. (2013)
<i>LCT13</i>	<i>LINE-1 chimeric transcripts 13</i>	Haut	-	Noyau	Métastases (-)	Inactivation épigénétique de la protéine anti-métastatique TFPI2 Régulation transcriptionnelle	TFPI2	Sein, colon	Cruickshanks et al. (2013)
<i>LET</i>	<i>LncRNA-low expression in tumor</i>	Bas	Négative	Cytoplasme	Invasion (-) Métastases (-)	Régulation transcriptionnelle Stabilité protéique Dégradation protéosomale Stabilisation de HIF1-α	NF90	Foie, colon, poumon, col, vessie, vésicule biliaire	Liu et al. (2016c)
<i>Linc00461/VISC2</i>	<i>long intergenic non-protein coding RNA 461</i>	-	-	Cytoplasme	Prolifération (-) Migration (+)	Précurseur de miRNA-9	CREB		Oliver et al. (2015)
<i>LincRNA-ROR</i>	<i>LincRNA-Regulator of reprogramming</i>	Haut	Positive	Noyau Cytoplasme	EMT (+) Invasion (+) Métastases (+)	Régulation transcriptionnelle Compétition endogène	miR-145 ZEB1, Nanog, Oct4, Sox2 p53, p27	Sein	Zhou et al. (2016)
<i>lncRNA-ATB</i>	<i>lncRNA-activated by TGF-β</i>	Haut	Positive	Noyau Cytoplasme	EMT (+) Invasion (+) Métastases (+)	Compétition endogène Stabilité des mRNAs	ZEB1 miR-200a E-cadhérine	Foie, colon, estomac, rein, prostate, sein	Shi et al. (2015b)
<i>LncRNA-HIT</i>	<i>HOXA transcript induced by TGF-β</i>	Haut	Positive	Noyau Cytoplasme	Prolifération (+) EMT (+) Migration (+) Invasion (+) Métastases (+)	-	E-cadhérine TGF-β	Sein	Richards et al. (2015)
<i>lncTCF7</i>	<i>lncRNA T-cell factor-7</i>	Haut	Positive	Noyau	EMT (+) Invasion (+)	Remodelage chromatinien SWI/SNF Régulation transcriptionnelle	Cible de IL-6/STAT3 EMT	Foie	Wu et al. (2015)
<i>MALAT1</i>	<i>Metastasis-associated lung adenocarcinoma transcript 1</i>	Haut	Positive	Noyau Cytoplasme	Prolifération (+) EMT (+) Migration (+) Invasion (+) Métastases (+)	Remodelage chromatinien Epissage alternatif Compétition endogène Activation de gènes impliqués dans l'EMT et l'évolution métastatique	CAMK2B, HMG2L1, ARHGEF1, B-MYB	Poumon, prostate, sein, rein, col, pancréas, oesophage,	Gutschner et al. (2013a)
<i>MEG3</i>	<i>Maternally expressed gene 3</i>	Bas	Négative	Noyau Cytoplasme	Prolifération (-) Migration (-) Invasion (-) Apoptose (+)	Remodelage chromatinien Compétition endogène	Voie Notch c-Myc p53 miR-1297	Estomac, poumon, thyroïde, colon, hypophyse	Guo et al. (2016)
<i>MVIH</i>	<i>Microvascular invasion in hepatocellular carcinoma</i>	Haut	Positive	Cytoplasme	Prolifération (+) Angiogénèse (+) Métastases (+)	Secrétion protéique	PGK1 miR-199a	Sein,foie, poumon	Lei et al. (2016)
<i>NBAT-1</i>	<i>Neuroblastoma associated transcript-1</i>	Bas	Négative	Noyau	Prolifération (-) Migration (-) Invasion (-)	Remodelage chromatinien	NRSF/REST	Sein, rein, neuroblastome	Xue et al. (2015)
<i>NEAT1</i>	<i>Nuclear enriched abundant transcript 1</i>	Haut	Positive	Noyau	Prolifération (+) Migration (+) Invasion (+)	Biogénèse des <i>paraspeckles</i>	ABCG2, miR- 204/ZEB1 miR-377-3p-E2F3, FUS miR-548, miR- 449b	Poumon, foie, cerveau, colon, prostate, vessie	Zeng et al. (2014)

<i>NKILA</i>	<i>NF-kB interacting lncRNA</i>	Bas	Négative	Cytoplasme	Métastases (-)	Régulation transcriptionnelle Stabilité protéique	IkB/NF-kB	Sein	Liu et al. (2015b)
<i>PACER</i>	<i>P50-associated COX-2 extragenic RNA</i>	Haut	-		Prolifération (+) Migration (+)	Régulation transcriptionnelle	COX-2	Os	Qian et al. (2016)
<i>PCAT1</i>	<i>Prostate cancer associated transcript 1</i>	Haut	-	Noyau	Prolifération (+) Migration (+)	Défaut de recombinaison homologue Augmentation des cassures double-brin Répression post-transcriptionnelle	BRCA2	Œsophage, foie, prostate	Wen et al. (2016)
<i>PCAT29</i>	<i>Prostate cancer-associated transcript 29</i>	Bas	Négative	-	Prolifération (-) Migration (-) Invasion (-) Métastases (-)	lncRNA réprimé par le récepteur aux androgènes	-	Prostate	Malik et al. (2014)
<i>PTENP1-AS</i>	<i>PTENP1 antisense RNA</i>	-	-	Noyau Cytoplasme	-	Régulation épigénétique PRC2 Répression post-transcriptionnelle	<i>PTENP1</i>		Deng and Suil. (2013)
<i>SCHLAP1</i>	<i>Second chromosome locus associated with prostate 1</i>	Haut	Positive	Noyau	Invasion (+) Métastases (+)	Remodelage chromatinien SWI/SNF	SNF5	Prostate, vessie	Atala. (2014)
<i>SPRY4-IT1</i>	<i>SPRY4 Intronic Transcript 1</i>	-	-	Noyau	Prolifération (+) Migration (+) Invasion (+)	Régulation de l'expression de la cadhérine-E et de la vimentine (EMT)	PRC2 ZNF703 SNAI1	Estomac, cerveau,	Shen et al. (2016)
<i>treRNA</i>	<i>Translational regulatory lncRNA</i>	Haut	Positive	Noyau Cytoplasme	Migration (+) Invasion (+) Métastases (+)	Enhancer RNA Contrôle traductionnel	E-cadhérine	Sein	Gumireddy et al. (2013)
<i>UCA1</i>	<i>Urothelial cancer associated-1</i>	Haut	Positive	Cytoplasme	Prolifération (+) Migration (+) Invasion (+)	Contrôle traductionnel	Voie Wnt-β caténine miR-507-FOXO1 miR-204-5p p27Kip1/CDK2	Vessie, langue, sein, peau, foie, oesophage, estomac	Li et al. (2014b)
<i>ZEB2-AS1 (ZEB2-NAT)</i>	<i>ZEB2- antisense RNA 1</i>	Haut	-	Noyau	EMT (+) Migration (+) Invasion (+)	Epissage alternatif Blocage du site de fixation ribosomal du mRNA de ZEB2	ZEB2	Colon, vessie	Zhuang et al. (2015)

Tableau 5 : Rôles biologiques et mécanismes d'action des principaux lncRNAs dans l'EMT, l'invasion et l'évolution métastatique (Adapté d'après Crea et al., 2014 ; Dhamija and Diederichs, 2016). Pages 123-124-125.

4 LONGS ARNs NON CODANTS ET CANCER DU SEIN

4.1 Dérégulation des lncRNAs dans les cancers du sein

4.1.1 Dérégulation globale des lncRNAs dans les carcinomes mammaires

Les lncRNAs peuvent réguler l'expression des gènes à de multiples niveaux par des mécanismes très variés, et une expression aberrante de ces lncRNAs peut contribuer significativement à l'initiation et à la progression des cancers du sein (**Tableau 6**).

Les oestrogènes, la progestérone et leurs récepteurs jouent un rôle crucial dans la progression du cancer du sein (Briskin, 2013). De nombreuses études utilisant les techniques de *global run-on sequencing* (GRO-seq) et de RNA-seq ont révélé que 25% du génome est régulé au niveau transcriptionnel par les récepteurs aux oestrogènes et ont identifié des milliers de lncRNAs régulés par les œstrogènes dans les cellules carcinomateuses mammaires (Hah et al., 2011 ; Hah et al., 2013a ; Li et al., 2013a ; Hah and Kraus, 2014a). Ces lncRNAs appartiennent à 2 groupes, l'un constitué d'eRNAs transcrits à partir d'*enhancers* adjacents à des gènes activés par les oestrogènes (Li et al., 2013a) et l'autre comportant des lncRNAs qui interagissent avec les récepteurs hormonaux et modulent leur activité. Les analyses à grande échelle ont également révélé l'expression aberrante de nombreux lncRNAs dans le cancer du sein. Des études ont permis d'identifier 220 lncRNAs anormalement exprimés dans les carcinomes mammaires par rapport au tissu mammaire normal et des centaines de lncRNAs spécifiquement dérégulés dans les tumeurs du sein par rapport à d'autres types de tumeurs, pouvant ainsi servir de biomarqueurs des tumeurs du sein (Gibb et al., 2011). Parmi ceux-ci, *HOTAIR* est surexprimé au niveau transcriptionnel par les oestrogènes dans les lignées de cancers du sein (Bhan et al., 2013).

Pour la description des principaux lncRNAs impliqués dans la carcinogénèse mammaire, nous prendrons uniquement en compte les lncRNAs oncogènes et oncosuppresseurs pour lesquels il existe des preuves de leur dérégulation dans le cancer du sein et/ou des preuves d'une fonction ou d'un mécanisme tumorigénique potentiel.

LncRNA	Chromosome	Expression aberrante	Localisation génomique	Oncogène Oncosuppresseur	Fonctions biologiques	Mécanismes d'action dans la cancérogénèse	Références
HIF-1 α -AS1	14q23.2	surexpression	antisens chevauchant	oncogène	régulation de HIF-1 α	induction de la prolifération inhibition de l'apoptose	Gutschner and Diederichs. (2012)
ANRASSF1	3p21.3	surexpression	antisens chevauchant	oncogène	régulation du suppresseur RASSF1 par recrutement de PRC2	inhibition de RASSF1 induction de la prolifération	Beckedorff et al. (2013)
ANRIL	9p21.3	surexpression	antisens chevauchant	oncogène	régulation du locus p16-CDKN2A/p15-CDKN2B/p14-ARF	locus p16-CDKN2A/p15-CDKN2B/p14-ARF instabilité génomique	Yap et al. (2010)
ARA	Xq23	surexpression	intronique	oncogène	régulation de la voie MAPK, du cycle cellulaire et de l'adhésion	résistance à l'adriamycine	Jiang et al. (2014)
ATB	14q11.2	surexpression	intergénique	oncogène	régulation de la voie TGF- β	EMT et métastase	Shi et al. (2015b)
BCAR4	16p13.13	surexpression	intergénique	oncogène	développement embryonnaire grossesse et placenta	résistance au trastuzumab	Xing et al. (2014)
BCYRN1/B200	2p21	surexpression	intronique	oncogène	régulation traductionnelle	répression de suppresseurs de métastases	Iacoangeli. (2013)
CCAT2	8q24.21	surexpression	intronique	oncogène	régulation transcriptionnelle	instabilité chromosomique activation de la voie Wnt (Myc)	Redis et al. (2013)
CYTOR	2p11.2	surexpression	intergénique	oncogène	régulation du cycle cellulaire	EMT et migration activation des voies MAPK, EGFR et mTOR	Van Grembergen et al. (2016)
EFNA3	1q21.22	surexpression		oncogène	adhésion et répulsion cellulaires par interactions avec les récepteurs Ephrine	extravasation des cellules tumorales et métastase	Gomez-Maldonado et al. (2015)
FAL1/FALEC	1	surexpression	intergénique	oncogène	stabilisation post-traductionnelle de BMI-1, inhibition de p21	prolifération, croissance indépendante de l'ancrage, pluripotence	Hu et al. (2014c)
HOTAIR	12q13.13	surexpression	antisens chevauchant	oncogène	régulation épigénétique, compétition endogène, stabilité protéique	répression multigénique en trans du locus HOXD via PRC2/LSD1	Gupta et al. (2010)
HOTAIRM1	7p15.2	surexpression	antisens	oncogène			
LSINCT5	5p15.33	surexpression		oncogène	régulation de MALAT1, NEAT1 et PSPCI induite par les stress	prolifération	Silva et al. (2011)
lncRNA-JADE	4q28.2	surexpression	intergénique	oncogène	réparation de l'ADN	prolifération et survie des cellules cancéreuses	Wan et al. (2013c)
LncRNA-Smad7	18q21.1	surexpression	antisens	oncogène	expression induite par TGF- β	fonction anti-apoptotique de TGF- β	Arase et al. (2014)
MALAT1	11q13.1	surexpression	intergénique	oncogène	régulation transcriptionnelle épissage alternatif	prolifération, métastase migration par compétition endogène avec cdc42 et miR-1	Gutschner et al. (2013a)
NEAT1	11q13.1	surexpression	intergénique	oncogène	architecture des paraspeckles régulation transcriptionnelle	séquestration de mRNAs et miRNAs	Knott et al. (2016)

PVT1	8q24	surexpression	intergénique	oncogène	répression épigénétique compétition endogène (CE) avec mir-200	coopération entre PVT1 et MYC perte de la CE	Cui et al. (2016)
SOX2OT	3q26.33	surexpression	chevauchant	oncogène	maintien de la pluripotence	prolifération indépendante de l'ancrage par induction de SOX2	Askarian et al. (2014)
SPRY4-IT1	5q31.3	surexpression	intronique	oncogène	régulation de la croissance et de l'apoptose	activation de la protéine oncogénique ZNF703	Shen et al. (2016)
TreRNA	20q13.13	surexpression	intergénique	oncogène	activateur transcriptionnel de SNAI1 inhibiteur traductionnel de l'E- cadherine	EMT, migration métastases	Gumireddy et al. (2013)
UCA1	19p13.12	surexpression	intergénique	oncogène	complexe ribonucléoprotéique fonctionnel avec hnRNPI	prolifération cellulaire via l'inhibition compétitive de p27.	Li et al. (2014b)
GAS5	1q25.1	sous-expression	chevauchant	oncosuppresseur	riborépression transcriptionnelle	perte de l'inhibition de la croissance	Pickard et al. (2014)
LIMT		sous-expression	intergénique	oncosuppresseur	suppresseur de métastase	inactivation par EGFR	Sas-Chen et al. (2016)
MEG3	14q32.2	sous-expression	intergénique	oncosuppresseur	régulation de la voie TGFβ	interaction avec EZH2 via des triplex RNA-DNA	Guo et al. (2016)
NKILA	20q13.31	sous-expression	antisens	oncosuppresseur	inhibition de la voie NF-kB	activation de la voie NF-kB et de gènes suppresseurs de métastases	Liu et al. (2015b)
PTENP1	9p13.3	sous-expression	antisens	oncosuppresseur	compétition endogène	inhibition de PTEN	Hesson et al. (2012)
Xist	Xq13.2	sous-expression	intergénique	oncosuppresseur	inactivation du chromosome X	duplication et perte de l'inactivation de l'autre chromosome X	Lee. (2009)
Zfas1	20q13.13	sous-expression	antisens chevauchant	oncosuppresseur	inhibition de la prolifération et de la différenciation mammaire	augmentation de la prolifération	Askarian et al. (2011)
H19	11p15.15	surexpression	chevauchant	controversé	empreinte génomique, méthylation de l'ADN, régulation transcriptionnelle, biogénèse de miRNAs, compétition endogène, inactivation traductionnelle	activation par C-MYC éponge moléculaire inhibant let-7 précurseur de miR675	Gabory et al. (2010)
SRA	5q31	surexpression	intergénique	controversé	co-activateur des récepteurs stéroïdiens	prolifération et différenciation	Novikova et al. (2012)

Tableau 6 : Principaux lncRNAs oncogènes et oncosuppresseurs impliqués dans la cancérogénèse mammaire (adapté d'après Shore and Rosen, 2014 ; Vikram et al., 2014). Pages 127-128.

4.1.2 LncRNAs oncogéniques et oncosuppresseurs

4.1.2.1 LncRNAs oncogéniques

4.1.2.1.1 Prolifération cellulaire et apoptose

- *UCA1*

UCA1 a été initialement identifié comme un lncRNA surexprimé dans les carcinomes urothéliaux et plus récemment dans les cancers du sein (Huang et al., 2014). Dans les lignées de carcinome mammaire MCF7, la surexpression d'*UCA1* s'accompagne d'une augmentation de la prolifération, alors que son inactivation la réduit. La transplantation de lignées MCF7 présentant une inactivation d'*UCA1* dans des modèles murins se traduit par une réduction de la prolifération et de la taille tumorales. En pathologie humaine, il existe une corrélation négative entre *UCA1* et p27 et de faibles niveaux d'expression de p27 ont été observés dans des carcinomes mammaires ER- et PR- de haut grade par rapport au tissu mammaire normal. L'hypothèse de sa fonction oncogénique a été confortée par la découverte de son mécanisme d'action moléculaire (Li and Hu, 2015). *UCA1* forme un complexe ribonucléoprotéique fonctionnel en interagissant avec hnRNPI (*heterogeneous nuclear RiboNucleoProtein 1*) et inhibe sa fonction d'activateur de la traduction de l'oncosuppresseur p27. *UCA1* présente donc des propriétés oncogéniques potentielles au cours de la cancérogénèse mammaire en favorisant la prolifération cellulaire *via* l'inhibition compétitive de p27.

- *LSINCT5*

LSINCT5 est un lncRNA antisens de 2.6 kb induit par le stress. Il est normalement localisé dans le noyau et est transcrit par l'ARN polymérase III (Dieci et al., 2007 ; Silva et al., 2011). *LSINCT5* est surexprimé dans les tumeurs mammaires et ovariennes par rapport aux tissus normaux. Son inactivation dans des lignées mammaires MCF7 entraîne une diminution de la prolifération cellulaire. L'analyse par *microarray* de l'inactivation de *LSINCT5* a permis d'identifier 95 cibles potentielles, dont les lncRNAs *NEAT1* et *MALAT1* et le gène *PSPC1* (Silva et al., 2011). Ce lncRNA est donc considéré comme potentiellement oncogénique en raison de son effet sur la prolifération cellulaire et sur l'expression post-transcriptionnelle des gènes.

- ***LncRNA-Smad7***

LncRNA-Smad7 a été récemment identifié comme un lncRNA localisé au voisinage du gène murin *Smad7*. L'expression de *LncRNA-Smad7* est induite par TGF- β dans les cellules épithéliales et dans les lignées tumorales mammaires. Son inactivation neutralise la fonction anti-apoptotique de TGF- β . Inversement l'expression ectopique de *LncRNA-Smad7* restaure l'apoptose induite par un inhibiteur du récepteur du TGF- β . Toutefois, l'implication de *LncRNA-Smad7* dans la cancérogénèse mammaire semble être limitée à l'apoptose puisque son inactivation n'affecte pas l'EMT induite par TGF- β , la phosphorylation de Smad2 ou l'expression du gène *Smad7* (Arase et al., 2014).

- ***LOC554202/MIR31HG***

LOC554202 est l'hôte du miRNA miR31 (Augoff et al., 2012). Son expression est augmentée dans les lignées et les tissus tumoraux mammaires (Shi et al., 2014). Le rôle de ce lncRNA dans la cancérogénèse mammaire est encore controversé. Son inactivation diminue la prolifération des cellules cancéreuses mammaires, induit l'apoptose et inhibe la migration et l'invasion *in vitro* et la tumorigénèse *in vivo*. Toutefois, une autre étude a montré que l'expression de *LOC554202* est inhibée dans les carcinomes mammaires de type triple-négatif (Augoff et al., 2012).

- ***LncRNA-JADE***

La transcription de ce lncRNA est induite par des altérations de l'ADN selon un mécanisme dépendant d'ATM (Wan et al., 2013c). Après la survenue de lésions de l'ADN, *lncRNA-JADE* se lie à *BRCA1* et amplifie la transcription du gène codant voisin *Jade1*, impliqué dans l'acétylation de H4 et les mécanismes de réparation de l'ADN (DDR) (Ciccia and Elledge, 2010). Dans les lignées de cancer du sein MCF7, l'inactivation de *lncRNA-JADE* réduit la prolifération cellulaire, tandis que sa surexpression l'augmente. Son inactivation favorise également l'apoptose induite par les lésions de l'ADN. En outre, la transplantation orthotopique de lignées inactivées pour le *lncRNA-JADE* dans le coussinet adipeux mammaire de souris réduit la croissance de la tumeur par rapport aux cellules témoins. En pathologie humaine, le *lncRNA-JADE* est exprimé à des niveaux élevés dans les tumeurs du sein par rapport aux tissus normaux. Ainsi, *lncRNA-JADE* est un lncRNA associé aux mécanismes de

réparation de l'ADN dont l'expression aberrante favorise la prolifération et la survie des cellules cancéreuses du sein.

- ***PVT1***

Le gène *PVT1* humain est un grand locus (> 30 kb) situé dans la région chromosomique 8q24.21. Cette région est un *hotspot* d'altérations génétiques qui contient l'oncogène *MYC* à environ 57 kb en amont de *PVT1*. Ce lncRNA intergénique possède une taille de 1716 nt et comporte plus de 20 isoformes. Il a été récemment démontré que *PVT1* participe tant aux processus physiologiques que pathologiques, dont le cancer. Dans des modèles murins, l'inactivation de *PVT1* provoque une létalité embryonnaire, soulignant le rôle essentiel de ce lncRNA dans le développement normal. *PVT1* semble se comporter comme un compétiteur endogène appartenant au réseau MMI (*miRNA mediated sponge interactions*) normal et interagit préférentiellement avec la famille miR-200. Ce lncRNA pourrait avoir un rôle déterminant dans la morphogénèse et le maintien de l'homéostasie tissulaire mammaire (Cui et al., 2016). Les niveaux d'expression de *PVT1* sont plus élevés (jusqu'à 100 fois) dans des lignées de cellules cancéreuses et dans différents types de cancers par rapport aux tissus non tumoraux adjacents, dont les tumeurs du sein. De plus, la surexpression de *PVT1* est un facteur prédictif d'agressivité des carcinomes mammaires et de pronostic défavorable dans la plupart des tumeurs malignes (Colombo et al., 2015). Les expériences de gain et de perte réalisées sur des lignées et tissus tumoraux du sein et de diverses autres tumeurs malignes, ont montré que *PVT1* présente des fonctions oncogéniques. Son inactivation inhibe la prolifération, la progression du cycle et la migration cellulaire *in vitro* et *in vivo* tandis que sa surexpression favorise la prolifération, la progression du cycle cellulaire, l'acquisition de propriétés de cellules souches et la migration. *PVT1* peut de plus constituer un partenaire de fusion dans divers types de cancers et code pour un grand nombre de ncRNAs (Cui et al., 2016). Dans les cancers du sein, certains isoformes surexprimés de *PVT1* sont dépourvus de *miRNAs response elements* (MREs) et ne peuvent plus assurer leur fonction de compétition endogène. Une coopération entre *PVT1* et *MYC* semble indispensable à la promotion tumorale mammaire en induisant la surexpression de *Rspo1*, un régulateur majeur de la voie Wnt- β -caténine (Sarver et al., 2016). De plus, des modèles murins *MMTV-neu* comportant un gain de copies *MYC* et *PVT1* présentent une accélération de leur progression tumorale, une augmentation de la réplication de l'ADN et de la phosphorylation de H2AX et une prolifération canalaire (Tseng and Bagchi, 2015). En outre, *PVT1* semble constituer une cible thérapeutique potentielle dans le traitement des cancers, car sa surexpression contribue à la

résistance aux chimiothérapies, tandis que son inhibition améliore la sensibilité des réponses aux thérapies anticancéreuses.

4.1.2.1.2 Invasion et évolution métastatique

- ***TreRNA***

TreRNA régule l'expression des gènes aux niveaux transcriptionnel et traductionnel. Dans le noyau, *treRNA* agit comme un activateur en *cis* et régule la transcription de *SNAIL* (Orom et al., 2010). Dans le cytoplasme, *treRNA* interagit avec un complexe ribonucléoprotéique et supprime la traduction de l'E-cadhérine (Gumireddy et al., 2013), favorisant aussi l'EMT. La surexpression de *treRNA* dans les lignées MCF7 se traduit par une réduction des niveaux d'expression de l'E-cadhérine et une surexpression de marqueurs mésenchymateux, suggérant un rôle de *treRNA* dans l'induction d'un phénotype d'EMT partielle. Sa surexpression dans les cellules MCF7 augmente également la migration cellulaire et l'invasion *in vitro* et stimule la survenue de métastases dans le poumon après xénogreffes orthotopiques chez la souris. L'hypothèse du rôle oncogénique de *treRNA* est également renforcée par la mise en évidence de sa surexpression dans des échantillons appariés de tumeurs mammaires primitives et de métastases ganglionnaires, suggérant que *treRNA* pourrait constituer une cible thérapeutique anti-métastatique potentielle.

- ***HOTAIR***

HOTAIR est un lncRNA intergénique de 2.2 kb appartenant au locus *HOXC*. Il est le premier lncRNA modulateur des structures chromatinienne à avoir été impliqué dans la carcinogénèse et le premier lncRNA identifié dont le mécanisme d'action s'effectue en *trans* (Rinn et al., 2007). *HOTAIR* est considéré comme un régulateur épigénétique constituant une plateforme et un guide pour les structures chromatinienne qui présente une remarquable capacité à se lier à des complexes répresseurs chromatinien de nature variée (Rinn et al., 2007 ; Tsai et al., 2011). Son extrémité 5' terminale interagit avec le complexe polycomb PRC2 tandis que son extrémité 3' terminale se lie à l'histone déméthylase LSD1, qui elle-même interagit avec les facteurs de transcription CoREST et REST (Kaneko et al., 2010). L'action de ces deux complexes est responsable d'une inactivation épigénétique *via* la marque répressive H3K27me3 (Gupta et al., 2010 ; Kogo et al., 2011). La cible de PRC2 est le locus

HOXD présent sur le chromosome 2 où PRC2 et *HOTAIR* favorisent la répression transcriptionnelle de 854 gènes disposés sur une distance de 40 kb, dont des gènes suppresseurs de métastases (*HOXD10*, *PR*, *PCDH10*, *PCDHB5* et *JAM2*) (Croce, 2010 ; Kogo et al., 2011). *HOTAIR* est actuellement considéré comme un lncRNA oncogénique associé à la prolifération, l'invasion et l'inhibition apoptotique. Si l'expression ectopique de *HOTAIR* augmente modestement la croissance tumorale, elle favorise l'invasion mais surtout la colonisation pulmonaire et la survenue de métastases *in vivo*. Sa déplétion dans les lignées cancéreuses mammaires diminue l'invasion de la matrice extra-cellulaire (Gupta et al., 2010 ; Kogo et al., 2011) et induit une reprogrammation des marqueurs chromatinien tels que la perte du marqueur répressif H3K27me3 au sein du locus *HOXD* (Rinn et al., 2007), des métalloprotéases et des facteurs de croissance vasculaire (Geng et al., 2011). Le profil d'expression de PRC2 dans les cellules tumorales mammaires surexprimant *HOTAIR* ressemble à celui des fibroblastes embryonnaires et *HOTAIR* pourrait reprogrammer les cellules cancéreuses en cellules de morphologie fibroblastique (Gupta et al., 2010). En pathologie humaine, *HOTAIR* est très fortement surexprimé (jusqu'à 2000 fois la normale) dans les métastases de cancer du sein alors que son expression est modérément élevée et de façon hétérogène dans les tumeurs mammaires primitives (Gupta et al., 2010). Il est également surexprimé dans diverses tumeurs primitives et métastatiques, telles que les carcinomes du sein (Gupta et al., 2010), du rein (Kogo et al., 2011), du foie (Geng et al., 2011), du tractus intestinal (Niinuma et al., 2012), du pancréas (Kim et al., 2012) et du poumon (Li et al., 2013a). En outre, son niveau d'expression est prédictif de la survenue de métastases et de la survie (Geng et al., 2011 ; Yang et al., 2011a ; Chisolm et al., 2012 ; Lu et al., 2012 ; Niinuma et al., 2012 ; Ishibashi et al., 2013 ; Liu et al., 2013 ; Hansji et al., 2014). Le niveau global de méthylation de l'ADN génomique est corrélé positivement au niveau d'expression de *HOTAIR*, suggérant que ce lncRNA pourrait être régulé au cours de la cancérogénèse mammaire par la méthylation de l'ADN intergénique (Lu et al., 2012). Globalement, ces résultats montrent que *HOTAIR* pourrait jouer un rôle important dans l'épigénome du cancer et représenter un biomarqueur d'intérêt et une cible thérapeutique potentielle pour le cancer du sein (Chisholm et al., 2012).

- ***BCYRN1/BC200***

BCYRN1/BC200 est un lncRNA de 200 pb, spécifique des primates, fonctionnant comme un répresseur traductionnel impliqué dans la synthèse des protéines neuronales au niveau des

synapses (Iacoangeli and Tiedge, 2013). *BCYRN1* est fortement surexprimé dans le cancer du sein mais aussi dans d'autres carcinomes d'origine ovarienne, cervicale et pulmonaire (Iacoangeli et al., 2004). *BCYRN1* pourrait jouer un rôle dans la progression tumorale des carcinomes mammaires *via* la répression traductionnelle de suppresseurs de métastases. En outre, il est également surexprimé dans les carcinomes canaux *in situ* de haut grade, suggérant son utilisation comme indicateur de leur potentiel invasif.

- ***EFNA3***

Les éphrines sont des protéines cellulaires de surface qui régulent divers processus biologiques modulant l'adhésion et la répulsion cellulaire en induisant des interactions avec des récepteurs de type tyrosine kinase, les récepteurs Eph (Pasquale et al., 2010). La présence au sein des carcinomes mammaires de régions hypoxiques est un facteur pronostique défavorable pour la survie des patientes (Semenza et al., 2012). L'hypoxie favorise l'expression de l'éphrine-A3 (EFNA3) *via* un mécanisme dépendant de HIF1 (Vihanto et al., 2005). En réponse à l'hypoxie, les niveaux d'expression du transcrite codant pour la protéine EFNA3 demeurent relativement stables, mais HIF1 induit également à partir du même locus la transcription du lncRNA *EFNA3* qui augmente l'expression et l'accumulation de la protéine EFNA3. La surexpression à la fois de la protéine EFNA3 et du lncRNA *EFNA3* majorerait le potentiel métastatique des cellules cancéreuses mammaires humaines, en favorisant l'extravasation des cellules tumorales dans les tissus environnants. En outre, une forte corrélation a été établie entre la surexpression du lncRNA *EFNA3* et une plus courte survie sans métastases chez les patientes atteintes de cancer du sein. L'hypoxie pourrait de ce fait contribuer à la propagation métastatique du cancer du sein *via* la surexpression médiée par HIF1 du lncRNA *EFNA3*, l'accumulation de la protéine EFNA3 et l'altération de l'angiogénèse (Gomez-Maldonado et al., 2015).

- ***NEATI***

NEATI est un lncRNA architectural localisé dans le noyau où il représente la matrice des *paraspeckles*. Il agit comme un régulateur transcriptionnel de nombreux gènes (Clemson et al., 2009). Le gène du lncRNA *NEATI* produit deux isoformes, l'un court de 3.7 kb (*NEATI-1*) et l'autre long de 23 kb (*NEATI-2*). *NEATI* est largement exprimé dans plusieurs tissus, bien que le niveau d'expression du long isoforme soit beaucoup plus faible que celui de

l'isoforme court (Hutchinson et al., 2007). *NEATI* pourrait servir de plateforme pour le recrutement des protéines DBHS (*drosophila behavior/human splicing*), telles SFPQ, NONO et PSPC1 (Clemson et al., 2009 ; Chen and Carmichael, 2009 ; Knott et al., 2016). La surexpression de *NEATI* est observée dans plusieurs types de tumeurs malignes et plus particulièrement dans les carcinomes mammaires (Hutchinson et al., 2007 ; Clemson et al., 2009 ; Sunwoo et al., 2009 ; Zeng et al., 2014). *NEATI* et *MALATI* sont les 2 gènes de lncRNAs les plus souvent mutés parmi les cancers du sein analysés par *whole-genome sequencing* (Nik-Zainal et al., 2016). *NEATI* est impliqué dans la survie des cellules cancéreuses mammaires *via* ses interactions avec l'oncoprotéine de fusion liant l'ARN *FUS-TLS* et miR-548. Ainsi, l'inactivation de *NEATI*, la sous-expression de *FUS-TLS* et la surexpression de miR-548 s'accompagnent d'une augmentation de l'apoptose des cellules tumorales mammaires (Ke et al., 2016).

- ***CCAT2***

Les études GWAS étudiant les SNPs ont conduit à la découverte de nombreuses régions génomiques associées au cancer. L'analyse du SNP rs6983267, associé à un risque accru de cancer colorectal, de la prostate, de l'ovaire et du sein inflammatoire, a permis d'identifier un lncRNA très conservé, *CCAT2*. Son rôle dans l'invasion et l'évolution métastatique a été analysé en combinant des essais *in vitro* et des modèles murins de xéno greffes de carcinomes coliques, révélant que la surexpression de *CCAT2* s'accompagne d'une plus grande incidence de tumeurs métastatiques (Ling et al., 2013). *CCAT2* favorise l'évolution métastatique et l'instabilité chromosomique par un mécanisme impliquant des facteurs de transcription, des oncogènes et des miRNAs (Ling et al., 2013). Il est impliqué dans la régulation de la signalisation Wnt et l'activation de la voie Wnt entraîne sa surexpression. *CCAT2* module à son tour l'expression de gènes cibles de la voie Wnt, tels *MYC*, ainsi que ses cibles d'aval, miR17HG et miR20a. Les lignées cellulaires surexprimant *CCAT2* présentent une augmentation de l'activité Wnt. Collectivement, ces résultats suggèrent que *CCAT2* agit en amplifiant les effets de la signalisation Wnt, favorisant ainsi le processus métastatique (Ling et al., 2013). En outre, l'amplification de la région génomique 8q24, où est localisé *CCAT2*, a été observée dans les cancers du sein (Courjal and Theuillet, 1997). Néanmoins, si *CCAT2* apparaît globalement surexprimé dans 2/3 des carcinomes mammaires, la valeur pronostique de son expression n'est cliniquement informative que dans le sous-groupe de patientes N+ ayant bénéficié d'une chimiothérapie (Redis et al., 2013).

4.1.2.1.3 Acquisition de caractères de cellules souches

- ***SOX2OT***

La séquence codante du gène *SOX2* est située dans l'intron du lncRNA multi-exonique *SOX2OT* (Amaral et al., 2008). Les expressions de *SOX2* et *SOX2OT* sont corrélées positivement dans le cancer du sein et *SOX2OT* s'exprime différemment dans les carcinomes mammaires RE+ et RE- (Askarian-Amiri et al., 2014). *SOX2* et *SOX2OT* sont tous deux surexprimés dans les cellules cancéreuses de type souche et *SOX2* est l'un des gènes clés impliqués dans le maintien de la pluripotence. En outre, l'expression ectopique de *SOX2OT* conduit à une importante surexpression de *SOX2* et à une augmentation de la croissance indépendante de l'ancrage. Le lncRNA *SOX2OT* présente donc un potentiel oncogénique dans le cancer du sein en induisant ou en maintenant l'expression de *SOX2*.

- ***FALI***

FALI est pourvu de propriétés potentiellement oncogéniques dans le cancer du sein. L'inactivation de *FALI* réduit significativement la clonogénicité et inhibe la prolifération et la croissance indépendante de l'ancrage des cellules cancéreuses mammaires. Cette inactivation supprime aussi la croissance des xénogreffes sous-cutanées issues des lignées cellulaires MDA-MB-231 (Hu et al., 2014c). *FALI* interagit directement avec la sous-unité BMI-1/PCGF4 du complexe polycomb PRC1, qui est impliquée dans l'acquisition et le maintien des propriétés de pluripotence de type souche des cellules cancéreuses mammaires (Godlewski et al., 2008). En outre, l'inactivation de *FALI* diminue significativement le niveau d'expression de la protéine BMI-1, mais non le niveau d'expression de son mRNA, ce qui suggère que *FALI* régule BMI-1 à un niveau post-translationnel en la stabilisant. Enfin, *FALI* a été impliqué dans la régulation de la transcription d'un grand nombre de gènes importants pour la prolifération, l'apoptose, la survie ainsi que les mouvements cellulaires et la dégradation des protéines. L'un de ces gènes est *p21* qui est un gène suppresseur de tumeur clé dans de nombreux types de carcinomes, dont le cancer du sein.

4.1.2.2 LncRNAs oncosuppresseurs

4.1.2.2.1 Prolifération et apoptose

- ***GAS5***

Le gène *GAS5* est situé au niveau du locus 1q25.1 et produit 2 variants d'épissage tandis que ses introns sont transcrits en plusieurs snoRNAs (Mourtada-Maarabouni et al., 2008 ; Kino et al., 2010). *GAS5* est largement surexprimé dans les tissus embryonnaires et adultes. Il fonctionne comme un suppresseur de tumeur en facilitant l'arrêt du cycle et l'apoptose *via* la riborépression de la transcription médiée par les récepteurs des glucocorticoïdes (Pickard et al., 2014). Il se fixe au domaine de liaison à l'ADN du récepteur des glucocorticoïdes (GR) et entre en compétition avec l'élément de réponse aux glucocorticoïdes (GRE), qui est une séquence de régulation génomique. Les GREs sont séquestrés et l'induction médiée par les glucocorticoïdes de plusieurs gènes de réponse est supprimée, incluant celle de l'inhibiteur de l'apoptose cellulaire 2 (cIAP2). *GAS5* favorise aussi l'inhibition de mTOR normalement impliqué dans la synthèse protéique, le cycle cellulaire et la prolifération (Mourtada-Maarabouni et al., 2010). Les niveaux d'expression de *GAS5* sont significativement diminués dans les cellules carcinomateuses mammaires par rapport aux cellules normales. A l'opposé, la surexpression de *GAS5* dans les cellules MCF7 et MCF10A majore l'apoptose induite par un stress génotoxique, l'irradiation UV et l'utilisation de cisplatine (Mourtada-Maarabouni et al., 2009). En outre, des lignées cellulaires soumises à un manque d'apport nutritif et de facteurs de croissance expriment des niveaux élevés de *GAS5* qui les sensibilisent à l'apoptose (Piao et al., 2012). *GAS5* est aussi régulé par le miRNA oncogène miR21, qui contrôle de nombreux gènes impliqués dans la croissance cellulaire et l'apoptose, et par le suppresseur de tumeur PTEN. La sous-expression de *GAS5* et de ses snoRNAs associée à des anomalies génétiques du locus 1q25.1 ont été observées dans diverse tumeurs de pronostic défavorable, dont les cancers du sein (Mourtada-Maarabouni et al., 2009), mais aussi les cancers du rein et de la vessie (Liu et al., 2013 ; Qiao et al., 2013), de la prostate (Pickard et al., 2013), de l'estomac (Sun et al., 2014a) et du col de l'utérus (Cao et al., 2014). Globalement, ces études suggèrent que le lncRNA *GAS5* pourrait être un suppresseur de tumeur favorisant l'inhibition de la croissance cellulaire induite par le stress et l'apoptose.

- ***PTENP1***

PTENP1 est un pseudogène du gène onco-suppresseur codant pour PTEN. Dans les lignées cellulaires de cancers, la surexpression de *PTENP1* en 3'UTR conduit à une augmentation des

niveaux d'expression de *PTEN* et à une inhibition de la prolifération, alors que l'inactivation de *PTENP1* diminue les niveaux d'expression de *PTEN* et favorise la prolifération cellulaire (Poliseno et al., 2010). De plus, les niveaux de *PTENP1* et de *PTEN* sont corrélés positivement dans les échantillons normaux et tumoraux, suggérant qu'ils sont co-régulés. En outre, *PTENP1* et *PTEN* contiennent des MREs communs, et pourraient agir par compétition endogène. Récemment, plusieurs lncRNAs antisens provenant du locus *PTENP1* ont été identifiés et mettent ainsi en évidence la complexité de la régulation de *PTEN*. Ces asRNAs régulent à la fois la transcription et la stabilité de *PTEN* et sont également susceptibles de jouer un rôle dans la tumorigénèse (Johnsson et al., 2013). Le locus *PTENP1* est focalement délété dans le cancer du sein (Poliseno et al., 2010) et le promoteur de *PTENP1* subit une hyperméthylation somatique dans des lignées cellulaires de cancers du sein, indiquant que la perte de *PTENP1* peut également contribuer à la progression des tumeurs mammaires *via* une inactivation épigénétique (Hesson et al., 2012).

- *Zfas1*

Zfas1 est un lncRNA antisens situé à l'extrémité 5' terminale du gène codant pour la protéine Zfx1. Il a été considéré initialement comme un lncRNA générant des snoRNAs. Ultérieurement, ce lncRNA a été observé dans les canaux et les structures ductulo-alvéolaires de la glande mammaire et semble impliqué dans son développement ainsi qu'au cours de la grossesse et de l'allaitement. *Zfas1* est un inhibiteur de la prolifération épithéliale mammaire et de la différenciation. L'inactivation de *Zfas1* dans les cellules épithéliales mammaires entraîne une augmentation de la prolifération cellulaire, et son expression dans des lignées tumorales est réduite par rapport aux cellules et tissus mammaires normaux (Askarian-Amiri et al., 2011). *Zfas1* est actuellement considéré comme un lncRNA pourvu d'un potentiel oncosuppresseur mais il est nécessaire de clarifier sa fonction et son mécanisme spécifique d'action dans la cancérogénèse mammaire.

4.1.2.2 Invasion et évolution métastatique

- *Xist*

Xist a été parmi les premiers lncRNAs identifiés et est exprimé essentiellement dans les cellules somatiques féminines (Lee, 2009). Il joue un rôle majeur dans l'inactivation du

chromosome X au cours du développement embryonnaire et dans le maintien de la stabilité génomique. Ce lncRNA de 17kb agit au niveau du chromosome X en recrutant les complexes polycomb afin de maintenir l'inactivation de plus de 1000 gènes sur ce chromosome. La dérégulation de *Xist*, conduisant à la perte ou au gain de chromosome X, est observée dans divers cancers féminins, masculins et non liés au sexe. Dans les cancers féminins, le chromosome X inactivé (corpuscule de Barr) est essentiellement absent dans les cellules malignes et parfois dupliqué. Les lignées de cancers féminins comportent de multiples copies du chromosome X actif, aboutissant à une ploïdie aberrante et à l'origine d'une chimiorésistance accrue (Kawakami et al., 2004). *Xist* est également surexprimé dans les tumeurs germinales testiculaires où la présence d'un nombre accru de chromosomes X semble jouer un rôle oncogénique (Looijenga et al., 1997 ; Kawakami et al., 2003). Le rôle de *Xist* dans la cancérogénèse mammaire a été largement étudié mais reste néanmoins encore actuellement controversé et incertain. Le gène oncosuppresseur *BRCA1* contrôle les niveaux des transcrits *Xist* sur le chromosome X inactif et joue un rôle essentiel dans le maintien des marqueurs de son inactivation (Ganesan et al., 2002). Cette observation suggère que la perte de *BRCA1* dans des cellules féminines pourrait conduire à une déstabilisation du chromosome X et à une réactivation de nombreux gènes normalement inactivés en raison de la perte de *Xist*. Toutefois, il a été ultérieurement constaté que les fonctions de *Xist* semblaient indépendantes de celles de *BRCA1* dans l'inactivation du chromosome X (Xiao et al., 2007). Une étude a récemment rapporté que l'expression de *Xist* est très altérée dans les carcinomes mammaires de type triple-négatif, dont 80% présentent des anomalies des chromosomes X, incluant la duplication et la perte de l'inactivation de l'autre chromosome X (Richardson et al., 2006). A l'opposé, seuls 10% des carcinomes mammaires appartenant aux autres sous-types tumoraux présentent une perte de l'inactivation de l'un de leurs chromosomes X. Ces anomalies ne semblent toutefois pas liées à *BRCA1* puisque la majorité des prélèvements tumoraux présentent un statut *BRCA1* sauvage.

- ***NKILA***

La voie NF- κ B représente un lien essentiel entre l'inflammation et le cancer et son activation constitutive dans les tumeurs malignes constitue un facteur clé de la progression tumorale. Seule une minorité des tumeurs présentant une activation aberrante de NF- κ B comporte des mutations génétiques activatrices affectant les composants de la voie NF- κ B. L'activation

constitutive de la voie NF- κ B dans les cellules malignes de nombreuses tumeurs solides semble être surtout favorisée par les cytokines inflammatoires du microenvironnement tumoral. *NKILA* se lie au complexe NF- κ B/I κ B et réprime la voie de signalisation NF- κ B et l'inflammation chronique associée au cancer (O'Dea et al., 2008). *NKILA* bloque directement la phosphorylation d'I κ B en formant un complexe ternaire stable avec NF- κ B/I κ B (Sue et al., 2011). De plus, *NKILA* est un régulateur négatif de NF- κ B, à la fois dans les cellules quiescentes et les cellules activées. Son expression peut réduire l'activité de NF- κ B dans les cellules épithéliales mammaires normales et malignes, tandis que son inhibition conduit à une activation aberrante de NF- κ B, même en l'absence des stimulations inflammatoires, suggérant que *NKILA* pourrait être un *gatekeeper* essentiel de la signalisation NF- κ B dans les cellules quiescentes. Dans les cancers du sein, une faible expression de *NKILA* est associée à une évolution métastatique plus fréquente et à un pronostic péjoratif. Par conséquent, *NKILA* pourrait interagir directement avec les domaines fonctionnels de protéines de signalisation, et constituer une classe de modulateurs de la voie NF- κ B et de suppresseurs de métastases (Liu et al., 2015b).

- ***LIMT***

LIMT (*LncRNA Inhibiting Metastasis*) a été récemment identifié comme un lncRNA très conservé au cours de l'évolution. *LIMT* est, avec *NEAT1* et *LINC00472*, un lncRNA inactivé par le facteur de croissance EGFR. Il agit comme un lncRNA suppresseur de métastase et inhibe l'invasion de la matrice extra-cellulaire *in vitro* et la formation de métastases *in vivo*. Dans les sous-groupes agressifs de cancers du sein, tels les carcinomes HER2+ et *basal-like*, l'activation plus fréquente d'EGFR résultant d'interactions réciproques entre le microenvironnement tumoral et les cellules cancéreuses, induit une inactivation de *LIMT* via la déacétylation de son promoteur et favorise l'évolution métastatique (Sas-Chen et al., 2016)

4.1.2.3 LncRNAs controversés dans la cancérogénèse mammaire.

- ***H19***

H19 est parmi les premiers lncRNAs à avoir été identifié comme un régulateur clé de l'expression des gènes (Brannan et al., 1990). Il est transcrit à partir du locus 11p15.5 soumis

à l’empreinte parentale. Ce locus où est situé le gène *H19*, contient également le gène *IGF2*. Ces deux gènes subissent l’empreinte parentale avec une expression différentielle, *H19* à partir de l’allèle maternel et *IGF2* à partir de l’allèle paternel. *H19* est un transcrite de 2.3 kb contrôlant l’inactivation du gène *IGF2* (Edwards and Ferguson-Smith, 2007). Il est fortement exprimé pendant l’embryogénèse mais est inactivé après la naissance dans la plupart des tissus. Il est reexprimé durant la cancérogénèse et représente l’un des lncRNAs les plus controversés quant à sa nature oncogénique ou oncosuppressive. Bien que l’inactivation de *H19* ne soit pas létale chez la souris, ce lncRNA semble présenter à la fois des propriétés oncogéniques et oncosuppressives (Matouk et al., 2014). Les expériences initiales réalisées *in vitro* à partir de lignées cancéreuses et les analyses plus récentes utilisant des modèles *in vivo* de tumeurs de souris, indiquent que *H19* semble agir en tant que suppresseur de tumeur dans des contextes cellulaires variés, en particulier *via* la perte de la croissance indépendante de l’ancrage (Yoshimizu et al., 2008 ; Gabory et al., 2010). Toutefois, d’autres études ont identifié des fonctions oncogéniques dans les cellules épithéliales mammaires (Matouk et al., 2013). La surexpression de *H19* dans la lignée mammaire MDA-MB-231 diminue le temps de latence tumorale et augmente la taille tumorale (Lottin et al., 2002). L’inactivation de *H19* s’accompagne de la diminution de la clonogénicité et de la suppression de la croissance indépendante de l’ancrage (Lottin et al., 2002). *H19* est une cible directe du facteur de transcription oncogénique c-MYC surexprimé dans les lignées cellulaires de cancer du sein, et les niveaux d’expression des transcrits de *H19* et c-MYC sont corrélés significativement dans les carcinomes mammaires (Barsyte-Lovejoy et al., 2006). C-MYC se lie au promoteur de *H19* et favorise l’initiation de sa transcription, alors qu’il inactive l’expression du gène *IGF* en se liant à son premier intron (Barsyte-Lovejoy et al., 2006). La fonction oncogénique de *H19* semble confortée *via* son rôle dans l’activation de gènes impliqués dans l’angiogénèse, la survie et la prolifération et dans l’établissement du phénotype tumoral des cancer du sein (Lottin et al., 2002 ; Berteaux et al., 2005), du foie (Fellig et al., 2005), du colon (Cui et al., 2002), du poumon (Matouk et al., 2010), du pancréas (Ma et al., 2014), de l’œsophage (Hibi et al., 1996), de la vessie (Luo et al., 2013) et du col utérin (Douc-Rasy et al., 1996). Toutefois, l’analyse par hybridation *in situ* de tumeurs du sein montre que *H19* est localisé principalement dans les cellules stromales plutôt que dans l’épithélium. Bien que sa fonction exacte dans les cellules mammaires soit encore méconnue, *H19* joue un rôle clé au cours de la cancérogénèse mammaire. Le mécanisme d’action de *H19* semble se situer en partie au niveau de la régulation traductionnelle des gènes. En effet, *H19* antagonise, à la manière d’une « éponge moléculaire », le miRNA let-7 qui inhibe l’autorenouvellement des cellules

cancéreuses souches et la tumorigénicité des cancers du sein (Kallen et al., 2013). En outre *H19* est le précurseur de miR-675, miRNA qui présente des fonctions antagonistes oncogéniques et oncosuppressives. MiR-675 supprime l'expression de pRB (Cai et al., 2007 ; Tsang et al., 2010) et inactive IGF1R, le récepteur d'IGF2 (Keniry et al., 2012). Une étude récente a révélé que *H19* est associé à la voie TGF- β , à l'EMT et à des protéines épigénétiques, telles SETD7, KDM1 et EZH1 (van Grembergen et al., 2016).

- ***SRA***

SRA est le premier lncRNA à avoir été identifié dans la cancérogénèse mammaire comme fonctionnant indépendamment de mécanismes épigénétiques ou catalytiques (Novikova et al., 2012). Il est transcrit à partir du locus 5q31.3 en un lncRNA de 2 kb fonctionnant comme un co-activateur de récepteurs stéroïdiens. Le gène *SRA1* joue un double rôle en codant à la fois pour une protéine (SRAP) et un lncRNA, *SRA*, obtenus par épissage alternatif (Kawashima et al., 2003 ; Hube et al., 2006 ; Cooper et al., 2011). La protéine SRAP et le lncRNA *SRA* transcrits à partir du gène *SRA1* sont susceptibles d'influencer la cancérogénèse et la progression tumorale en modulant les niveaux d'expression des différents gènes cibles (Gutschner et al., 2012). *SRA* a été initialement identifié comme un nouveau régulateur de l'ARN du récepteur de la progestérone (Lanz et al., 1999), mais il peut également interagir avec de multiples récepteurs nucléaires, des co-régulateurs, ainsi que d'autres facteurs de transcription non nucléaires (Colley and Leedman, 2011a). Ainsi, *SRA* est actuellement considéré comme un lncRNA servant de plateforme multimoléculaire assurant les interactions de plusieurs régulateurs transcriptionnels au niveau de loci génomiques cibles afin d'y moduler l'expression des gènes. *SRA* est localisé dans le noyau et le cytoplasme. Il régule l'expression génique des récepteurs stéroïdiens en interagissant avec des protéines contenant également un co-activateur du récepteur des stéroïdes (SRC-1). Les données expérimentales actuelles fournissent des arguments pour un rôle à la fois oncogénique et oncosuppresseur. La surexpression de *SRA* favorise également la prolifération et la différenciation cellulaire mammaire (Lanz et al., 2003). En pathologie humaine, *SRA* est exprimé à des niveaux très faibles dans le sein normal et son expression est augmentée dans les tumeurs du sein, en particulier dans les carcinomes RP+, ainsi que dans d'autres cancers (Leygue et al., 1999 ; Kawashima et al., 2003 ; Chooniedass-Kothari et al., 2006 ; Cooper et al., 2009). Enfin, le rôle de *SRA* dans la tumorigénèse pourrait être dépendant du contexte tissulaire ou des niveaux d'expression différentiels des isoformes codants et non codants (Cooper et al., 2009).

4.2 Rôles des lncRNAs au cours de la progression tumorale

Bien qu'encore très largement méconnue, l'expression différentielle des lncRNAs à chaque étape de l'initiation et de la promotion tumorale mammaire peut être prédite sur la base de leurs fonctions tumorigènes (**Fig. 27**).

Les lncRNAs oncosuppresseurs, tels *Zfas1* (Askarian-Amiri et al., 2011), *MEG3* (Braconi et al., 2011), *PTENP1* (Johnsson et al., 2013) et *GAS5* (Brisken, 2013), sont exprimés par les cellules épithéliales mammaires normales et agissent principalement comme des inhibiteurs de croissance.

Les lncRNAs oncogènes, tels *UCA1* (Huang et al., 2014), *LSINCT5* (Silva et al., 2011), *lncRNA-JADE* (Wan et al., 2013c), *H19* (Matouk et al., 2013) , *BC200* (Iacoangeli and Tiedge, 2013) et *SRA* (Novikova et al., 2012), favorisent la prolifération des cellules épithéliales mammaires et sont surexprimés dans les hyperplasies canalaire atypiques et les carcinomes *in situ*.

Les lncRNAs pro-métastatiques, tels *treRNA*, *MALAT1* et *HOTAIR*, sont impliqués dans la progression tumorale et la dissémination systémique (Dhamija and Diederichs, 2016).

Structure canalaire normale

Hyperplasie canalaire atypique 1

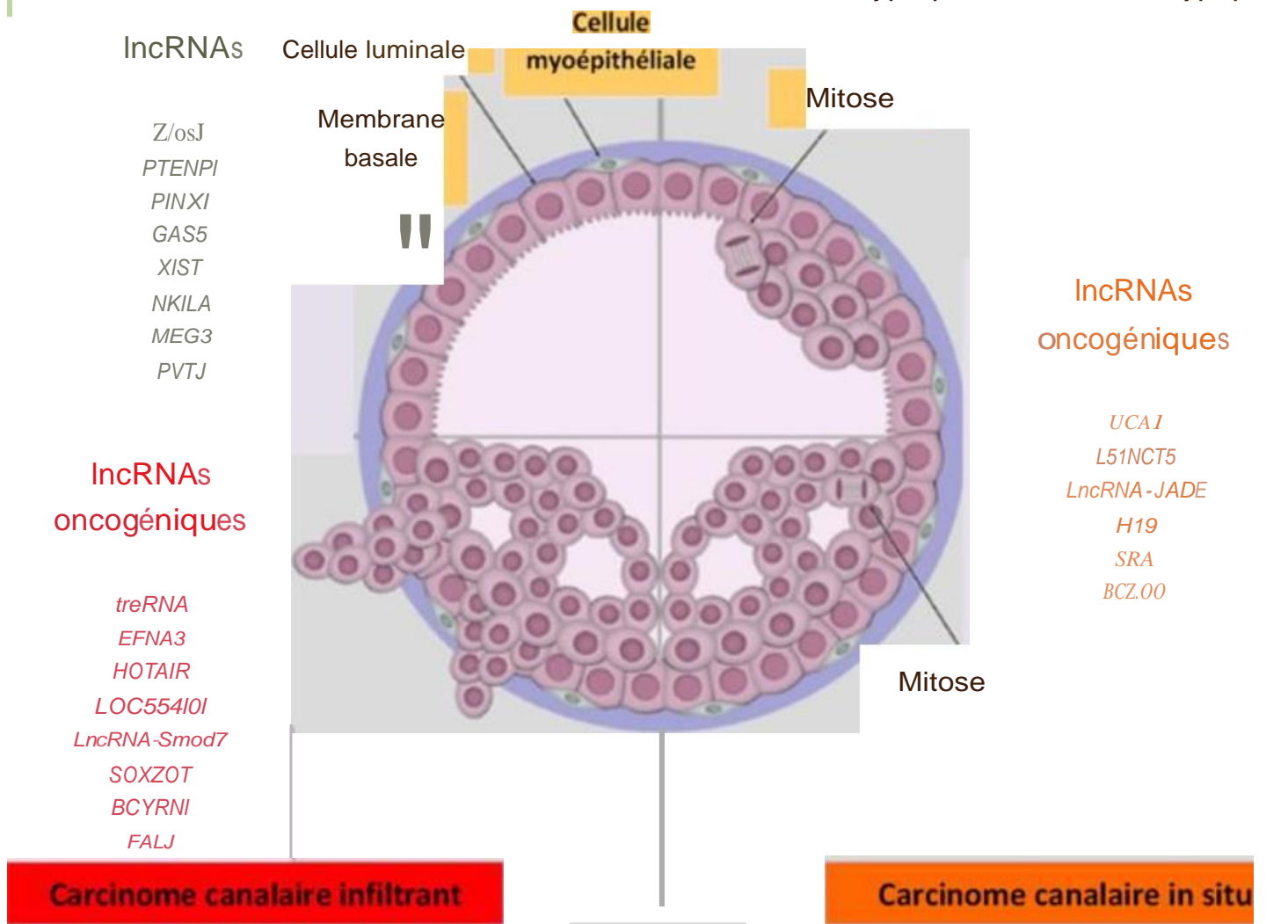


Figure 27 : Principaux lncRNAs oncogènes et oncosuppresseurs impliqués dans la survenue des lésions d'hyperplasie canalaire atypique, de carcinomes canaux in situ et de carcinomes canaux infiltrants (Adapté de Shore and Rosen, 2014)

4.3 Classifications des cancers du sein intégrant des lncRNAs

Les principales classifications moléculaires ont été jusqu'à présent essentiellement basées sur l'intégration de données concernant les gènes codant pour des protéines. Cependant, les tumeurs d'un même sous-type moléculaire peuvent répondre différemment aux thérapies et présenter des évolutions variées. La compréhension des mécanismes moléculaires responsables de l'hétérogénéité tumorale demeure un enjeu majeur de manière à améliorer les diagnostics et les traitements. La relative haute spécificité des lncRNAs les font considérer comme des *drivers* oncogéniques potentiels et des biomarqueurs de cancers. Des études ont révélé l'intérêt diagnostique et pronostique croissant de l'intégration des lncRNAs dans les classifications des cancers, dont les tumeurs du colon (Hu et al., 2014b), de l'ovaire et du poumon (Du et al., 2013), de l'endomètre (Jiang et al., 2015), du rein (Malouf et al., 2015) et des gliomes (Li et al., 2014d).

L'étude des profils d'expression des lncRNAs dans le cancer du sein, a permis l'identification de nouveaux lncRNAs associés aux diverses classifications cliniques, aux sous-types moléculaires et au pronostic.

- L'étude des lncRNAs dérégulés entre les différents sous-types de cancers du sein a permis d'identifier 20 gènes de lncRNAs significativement corrélés avec les sous-types moléculaires de cancers mammaires, 14 lncRNAs avec les grades et 3 lncRNAs avec l'évolution clinique et la survie des patientes (Zhao et al., 2014).
- L'analyse du profil d'expression de 1623 lncRNAs dans une série de 658 tumeurs mammaires a conduit à l'identification de 4 *clusters* corrélés à la classification PAM50. L'analyse non supervisée a permis d'identifier 4 groupes de lncRNAs de pronostic différent : le groupe 1 lié aux tumeurs de type *basal-like*, le groupe 2 associé aux tumeurs HER2+, le groupe 3 lié aux tumeurs lumineales A et le groupe 4 associé aux tumeurs lumineales B (Fig. 28) (Su et al., 2014).
- L'étude des profils transcriptomiques des lncRNAs dans une série de 995 tumeurs mammaires a mis en évidence 215 lncRNAs dérégulés par rapport au tissu normal. Le *clustering* non supervisé basé sur les lncRNAs a révélé 4 sous-groupes corrélés aux sous-types de la signature PAM50 définis par des mRNAs. Le profil des lncRNAs permet de distinguer les tumeurs RH+ des tumeurs RH-. Les lncRNAs dérégulés sont

impliqués dans les principales voies de signalisation oncogéniques. Ainsi, des corrélations significatives ont été observées entre (i) le sous-groupe de lncRNAs spécifiques du sous-type luminal A et l'activation de PI3K, FGF et TGF β , (ii) entre le sous-groupe *basal-like* et EGFR, IL6 et l'EMT, (iii) le sous-groupe HER2 et mTOR, p53, PI3K et (iv) le sous-groupe luminal B et MYC, MAPK et RAF. Les sous types tumoraux contenant des lncRNA ou des mRNA activent ou inhibent globalement les mêmes voies à l'exception d'EGFR. Une signature pronostique comportant 27 lncRNAs a été identifiée (van Grembergen et al., 2016).

Les différents sous-types moléculaires des carcinomes définis soit par des lncRNAs, soit par des mRNAs, sont globalement associés à l'activation ou à l'inhibition des mêmes voies de signalisation, à l'exception des lncRNAs spécifiques du sous-type *basal-like* des carcinomes mammaires impliqués dans l'activation de la voie EGFR (van Grembergen et al., 2016).

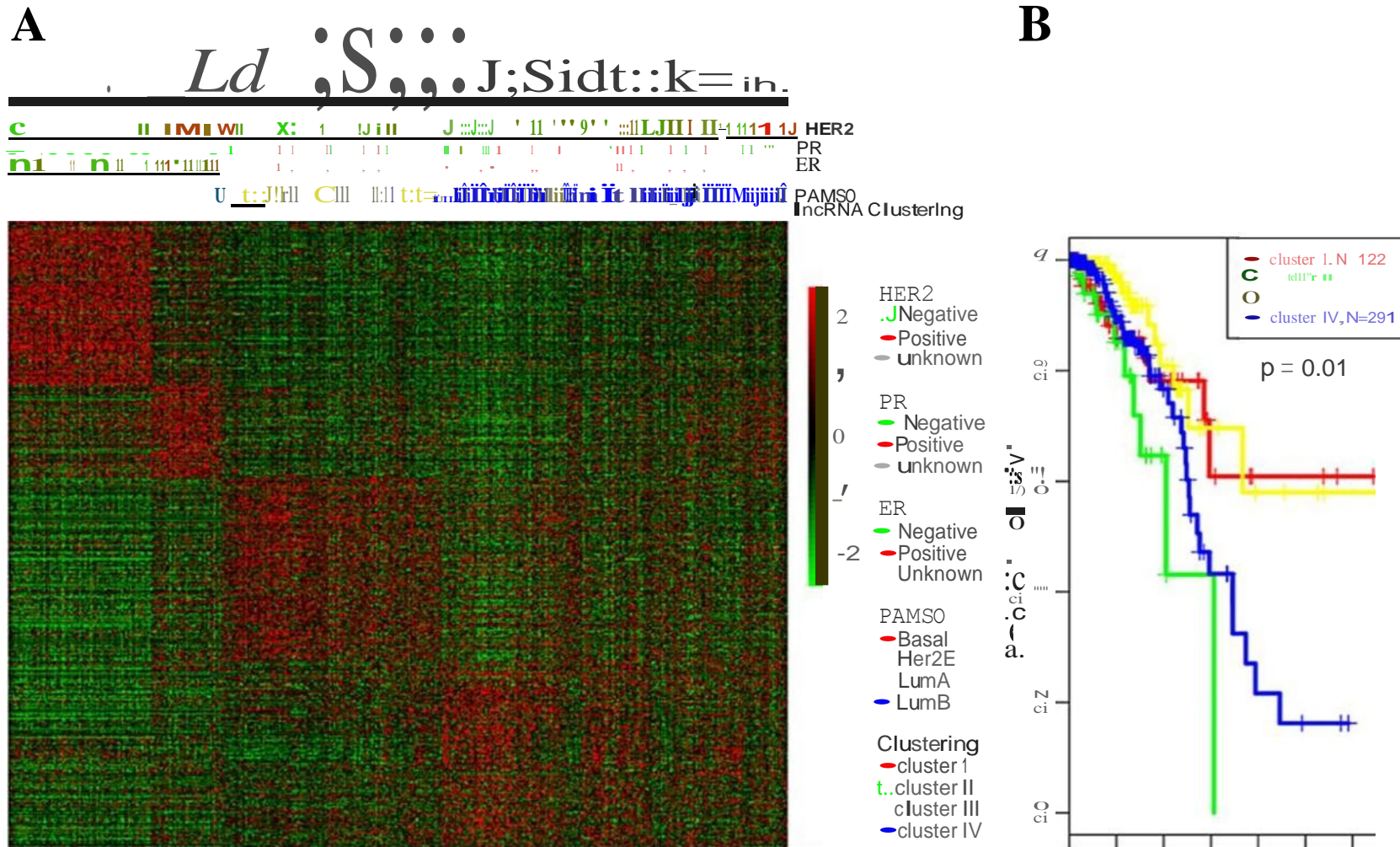


Figure 28 : Classification moléculaire intégrant des longs ARN non codants. A: Clustering non supervisé identifiant les différents sous-groupes; B: courbe de survie (D'après Su et al., 2014)

5 LES LONGS ARNs NON CODANTS COMME BIOMARQUEURS ET CIBLES THERAPEUTIQUES POTENTIELS

5.1 Les lncRNAs comme biomarqueurs pronostiques et prédictifs

5.1.1 Généralités

La découverte d'une plus grande spécificité cellulaire et tissulaire des lncRNAs ainsi que leur capacité à établir de multiples interactions rend possible leur utilisation comme biomarqueurs et cibles thérapeutiques potentielles (**Tableau 7**). L'utilisation de ncRNAs comme biomarqueurs diagnostiques et pronostiques présente des avantages par rapport aux mRNAs traduits en protéines (Wapinski and Chang, 2011 ; Wahlestedt, 2013). Par rapport aux mRNAs dont les niveaux d'expression ne sont que le reflêt indirect de l'activité fonctionnelle de la protéine active, le niveau d'expression des lncRNAs mesure exactement l'activité de ces molécules. A l'inverse des mRNAs qui sont exprimés par de nombreux sous-types de tumeurs malignes, les lncRNAs semblent présenter une plus grande spécificité cellulaire et tissulaire. De plus, les lncRNAs sécrétés ou libérés par les cellules tumorales apoptotiques ou nécrotiques peuvent être détectés dans le sang, le plasma, la salive et l'urine (Tong and Lo, 2006 ; Ayers, 2013). En outre, les lncRNAs sont souvent stables dans le sérum humain et peuvent de ce fait être mesurés par les techniques non-invasives individuelles (qPCR) ou pangénomiques (RNA-seq). Enfin, il semble exister une corrélation significative entre les niveaux d'expression des lncRNAs circulants et les altérations génomiques, épigénétiques ou transcriptionnelles observées dans les tumeurs malignes.

5.1.2 Les lncRNAs comme biomarqueurs dans divers types de cancers

- **Sein** : Hormis *ANRIL* et *MALAT1*, les principaux autres lncRNAs oncogéniques dérégulés dans les cancers du sein sont *UCA1*, *LSINCT5*, *LncRNA-Smad7*, *LOC554202*, *lncRNA-JADE*, *PTV1*, *CYTOR*, *treRNA*, *HOTAIR*, *HOTAIRM1*, *BCYRN1/BC200*, *EFNA3*, *NEAT1*, *CCAT2*, *SOX2OT*, *ANRASSF1*, *MIR31HG*, *SPRY4-IT1*, *ARA*, *DSCAM-AS1*, *α -HIF*, *BCAR4* et *FAL1*. Les lncRNAs oncosuppresseurs dérégulés comprennent *GAS5*, *PTENP1*, *Zfas1*, *Xist*, *NKILA* et *LIMT*. Les lncRNAs controversés dans la cancérogénèse mammaire sont *H19* et *SRA*. La plupart des

lncRNAs associés spécifiquement au sous-type *basal-like* apparaissent sous-exprimés (van Grembergen et al., 2016). *HOTAIR* est significativement surexprimé dans les cancers du sein HER2+ et *HOTAIRM1* dans les cancers *basal-like* (Su et al., 2014). La surexpression de LINC00472 est significativement corrélée avec un risque moindre de récurrence et de mortalité (Shen et al., 2015). La surexpression de *NKILA* est associée à un bon pronostic (Liu et al., 2015b) et la sous-expression de *SNHG17* est corrélée à une plus longue survie (Zhao et al., 2014). La surexpression d'*a-HIF* est un facteur de mauvais pronostic dans les cancers du sein (Gutschner and Diederichs, 2012).

- **Œsophage** : Les lncRNAs *CCAT2* et *PCAT1* sont surexprimés dans les carcinomes œsophagiens et corrélés à un pronostic défavorable (Chen et al., 2014b).
- **Estomac** : *HOTAIR*, *CCAT1*, *GHEI1*, *MALAT1*, *PVT1*, *ANRIL*, *HULC*, *GACAT1/2/3*, *H19* et *SUMO1P3* sont surexprimés dans les carcinomes gastriques et les lncRNAs oncosuppresseurs *MEG3* et *GAS5* sont sous-exprimés. Ce profil est souvent associé à l'existence de métastases ganglionnaires et viscérales (Hajjari et al., 2013a ; Sun et al., 2013 ; Mei et al., 2013 ; Fang et al., 2015).
- **Intestin** : Les principaux lncRNAs impliqués la cancérogénèse colique sont *MALAT1*, *H19*, *CCAT1*, *uc.73*, *CCAT2*, *HOTAIR*, *PRNCRI*, *PVT1*, *CRNDE*, *GAS5* et *lincRNA-p21* (Hu et al., 2014b). La surexpression de *HOTAIR*, *MALAT1* et *uc.73* est observée dans les carcinomes colorectaux de stade avancé présentant des métastases hépatiques (Sana et al., 2012).
- **Foie** : *HULC*, *H19*, *MALAT-1* et *HOTAIR* sont les principaux lncRNAs surexprimés dans les hépatocarcinomes (CHC) (Xie et al., 2013 ; Hu et al., 2014a). De plus, *HOTAIR* est un biomarqueur pronostique de récurrence après transplantation hépatique (Yang et al., 2011c). *HOTTIP* et *HOXA13* sont également surexprimés et leurs niveaux d'expression sont corrélés à la progression tumorale, la survenue de métastases et la survie (Quagliata et al., 2014b). *HEIH* est un autre lncRNA surexprimé dans les CHC et un facteur pronostique indépendant associé à la récurrence (Yuan et al., 2012).

- **Poumon** : *MALAT1* est un biomarqueur pronostique clé d'évolution métastatique dans les adénocarcinomes pulmonaires (Gutschner et al., 2013a,b). *HOTAIR* est fortement surexprimé dans les carcinomes pulmonaires non à petites cellules (CPNPC) de stade avancé ou comportant des métastases ganglionnaires (Chen et al., 2014a). *BANCR* et *TUG1* sont sous-exprimés dans les CPNPCs et leurs faibles niveaux d'expression sont associés à un stade TNM avancé, une grande taille tumorale et une survie globale réduite (Sun et al., 2014b ; Zhang et al., 2014c).
- **Glioblastome** : Le lncRNA oncogénique le plus surexprimé est *CRNDE* et le lncRNA oncosuppresseur *MEG3* est sous-exprimé dans les glioblastomes (Graham et al., 2011 ; Ellis et al., 2012). *H19* et le miRNA miR-675 appartiennent à une voie de signalisation majeure impliquée dans la progression tumorale des glioblastomes et la surexpression de *H19* est observée dans les gliomes de haut grade (Shi et al., 2014).
- **Vessie** : *UCA1*, *H19*, *TUG1*, *Linc-UBC1*, *MALAT1*, *CUDR*, *HIF1 α -AS* et *linc-UBC1*, sont surexprimés dans les carcinomes de vessie et activent les voies de signalisation PI3K-Akt et Wnt/ β -caténine (Zhang et al., 2013 ; Wang et al., 2014d). Le lncRNA oncosuppresseur *MEG3* est sous exprimé. Le dosage de *UCA1* dans les urines s'est avéré être particulièrement utile pour les patients atteints de tumeurs vésicales superficielles de grade G2/G3 et présentant un risque élevé d'invasion musculaire.
- **Prostate** : *PCA3*, *PCGEM1*, *PCAT1* et *ANRIL* sont fortement surexprimés dans les carcinomes de la prostate (Walsh et al., 2014 ; Crea et al., 2014b ; Fang et al., 2016a). *PCA3* peut être détecté dans les tissus tumoraux prostatiques, l'urine et/ou les sédiments urinaires. Sa spécificité est plus élevée que celle du biomarqueur classique PSA (Ploussard et al., 2011). Le dosage de *PCA3* pourrait être utilisé comme un test urinaire non invasif destiné au dépistage à grande échelle des carcinomes prostatiques et à la prédiction de leur agressivité (Roobol et al., 2010).
- **Peau** : *BANCR* est surexprimé dans les mélanomes malins et favorise la prolifération et l'invasion des cellules mélanocytaires atypiques *via* l'activation des voies ERK1/2 et JNK MAPK. Sa surexpression est corrélée à un taux de survie plus faible (Li et al., 2014c).

- **Ovaire** : *HOTAIR*, *HOST2*, *MALAT1* et *LSINCT5* sont surexprimés dans les tumeurs ovariennes et fortement corrélés à un stade FIGO avancé, un grade histologique élevé, la survenue de métastases ganglionnaires et une survie globale réduite (Qiu et al., 2014a).
- **Pancréas** : *SNHG8*, *PVT1*, *H19*, *HOTAIR*, *MIAT*, *GAS5*, *LINC00261*, *LINC00152*, *HNFI1A-AS1* et *AFAP1-AS1* apparaissent dérégulés dans les cancers pancréatiques (Müller et al., 2015).

LncRNA	Taille	Locus	Mécanismes et fonctions	Nature	Cancer	Références
ANRIL	3.9 kb	9q21.3	répression en cis du locus CDKN2A/2B via PRC1/PRC2	oncogène	prostate, sein, poumon colon, ovaire, foie	Aguilo et al. (2016)
HOTAIR	2.2 kb	12q13.13	répression multigénique en trans via PRC2/LSD1	oncogène	sein, foie, poumon, estomac, colon	Hansji et al. (2014)
MALAT1	8 kb	11q13.1	modulation de l'épissage alternatif	oncogène	sein, colon	Gutschner et al. (2013a)
H19	2.3 kb	11p15.5	empreinte génomique	oncogène	sein, vessie, col, colon, foie, pancréas	Matouk et al. (2014)
UCA1	1.4 kb, 2.2 kb, 2.7 kb	19p13.12	régulation du cycle cellulaire via CREB	oncogène	sein, vessie	Li et al. (2014b)
NEAT1	3.7 kb, 23 kb	11q13.1	régulation transcriptionnelle et post-transcriptionnelle	oncogène	prostate, sang	Zeng et al. (2014)
HULC	0.5 kb	6p24.3	interacteur de CREB	oncogène	foie, colon, estomac, pancréas	Zhang et al. (2015)
KCNQ1OT1	91.5 kb	11p15.5	empreinte génomique	oncogène	estomac, colon	Kanduri, 2011
CCAT1	2.6 kb, 5.2 kb	8q24	régulation de Myc	oncogène	foie, estomac, colon	Deng et al. (2015)
CCAT2	0.34 kb	8q24.21	Régulation transcriptionnelle	oncogène	sein, poumon, colon	Ling et al. (2013)
BC200	0.2 kb	2p21	modulation de la traduction	oncogène	multiples	Iacoangeli and Tiedge. (2013)
SRA	2 kb	5q31.3	régulation via les hormones stéroïdes et PRC2	oncogène	prostate, sein, ovaire	Novikova et al. (2012)
TUG1	6.7 kb	22q12.2	interaction avec PRC2 et répression génique	oncogène	vessie, poumon	Nakagawa et al. (2012)
anti-NOS2A	1.9 kb	17q23.2	répression de Nos2A	oncogène	cerveau	Korneev et al. (2008)
lncRNA-ATB	2.4 kb	14q11.2	activation par TGF- β	oncogène	foie	Shi et al. (2015b)

CRNDE	10.3 kb	16q12.2	interaction avec PRC2 et CoREST régulation par l'insuline et IGF	oncogène	estomac, colon, cerveau, ovaire	Liu et al. (2016b)
CUDR	2.2 kb	19p13.12	résistance médicamenteuse	oncogène	poumon, vessie, col, estomac, colon	Pu et al. (2015)
HEIH	1.6 kb	5q34.3	répression des gènes cibles de PRC2 via EZH2	oncogène	foie	Zhang et al. (2015)
ncRAN	2 kb	17q25.1	interaction avec N-myc	oncogène	vessie, estomac, colon, cerveau	Buechner et al. (2012)
PCA	2, 4 et 23 kb	9q21.2	signalisation du RA	oncogène	prostate	Crea et al. (2014b)
PCAT1	7.8 kb	8q24.21	inhibition de la recombinaison homologue	oncogène	prostate	Wen et al. (2016)
PCGEM	27 kb	2q32	activation des gènes régulés par le RA	oncogène	prostate	Prensner et al. (2014b)
PCNA-AS1	384 bp	20p12.3	accroissement de la stabilité du mRNA de PCNA	oncogène	foie, estomac	Yuan et al. (2014)
PlncRNA-1	245 pb	21q22.12	interaction avec le RA	oncogène	prostate	Fang et al. (2016)
PRNCR1	13 kb	8q24.22	activation des gènes régulés par le RA	oncogène	prostate	Prensner et al. (2014b)
PVT1	>300 kb	8q24	Répression épigénétique, MYC, miR-200, Rspo1	oncogène	foie, sein, ovaire, estomac, poumon	Colombo et al. (2015)
SChLAP1	224.8 kb	2q31.3	inhibition de la fixation aux complexes SWI/SNF	oncogène	prostate	Atala. (2014)
AK126698	3.8 kb	1q24.2.	résistance au cisplatine via la voie Wnt	oncogène	poumon	Yang et al. (2013b)
BANCR	693 bp	9q21.11	régulation de la voie MAPK	oncosuppresseur	peau, rétine, poumon	Liu et al. (2016a)
GAS5	lncRNAs et snoRNAs	1q25.1	régulation hormonale	oncosuppresseur	sein, prostate, estomac, col, rein	Sun et al. (2014a)
LET	2.6 kb	15q24.1	répression par HDAC3 sous hypoxie	oncosuppresseur	poumon, foie, estomac, colon	Liu et al. (2016c)
LincRNA-p21	3 kb	6p21.2	répression de gènes induits par TP53 via hnRNP-K	oncosuppresseur	poumon, estomac, colon, sang	Tang et al. (2015)

MEG3	1.6 kb	14q32.2	activation de la voie p53	oncosuppresseur	prostate, vessie, hypophyse, cerveau	Guo et al. (2016)]
PTEN-P1	4 kb	9p13.3	augmentation de l'expression de PTEN	oncosuppresseur	prostate, TD	Deng and Sui. (2013)
XIST	19 kb	Xq13.2	empreinte génomique, inactivation du chromosome X, fixation à PRC2	oncosuppresseur	sein, ovaire, col	Lee. (2009)

Tableau 7 : Principaux lncRNAs pouvant représenter des biomarqueurs et des cibles thérapeutiques potentiels (D'après Wahlestedt, 2013 ; Roshan et al. ; 2013).
Pages 152-153-154.

5.2 Les lncRNAs comme cibles thérapeutiques potentielles

5.2.1 Généralités

En raison de leur rôle central dans la régulation de l'expression des gènes et de leur spécificité cellulaire, les lncRNAs pourraient représenter des cibles thérapeutiques potentielles prometteuses. Le ciblage des lncRNAs peut être réalisé en utilisant plusieurs approches, incluant l'inactivation médiée par les siRNAs, le blocage fonctionnel utilisant de petites molécules, des inhibiteurs oligonucléotidiques empêchant les interactions entre les lncRNAs et des protéines et la désorganisation de leur structure par de petites molécules ou des inhibiteurs oligonucléotidiques modifiant ou mimant leur structure secondaire de manière à les empêcher de se fixer à leurs partenaires (Sanchez and Huarte, 2013; Fatemi et al., 2014; Takahashi and Caminci, 2014a). Les stratégies actuellement employées pour cibler les lncRNAs sont issues de procédés utilisant des acides nucléiques. Parmi celles-ci, les techniques basées sur l'interférence par l'ARN (ARNi) sont sans doute les méthodes les plus couramment utilisées pour inhiber les lncRNAs dans les cellules cancéreuses. Les siRNAs et shRNAs présentent habituellement une grande sélectivité et sont dans l'ensemble efficaces dans l'inactivation des RNAs. Cependant, d'autres méthodes incluant les oligonucléotides antisens (ASO), les ribozymes, les aptamères et les petites molécules inhibitrices peuvent également être efficaces pour moduler les lncRNAs et présentent des avantages par rapport aux siRNAs. Néanmoins, si certains essais thérapeutiques sur des modèles murins laissent entrevoir des espoirs, de nombreux lncRNAs sont spécifiques des primates ou des humains et ne peuvent être étudiés *in vivo* par des modèles classiques de *knockdown/knockout*. En outre, s'il est actuellement bien établi que la dérégulation des lncRNAs est associée à de nombreux cancers, la question demeure quant à la signification de cette dérégulation, cause ou conséquence de la maladie.

5.2.2 Techniques de ciblage thérapeutique des lncRNAs

5.2.2.1 SiRNAs

Les petits ARNs interférents (siRNAs) sont de courtes séquences (19-30 nt) d'ARN double brin qui ciblent les molécules d'ARN par leur complémentarité avec des séquences non appariées de lncRNAs. Les siRNAs sont totalement complémentaires de leurs cibles d'ARN

qui sont ensuite clivées au niveau d'une simple liaison phosphodiester située près du centre de la séquence complémentaire de celle des siRNAs. Ils présentent un rendement d'inactivation élevé vis-à-vis de nombreux lncRNAs oncogéniques surexprimés dans les cellules cancéreuses et induisent des effets antitumoraux à la fois *in vitro* et *in vivo*.

- **HOTAIR** est un biomarqueur diagnostique et pronostique potentiel surexprimé dans les cancers du sein, du poumon, du foie et du colon (Yao and Yang, 2014). Son inhibition par des siRNAs diminue le caractère invasif des cellules cancéreuses mammaires et inhibe la croissance tumorale des xénogreffes de carcinomes pancréatiques (Bhan et al., 2013 ; Kim et al., 2013b). L'inhibition de **HOTAIR** est également associée à une sensibilité accrue des cellules cancéreuses au TNF α induit par la réponse immune et au cisplatine et à la doxorubicine (Yang et al., 2011c).
- **MALAT1** est un biomarqueur pronostique potentiel pour divers types de cancers et son inhibition par des siRNAs peut s'accompagner d'une réduction partielle de la viabilité et de la migration des cellules cancéreuses et d'un accroissement de l'apoptose (Lai et al., 2012).
- **PCA3**, **PCAT1**, **PCGEM** et **PRNCRI** sont considérés comme des biomarqueurs potentiels surexprimés dans les carcinomes prostatiques agressifs. Leur inactivation par des siRNAs induit un arrêt de la prolifération et provoque une apoptose *via* l'inhibition du récepteur androgénique (Cui et al., 2013 ; Crea et al., 2014b ; Yang et al., 2014b).
- **H19**, **HULC**, **HEIH** et **MVIH** apparaissent surexprimés dans les hépatocarcinomes et leur inhibition par des siRNAs se traduit par une réduction de la croissance tumorale des xénogreffes (Yang et al., 2011a ; Du et al., 2012).
- **H19**, **CUDR** et **HIF1A-AS** sont des biomarqueurs des carcinomes de vessie dont l'activité oncogénique est réduite suite à l'utilisation de siRNAs (Dorsett et al., 2004 ; Bertozzi et al., 2011 ; Wang et al., 2012).

Des essais cliniques de phase I et II utilisant des siRNAs sont en cours d'évaluation, de même que des thérapies combinant des siRNAs et des chimiothérapies conventionnelles (Dorsett et

al., 2004 ; Burnett et al., 2011). Toutefois, le principal obstacle des thérapies utilisant les siRNAs reste leur faible biodisponibilité. En particulier, les siRNAs utilisant des acides ribonucléiques sont sensibles à la dégradation par la nucléase et ont de ce fait une pharmacocinétique peu performante.

5.2.2.2 ASO

Les oligonucleotides antisens (ASO) sont de courtes molécules d'ADN ou d'ARN simple brin (8 à 50 nt) dotées de séquences spécifiques destinées à cibler entre autres des lncRNAs. Les ASOs s'hybrident directement aux transcrits des gènes codant pour les lncRNAs par appariement des bases et les hybrides ainsi formés sont ensuite reconnus par la RNase H1 endogène, résultant en un clivage des molécules de lncRNAs (Bennett and Swayze, 2010). Les ASOs sont plus aisés à utiliser que les siRNAs qui sont plus instables et dont les capacités de ciblage sont moins performantes *in vivo* (Veedu and Wengel, 2010). Parmi les ASOs, les antagoNATs sont dirigés contre les lncRNAs antisens et utilisés pour majorer l'expression de mRNAs spécifiques *via* l'inhibition du lncRNA antisens correspondant (Modarresi et al., 2012). L'inhibition de *MALATI* par les ASOs atténue divers phénotypes malins *via* l'arrêt du cycle cellulaire dans les cellules cancéreuses du col utérin (Tripathi et al., 2013). Dans des modèles xéno greffés de cancers pulmonaires, l'injection d'ASOs dans des tumeurs sous-cutanées de souris *nude* inhibe efficacement *MALATI in vivo* et bloque l'apparition de métastases (Gutschner et al., 2013a). Certains ASOs sont l'objet de premiers essais cliniques (Watts et al., 2012). Globalement, les ASOs peuvent être utilisés contre les NATs (AntagoNATs), les interactions lncRNAs-complexes de remodelage chromatinien (PRC2), les eRNAs et les facteurs de transcription.

5.2.2.3 Ribozymes

Les ribozymes sont des molécules d'ARN produites naturellement qui présentent des fonctions intracellulaires catalytiques, parmi lesquelles la dégradation des molécules d'ARN (Citti and Rainaldi, 2005). Parmi les différents types de ribozymes, les ribozymes « en tête de marteau » (*Hammerhead ribozyme, HamRz*) semblent présenter un intérêt thérapeutique majeur car ils possèdent des actions inhibitrices efficaces tout en ayant le plus petit motif d'ARN endoribonucléolytique (Ruffner et al., 1990 ; Tedeschi et al., 2009).

5.2.2.4 Aptamères

Un obstacle à l'utilisation des thérapies antisens est la présence de structures secondaires étendues au sein des molécules de lncRNAs, qui réduisent l'efficacité des siRNAs, des ASOs et des ribozymes (Luo and Chang, 2004 ; Gredell et al., 2008). Les aptamères sont de courts oligonucléotides d'ADN, d'ARN ou de peptides qui ont une structure tridimensionnelle stable *in vivo* (Watrin et al., 2009). Ils présentent un large spectre de cibles moléculaires dont les protéines, l'ARN et de petites molécules (Mayer et al., 2007). Ils se lient spécifiquement aux lncRNAs *via* leur structure tridimensionnelle. Des données récentes suggèrent que les aptamères pourraient constituer des agents thérapeutiques prometteurs par la dégradation ou l'inhibition des fonctions des lncRNAs (Darfeuille et al., 2006 ; Kotula et al., 2012).

5.2.2.5 Petites molécules inhibitrices modulant l'activité des lncRNAs

Ces petites molécules sont synthétisées de manière à se lier de façon compétitive aux lncRNAs et à bloquer les interactions avec leurs partenaires protéiques (Fatemi et al., 2014). L'action de ces petites molécules peut également induire un changement conformationnel dans les lncRNAs et perturber la formation de structures secondaires dotées de fonctions importantes. Ainsi, l'inhibition de l'interaction de *HOTAIR* avec PRC2/LSD1 par ces molécules pourrait réduire la survenue de métastases dans les carcinomes mammaires (Tsai et al., 2011).

5.2.2.6 Plasmides

Une nouvelle approche thérapeutique non spécifique utilisant un plasmide, BC-819/DTAH19, permet d'injecter aux cellules tumorales une sous-unité de la toxine diphtérique et de la placer sous le contrôle du promoteur du lncRNA *H19*. La production de toxines diphtériques dans les cellules cancéreuses urothéliales, pulmonaires, coliques, ovariennes et pancréatiques s'accompagne plutôt d'une réduction de la taille des tumeurs que de l'inhibition de lncRNAs (Mizrahi et al., 2009 ; Gibb et al., 2011).

5.2.2.7 Thérapie génique

Des essais de thérapie génique pourraient être proposés pour compenser la baisse du niveau d'expression de certains lncRNAs dont l'activité onco-suppressive est inhibée au cours de la cancérogénèse (Wahlestedt, 2013): *PTCSC3* dans les carcinomes papillaires de la thyroïde (Fan et al., 2013), *PTENP1* dans les carcinomes coliques (Poliseno et al., 2010), *MEG3* dans les méningiomes et les glioblastomes (Zhang et al., 2010), *LincRNA-p21* dans les carcinomes colo-rectaux, les carcinomes pulmonaires et les lymphomes (Huarte et al., 2010).

5.2.2.8 LncRNAs circularisés

La dérégulation du niveau d'expression de nombreux lncRNAs circulaires dans les tissus tumoraux par rapport au tissu sain, a été récemment observée au cours de la cancérogénèse (surexpression de ciRS-7 dans les neuroblastomes, les astrocytomes, les carcinomes rénaux et pulmonaires ; sous-expression de cir-ITCH dans les carcinomes oesophagiens et colorectaux). Certains circRNAs sont associés à des miRNAs liés au cancer (ciRS-7 et/ miR-7/miR671 dans les cancers du sein, les hépatocarcinomes et les cancers du col utérin). Les circRNAs régulent des voies de signalisation associées au cancer (ciRS-7 et voie EGFR) et la majorité d'entre eux sont régulés par l'EMT. Des translocations génèrent des circRNAs de fusion oncogéniques (F-circRNAs) en plus des oncoprotéines de fusion et contribuent à la transformation cellulaire et à la résistance aux thérapies (MLL/AF9 et f-circM9) (Guarnerio et al., 2016). Les circRNAs pourraient représenter des biomarqueurs pertinents (circRNAs circulants dans le sérum, le plasma, la salive, les urines) et des vecteurs thérapeutiques innovants (circularisation des « éponges » à miRNAs).

6 PERSPECTIVES

Les lncRNAs représentent une nouvelle famille de molécules impliquées dans l'édification de l'architecture nucléaire, la coordination de l'expression des gènes, la modulation des grandes fonctions physiologiques et l'équilibre homéostatique cellulaire et tissulaire. Les conséquences d'une expression aberrante des lncRNAs au cours de la carcinogénèse peuvent diversement associer une désorganisation des structures intranucléaires, la dérégulation de

l'expression des gènes et l'activation/inactivation de multiples voies de signalisation. Leur utilisation potentielle comme biomarqueurs et cibles thérapeutiques est prometteuse.

Toutefois, de nombreuses questions demeurent concernant les lncRNAs. Leur définition demeure actuellement ambiguë et il n'existe pas de réelle classification fonctionnelle. Les lncRNAs sont des molécules très hétérogènes présentant de nombreux variants d'épissage. Les domaines structuraux et fonctionnels des lncRNAs et la nature de leurs interactions avec l'ADN, l'ARN et les protéines, sont largement méconnus. Bien que leur implication dans le cancer soit établie et leurs mécanismes d'action mieux connus, il n'est pas encore formellement exclu que les lncRNAs soient en partie des produits situés en aval de certaines structures chromatinienne ou de processus transcriptionnels dérégulés. En outre, leur fonctionnalité est actuellement encore débattue en raison de leur faible niveau d'expression et de conservation entre les espèces. En raison d'une divergence évolutive survenue récemment, la majorité des lncRNAs sont observés chez les primates, avec pour conséquence une moindre pertinence des modèles murins pour leurs investigations et donc la nécessité d'utiliser des modèles humanisés, des xéno greffes et des cultures organoïdes. De plus, les mécanismes moléculaires sous-tendant leurs fonctions moléculaires ne sont pas encore totalement élucidés, les lncRNAs possédant souvent plus d'un mécanisme d'action et leurs fonctions semblant modifiées dans les tissus cancéreux par rapport aux tissus sains (*Xist*). En outre, de nouvelles fonctions ont été très récemment découvertes, même pour les lncRNAs déjà connus et analysés (*H19* et *Xist*), d'où l'importance d'utiliser des méthodes innovantes, telles le CRISPr, les études de perte de fonction *in vivo* et de délétion du locus ou du promoteur. Enfin, les lncRNAs interagissent dans le cadre de réseaux très complexes incluant des mRNAs, des miRNAs, des lncRNAs et des facteurs de transcription, qui sont dérégulés au cours de la cancérogénèse.

Ainsi, les principaux défis durant les prochaines années concerneront l'identification et la cartographie de tous les lncRNAs appartenant au génome humain, leur caractérisation fonctionnelle et l'identification de leurs réseaux d'interactions. Ces défis ne pourront être atteints que grâce aux efforts combinés de la génomique fonctionnelle, de l'épigénomique et de la bioinformatique.

PROBLEMATIQUE DE LA RECHERCHE

L'importante hétérogénéité des cancers du sein complique le pronostic, l'approche thérapeutique et la compréhension des mécanismes de résistance aux traitements.

Actuellement, les récepteurs des estrogènes (ER) et de la progestérone (PR), le récepteur 2 du facteur de croissance épidermique humain (HER2), l'index Ki67 et la signature PAM50 sont pris en compte dans la classification histologique et l'orientation thérapeutique. La recherche de nouveaux biomarqueurs et de nouvelles cibles thérapeutiques s'avère donc impérative et est l'objet de recherches intensives.

Les études pangénomiques systématiques réalisées durant la dernière décennie ont révélé un large spectre de lncRNAs impliqués dans la cancérogénèse.

La fonction globale des lncRNAs est de coordonner l'expression des gènes codant pour les protéines aux niveaux épigénétique, transcriptionnel et post-transcriptionnel. Ils ont été également impliqués dans la régulation de nombreuses fonctions cellulaires fondamentales, dont la croissance, l'apoptose, le cycle, la réparation de l'ADN, le métabolisme, le maintien de la pluripotence, l'angiogénèse et la migration. En raison de leur plus grande spécificité tissulaire, les lncRNAs pourraient aussi servir de biomarqueurs cliniques très performants à des fins diagnostiques et pronostiques mais aussi comme cibles thérapeutiques potentielles.

Dans le cancer du sein, plusieurs lncRNAs ont été très récemment impliqués dans la progression tumorale des carcinomes mammaires.

La plupart ont été identifiés comme des oncogènes et plus rarement comme des suppresseurs de tumeur. En outre, certains lncRNAs semblent sécrétés par les cellules cancéreuses dans des exosomes, suggérant qu'ils pourraient également jouer un rôle crucial dans la dissémination métastatique *via* leurs interactions avec les populations cellulaires des niches pré-métastatiques. Cependant, le rôle exact des lncRNAs dans la cancérogénèse mammaire et leur intérêt réel dans la pratique clinique demeure encore largement méconnu et se doit d'être clarifié.

L'objectif principal du projet vise à analyser l'expression des lncRNAs, leur fonctionnalité et leur rôle dans l'oncogénèse mammaire

en recherchant plus particulièrement l'intérêt potentiel de ces molécules comme biomarqueurs à visée diagnostique, pronostique ou thérapeutique.

La première partie de la thèse s'est focalisée sur l'étude de deux lncRNAs, ANRIL et MALAT1, fortement impliqués dans de nombreux types de cancers, mais dont les mécanismes d'action biologiques et la signification clinique au cours de la cancérogénèse mammaire sont encore controversés et/ou méconnus.

ANRIL

Localisation génomique

Le lncRNA *CDKN2B-AS1*, communément appelé *Antisense Non-coding RNA in the INK4 Locus (ANRIL)*, est un lncRNA naturel antisens situé chez l'homme dans la région chromosomique 9p21.3 au niveau du gène *INK4b*. *ANRIL* chevauche une région critique de 35 kb englobant trois principaux suppresseurs de tumeur juxtaposés et réunis en un *cluster p16-CDKN2A/p15-CDKN2B/p14-ARF*, ainsi que le gène de la méthyl-thioadénosine phosphorylase (*MTAP*). *ANRIL* a été initialement identifié au sein d'une délétion germinale de 403 kb, observée parmi les membres d'une famille française porteurs de mélanomes et de tumeurs du système nerveux (Pasmant et al., 2007).

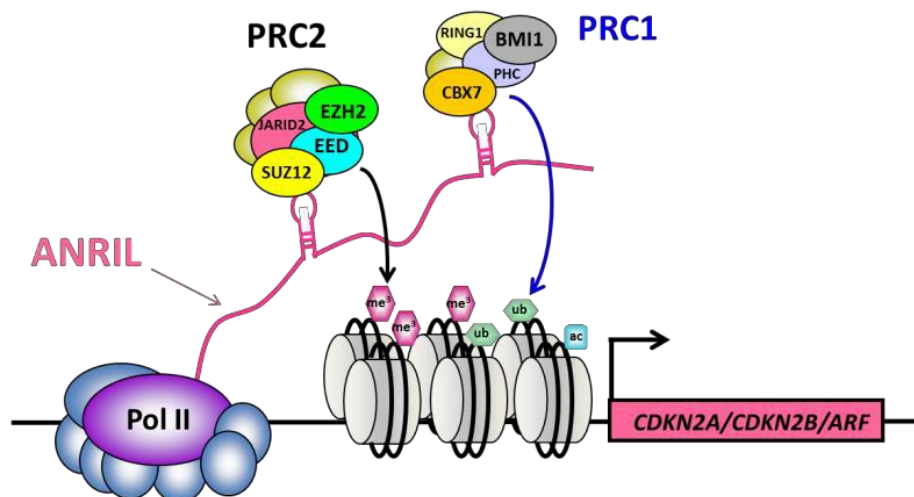


Figure 29 : Interactions du lncRNA ANRIL et des complexes polycomb PRC2/PRC1 au niveau du locus *p16-CDKN2A/p15-CDKN2B/p14-ARF* (D'après Congrains et al., 2013)

Caractéristiques moléculaires

ANRIL est transcrit par l'ARN polymérase II en un lncRNA de 3.834 pb à partir du brin opposé au *cluster p16-CDKN2A/p15-CDKN2B/p14-ARF* (Yu et al., 2008). Il partage un promoteur bidirectionnel avec *p14-ARF*, l'extrémité 5' du premier exon d'*ANRIL* étant situé à 300 pb en amont du site d'initiation de la transcription (TSS) du gène *p14-ARF*. En conséquence, l'expression des deux gènes est coordonnée et l'activation de la transcription de ce promoteur divergent est assurée par le facteur de transcription E2F1 et l'insulateur CTCF (Sato et al., 2010 ; Rodriguez et al., 2010). *ANRIL* contient 20 exons dont la plupart consistent en des séquences répétitives de type LINE, SINE et Alu pouvant être épissées alternativement (Jarinova et al., 2009). Les nombreux transcrits d'*ANRIL* peuvent être classés en 4 grands sous-types d'isoformes épissés linéaires et circulaires et sont exprimés à des niveaux habituellement très faibles. Parmi les isoformes linéaires, les plus abondants sont les deux isoformes courts se terminant avec l'exon 13 polyadénylé et l'isoforme long dépourvu d'exon 13 et se terminant par l'exon 20 polyadénylé. Des isoformes circulaires non polyadénylés d'expression plus abondante ont été récemment décrits, dont 2 lncRNAs mesurant 3.9kb et 34.8kb, résultant d'un mécanisme d'*exon skipping* survenu lors de l'épissage alternatif (Yu et al., 2008 ; Folkensen et al., 2009 ; Kotake et al., 2011). De nombreux isoformes d'*ANRIL* peuvent en réalité coexister dans un même type de cellules, ou au contraire être spécifiques de tissus (Folkensen et al., 2009), augmentant ainsi la complexité de ses mécanismes d'action et de régulation.

Mécanismes d'action

Des études mécanistiques ont révélé qu'*ANRIL* peut réguler ses gènes voisins immédiats en *cis* en servant de plateforme pour les complexes modifiant la chromatine, tels PRC1 et PRC2, et médiant la répression en *cis* du locus *p16-CDKN2A/p15-CDKN2B/p14-ARF* (Yap et al., 2010 ; Kotake et al., 2011) (**Fig. 29**). Un mécanisme d'action possible pour *ANRIL* serait que plusieurs de ses variants d'épissage interagiraient spécifiquement avec le complexe PRC1 *via* CBX7 à proximité du promoteur du gène *p16-INK4a* afin de réprimer sélectivement ce gène (Aguilo et al., 2016). L'épissage alternatif pourrait ainsi modifier la structure d'*ANRIL*, conduisant à des changements dans la régulation du locus *p16-CDKN2A/p15-CDKN2B/p14-ARF*, médiée par les protéines des complexes polycomb. La présence de multiples *enhancers*

dans cette région suggère que l'expression du locus *p16-CDKN2A/p15-CDKN2B/p14-ARF* subit une régulation temporo-spatiale.

Le locus *p16-CDKN2A/p15-CDKN2B/p14-ARF* comporte les gènes *p14-ARF* (*p19-ARF* chez la souris), *p15-INK4b* et *p16-INK4a*. Ces trois gènes suppresseurs de tumeur jouent un rôle central dans l'arrêt du cycle cellulaire et sont impliqués dans la sénescence, l'apoptose et l'auto-renouvellement des cellules souches en favorisant les activités anti-prolifératives des deux gènes *RB1* et *TP53* (Popov and Gil, 2010). Les suppresseurs p15-INK4b et p16-INK4a ciblent les kinases CDK4 et CDK6 dépendantes des cyclines, empêchant la liaison de ces protéines aux cyclines de type D, et induisant l'inhibition de la phosphorylation (inactivation) médiée par CDK4/6, des membres de la famille RB1. La protéine p14-ARF agit principalement en se liant à l'E3 ubiquitine protéine ligase MDM2, favorisant sa dégradation et donc l'abrogation médiée par MDM2 de l'activité de p53 (Popov and Gil, 2010).

Les deux complexes macromoléculaires répressifs sont le complexe polycomb répressif 1 (PRC1) et le complexe polycomb répressif 2 (PRC2). Le complexe PRC2 se compose de trois sous-unités : *embryonic ectoderm development* (EED), *suppressor of zeste 12* (SUZ12) et *enhancer of zeste 2 or 1* (EZH2/1), qui catalysent les mono-, di- et triméthylation de la lysine 27 de l'histone H3 (H3K27me1, H3K27me2 et H3K27me3) (Margueron et al., 2008). Le complexe PRC1 est plus hétérogène et sa composition dépend du contexte cellulaire. Il associe de façon variable des protéines de la famille PCGF (PCGF1-PCGF6), HPH (HPH1-HPH3), des protéines à domaine *chromobox* (CBX) et des protéines RING1a/1b qui catalysent la monoubiquitination de H2A sur K119 (H2AK119ub1), de manière à maintenir fermement l'inactivation de la chromatine (Wang et al., 2004).

ANRIL se lie spécifiquement à deux protéines polycomb, SUZ12 puis CBX7, afin de modifier le code histonique du locus *p16-CDKN2A/p15-CDKN2B/p14-ARF* (Popov and Gil, 2010). Il interagit initialement avec la sous-unité SUZ12 de PRC2 et favorise sa liaison au niveau du locus *p15-INK4b*, la triméthylation de la lysine 27 de l'histone H3 et l'inactivation du gène *p15-INK4b*. H3K27me3 est une marque épigénétique répressive constituant un signal de reconnaissance pour les protéines à domaines *chromobox* CBX et le recrutement du deuxième complexe polycomb PRC1. *ANRIL*, en s'associant avec la sous-unité protéique CBX7 (*chromodomain of chromobox homolog 7*) du complexe PRC1, favorise la reconnaissance de H3K27me3 par CBX7 et l'inactivation du *cluster p16-CDKN2A/p15-CDKN2B/p14-ARF* (Yap et al., 2010). CBX7 stabilise les complexes PRC1/PRC2 au niveau de la chromatine,

maintient l'inactivation des gènes cibles et la compaction de la chromatine au niveau du locus, en interagissant directement avec H3K27me3 *via* ses chromodomaines. Ainsi, PRC2 et PRC1 agissent de concert et successivement pour déposer les marques répressives H3K27me3 et H2AK119Ub1, assurant ainsi une répression transcriptionnelle ferme et pérenne. En outre, EZH2 recrute et interagit également avec des ADN méthyltransférases, assurant ainsi la méthylation des îlots CpG des promoteurs de gènes cibles et l'établissement de structures chromatiniennes répressives. Il est aussi un régulateur clé du contrôle de l'engagement et de la différenciation des cellules souches embryonnaires et adultes. De même, CBX7 joue également un rôle essentiel dans l'auto-renouvellement des cellules souches et le maintien de la pluripotence.

Fonctions

ANRIL est impliqué dans le contrôle des *checkpoints* du cycle cellulaire, l'apoptose, la sénescence et la réparation de l'ADN. Des niveaux transcriptionnels accrus d'*ANRIL* sont observés à la phase tardive de la réparation cellulaire des altérations de l'ADN. Ils s'accompagnent de l'inactivation transitoire de l'expression des gènes *p14-ARF*, *p15-INK4b* et *p16-INK4a*, afin de favoriser le retour à une cinétique cellulaire normale. *ANRIL* est également impliqué dans la sénescence induite par le stress et par les oncogènes (*oncogene-induced senescence*, OIS). Les mécanismes de l'OIS dépendent en partie de trois lncRNAs : *ANRIL*, *UCA1* et un transcrit non codant du gène *MIR31HG* situé à 400 kb en amont du locus *p16-CDKN2A/p15-CDKN2B/p14-ARF*.

Régulation

ANRIL est régulé par la voie ATM-E2F1 et est activé au niveau transcriptionnel par le facteur de transcription E2F1 à la suite de lésions de l'ADN. La voie Ras/Raf/MEK/ERK est également impliquée dans la régulation d'*ANRIL* et du *cluster p16-CDKN2A/p15-CDKN2B/p14-ARF* (Malumbres et al., 2000). La protéine oncogénique Ras inhibe l'expression d'*ANRIL*, favorise l'activation de *p15INK4b* et induit le mécanisme protecteur de l'OIS. TET2 (*Ten to eleven translocation 2*) se lie à la région promotrice d'*ANRIL* et inhibe son expression, favorisant ainsi l'apoptose.

Rôles d'*ANRIL* dans la cancérogénèse

Polymorphismes et altérations structurales

Les polymorphismes d'*ANRIL* et ses altérations structurales peuvent affecter l'homéostasie cellulaire et contribuer à l'apparition de diverses pathologies.

Des nombreuses études d'association pangénomique (GWAS) ont identifié des régions spécifiques du génome humain présentant une forte corrélation non aléatoire avec des facteurs de prédisposition à diverses maladies (de los Campos et al., 2010). Ces études ont permis de préciser que la région 9p21.3 est le siège de multiples altérations génétiques associées à un nombre croissant de pathologies et comporte des polymorphismes nucléotidiques simples (SNPs), correspondant à une sensibilité accrue vis à vis de diverses maladies humaines. Les SNPs ont été identifiés au niveau du locus *p16-CDKN2A/p15-CDKN2B/p14-ARF* et sont étroitement associés aux maladies cardiovasculaires, aux accidents vasculaires cérébraux, aux anévrismes de l'aorte, au diabète de type II et au cancer (Yap et al., 2010 ; de los Campos et al., 2010). Les SNPs présents au sein du locus d'*ANRIL* affectent ses niveaux d'expression et modifient les sites de liaison aux facteurs de transcription et l'épissage alternatif, à l'origine de nouveaux transcrits qui perturbent l'expression du locus *p16-CDKN2A/p15-CDKN2B/p14-ARF* (Harismendy et al., 2011). Ainsi, le SNP rs1063192-C est fortement corrélé à la survenue de gliomes, le SNP rs1011970-T à la survenue de mélanomes (Cunnington et al., 2010), et le SNP Rs564398 augmente le risque de leucémie lymphoblastique (Iacobucci et al., 2011).

Les altérations structurales d'*ANRIL*, telles les délétions et les translocations, peuvent affecter son efficacité dans la répression du locus *p16-CDKN2A/p15-CDKN2B/p14-ARF* (Burd et al., 2010) et sont associées à une diminution de la viabilité cellulaire, une augmentation de la prolifération et de l'adhérence, et une diminution de l'apoptose et de la sénescence. Ces modifications peuvent conduire à la dérégulation de l'expression des gènes en *cis*. Les gènes du *cluster p16-CDKN2A/p15-CDKN2B/p14-ARF* sont le siège d'une délétion homozygote ou d'une inactivation transcriptionnelle dans environ 40% des cancers humains (Iacobucci et al., 2011). *ANRIL* apparaît ainsi comme l'un des lncRNAs plus fréquemment altérés dans les cancers, dont les carcinomes de l'ovaire, du sein, du nasopharynx, de la peau, les neurofibromes et les gliomes ainsi que les leucémies lymphoblastiques (Pasmant et al., 2011b). Des transcrits de fusion *MTAP-ANRIL* ont été récemment identifiés dans des lignées

cellulaires de mélanomes et les tumeurs mélanocytaires malignes et sont associés à une délétion focale au niveau du locus *p16-CDKN2A/p15-CDKN2B/p14-ARF* (Xie et al., 2016). L'oncoprotéine traduite à partir du transcrit de fusion *MTAP-ANRIL* est analogue à la protéine *MTAP* tronquée, car *ANRIL* contribue seulement à trois résidus d'acides aminés. Un transcrit de fusion similaire a été identifié dans une tumeur gliale, résultant de la fusion de l'exon 4 de *MTAP* avec l'exon 2 d'*ANRIL* et la protéine produite contient 27 acides aminés issus du gène non codant. Les altérations structurales d'*ANRIL* conduisent également à une dérégulation de l'expression des gènes en *trans*. La surexpression des exons 13-19 d'*ANRIL* dans les cellules HeLa aboutit à la répression de nombreux gènes impliqués dans le remodelage de l'architecture chromatinienne, dont la protéine centrosomale CEP290, p300 et TCF7L1 (Sato et al., 2010).

Mécanismes moléculaires

La dérégulation d'*ANRIL* dans le cancer se traduit par sa surexpression aberrante, une moindre efficacité dans la répression en *cis* du locus *p16-CDKN2A/p15-CDKN2B/p14-ARF* et une capacité inattendue à réguler de nombreux gènes en *trans*. La liaison directe d'*ANRIL* avec les complexes répressifs polycomb favorise la compaction chromatinienne et la répression permanente en *cis* des gènes *CDKN2A*, *CDKN2B* et *ARF*. La liaison avec l'insulateur CTCF, qui est nécessaire afin de maintenir le locus *p16-CDKN2A/p15-CDKN2B/p14-ARF* dans une conformation inductible, est abrogée lors de la méthylation de l'ADN, avec des conséquences sur la progression tumorale (Rodriguez et al., 2010). Les altérations de l'ADN favorisent l'activation de la voie ATM et E2F1 et induisent la surexpression d'*ANRIL*. Sa surexpression provoque une inefficacité des mécanismes de réparation des lésions de l'ADN, conduisant à une instabilité génomique et à la progression tumorale (Wan et al., 2013a). Le locus *p16-CDKN2A/p15-CDKN2B/p14-ARF* responsable de l'OIS est inactivé dans les cellules cancéreuses. Au cours de la cancérogénèse gastrique, *ANRIL* inhibe des miRNAs en se liant à EZH2. En inactivant épigénétiquement miR-49a et miR-449A, *ANRIL* favorise l'activation des voies mTOR et CDK6/E2F1 et un mécanisme de *feed-back* positif, qui augmente l'expression d'*ANRIL* (Aguilo et al., 2016). MiR-449a inhibe l'expression de *bcl2* et induit l'arrêt du cycle et la sénescence dans les cellules cancéreuses prostatiques. MiR-99a inhibe la prolifération des cellules cancéreuses hépatiques et gastriques et la malignité des cellules cancéreuses mammaires en inhibant IGF-1R (Zhang et al., 2014d).

ANRIL favorise la croissance cellulaire en inhibant la voie de signalisation TGF β /Smad dans les carcinomes malpighiens de l'œsophage (Chen et al., 2014c). L'expression de *KLF2* peut être inhibée par *EZH2*, qui est recruté par *ANRIL* au niveau la région promotrice de *KLF2* dans les hépatocarcinomes (Huang et al., 2015). La surexpression d'*ANRIL* favorise l'inhibition de p21. C-Myc se lie directement au promoteur d'*ANRIL* et favorise son expression. La régulation d'*ANRIL* par l'enzyme TET2 qui induit normalement l'apoptose, est altérée avec une sous-expression de TET2 dans les mélanomes et les carcinomes mammaires et gastriques (Deng et al., 2016c). *ANRIL* est corrélé positivement à TNF- α , NF-kB et aux cytokines inflammatoires. La surexpression d'*ANRIL* dans des modèles murins de carcinomes pulmonaires induit l'augmentation de l'expression de VEGF-C et VEGFR-3, une lymphangiogénèse et des métastases ganglionnaires (Sun et al., 2016 ; Wang et al., 2016).

Analyses fonctionnelles

Les expériences de gain et de perte de fonction réalisées *in vitro* et *in vivo* suggèrent qu'*ANRIL* est un lncRNA oncogénique favorisant la progression du cycle cellulaire et inhibant l'apoptose et la sénescence pour stimuler la prolifération tumorale.

La surexpression d'*ANRIL* (i) provoque d'importants changements dans l'expression des gènes distribués à distance dans le génome, favorise l'activité métabolique et majore la prolifération et l'adhésion cellulaire tout en inhibant l'apoptose, (ii) promeut la prolifération cellulaire dans les cellules cancéreuses du colon et (iii) augmente la prolifération, favorise la progression du cycle cellulaire et inhibe l'apoptose et la sénescence des cellules carcinomateuses ovariennes, (iv) provoque la sous-expression de plusieurs gènes impliqués dans des mécanismes de remodelage chromatinien, dont p300 et (iv) induit l'hyperméthylation de l'ADN du locus *p16-CDKN2A/p15-CDKN2B/p14-ARF* dans les cellules différenciées.

La déplétion d'*ANRIL* (i) s'accompagne de la surexpression de *CDKN2B* mais pas de celle de *CDKN2A* ou d'*ARF* dans des lignées cellulaires de fibroblastes foetaux. Cette inactivation réduit de plus les niveaux de H3K27me3 dans le cluster *p16-CDKN2A/p15-CDKN2B/p14-ARF*, (ii) inhibe la croissance des tumeurs de l'ovaire chez la souris *nude* et induit une apoptose *in vitro* et *in vivo* dans le cancer de l'ovaire, (iii) diminue la croissance cellulaire et augmente le nombre de cellules sénescents dans les fibroblastes humains, (iv) supprime la prolifération, la migration et l'invasion des cellules cancéreuses gastriques, pulmonaires et

vésicales, et favorise l'apoptose des cellules cancéreuses *in vivo* et *in vitro*, (v) supprime l'invasion, la prolifération et l'apoptose des cellules hépatocarcinomeuses à la fois *in vivo* et *in vitro* et (vi) induit un arrêt du cycle cellulaire en phase G2/M des cellules carcinomeuses du col de l'utérus.

L'inactivation complète d'*ANRIL* s'accompagne d'une réduction de la marque répressive H3K27me3, d'une diminution des liaisons de CBX7 et EZH2 au niveau du TSS de *p16-CDKN2A* et d'une expression accrue de *p15-CDKN2B* et *p16-CDKN2A*, provoquant la suppression de la prolifération cellulaire et la sénescence (Yap et al., 2010).

Pathologie humaine

ANRIL est surexprimé dans de nombreux tissus pré-néoplasiques et néoplasiques par rapport aux tissus non tumoraux (Yap et al., 2010), dont les cancers de l'estomac (Zhang et al., 2014d), de la prostate, du rein, de l'œsophage (Chen et al., 2014b), du côlon, de l'ovaire, du poumon ainsi que les leucémies (Yu et al., 2008 ; Cunnington et al., 2010). Pour ces cancers, les analyses statistiques et les méta-analyses ont révélé que la surexpression d'*ANRIL* est corrélée positivement à un stade TNM avancé, la survenue de métastases ganglionnaires et un facteur prédictif indépendant de survie globale défavorable (Liu et al., 2016a).

Conclusion

ANRIL apparaît comme un *driver* oncogénique inhibant l'apoptose et la sénescence et favorisant la prolifération tumorale et la lymphangiogénèse. Il pourrait constituer un biomarqueur pertinent et une cible thérapeutique potentielle en raison de son implication dans de nombreux cancers. Toutefois, en dépit d'importants progrès réalisés dans la compréhension de son fonctionnement, ses fonctions physiologiques ainsi que les mécanismes de sa dérégulation au cours de la carcinogénèse mammaire restent encore largement méconnus. En outre, des études successives suggèrent que certaines sous-unités polycomb, dont CBX7, pourraient agir tantôt comme des oncogènes, tantôt comme des oncosuppresseurs, en fonction de la localisation tumorale primitive. Une approche nécessairement plus large de ses mécanismes d'action physiologiques et pathologiques, de ses

voies de régulation et d'épissage et de ses réseaux d'interactions moléculaires, constituera une étape décisive pour son utilisation thérapeutique potentielle.

MALATI

Localisation génomique

MALATI (*Metastasis Associated Lung Adenocarcinoma Transcript 1*), également connu sous le nom de *NEAT2* (*Nuclear Enriched Abundant Transcript 2*), est un lncRNA intergénique transcrit à partir d'un locus situé en 11q13.1. Le transcrit du gène *MALATI* a été initialement décrit en 1997 comme un transcrit "alpha" dans une étude ayant identifié de multiples ARNs issus du locus associé aux MEN de type 1. En 2003, *MALATI* a été identifié comme un transcrit associé à la survenue de métastases chez des patients atteints de cancers pulmonaires non à petites cellules de stade précoce. En 2007, il a été localisé dans les *speckles*, domaines nucléaires destinés au stockage et à l'épissage alternatif des pré-mRNAs.

Caractéristiques moléculaires

A la différence de la plupart des lncRNAs, *MALATI* est ubiquitaire et présente un haut niveau d'expression dans de nombreux tissus humains normaux (Spector and Lamond, 2011), particulièrement dans le cerveau (Bernard et al., 2010), ainsi qu'un degré de conservation inter-espèces élevé (Ji et al., 2003 ; Hutchinson et al., 2007), en faveur de fonctions biologiques importantes (Ulitsky et al., 2011). La séquence de *MALATI* est en effet conservée parmi 33 espèces de mammifères et de putatifs orthologues pourvus d'une faible homologie de séquence ont été identifiés au niveau de chromosomes synténiques chez les vertébrés inférieurs. *MALATI* est l'un des plus abondants lncRNAs et sa localisation est essentiellement intranucléaire (Tripathi et al., 2010). *MALATI* subit un processus de maturation non canonique aboutissant à la création de 2 isoformes. L'isoforme le plus long mesure 8kb chez l'homme, possède un signal de polyadénylation conservé en 3', est entièrement localisé dans le noyau et comporte un faible niveau d'expression. Le deuxième isoforme mesure 7.5kb et

son extrémité 3' ne dispose pas d'une queue polyA+, mais d'une structure de type *tRNA-like*, qui est synthétisée *via* la voie de maturation des tRNAs par des endonucléases de type RNase P et RNase Z. Ce clivage génère un transcrit stable intra-nucléaire très abondant et un court RNA de type tRNA cytoplasmique (*malat1 associated small cytoplasmic RNA, mascRNA*) dont la fonction est inconnue (Wilusz et al., 2008). L'extrémité 3' forme une structure tri-hélicoïdale stable U-A-U, qui confère au lncRNA *MALATI* une résistance aux nucléases et favorise une traduction plus efficace quand il est exporté vers le cytoplasme (Brown et al., 2012). L'homologie de séquences de *MALATI* entre le poisson zèbre et les mammifères, est limitée à son extrémité 3', ce qui souligne son importance biologique. La demi-vie de *MALATI* dans les cellules humaines est de 9 à 12 heures, relativement plus longue que celles des autres lncRNAs, en partie due à la structure en triple hélice située à son l'extrémité 3'. Bien que *MALATI* soit localisé spécifiquement dans des organelles intranucléaires, il n'est pas considéré comme un lncRNA architectural puisque sa déplétion n'entraîne pas la dispersion des corps nucléaires et il ne semble pas non plus participer à l'organisation des complexes d'épissage (Tripathi et al., 2010).

Régulation

Au niveau transcriptionnel, l'expression de *MALATI* est régulée pendant la différenciation cellulaire. Ainsi, la myostatine, un régulateur négatif de la myogénèse, inhibe la transcription de *MALATI* et régule ainsi le développement et la croissance musculaire squelettique. La déplétion de *MALATI* conduit à l'activation de p53 (Tripathi et al., 2013).

Au niveau post-transcriptionnel, la voie Drosha-DGCR8-microprocesseur requise pour la biogénèse des miRNAs est impliquée dans la régulation de *MALATI* (Macias et al., 2012). La déplétion de Drosha augmente la transcription de *MALATI*, suggérant que ce dernier est reconnu comme un substrat par le complexe microprocesseur et transformé en petits RNAs. Les transcrits de *MALATI* interagissent avec des miRNAs. Ago2, un composant du complexe répressif médié par les miRNAs, se lie à *MALATI* et les miRNAs miR-9, miR-102, miR-217 et miR-205 le ciblent directement. L'expression des transcrits de *MALATI* est également régulée par des protéines. TDP-43 est une protéine de liaison à l'ADN qui se lie à la séquence TAR du virus de l'immunodéficience humaine 1 (VIH-1) et la déplétion de TDP-43 dans le cerveau de la souris réduit l'expression de *MALATI*. Les modifications post-transcriptionnelles des nucléosides, telles la N6-méthyladénosine (m6A), agissent comme des

commutateurs moléculaires favorisant la liaison de hnRNP C1/C2 à *MALAT1* et contrôlant la progression du cycle cellulaire en phase G2/M.

Mécanismes d'action

Les mécanismes d'action de *MALAT1* sont encore largement méconnus mais pourraient être liés à sa capacité d'interagir avec de multiples réseaux protéiques dont l'expression varie dans différents contextes cellulaires. A ce jour, plus de 20 protéines interagissent avec *MALAT1*, parmi lesquelles SRSF1, SRSF2, SRSF3, U2AF65, IBP160, RNPS1, SRSF4, TDP-43, DGCR8, Ago2, hnRNP C1/C2, LSD1 et SETDB2. De plus, *MALAT1* agit comme un compétiteur endogène dans la myogénèse *via* son interaction avec miRNA-133 (Han et al., 2015).

Fonctions

MALAT1 est impliqué dans la régulation transcriptionnelle de l'expression génique et dans l'épissage alternatif avec d'importantes conséquences dans le contrôle du cycle cellulaire, l'apoptose, la plasticité et la mobilité cellulaire (**Fig. 30**).

MALAT1 est localisé dans les *speckles* qui constituent des corps intranucléaires contenant divers facteurs impliqués dans l'épissage des pré-mRNAs, dont les complexes U snRNPs et la famille des régulateurs d'épissage SR. *MALAT1* semble réguler l'épissage alternatif des pré-mRNAs en modulant le recrutement, les niveaux d'expression et la phosphorylation des facteurs actifs d'épissage sérine/arginine, telles les protéines SRSF1, SRSF2 et SRSF3 (Hutchinson et al., 2007 ; Tripathi et al., 2010) et en jouant le rôle d'« éponge moléculaire » par ses multiples sites de liaison à la protéine TDP-43, impliquée dans la fixation à de nombreux pré-mRNAs (Tollervey et al., 2011). L'inactivation de *MALAT1* s'accompagne d'altérations de l'épissage et de modifications de l'expression des gènes (Tano et al., 2010 ; Tripathi et al., 2013). La déplétion de *MALAT1* provoque des anomalies du positionnement de certains facteurs d'épissage (SF1, U2AF65 et SF3a60) au sein des *speckles*, et une modification de la phosphorylation des protéines SR. Ces résultats suggèrent que *MALAT1* contrôle indirectement l'épissage alternatif en changeant la distribution des régulateurs d'épissage dans les *speckles*. Ces effets sur l'épissage alternatif sont cependant absents dans

les modèles murins *KO* pour *MALATI*. Toutefois, une étude concernant des modèles murins déplétés en *MALATI* n'a pu révéler de modification des niveaux d'expression des facteurs d'épissage (Nakagawa et al., 2012).

Ces résultats apparemment discordants pourraient être expliqués si l'on considère que le rôle de *MALATI* n'est pas indispensable à la régulation de l'épissage alternatif dans des conditions normales, mais devient essentiel dans des environnements spécifiques. De plus, l'absence de *MALATI* pourrait être compensée par un mécanisme de rétrocontrôle encore inconnu pour maintenir un épissage dans les modèles *KO* animaux. En outre, l'extrême abondance des transcrits de *MALATI* pourraient provoquer la libération rapide des facteurs d'épissage et simuler les effets d'une surexpression aigüe de ces facteurs.

Bien que les *speckles* aient été initialement considérés comme les lieux dédiés au stockage des facteurs d'épissage, la localisation de la sérine-2 phosphorylé ARN polymérase II dans ces corps intranucléaires suggère l'existence d'une transcription *de novo*. En accord avec cette hypothèse, la déplétion de *MALATI* induite par les siRNAs affecte l'expression d'un ensemble de gènes impliqués dans la motilité cellulaire. *MALATI* pourrait ainsi activer de multiples gènes en interagissant avec la protéine Pc2 (*polycomb protein 2*), un composant du complexe répressif PRC1 et un lecteur du code histone *via* son chromodomaine CBX4. *MALATI* régule la localisation de Pc2 en coopérant avec un autre lncRNA, *TUG1*. La forme non méthylée de Pc2 interagit préférentiellement avec *MALATI* dans les *speckles*, alors que la forme méthylée de Pc2 interagit surtout avec *TUG1* au sein de corps intranucléaires voisins, les corps polycomb (CP). La méthylation/déméthylation de Pc2/CBX4 coordonne la relocalisation des gènes contrôlant la prolifération entre les *speckles* et les CPs en réponse à des signaux de croissance. *MALATI* agit donc comme une plateforme moléculaire permettant l'expression des gènes en favorisant les interactions entre la forme non méthylée de Pc2, le facteur de transcription E2F, les marques histoniques et le complexe co-activateur de la transcription. En outre, la liaison de *MALATI* à la forme non méthylée de Pc2/CBX4 favorise la sumoylation d'E2F1 conduisant à l'activation du programme génique de contrôle du cycle et de la croissance cellulaire (Nakagawa et al., 2012). L'expression génique peut ainsi être régulée par un déplacement entre deux corps nucléaires dont l'activité antagoniste est médiée par deux lncRNAs. *MALATI* interagit également avec les loci voisins et contrôle leur expression, suggérant un mécanisme de *cis*-activation transcriptionnelle. Ainsi, l'expression de *NEAT1*, située à 30 kb en 5' de *MALATI* est réduite dans les modèles murins *KO* pour *MALATI* et dans les cultures cellulaires soumises à des siRNAs inactivant *MALATI*.

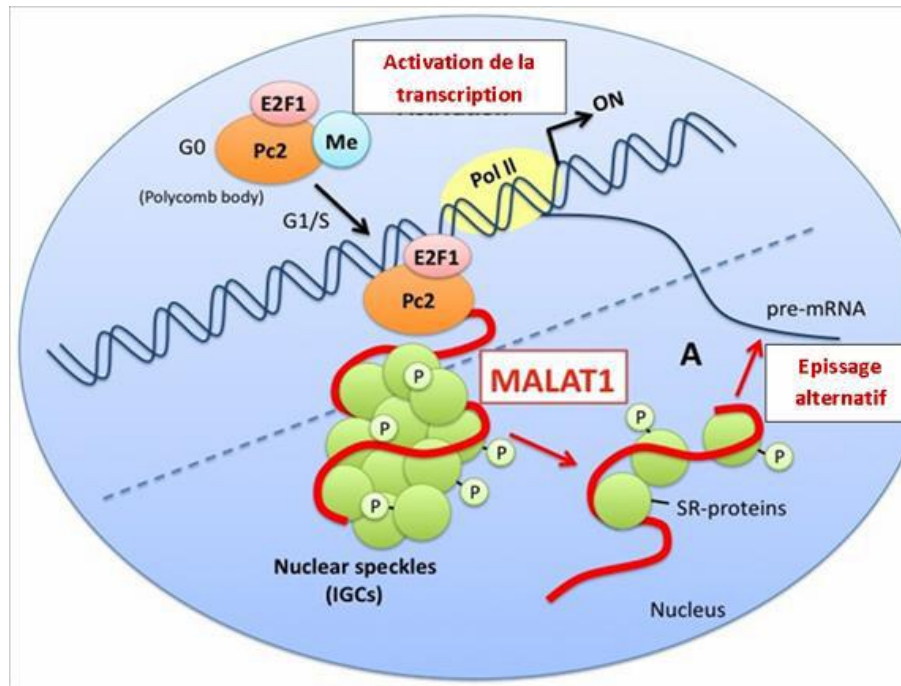


Figure 30 : Mécanismes d'action du lncRNA *MALAT1* au niveau des *speckles* intra-nucléaires (D'après Tano and Akimitsu, 2012)

Rôles de *MALAT1* dans la cancérogénèse

Mécanismes moléculaires

Des altérations ont été identifiées dans divers types de cancers, et en particulier les tumeurs mammaires, au niveau du locus de *MALAT1*, dont des translocations chromosomiques, des amplifications et des mutations ponctuelles (Ellis et al., 2012 ; Nik-Zainal et al., 2016). Une translocation *MALAT1-TFEB* a été observée dans les carcinomes du rein avec création d'un transcrite de fusion oncogénique. Dans les cellules de neuroblastomes humains, l'ocytocine élève le niveau d'expression de *MALAT1* via la liaison du facteur de transcription interagissant avec l'AMPc au promoteur du gène *MALAT1*. Dans le myélome multiple, *MALAT1* recrute le facteur de transcription Sp1 au niveau de *LTBP3* (*latent transforming growth factor beta binding protein 3*), favorisant ainsi l'activation et la sécrétion du facteur de croissance TGF- β . JMJD1A, une histone H3K9 déméthylase, se lie au promoteur du gène de *MALAT1* et déméthyle l'histone H3K9 au niveau du promoteur, conduisant à la surexpression

de *MALATI*. Dans les cancers de l'endomètre, la méthylation du promoteur de la proto-cadhérine 10 (PCDH10), un gène oncosuppresseur potentiel, induit la transcription de *MALATI* via l'activation de la voie Wnt/ β -caténine. Dans les cellules de cancers du sein exprimant le RE, le traitement par le 17 β -oestradiol régule négativement la transcription de *MALATI*, conduisant à l'inhibition de la prolifération cellulaire, de la migration et de l'invasion. Le facteur de transcription Sox17, fréquemment inactivé par hyperméthylation dans les cellules cancéreuses, inhibe l'expression du gène *MALATI* dans le cancer de l'œsophage. Dans les cellules endothéliales humaines, l'hypoxie induit la transcription de *MALATI*, conduisant à la prolifération du système vasculaire. Dans le cancer de la vessie, *MALATI* s'associe avec SUZ12, une sous-unité de PRC2 et induit l'expression de la N-cadhérine et une réduction concomitante de la E-cadhérine. Le TGF- β induit la transcription de *MALATI* dans les cellules cancéreuses et favorise l'EMT. Dans les cancers colorectaux et les ostéosarcomes, *MALATI* favorise la croissance tumorale et les métastases en se liant à PSF, une protéine de liaison à l'ARN multifonctionnelle (Yoshimoto et al., 2016)

Analyses fonctionnelles

Les études de gain et de perte de fonction conduites *in vitro* et *in vivo*, ont révélé que *MALATI* est particulièrement impliqué dans les mécanismes fondamentaux de la carcinogénèse. *MALATI* apparaît surexprimé dans de nombreux cancers et confère aux cellules tumorales un phénotype prolifératif et métastatique (Gutschner et al., 2013). Cependant, alors que la surexpression de *MALATI* conduit à des conséquences délétères, son inactivation chez la souris n'est pas létale (Zhang et al., 2012). *MALATI* favorise l'acquisition par les cellules tumorales de différents *hallmarks* du cancer qui peuvent être annulés par son inhibition. En effet, l'inactivation de *MALATI* par des shRNAs inhibe la croissance, la progression du cycle cellulaire et l'invasion des cellules cancéreuses du col utérin via la dérégulation de l'expression de multiples gènes, dont les gènes codant pour les caspases 3 et 8, Bax, Bcl-2, et Bcl-xL (Guo et al., 2010). Dans les adénocarcinomes pulmonaires, la répression de *MALATI* par des siRNAs conduit à l'inactivation des gènes *CTHRC1*, *CCT4*, *HMMR* ou *RODI*, et contribue à l'inhibition de la motilité cellulaire (Tano et al., 2010). De plus, la déplétion de *MALATI* inhibe la prolifération cellulaire, augmente l'apoptose et diminue la migration des cellules cancéreuses dans les carcinomes hépatocellulaires et les cancers de vessie (Han et al., 2013). *MALATI* semble également impliqué dans l'EMT et dans

l'invasion des cellules malignes. L'inactivation de *MALAT1* provoque une diminution du niveau d'expression des gènes liés à l'EMT, dont *ZEB1*, *ZEB2* et *Slug*, et une augmentation de l'expression de la E-cadhérine (Ying et al., 2012). Dans des lignées cellulaires de cancer du sein, l'inactivation de *MALAT1* induit une EMT *via* l'activation de la voie PI3K/Akt/mTOR. L'inhibition de *MALAT1* induit un *switch* des cellules endothéliales avec acquisition d'un phénotype anti-prolifératif et pro-migratoire (Michalik et al., 2014). Enfin, l'inactivation de *MALAT1* dans les cellules cancéreuses du poumon réduit considérablement le niveau d'expression de plusieurs gènes de métastases, dont *GPC6* et *CXCL5*. Plus récemment, il a été observé que la surexpression de *MALAT1* inhibe la voie ATM-CHK2 et que *MALAT1* est régulé au niveau post-transcriptionnel par les miRNAs suppresseurs miR-101 et miR-217 dans les cellules de carcinomes oesophagiens (Hu et al., 2015). En outre, l'inactivation de *MALAT1* comporte des effets anticancéreux *in vivo*. Les cellules des carcinomes pulmonaires non à petites cellules présentant des niveaux d'expression faibles de *MALAT1*, ont une moindre capacité à former des tumeurs et à proliférer chez les souris *nude*. Les modèles murins pourvus d'un phénotype d'inactivation stable de *MALAT1* présentent une inhibition de la croissance métastatique des carcinomes du poumon et une réduction du nombre de nodules tumoraux (Gutschner et al., 2013a,b). Plus récemment, la perte de fonction de *MALAT1* induite par *ASO* dans un modèle de carcinome mammaire MMTV-PyMT-HER2 amplifié a révélé une réduction de la croissance tumorale, une augmentation de l'adhésion intercellulaire et une perte de la fonction migratoire. Au niveau moléculaire, cette perte de fonction s'accompagne d'altérations dans l'expression et l'épissage alternatif de gènes impliqués dans la différenciation et les voies de signalisation pro-oncogéniques (Arun et al., 2016). Ainsi, l'inactivation de *MALAT1* par *ASO* représente une thérapie potentielle innovante dans l'inhibition de la progression des carcinomes mammaires (Mendell, 2016).

Pathologie humaine

La surexpression de *MALAT1* a été identifiée initialement dans les cancers du poumon non à petites cellules et constitue un facteur pronostique péjoratif chez les patients atteints de tumeurs de stade I (Gutschner et al., 2013a,b). Des études successives ont depuis révélé que *MALAT1* est également surexprimé dans de nombreux cancers, dont les cancers du poumon, de l'ovaire, du rein, de la prostate, de la vessie, du sein, du col, du colon, du pancréas, de l'endomètre, de l'œsophage, de l'estomac, du foie, les mélanomes, les lymphomes et

leucémies et certains sarcomes (Ji et al., 2003 ; Guffanti et al., 2009 ; Guo et al., 2010 ; Tano et al., 2010 ; Lai et al., 2012 ; Ying et al., 2012 ; Han et al., 2013 ; Ren et al., 2013 ; Ji et al., 2014 ; Liu et al., 2014a). De plus, la surexpression de *MALAT1* peut être utilisée comme un biomarqueur pronostique et l'étude des données fournies par une méta-analyse suggère que la surexpression de *MALAT1* est significativement corrélée avec une survie globale réduite dans la majorité des cancers. En plus de son importance comme biomarqueur, *MALAT1* peut être une cible thérapeutique potentielle pertinente dans les tumeurs métastatiques.

Conclusion

Bien que de nombreuses études aient souligné l'implication de *MALAT1* dans divers cancers, les mécanismes moléculaires impliquant ce lncRNA dans la cancérogénèse ne sont que très partiellement dévoilés. Une meilleure compréhension de sa structure secondaire, de ses domaines fonctionnels, des variants d'épissage et de sa régulation permettra d'établir le lien entre ses fonctions encore largement méconnues et les mécanismes de sa dérégulation dans le cancer.

La deuxième partie du sujet de thèse a consisté en une analyse préliminaire du transcriptome non codant de carcinomes mammaires infiltrants.

Cette analyse transcriptomique préliminaire a comporté un volet *discovery* et un volet *gene candidate*, qui n'a pu être totalement achevé avant la soumission de la thèse.

La stratégie discovery a consisté en un *screening* par RNAseq (*Illumina 2500 Whole Genome*) d'une cohorte initiale de tumeurs malignes appartenant aux différents sous-types moléculaires de carcinomes mammaires infiltrants : luminal A (n=2), HER2 (n=4) et triple négatif (n=2) et de 4 tissus mammaires sains prélevés à distance (>10mm) de chaque foyer tumoral. Un premier travail de faisabilité a consisté à analyser un cas de chacun de ces quatre sous- types, puis, après validation, à analyser les autres cas de chaque sous-type, soit un total de 8 tumeurs mammaires. L'objectif à court terme de cette stratégie est d'obtenir une liste de lncRNAs, incluant de nouveaux lncRNAs antisens et intergéniques, dont l'expression est

différente d'une part entre les tumeurs et les tissus mammaires sains et d'autre part entre les différents sous-types de carcinomes mammaires.

Ces 2 stratégies ont permis l'identification d'un set de gènes d'intérêt. L'analyse de l'expression de ce *set* de gènes sélectionnés sera réalisée pour tous les lncRNAs par RT-PCR sur une large série de 456 carcinomes mammaires corrélée aux données cliniques et pronostiques de manière à identifier des profils d'expression caractéristiques des différents sous-types moléculaires analysés et des signatures pronostiques indépendantes de ces sous types.

RESULTATS

ARTICLE N°1 : Analyse de l'expression des gènes appartenant au complexe *ANRIL* - Polycomb - *CDKN2B* / *CDKN2A* / *ARF* dans les tumeurs du sein : identification d'une signature pronostique indépendante de deux gènes *EZH2* et *CBX7*.

7.1 Objectif

Les carcinomes canaux infiltrants du sein (CCI) sont des tumeurs très hétérogènes, et le sous-type tumoral de type basal ou triple négatif (CSTN) reste une cause majeure de décès chez les femmes. Ainsi, l'identification de nouveaux facteurs pronostiques pertinents et de cibles thérapeutiques potentielles apparaît cruciale pour améliorer les traitements des patientes. Plusieurs études récentes ont impliqué la dérégulation des lncRNAs dans les mécanismes de la cancérogénèse. Parmi les plus importants lncRNAs dérégulés, *ANRIL* possède un rôle direct dans le recrutement des complexes polycomb PRC1 et PRC2 qui répriment les gènes appartenant au *cluster p16-CDKN2A/p15-CDKN2B/p14-ARF*. Bien que des niveaux d'expression élevés d'*ANRIL* aient été observés dans les cancers prostatiques, le rôle fonctionnel et les mécanismes d'action de ce lncRNA dans la carcinogénèse mammaire demeurent controversés. Le but de cette étude a consisté à analyser l'expression d'*ANRIL*, des complexes polycomb PRC1/2 et du locus *p16-CDKN2A/p15-CDKN2B/p14-ARF* aux niveaux ARN et protéique dans les CCIs et leur signification clinique. L'évaluation des niveaux d'expression des gènes *ANRIL*, *EZH2*, *SUZ12*, *EED*, *JARID2*, *CBX7*, *BMI-1*, *p16*, *p15* et *p14/ARF* a été effectuée en utilisant la technique de qRT-PCR dans une large série de 456 CCIs. L'analyse de l'expression des protéines *CBX7*, *EZH2*, *p14*, *p15*, *p16*, *H3K27me3* et *H3K27ac* a été réalisée par une étude immunohistochimique dans une série de 80 CCIs.

Ce travail a fait l'objet de l'article suivant :

Expression of *ANRIL* – Polycomb complexes - *CDKN2B/CDKN2A/ARF* genes in breast tumors: identification of a two-gene (*EZH2/CBX7*) expression signature of independent prognostic value.

Didier Meseure, Sophie Vacher, Kinan Drak Alsibai, Andre Nicolas, Walid Chemlali, Martial Caly, Rosette Lidereau, Eric Pasmant, Celine Callens and Ivan Bieche.

Article publié dans la revue *Molecular Cancer Research* 2016;14:623-33.

7.2 Résumé

Le lncRNA *ANRIL* recrute les complexes polycomb répressifs PRC1 et PRC2 afin de réguler l'expression du locus *p16-CDKN2A/p15-CDKN2B/p14-ARF*. Nous avons évalué les niveaux d'expression des gènes *ANRIL*, *EZH2*, *SUZ12*, *EED*, *JARID2*, *CBX7*, *BMI-1*, *p16*, *p15* et *p14/ARF* par qRT-PCR dans une large série de 456 CCI. Les niveaux d'expression des protéines *CBX7*, *EZH2*, *p15*, *p16*, *H3K27me3* et *H3K27ac*, ont été évalués semi-quantitativement par analyse immunohistochimique (IHC). Nous avons observé une surexpression d'*ANRIL* dans 19.7% des CCI, une surexpression d'*EZH2* dans 77.0% des CCI et une sous-expression de *CBX7* dans 39.7% des CCI. Des corrélations significatives ont été identifiées entre les niveaux d'expression de ces différents gènes et divers paramètres cliniques et pathologiques, les sous-types moléculaires et les marqueurs de prolifération, d'EMT et de cellules souches mammaires. L'analyse multivariée a révélé que le rapport *EZH2*/*CBX7* est un facteur pronostique indépendant. Des miRNAs corrélés négativement à la sous-expression de *CBX7* et à la surexpression d'*EZH2* ont également été identifiés.

7.3 Résultats

Nous avons observé une surexpression d'*ANRIL* dans 19.7% des 456 CCI et plus particulièrement dans le sous-type agressif triple négatif. Nous avons également mis en évidence une surexpression marquée d'*EZH2* dans 77.0% des CCI et une sous-expression de *CBX7* dans 39.7% des tumeurs. Nous avons confirmé ces données au niveau protéique dans une série de 80 CCI en identifiant un score nucléaire d'intensité modérée et intense (score 2 et 3) à l'aide d'un anticorps anti-*EZH2* dans 82.0% des CCI et une absence ou un score nucléaire faible (score 0 et 1) à l'aide d'un anticorps anti-*CBX7* dans 43.0% des CCI. A la différence du cancer de prostate où des corrélations positives ont été observées entre la surexpression de *CBX7*, la surexpression d'*ANRIL* et la sous-expression du locus *p16-CDKN2A/p15-CDKN2B/p14-ARF*, nous avons mis en évidence une sous-expression marquée de *CBX7* et de façon inattendue une forte corrélation positive entre la surexpression d'*ANRIL* et la surexpression de *p16-CDKN2A/p15-CDKN2B/p14-ARF* dans notre série de 456 CCI. La sous-expression de *CBX7* est fortement corrélée avec un haut grade histologique SBR, un statut RE- et RP- et l'expression de marqueurs d'EMT et de cellules souches cancéreuses mammaires. La surexpression d'*EZH2* est également fortement associée à un grade SBR élevé et un statut RE- et RP-. Toutefois, à la différence de *CBX7*, la surexpression d'*EZH2* est étroitement liée au marqueur de prolifération MKI67, suggérant un rôle différent de ces deux

protéines dans la tumorigénèse mammaire. Une analyse par le test du log-rank a été effectuée afin d'identifier des corrélations entre les niveaux d'expression en ARNm des 10 gènes étudiés appartenant à la voie de signalisation *ANRIL* et la survie sans métastase (MFS). Cette étude a montré que seules la surexpression d'*EZH2* et la sous-expression de *CBX7* sont significativement associées à une plus courte survie (*EZH2* : P=0.000045 ; *CBX7* : P=0.0051, respectivement). L'analyse multivariée a révélé que le rapport *EZH2/CBX7* est un facteur pronostique indépendant (P=0.001). Des études récentes ont suggéré l'existence d'interactions complexes entre les complexes polycomb et les miRNAs lors de la progression tumorale. Parmi les 17 miRNAs sélectionnés, les niveaux d'expression de miR-181B1, miR-181A2 et miR-203 sont significativement augmentés dans les CCI et inversement corrélés à la sous-expression de *CBX7*. De plus, les niveaux d'expression de miR-26A1 et miR-125B2 sont significativement diminués dans les CCI avec surexpression d'*EZH2*. En outre, nous avons identifié un score nucléaire d'intensité modérée à intense (score 2 et 3) avec l'anticorps anti-H3K27ac dans 73.8% des carcinomes canaux infiltrants associé à un score nul ou faible (score 0 et 1) avec l'anticorps anti-H3K27me3 dans 76.3% des CCI.

7.4 Conclusion

Cette étude révèle l'existence d'interactions complexes entre *ANRIL*, le cluster *p16-CDKN2A/p15-CDKN2B/p14-ARF*, les complexes polycomb PRC2/PRC1 et des miRNAs oncogéniques dans les CCI. Parmi les 10 gènes étudiés, nos résultats soulignent qu'*EZH2* et *CBX7* sont les gènes les plus pertinents pour d'éventuelles applications cliniques et de nouvelles approches thérapeutiques des cancers du sein. Cette étude suggère également que c'est le profil global d'expression des sous-unités des complexes polycomb qui doit être pris en compte lors de l'évaluation de la fonctionnalité répressive de ces complexes et du choix des thérapies moléculaires ciblant ces gènes. Ainsi, l'utilisation d'un score d'activation des complexes PRC2/PRC1 dans les CCI pourrait se révéler pertinente pour guider l'indication des inhibiteurs des bromodomains ciblant les protéines BET. En outre, l'inhibition des oncoMIRs miR-181B1 et miR-181A2 (ainsi que la restauration des oncosuppresseurs miR-26A1 et miR-125B) pourraient représenter une nouvelle stratégie dans le traitement des cancers du sein.

7.5 Article N°1: Expression of *ANRIL* – Polycomb complexes - *CDKN2B/CDKN2A/ARF* genes in breast tumors: identification of a two-gene (*EZH2/CBX7*) expression signature of independent prognostic value.

Expression of ANRIL–Polycomb Complexes–CDKN2A/B/ARF Genes in Breast Tumors: Identification of a Two-Gene (EZH2/CBX7) Signature with Independent Prognostic Value

Didier Meseure^{1,2}, Sophie Vacher¹, Kinan Drak Alsibai², Andre Nicolas², Walid Chemlali¹, Martial Caly³, Rosette Lidereau¹, Eric Pasmant⁴, Celine Callens¹, and Ivan Bieche^{1,4}

Abstract

ANRIL, a long noncoding RNA (lncRNA), has recently been reported to have a direct role in recruiting polycomb repressive complexes PRC2 and PRC1 to regulate the expression of the p15/CDKN2B-p16/CDKN2A-p14/ARF gene cluster. Expression analysis of ANRIL, EZH2, SUZ12, EED, JARID2, CBX7, BMI1, p16, p15, and p14/ARF genes was evaluated in a large cohort of invasive breast carcinomas (IBC, n = 456) by qRT-PCR and immunohistochemistry (IHC) was performed on CBX7, EZH2, p14, p15, p16, H3K27me3, and H3K27ac. We observed significant overexpression in IBCs of ANRIL (19.7%) and EZH2 (77.0%) and an underexpression of CBX7 (39.7%). Correlations were identified between these genes, their expression patterns, and several classical clinical and pathologic parameters, molecular subtypes, and patient outcomes, as well as with proliferation, epithelial–mesenchymal transition, and breast cancer stem cell

markers. Multivariate analysis revealed that combined EZH2/CBX7 status is an independent prognostic factor ($P < 0.001$). In addition, several miRNAs negatively associated with CBX7 underexpression and EZH2 overexpression. These data demonstrate a complex pattern of interactions between lncRNA ANRIL, several miRNAs, PRC2/PRC1 subunits, and p15/CDKN2B-p16/CDKN2A-p14/ARF locus and suggest that their expression should be considered together to evaluate antitumoral drugs, in particular the BET bromodomain inhibitors.

Implications: This study suggests that the global pattern of expression rather than expression of individual family members should be taken into account when defining functionality of repressive Polycomb complexes and therapeutic targeting potential. *Mol Cancer Res*; 14(7): 623–33. ©2016 AACR.

Introduction

Invasive breast carcinomas (IBC) are highly heterogeneous tumors, and triple-negative subtype (TNC) remains a major cause of death in women. Although more than 90% of breast tumors are clinically localized, about 50% of them relapse within 5 years. Thus, identification of new prognostic factors and potential targeted therapies is crucial to improve outcome of patients with IBC.

Changes in the epigenetic landscape, including histone modifications, DNA methylation, and noncoding RNAs (ncRNA,) altered expression are now considered as a new hallmark of cancer (1). Polycomb repressive complexes (PRC) play impor-

tant roles in chromatin remodeling and inhibition of transcription, suppressor genes silencing, and interconnection with major signaling pathways (2). Polycomb group protein subunits are essential regulators of embryonic stem cell pluripotency and early developmental cell fate decisions that are often deregulated in cancer. Two multiprotein complexes, PRC2 and PRC1, act concertedly and sequentially in transcriptional repression via two distinct histone modifications, trimethylation of lysine 27 on histone H3 (H3K27me3) and monoubiquitination of histone H2A (H2AK119ub). PRC2 contains mainly four subunits (EZH2, SUZ12, EED, and JARID2) and is involved in initiation of silencing by catalyzing methylation of histone H3 (3). PRC1 also comprises mainly four subunits (BMI1, PHC, CBX7, and RING1) and is implicated in maintenance of silencing by catalyzing monoubiquitination of histone H2A (ref. 4; Supplementary Fig. S1).

PRC2 and PRC1 have been described to interact with long noncoding RNAs (lncRNA). LncRNAs are defined by a length ranged from 200 bp to 100 kbp and their number is in constant increase: 48,680 lncRNAs are at present identified (5–7). LncRNAs are important regulators coordinating expression of protein-coding genes nearby (cis-regulation) or at distance (trans-regulation). LncRNAs function as guide, tethers, scaffolds, and decoys and their mechanisms of action are located at epigenetics, translational, and posttranslational levels (8). Accumulating data suggest that deregulation of lncRNAs is

¹Unit of Pharmacogenomics, Department of Genetics, Curie Institute, Paris, France. ²Platform of Investigative Pathology, Curie Institute, Paris, France. ³Department of Biopathology, Curie Institute, Paris, France. ⁴Faculty of Pharmaceutical and Biological Sciences, Paris Descartes University, Paris, France.

Note: Supplementary data for this article are available at Molecular Cancer Research Online (<http://mcr.aacrjournals.org/>).

Corresponding Author: Ivan Bieche, Unit of Pharmacogenomics, Department of Genetics, Institut Curie, 26 rue d'Ulm, Paris 75005, France. Phone: 331-7238-9363. Fax: 331-5310-2665; E-mail: ivan.bieche@curie.fr

doi: 10.1158/1541-7786.MCR-15-0418

©2016 American Association for Cancer Research.

pivotal in cancer initiation, progression, and metastatic spread. Among important lncRNAs deregulated in cancer, ANRIL (antisense noncoding RNA in the INK4 locus) is encoded in the chromosome 9p21 region (9) and has been reported to have a direct role in recruiting PRC2 and PRC1 complexes to specific loci and in repressing gene expression (10). ANRIL is transcribed as a 3.8-kb lncRNA in the opposite direction from the p15/CDKN2B-p16/CDKN2A-p14/ARF gene cluster. Common disease genome-wide association studies (GWAS) have identified the ANRIL gene as a shared genetic susceptibility locus to coronary disease, intracranial aneurysm, type 2 diabetes, and numerous cancers (11). Increased ANRIL expression levels were observed in prostate carcinomas and involved in repression of the CDKN2B/CDKN2A/ARF gene cluster in cis by directly binding to PRC complexes (10). EZH2 (PRC2) and CBX7 (PRC1) act as transcriptional repressors of many genes and are particularly implicated in silencing via ANRIL of the CDKN2B/CDKN2A/ARF locus (12). However, the functional role and underlying mechanism of action of ANRIL in breast carcinogenesis still remain unclear (13).

The aim of the present study was to analyze at RNA and protein levels, the ribonucleoprotein network composed of the lncRNA ANRIL, the two polycomb complexes PRC2/PRC1, and the downstream CDKN2B/CDKN2A/ARF gene cluster, as well as several miRNAs targeting PRC units, in IBCs and to evaluate their clinical significance.

Materials and Methods

Patients and samples

Samples of 456 primary unilateral invasive primary breast tumors excised from women managed at Institut Curie (Saint-Cloud, France) from 1978 to 2008 have been analyzed. All patients cared in our institution before 2007 were informed that their tumor samples might be used for scientific purposes and had the opportunity to decline. Since 2007, patients treated in our institution have given their approval by signed informed consent. This study was approved by the local ethics committee (Breast Group of Rene Huguenin Hospital). Samples were immediately stored in liquid nitrogen until RNA extraction. A tumor sample was considered suitable for this study if the proportion of tumor cells exceeded 70%. The patients (mean age, 61.7 years; range, 31–91 years) all met the following criteria: primary unilateral nonmetastatic breast carcinoma for which complete clinical, histologic, and biologic data were available; no radiotherapy or chemotherapy before surgery; and full follow-up at Institut Curie. Treatment consisted of modified radical mastectomy in 283 cases (63.9%) and breast-conserving surgery plus locoregional radiotherapy in 160 cases (36.1%). The patients had a physical examination and routine chest radiography every 3 months for 2 years and then annually. Mammograms were done annually. Adjuvant therapy was administered to 369 patients, consisting of chemotherapy alone in 91 cases, hormone therapy alone in 176 cases, and both treatments in 102 cases. The histologic type and the number of positive axillary nodes were established at the time of surgery. The malignancy of infiltrating carcinomas was scored according to Scarff Bloom Richardson (SBR) histoprosthetic system. Estrogen receptor (ER α), progesterone receptor (PR), and human EGF receptor 2 (ERBB2) status was determined at the protein level by using biochemical methods [dextran-coated charcoal method,

enzyme immunoassay, or immunohistochemistry (IHC)] and confirmed by real-time quantitative RT-PCR assays (14, 15). The population was divided into 4 groups according to HR (ER α and PR) and ERBB2 status, as follows: two luminal subtypes [HR β (ER $\alpha\beta$ or PR β)/ERBB2 β (n ¼ 54)] and [HR β (ER $\alpha\beta$ or PR β)/ERBB2– (n ¼ 290)]; an ERBB2 β subtype [HR– (ER α – and PR–)/ERBB2 β (n ¼ 45)]; and a triple-negative subtype [HR– (ER α – and PR–)/ERBB2– (n ¼ 69)]. Standard prognostic factors are shown in Supplementary Table S1. Patients (n ¼ 169) developed metastasis during a median follow-up of 8.9 years (range, 6 months to 29 years). Ten specimens of adjacent normal breast tissue from patients with breast cancer (n ¼ 7) and normal breast tissue from women undergoing cosmetic breast surgery (n ¼ 3) were used as sources of normal RNA.

RNA extraction

Total RNA was extracted from breast tumor samples by using acid-phenol guanidium as previously described (16). RNA quality was determined by electrophoresis through agarose gels, staining with ethidium bromide, and visualization of the 18S and 28S RNA bands under ultraviolet light.

Real-time RT-PCR

lncRNA and protein-coding genes expression analysis. ANRIL, EZH2, SUZ12, EED, JARID2, CBX7, BMI1, P16/CDKN2A, P15/CDKN2B, and P14/ARF gene mRNA expression levels were quantified by using real time RT-PCR in a retrospective series of 456 well-characterized tumors. Quantitative values were obtained from the cycle number (C_t value) at which the increase in the fluorescence signal associated with exponential growth of PCR products started to be detected by the laser detector of the ABI Prism 7900 Sequence Detection System (Perkin-Elmer Applied Biosystems), using PE Biosystems analysis software according to the manufacturer's manuals. The precise amount of total RNA added to each reaction mix (based on optical density) and its quality (i.e., lack of extensive degradation) are both difficult to assess. We therefore also quantified transcripts of the TBP gene (Genbank accession: NM_003194) encoding the TATA box-binding protein (a component of the DNA-binding protein complex TFIID) as an endogenous RNA control and normalized each sample on the basis of its TBP content. We selected TBP as an endogenous control because the prevalence of its transcripts is moderate and because there are no known TBP retropseudogenes (retropseudogenes lead to coamplification of contaminating genomic DNA and thus interfere with RT-PCR, despite the use of primers in separate exons; ref. 14). Results, expressed as N-fold differences in target gene expression relative to the TBP gene and termed " N_{target} ," were determined as $N_{target} \frac{1}{2} 2^{DC_t sample}$, where the DC_t value of the sample was determined by subtracting the average C_t value of the target gene from the average C_t value of the TBP gene. The N_{target} values of the samples were subsequently normalized such that the median of the N_{target} values for the 10 normal breast tissues was 1. The primers for TBP, ANRIL, and the 9 protein-coding genes were chosen with the assistance of the Oligo 6.0 program (National Biosciences). We scanned the dbEST and nr databases to confirm the total gene specificity of the nucleotide sequences chosen for the primers and the absence of SNPs. To avoid amplification of contaminating gDNA, 1 of the 2 primers was placed at the junction between 2 exons or on 2 different exons. Agarose gel electrophoresis was used to verify the specificity of PCR amplicons. The conditions of cDNA synthesis and PCR were as described (17).

miRNAs expression analysis. miRNAs were isolated with the RNA extraction procedure used for the protein-coding transcripts (total RNA extraction). Reverse transcription was performed with the QIAGEN miScript Reverse Transcription Kit, according to the manufacturer's protocol (QIAGEN, GmbH). Specific miRNAs were quantified by real-time PCR with the QIAGEN miScript SYBR Green PCR Kit (QIAGEN). The small nucleolar RNU44 was used as an internal control. The relative expression level of each miRNA, expressed as N-fold difference in target miRNA expression relative to RNU44, and termed " N_{target} ", was calculated as follows: $N_{\text{target}} = 2^{\text{DCt}_{\text{sample}} - \text{DCt}_{\text{RNU44}}}$. The DCt value of a given sample was determined by subtracting the Ct value of the target miRNA from the Ct value of RNU44. The N_{target} values of the samples were subsequently normalized so that the median N_{target} value of samples with low level of EZH2 and CBX7 transcripts was 1.

IHC

We performed IHC assay by using CBX7 (Novus, polyclonal, rabbit, NBP1-79042, 1/250, pH6), EZH2 (Novocastra, monoclonal mouse, 6A10, 1/100, pH6), p15 (Abcam, polyclonal, rabbit, ab53034, 1/50, pH6), p16 (CINtec p16, Ventana 705-4713, mouse, monoclonal, E6H4 prediluted, 1 mg/mL), H3K27me2/3 [Abcam, anti-histone H3 (di- p tri-methyl K27), ab6147, mouse, monoclonal, 1/200, pH6], and H3K27ac [Abcam, ab 4729, Rabbit polyclonal to Histone H3 (acetyl K27), 1/400, pH6] antibodies in a series of 80 IBCs. Paraffin-embedded tissue blocks, obtained at the time of the initial diagnosis, were retrieved from the archives of the Department of Biopathology, Institut Curie. Sections of 3 mm were cut with a microtome from the paraffin-embedded tissue blocks of normal breast tissue, preinvasive lesions, and IBCs. Tissue sections were deparaffinized and rehydrated through a series of xylene and ethanol washes. Briefly, key figures included: (i) antigen retrieval in 0.1 mol/L citrate buffer, pH 6 (Biocare) in a pressure cooker (4 minutes); (ii) blocking of endogenous peroxidase activity by immersing sections in 3% hydrogen peroxide in methanol for 15 minutes and subsequently rinsing them in water and PBS; (iii) incubation with primary antibodies against the targeted antigen; (iv) immunodetection with a biotin-conjugated secondary antibody formulation that recognizes rabbit and mouse immunoglobulins, followed by peroxidase-labeled streptavidin and linking with a rabbit biotinylated antibody against mouse immunoglobulin G (DAKO SA); and (v) chromogenic revelation with AEC and counterstaining with Mayer's hematoxylin. All immunostaining was processed by using a DAKO automated immunostaining device. The specificity of the antibodies was confirmed by doing IHC studies with the same protocol on paraffin-embedded human tissue sections containing lymphocytes. A semiquantitative histologic score (H-score = intensity \times frequency) was performed (score 0 = negative staining, score 1 = weak staining, score 2 = moderate staining, score 3 = strong staining).

Statistical analysis

The distributions of target mRNA levels were characterized by their median values and ranges. Relationships between mRNA levels of the different target genes and between mRNA (and protein) levels and clinical parameters were identified by using nonparametric tests, namely, the χ^2 test (relation between 2 qualitative parameters), the Mann–Whitney U test (relation between 1 qualitative and 1 quantitative parameter), and the Spearman rank correlation test (relation between 2 quantitative

parameters). Differences were considered significant at confidence levels greater than 95% ($P < 0.05$). Metastasis-free survival (MFS) was determined as the interval between initial diagnosis and detection of the first metastasis. Survival distributions were estimated by the Kaplan–Meier method, and the significance of differences between survival rates was ascertained with the log-rank test. The Cox proportional hazards regression model was used to assess prognostic significance, and the results are presented as HRs and 95% confidence intervals (CI). Hierarchical clustering was performed with GenANOVA software (18).

Results

ANRIL and EZH2 are upregulated, whereas CBX7 is downregulated in IBCs

We measured mRNA levels of the lncRNA ANRIL and 6 genes belonging to the two repressive polycomb PRC2 and PRC1 complexes (PRC2: EZH2, SUZ12, EED, and JARID2; PRC1: BMI1 and CBX7) in a series of 456 patients with unilateral invasive breast tumors and in a series of 10 normal breast tissues (including adjacent normal breast tissue from patients with breast cancer and normal breast tissue from women undergoing cosmetic breast surgery) by using quantitative RT-PCR assays (Table 1). We did not observe significant differences concerning the expression profiles of the 7 genes tested between the adjacent normal breast tissue from patients with breast cancer (7 samples) and normal breast tissue from women undergoing cosmetic breast surgery (3 patients; data not shown). The mRNA values of the breast cancer samples were normalized such that the median of the 10 normal breast tissue mRNA values was 1. To determine the cutoff point for altered target genes expression in breast cancer tissues, the N_{target} value, calculated as described in Materials and Methods, was determined for the 10 normal breast RNA samples. As these normal values were consistently between 0.39 (CBX7) and 2.45 (ANRIL), values of 0.33 or less were considered to represent underexpression, and values of 3 or more to represent overexpression of these genes in tumor samples. We have previously used the same approach to determine cutoff points for tumor gene altered expression (14, 19). In our series of 456 breast carcinomas, we observed overexpression of ANRIL in 19.7% of IBCs (median = 1.58; minimum = 0.05; maximum = 13.5). Concerning Polycomb subunits, we observed a marked overexpression of EZH2 in 77.0% of IBCs (median = 5.09; minimum = 0.76; maximum = 48.3) and to a lesser extent of BMI1 (median = 1.23; minimum = 0.00; maximum = 14.0) in 10.1% of IBCs and an unexpected underexpression of CBX7 in 39.7% of IBCs (median = 0.38; minimum = 0.00; maximum = 2.59). SUZ12, EED, and JARID2 were overexpressed only in 2.6%, 1.8%, and 5.1% of IBCs, respectively.

We performed an IHC analysis by using anti-EZH2 and anti-CBX7 antibodies in a series of 80 IBCs (standard prognostic factors are shown in Supplementary Table S2). We identified a moderate-to-intense nuclear staining of tumor cells (H-score = 2–3) with EZH2 antibodies in 75.9% of IBCs (n = 63 of 80) and an absence or a slight nuclear and cytoplasmic staining of tumor cells (H-score = 0–1) with CBX7 antibodies in 63.8% of IBCs (n = 51 of 80; Fig. 1).

The p15/CDKN2B-p16/CDKN2A-p14/ARF cluster is unexpectedly not silenced in IBCs

We measured mRNA levels of p15/CDKN2B, p16/CDKN2A, and p14/ARF in the series of 456 patients. We showed marked

Table 1. ANRIL pathway genes mRNA expression levels in the series of 456 breast tumors

Genes	Median C _t of normal breast tissue (n ¼ 10)	Normal breast tissue (n ¼ 10)	Breast tumors (n ¼ 456)	Percentage of underexpressed tumors (N _{target} < 0.33)	Percentage of normal expressed tumors	Percentage of overexpressed tumors (N _{target} > 3)
ANRIL	34.07 (31.91–35.96) ^a	1.0 (0.35–2.45) ^b	1.58 (0.05–13.5) ^b	0.9 ^c	79.4 ^c	19.7 ^c
Polycomb complex PRC2 genes (n ¼ 4)						
EZH2	32.39 (31.23–34.30)	1.0 (0.48–2.32)	5.09 (0.76–48.3)	0	23.0	77.0
SUZ12	28.26 (27.17–29.98)	1.0 (0.56–1.25)	1.23 (0.43–59.2)	0	97.4	2.6
EED ^d	29.77 (28.70–33.16)	1.0 (0.58–1.84)	1.21 (0.03–7.74)	3.1	95.1	1.8
JARID2 ^e	28.63 (27.3–30.55)	1.0 (0.83–1.46)	1.37 (0.34–8.53)	0	94.9	5.1
Polycomb complex PRC1 genes (n ¼ 2)						
BMI1	28.40 (26.91–30.22)	1.0 (0.59–1.47)	1.23 (0.00–14.0)	4.2	85.8	10.1
CBX7	28.99 (27.78–31.25)	1.0 (0.39–1.62)	0.38 (0.00–2.59)	39.7	60.3	0
ANRIL regulated genes (n ¼ 3)						
ARF	32.93 (31.69–34.75)	1.0 (0.21–2.41)	2.28 (0.00–105)	3.3	56.8	39.9
P16 ^f	37.55 (34.64–50)	0.0 (0.0–6.95)	0.65 (0.0–63.4)	0	84.2	15.8
P15	30.89 (29.21–33.74)	1.0 (0.11–2.91)	1.62 (0.00–40.5)	7.0	71.3	21.7

^aMedian (range) of gene C_t values (cycle threshold).

^bMedian (range) of gene mRNA levels; the mRNA values of the samples were normalized such that the median of the 10 normal breast tissues mRNA values was 1.

^cPercentage of underexpressing, normal, and overexpressing tumors using cutoffs of N_{target} < 0.33 and N_{target} > 3.

^dData available in 446 samples.

^eData available in 454 samples.

^fThe mRNA values of the samples were normalized such that a C_t value of 35 was 1.

overexpression of p16, p14/ARF, and p15, respectively, in 15.8%, 39.9%, and 21.7% of IBCs (Table 1).

We studied protein expression of p14, p16, and p15 in IBCs by using IHC in our series of 80 breast tumors. We observed nuclear and cytoplasmic overexpression of cancer cells (H-score ¼ 2–3) with anti-p14, p16, and p15 antibodies in, respectively, 41.2% (n ¼ 33 of 80), 51.3% (n ¼ 41 of 80), and 37.5% (n ¼ 30 of 80) of IBCs (Supplementary Fig. S2).

PRC2/PRC1 activity status revealed decrease of H3K27me3 expression levels and increase of H3K27ac expression levels in IBCs

To evaluate PRC2/PRC1 repressive activity status in breast tumors, we performed an IHC analysis in our series of 80 IBCs by using anti-H3K27me3 and anti-H3K27ac antibodies. We could identify an absence or a slight nuclear staining of cancer cells (H-score ¼ 0–1) with H3K27me3 antibody in 76.3% of IBCs (n ¼ 61 of 80) and a moderate-to-intense nuclear staining of tumor cells (H-score ¼ 2–3) with H3K27ac antibody in 73.8% of IBCs (n ¼ 59 of 80; Supplementary Fig. S3). An expected negative correlation was observed between staining of H3K27me3 and H3K27ac (P ¼ 0.0000073).

ANRIL, PRC2/PRC1 subunits and p15/CDKN2B-p16/CDKN2A-p14/ARF cluster expression levels are associated with clinicopathologic parameters, molecular subtypes, and patient outcome

Clinicopathologic parameters. ANRIL mRNA expression level was exclusively and weakly correlated with ERa (P ¼ 0.038) and PR (P ¼ 0.011) status (Supplementary Table S3).

Concerning the PRC2 complex, EZH2 mRNA expression level was strongly associated with SBR histologic grade (P < 10⁻⁷), ERa status (P ¼ 3 × 10⁻⁷), and PR status (P ¼ 3 × 10⁻⁷), moderately associated with ERBB2 status (P ¼ 0.00020) and tumor size (P ¼ 0.0062; Supplementary Table S4). Expression of SUZ12 was exclusively associated with lymph node status (P < 10⁻⁷; Supplementary Table S5), EED with SBR histologic grade (P ¼ 0.0011), and PR status (P ¼ 0.0028; Supplementary

Table S6) and JARID2 with ERa status (P ¼ 0.0027) and ERBB2 status (P ¼ 0.0035; Supplementary Table S7).

Concerning the PRC1 complex, CBX7 mRNA expression level was highly significantly correlated with SBR histologic grade (P ¼ 0.000025), ER status (P ¼ 0.000011), and PR status (P ¼ 0.0000013; Supplementary Table S8). BMI1 was weakly associated with ERa status (P ¼ 0.036) but not with any of other clinical predictive factors (Supplementary Table S9).

Finally, ARF/p14 and p16 mRNA expression levels were highly correlated with SBR histological grade (P < 10⁻⁷ and P ¼ 0.000019, respectively), ERa status (P ¼ 1.8 × 10⁻⁷ and P < 10⁻⁷, respectively), and PR status (P ¼ 1.3 × 10⁻⁵ and P ¼ 1.8 × 10⁻⁶, respectively; Supplementary Table S10), whereas p15 was moderately correlated with ERa status (P ¼ 0.009) and PR status (P ¼ 0.0008), but not with SBR histologic grade, like ARF/p14 and p16 (Supplementary Tables S11 and S12).

It is also noteworthy that none of these 10 gene expressions showed significant link with the adjuvant treatment status including hormone therapy alone (176 patients), chemotherapy alone (91 patients), both hormone therapy and chemotherapy (102 patients), and no adjuvant therapy (87 patients; Supplementary Tables S3–S12).

Breast cancer molecular subtypes. By using HR (ERa and PR) and ERBB2 status, we subdivided our total population of IBCs (n ¼ 456) into 4 subgroups: HR⁺/ERBB2⁺ (n ¼ 54), HR⁺/ERBB2⁻ (n ¼ 289), HR⁻/ERBB2⁺ (n ¼ 45), and HR⁻/ERBB2⁻ (n ¼ 68). ANRIL, PRC1/2 genes, and p14/p15/p16 mRNA levels of expression in the different breast cancer molecular subtypes are shown in Supplementary Table S13. A strong correlation was observed between molecular subgroups and ANRIL (P ¼ 0.00013), EZH2 (P < 10⁻⁷), JARID2 (P ¼ 0.00090), CBX7 (P ¼ 0.000020), and p14/ARF and p16 (P < 10⁻⁷) mRNA levels. High frequencies of ANRIL, p14/ARF, and p16 deregulated expressions (under- or overexpressions) were mainly observed in triple-negative subtypes, EZH2 overexpression, and CBX7 underexpression in the two HR⁻ subtypes, and JARID2 overexpression in HR⁻/ERBB2⁺ subtypes (Supplementary Table S13).

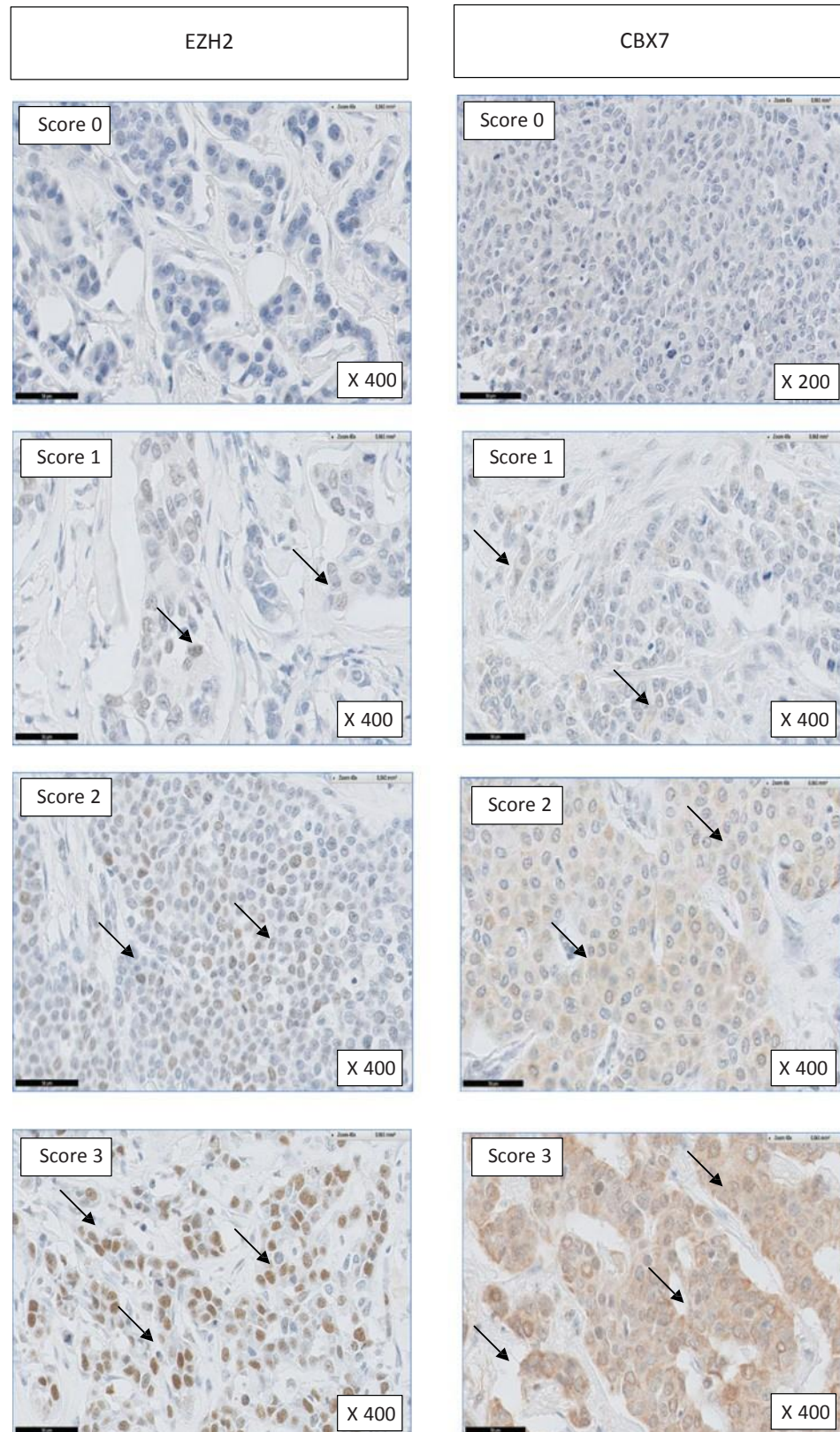


Figure 1. IHC staining for EZH2 and CBX7. Nuclear overexpression of EZH2 (H-score ≥ 2 and 3) in cancer cells was observed in 75.2% of IBCs and nuclear EZH2 underexpression in 24.8% of IBCs. Nuclear and cytoplasmic under expression of CBX7 (H-score ≤ 0 and 1) was observed in 63.8% of IBCs and nuclear and cytoplasmic CBX7 overexpression in 36.2% of IBCs. Score 0, no cytoplasmic staining for CBX7 or nuclear staining for EZH2 is seen in tumor cells; Score 1, weak cytoplasmic staining for CBX7 and weak nuclear staining for EZH2 is seen in a subset of tumor cells; Score 2, moderate cytoplasmic staining for CBX7 and moderate nuclear staining for EZH2 is seen in a subset of tumor cells; Score 3: strong and diffuse cytoplasmic staining for CBX7, and strong nuclear staining for EZH2 is seen in most of tumor cells.

Table 2. Relationship between ANRIL pathway gene mRNA levels and MFS in the 456 breast tumors

Gene mRNA expression	Population (%)	number or metastases (%)	P ^b
Total population (%)	456 (100.0)	169 (37.1)	
ANRIL			
Non-overexpression	366 (80.3)	129 (76.3)	0.062 (NS)
Overexpression	90 (19.7)	40 (23.7)	
EZH2			
Non-overexpression	105 (23.0)	20 (11.8)	0.000045
Overexpression	351 (77.0)	149 (88.2)	
SUZ12			
Non-overexpression	444 (97.4)	163 (96.4)	0.71 (NS)
Overexpression	12 (2.6)	6 (3.6)	
EED ^b			
Underexpression	14 (3.1)	3 (1.8)	0.20 (NS)
Nonunderexpression	432 (96.9)	161 (98.2)	
JARID2 ^c			
Non-overexpression	431 (94.9)	161 (96.4)	0.80 (NS)
Overexpression	23 (5.1)	6 (3.6)	
BMI1			
Non-overexpression	410 (89.9)	148 (87.6)	0.49 (NS)
Overexpression	46 (10.1)	21 (12.4)	
CBX7			
Underexpression	181 (39.7)	79 (46.7)	0.0051
Non-underexpression	275 (60.3)	90 (53.3)	
ARF			
Non-overexpression	274 (60.1)	96 (56.8)	0.086 (NS)
Overexpression	182 (39.9)	73 (43.2)	
P16			
Normal expression	384 (84.2)	137 (81.1)	0.025
Overexpression	72 (15.8)	32 (18.9)	
P15			
Non-overexpression	357 (78.3)	128 (75.7)	0.090 (NS)
Overexpression	99 (21.7)	41 (24.3)	

Abbreviation: NS, not significant.

^aLog-rank test.

^bInformation available for 446 patients.

^cInformation available for 454 patients.

Patient outcome. We assessed the impact of variations of ANRIL, EZH2, SUZ12, EED, JARID2, CBX7, BMI1, p14/ARF, p15, and p16 mRNA levels on patient outcome. We used a log-rank test to identify relations between MFS and mRNA levels of these 10 genes (Table 2). EZH2 overexpression and CBX7 under-expression were significantly associated with shorter MFS

($P = 0.000045$ and $P = 0.0051$, respectively). We observed also a light significant difference in MFS among patients with p16 overexpression ($P = 0.025$). MFS was not influenced by expression status of ANRIL and other genes analyzed (i.e. SUZ12, EED, JARID2, BMI1, ARF, and p15).

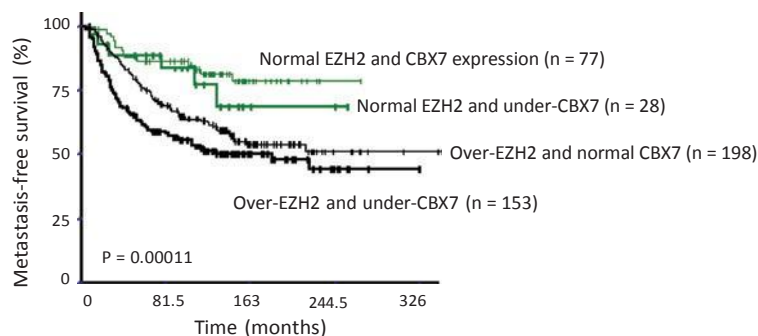
Two molecular markers may provide a more accurate prediction of patient outcome when combined than when considered in isolation. By combining EZH2 and CBX7 status, we identified four separate prognostic groups with significantly different MFS curves ($P = 0.00011$; Fig. 2). The patients with the poorest prognosis had EZH2 overexpression and CBX7 underexpression [5-year MFS: 61.7% (57.7%–65.7%); 10-year MFS: 51.4% (47.2%–55.6%)], whereas those with the best prognosis had EZH2 and CBX7 normal expression [5-year MFS: 86.4% (82.4%–90.4%); 10-year MFS: 81.2% (76.5%–85.9%)].

Multivariate analysis (Cox proportional hazards model) showed that lymph node status, tumor size, and combined EZH2/CBX7 status were independent variables predictive of MFS ($P = 0.00021$, $P = 0.014$, and $P = 0.001$, respectively; Supplementary Table S14).

There is a complex interplay between ANRIL, PRC1/PRC2 subunits and p15/CDKN2B-p16/CDKN2A-p14/ARF genes cluster

By using the Spearman rank correlation test for continuous variables, we could find a strong positive correlation between mRNA expression levels of the following genes belonging to the ANRIL pathway: (i) ANRIL and both ARF ($r = 0.594$, $P < 10^{-7}$), p15 ($r = 0.381$, $P < 10^{-7}$), and p16 ($r = 0.323$, $P < 10^{-7}$); (ii) EZH2 and both EED ($r = 0.295$, $P < 10^{-7}$), JARID2 ($r = 0.278$, $P < 10^{-7}$), and BMI1 ($r = 0.229$, $P = 0.0000031$); and (iii) SUZ12 and CBX7 ($r = 0.317$, $P < 10^{-7}$), reflecting very complex interplay across this gene set (Supplementary Table S15).

Hierarchical clustering of ANRIL pathway genes according to their expression patterns in the series of 456 IBCs represented in dendrogram format, subdivided genes into two subgroups (Supplementary Fig. S4). Group I included ANRIL, p14/ARF, p15, p16, EZH2, EED, JARID2, and BMI1 and contained 4 genes, ANRIL and the p15/CDKN2B-p16/CDKN2A-p14/ARF gene cluster, highly related. Group II included SUZ12 (from the PRC2 complex) and CBX7 (from the PRC1 complex).



Patients at risk (n)	0	81.5	163	244.5	326
Normal EZH2 and CBX7	59	15	2	0	0
Normal EZH2 and under-CBX7	17	4	2	0	0
Over-EZH2 and normal CBX7	114	37	10	3	3
Over-EZH2 and under-CBX7	78	27	8	1	1

Figure 2. Survival curves of 4 patients groups according to EZH2 mRNA overexpression (over-EZH2) and CBX7 mRNA underexpression (under-CBX7) in the series of 456 breast tumors.

Table 3. Correlations between ANRIL pathway and EMT, proliferation and breast cancer stem cell genes mRNA expression levels in the series of 456 breast tumors

	EMT		Proliferation	Stem cell		
	VIM	TWIST1	MKI67	ALDH1A1	ALDH1A3 ^c	CD133 ^d
ANRIL	ρ 0.134 ^a 0.0043	ρ 0.117 0.012	ρ 0.187 0.0001	ρ 0.182 0.00015	ρ 0.050 0.29 (NS)	-0.098 0.038
EZH2	-0.068 0.14 (NS)	-0.010 0.83 (NS)	ρ 0.738 <0.0000001	-0.164 0.00057	ρ 0.109 0.020	ρ 0.043 0.37 (NS)
SUZ12	ρ 0.165 0.00052	ρ 0.131 0.0053	ρ 0.191 0.000075	ρ 0.049 0.30 (NS)	ρ 0.013 0.78 (NS)	-0.128 0.0073
EED ^b	ρ 0.115 0.014	ρ 0.123 0.0092	ρ 0.208 0.000023	ρ 0.123 0.0093	ρ 0.295 <0.0000001	ρ 0.155 0.0015
JARID2 ^c	ρ 0.214 0.000012	ρ 0.116 0.013	ρ 0.276 <0.0000001	ρ 0.061 0.19 (NS)	ρ 0.233 0.0000022	ρ 0.088 0.064 (NS)
BMI1	-0.141 0.0028	-0.081 0.080 (NS)	ρ 0.176 0.00023	-0.122 0.0087	-0.071 0.13 (NS)	-0.199 0.0000053
CBX7	ρ 0.300 <0.0000001	ρ 0.328 <0.0000001	-0.261 0.00000016	ρ 0.464 <0.0000001	ρ 0.190 0.000082	-0.028 0.57 (NS)
ARF	-0.032 0.50 (NS)	ρ 0.059 0.20 (NS)	ρ 0.351 <0.0000001	-0.029 0.54 (NS)	ρ 0.126 0.0071	ρ 0.043 0.37 (NS)
P16	ρ 0.035 0.47 (NS)	ρ 0.154 0.0011	ρ 0.114 0.015	ρ 0.001 0.97 (NS)	ρ 0.185 0.00013	ρ 0.160 0.00093
P15	ρ 0.155 0.0011	ρ 0.279 <0.0000001	ρ 0.025 0.60 (NS)	ρ 0.178 0.00021	ρ 0.282 <0.0000001	ρ 0.034 0.48 (NS)

Abbreviation: NS, not significant.

^aSpearman rank correlation test.

^bData available in 446 samples.

^cData available in 454 samples.

^dData available in 438 samples.

Correlations of ANRIL and PRC2/PRC1 genes expression levels with epithelial–mesenchymal transition, proliferation, and breast cancer stem cell markers

Data from the literature have suggested that Polycomb complexes are implicated in proliferation, epithelial–mesenchymal transition (EMT) induction, and stem cell differentiation. In consequence, we tested the possible links between ANRIL pathway gene expression and mRNA levels of various markers of proliferation (i.e., MKI67), of EMT (i.e., VIM and TWIST1), and of breast cancer stem cell (i.e., ALDH1A1, ALDH1A3, and CD133) in our series of 456 IBCs (Table 3). Marked positive associations were observed between CBX7 and two EMT markers (VIM and TWIST1; $P < 10^{-7}$ and $P < 10^{-7}$, respectively). A significant positive correlation was also identified between CBX7 and one breast cancer stem cell marker (ALDH1A1; $P < 10^{-7}$). EZH2 was highly linked to MKI67 ($P < 10^{-7}$), whereas EED was linked to two breast cancer stem cell markers: ALDH1A3 and CD133 ($P < 10^{-7}$ and $P \frac{1}{4} 0.000082$, respectively).

As previous studies suggested a role of ANRIL in EMT (13), we further investigated the correlations between the expression levels of ANRIL and 9 different biomarkers of EMT in a series of 60 breast carcinomas (30 tumors with high expression level for ANRIL and 30 tumors with low expression level for ANRIL). We confirmed the absence of significant associations between ANRIL and various EMT-associated markers including TWIST1, ZEB1, ZEB2, SNAI1/Snail, SNAI2/Slug, VIM/Vimentin, CDH2/N-cadherin, CDH1/E-cadherin, and ZO-1 (Supplementary Table S16).

miRNAs are implicated in posttranscriptional regulation of PRC1/PRC2

Previous studies suggested complex regulations between Polycomb subunit genes and miRNAs in carcinomas (20, 21). To investigate roles of miRNAs in posttranscriptional deregu-

lation of the ANRIL/PRC pathway in IBCs, we selected 17 miRNAs from the literature. Among them, 4 miRNAs (miR-26A1, miR-101, miR-125B2, and miR-214) are known to target directly EZH2 transcript (22–24) and 2 miRNAs (miR-181B1, miR-181A2) to target directly CBX7 transcript (25). The 11 others miRNAs also target putatively EZH2 and CBX7 (26). We analyzed expression of these 17 miRNAs, using real-time quantitative RT-PCR, in IBCs with normal ($n \frac{1}{4} 20$) and marked low ($n \frac{1}{4} 20$) levels of CBX7 transcripts and with normal ($n \frac{1}{4} 20$) and marked high ($n \frac{1}{4} 20$) levels of EZH2 transcripts. We showed that (i) miR-181B1, miR-181A2, and miR-203 levels were marked significantly increased ($P < 0.01$) in CBX7-underexpressed tumors (Table 4) and that (ii) miR-26A1 and miR-125B1 levels were significantly decreased in EZH2-overexpressed tumors (Table 5).

These results of microRNA levels, shown in Tables 4 and 5, were standardized versus RNU44 as endogenous microRNA control. Similar results (significantly links between microRNA and CBX7 and EZH2 expressions) were obtained using a second well-known endogenous microRNA control (i.e., miR-191).

Discussion

Among the major discovered lncRNAs deregulated in carcinogenesis, ANRIL has been shown to regulate its neighbor tumor suppressor genes CDKN2B, CDKN2A, and ARF by epigenetic mechanisms and thereby control cell proliferation and senescence (27). ANRIL, acting in cis, specifically binds various Polycomb proteins, in particular EZH2 (PRC2) and CBX7 (PRC1), resulting in histone modification in the CDKN2A/CDKN2B/ARF locus. EZH2 is a histone methyltransferase and a member of PRC2 that not only methylates histone H3 on Lysine 27 but also interacts with and recruits DNA methyltransferases to methylate CpG of certain EZH2 target genes to

Table 4. Statistical analysis of microRNAs expressions relative to CBX7 mRNA expression level

MiRNAs	Breast cancer with underexpression	Breast cancer with normal expression	P ^a	ROC-AUC ^b
	of CBX7 (n = 20)	of CBX7 (n = 20)		
MIRN181B1	1.0 (0.22–2.29) ^a	0.34 (0.12–1.30) ^b	0.000024	0.11
MIRN181A2	1.0 (0.33–4.83)	0.39 (0.13–1.92)	0.00011	0.142
MIRN203	1.0 (0.25–4.72)	0.42 (0.16–1.02)	0.0058	0.245
MIRN200C	1.0 (0.22–3.34)	0.45 (0.11–1.77)	0.02	0.285
MIRN98	1.0 (0.17–7.66)	0.76 (0.14–3.35)	0.048	0.317
MIRN200B	1.0 (0.20–3.33)	0.73 (0.24–1.94)	0.11 (NS)	0.352
MIRN21	1.0 (0.25–3.62)	0.97 (0.13–6.49)	0.30 (NS)	0.405
MIRN192	1.0 (0.20–7.00)	0.75 (0.11–84.4)	0.37 (NS)	0.418
MIRN219-1	1.0 (0.00–4.05)	0.55 (0.10–7.50)	0.59 (NS)	0.45
MIRN200A	1.0 (0.02–3.92)	0.79 (0.14–6.33)	0.65 (NS)	0.458
MIRN183	1.0 (0.30–4.93)	0.55 (0.07–3.84)	0.70 (NS)	0.465
MIRN138-2	1.0 (0.09–33.9)	2.10 (0.06–41.0)	0.89 (NS)	0.512
MIRN217	1.0 (0.22–2.96)	1.32 (0.19–7.49)	0.23 (NS)	0.611
MIRN214	1.0 (0.29–2.12)	1.51 (0.26–3.22)	0.088 (NS)	0.658
MIRN101-1	1.0 (0.17–4.02)	1.95 (0.26–13.4)	0.041	0.689
MIRN125B1	1.0 (0.43–4.51)	2.12 (0.27–12.1)	0.019	0.718
MIRN26A1	1.0 (0.24–4.64)	3.05 (0.57–21.4)	0.015	0.725

Abbreviations: AUC, area under curve; NS, not significant; ROC, receiver operating characteristics.

^aKruskal–Wallis H test.

^bMedian (range) of microRNA levels.

establish firm repressive chromatin structures, contributing to tumor progression. CBX7 is a Chromobox family protein and a member of the PRC1 and is directly involved in regulation of genes frequently silenced in cancer.

To assess expression of the lncRNA ANRIL and the Polycomb subunits of PRC2/PRC1 as well as CDKN2A/CDKN2B/ARF locus in IBCs, we used qRT-PCR in a large series of 456 IBCs from patients with known clinical and pathologic status and long-term outcome. Evaluation at mRNA level was combined with an IHC study in a series of 80 IBCs. Indeed, these two techniques of expression analysis (qRT-PCR and IHC) give complementary information, the technique of qRT-PCR with an extraction in homogeneous solution quantified transcripts levels both in tumor cells and in stroma cells, while IHC study allows only a semiquantitative analysis of protein levels, but on an individual

cell basis (allowing distinction between epithelial tumor cells and stroma cells).

In the present study, we observed ANRIL overexpression in 19.7% of IBCs. ANRIL depletion has been associated with reduced proliferation (28), suggesting a procarcinogenic role supported by overexpression of this lncRNA in various others cancer types (29–31). In our IBC series, ANRIL overexpression was observed more particularly in the aggressive TNC subtype.

We also observed a marked overexpression of EZH2 and to a lesser extent of BMI1, respectively, in 77.0% and 10.1% of IBCs and CBX7 underexpression in 39.7% of the tumors. We confirmed these data at the protein level in the series of 80 IBCs by identifying a moderate and intense nuclear staining with an anti-EZH2 antibody in 82% of IBCs and an absence or a slight nuclear and cytoplasmic staining of cancer cells with a CBX7 antibody in 43%

Table 5. Statistical analysis of microRNAs expressions relative to EZH2 mRNA expression level

MicroRNA s	Breast cancer with normal	Breast cancer with overexpression	P ^a	ROC-AUC
	expression of EZH2 (n = 20)	of EZH2 (n = 20)		
MIRN26A1	1.0 (0.17–6.46) ^b	0.29 (0.04–5.76) ^b	0.0027	0.222
MIRN125B1	1.0 (0.09–5.74)	0.37 (0.14–1.60)	0.0063	0.247
MIRN214	1.0 (0.11–2.35)	0.44 (0.20–1.42)	0.016	0.277
MIRN101-1	1.0 (0.06–7.72)	0.51 (0.13–1.99)	0.37 (NS)	0.417
MIRN217	1.0 (0.14–4.25)	0.45 (0.18–465)	0.69 (NS)	0.464
MIRN138-2	1.0 (0.12–36.3)	0.67 (0.07–25.7)	0.70 (NS)	0.465
MIRN192	1.0 (0.09–97.0)	0.91 (0.26–7.65)	0.85 (NS)	0.482
MIRN21	1.0 (0.08–3.86)	0.85 (0.21–5.98)	0.98 (NS)	0.498
MIRN203	1.0 (0.48–3.16)	1.59 (0.29–8.84)	0.76 (NS)	0.529
MIRN200A	1.0 (0.14–3.75)	1.66 (0.25–13.7)	0.30 (NS)	0.595
MIRN98	1.0 (0.43–3.47)	1.50 (0.26–13.2)	0.088 (NS)	0.657
MIRN200C	1.0 (0.45–2.37)	1.59 (0.28–8.50)	0.055 (NS)	0.677
MIRN200B	1.0 (0.17–1.77)	1.23 (0.40–8.87)	0.048	0.682
MIRN183	1.0 (0.05–4.35)	1.06 (0.26–10.8)	0.04	0.69
MIRN181A2	1.0 (0.20–2.28)	1.54 (0.46–5.97)	0.027	0.705
MIRN219-1	1.0 (0.16–3.56)	1.92 (0.32–14.0)	0.019	0.716
MIRN181B1	1.0 (0.41–2.37)	1.94 (0.44–5.30)	0.011	0.738

NOTE: miRNA levels were standardized versus U44 control microRNA.

Abbreviations: AUC, area under curve; NS, not significant; ROC, receiver operating characteristics.

^aKruskal–Wallis H test.

^bMedian (range) of microRNA levels.

of IBCs. In malignant tumors, ANRIL is a natural antisense transcript which usually recruits the two Polycomb repressor complexes PRC2 and PRC1, resulting in chromatin reorganization with silencing of the INK4A/ARF/INK4B locus. For example, positive links were observed between CBX7 overexpression, ANRIL overexpression, and CDKN2A/CDKNB/ARF underexpression in prostate cancer (10). Conversely, in our series of breast tumors, we showed marked CBX7 underexpression and a non-expected high positive link between ANRIL overexpression and CDKN2A/CDKNB/ARF overexpression, and more particularly with p14/ARF. Although these results seem to be contradictory, it can be hypothesized that CBX7 can exhibit both oncosuppressive and oncogenic functions, depending on type of cancer, microenvironment, and presence of interacting proteins (32–35).

CBX7 underexpression was strongly correlated with high SBR histologic grade, negative ER α and PR status, and EMT and breast cancer stem cell markers. Apart from few exceptions (as prostate cancer), CBX7 expression is lost in human malignant neoplasias, and a clear correlation between its downregulated expression and a cancer aggressiveness and poor prognosis has been observed (36).

EZH2 overexpression was also strongly associated with high SBR histologic grade and negative ER α and PR status. But, unlike CBX7, EZH2 overexpression was highly linked to the proliferation marker MKI67, suggesting a different role of these two proteins in breast tumorigenesis (37). Numerous studies point on the widespread roles of PRC2 (as well as H3K27 methylation) in developmental and differentiation processes of multicellular organisms, and on their implication in fundamental chromatin mechanisms that underlie stem cell regulatory circuits and cancer progression (38). Thus, it is not surprising that increasing evidences indicate that EZH2 deregulation is frequently observed in a variety of cancers (39–41). EZH2 overexpression is mainly found in solid tumors (as in our present study in breast cancer), whereas activating or inactivating mutations are identified in hematologic malignancies (42).

Using log-rank analysis to identify relations between MFS and mRNA levels of the 10 ANRIL pathway genes, only EZH2 overexpression and CBX7 underexpression were significantly associated with shorter MFS. Multivariate analysis showed that combined EZH2/CBX7 status was an independent prognostic factor.

Results from previous studies suggest complex interactions between Polycomb complexes and miRNAs in cancer progression. As no mutation or allelic loss of CBX7 located in 22q13.1 were identified in large genomic studies (43) and focal amplification of EZH2 located in 7q36 was only observed in 1% of IBCs (43), we suspected epigenetic mechanisms implicating miRNAs could account for CBX7 underexpression and EZH2 overexpression in IBCs. Among the 17 miRNAs selected, (i) miR-181B1, miR-181A2, and miR-203 expressions were significantly increased in IBCs with CBX7 underexpression and (ii) miR-26A1 and miR-125B2 expressions were significantly decreased in IBCs with EZH2 overexpression, in agreement with the literature (44, 45). Our results thus showed that EZH2 is targeted by 3 oncosuppressive miRNAs (miR-26A1, miR-125B, and miR-214) that are decreased in IBCs, suggesting an epigenetic mechanism of EZH2 overexpression in IBCs. Concerning CBX7, our results were also consistent with recent data in the literature, suggesting that the 2 oncomiRs miR-181B1 and

miR-181A2 could be induced by the oncogene HMGA1, resulting in CBX7 underexpression (46). Indeed, we observed a positive link between HMGA1 mRNA levels and expression of both miR-181B1 ($r = 0.340$, $P = 0.0038$) and miR-181A2 ($r = 0.264$, $P = 0.025$) in 71 tumors of our IBC series (data not shown).

Recent data revealed that histone modifier complexes PRC2 and PRC1 have a bivalent role in cancer, bearing both oncogenic and tumor-suppressive properties depending on cell tumor type (47). Complex deficiency in PRC2 and PRC1 components could decrease expression levels of epigenetic repressive mark H3K27 trimethylation, thereby contributing to H3K27 acetylation and resulting in increased transcription and activation of oncogenic pathways. Thus, defining a PRC2/PRC1 activity status in IBCs may be pivotal in combining or reconsidering therapeutic options. We identified a moderate-to-intense nuclear staining of tumor cells with H3K27ac antibodies in 73.8% of IBCs associated to an absence or a slight nuclear staining of cancer cells with H3K27me3 antibody in 76.3% of IBCs. Collectively, these data suggest that loss of Polycomb repressive function in IBCs might be due to epigenetic alteration of PRC1 and PRC2 by several microRNA targeting CBX7 and EZH2, resulting in altered H3K27 trimethylation and H3K27 acetylation. Development of therapies targeting epigenetic processes in cancer has gained increasing focus with study of HDAC and DNA methyltransferase inhibitors and more recently of BET bromodomain inhibitors (48). Indeed, recent studies of malignant peripheral nerve sheath tumors (MPNST) have showed that alterations of PRC2 subunits lead to the amplification of Ras-driven transcription and confer sensitivity to BET bromodomain inhibitors (49, 50).

In conclusion, our results revealed a complex pattern of interactions between ANRIL, PRC2/PRC1 and suppressive and oncogenic miRNAs in IBCs. Our findings also point to EZH2 and CBX7 as the most promising of the 10 genes investigated for clinical applications and therapeutic approaches to breast cancer. This study suggest that it is the global pattern of expression, rather than expression of individual family members, that should be taken into account when defining functionality of repressive Polycomb complexes and evaluating antitumor drugs targeting these genes. Thus, establishment of a PRC2/PRC1 activity status in IBCs could be contributive in targeting BET proteins by using bromodomain inhibitors when levels of H3K27me3 are decreased and H3K27ac increased. Moreover, inhibition of oncomiRNAs miR-181B1 and miR-181A2 (as well as restoration of oncosuppressive miR-26A1, miR-125B, and miR-214) may represent another attractive combined strategy for breast cancer-targeted therapy.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: D. Meseure, E. Pasmant, I. Bieche
Development of methodology: D. Meseure, S. Vacher, I. Bieche
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): D. Meseure, S. Vacher, K. Drak Alsibai, A. Nicolas, I. Bieche
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): D. Meseure, S. Vacher, K. Drak Alsibai, W. Chemlali, R. Lidereau, I. Bieche

Writing, review, and/or revision of the manuscript: D. Meseure, E. Pasmant, C. Callens, I. Bieche

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): K. Drak Alsibai, M. Caly, C. Callens, I. Bieche
Study supervision: I. Bieche

Grant Support

This work was supported by grant INCa-DGOS-4654.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received October 15, 2015; revised March 23, 2016; accepted April 8, 2016; published OnlineFirst April 21, 2016.

References

- Piao HL, Ma LJ. Non-coding RNAs as regulators of mammary development and breast cancer. *J Mammary Gland Biol Neoplasia* 2012;17:33–42.
- Kerppola TK. Polycomb group complexes: many combinations, many functions. *Trends Cell Biol* 2009;19:692–704.
- Margueron R, Reinberg D. The Polycomb complex PRC2 and its mark in life. *Nature* 2011;469:343–9.
- Luis NM, Morey L, Di Croce L, Benitah SA. Polycomb in stem cells: PRC1 branches out. *Cell Stem Cell* 2012;11:16–21.
- Iyer MK, Niknafs YS, Malik R, Singhal U, Sahu A, Hosono Y, et al. The landscape of long noncoding RNAs in the human transcriptome. *Nat Genet* 2015;47:199–208.
- Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res* 2012;22:1775–89.
- Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, et al. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* 2012;22:1760–1774.
- Wang KC, Chang HY. Molecular mechanisms of long noncoding RNAs. *Mol Cell* 2011;43:904–14.
- Pasmant E, Laurendeau I, Héron D, Vidaud M, Vidaud D, Bieche I. Characterization of a germ-line deletion, including the entire INK4/ARF locus, in a melanoma-neural system tumor family: identification of ANRIL, an antisense noncoding RNA whose expression coclusters with ARF. *Cancer Res* 2007;67:3963–9.
- Yap KL, Li S, Munoz-Cabello AM, Raguz S, Zeng L, Mujtaba S, et al. Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. *Mol Cell* 2010;38:662–74.
- Pasmant E, Sabbagh A, Masliah-Planchon J, Ortonne N, Laurendeau I, Melin L, et al. Role of noncoding RNA ANRIL in genesis of plexiform neurofibromas in neurofibromatosis type 1. *J Natl Cancer Inst* 2011;103:1713–22.
- Popov N, Gil J. Epigenetic regulation of the INK4b-ARF-INK4a locus: in sickness and in health. *Epigenetics* 2010;5:685–90.
- Li CH, Chen Y. Targeting long non-coding RNAs in cancers: progress and prospects. *Int J Biochem Cell Biol* 2013;45:1895–910.
- Bieche I, Onody P, Laurendeau I, Olivi M, Vidaud D, Lidereau R, et al. Real-time reverse transcription-PCR assay for future management of ERBB2-based clinical applications. *ClinChem* 1999;45:1148–56.
- Bieche I, Parfait B, Laurendeau I, Girault I, Vidaud M, Lidereau R. Quantification of estrogen receptor alpha and beta expression in sporadic breast cancer. *Oncogene* 2001;20:8109–15.
- Chomczynski P, Sacchi N. Single-step method of RNA isolation by acid guanidiniumthiocyanate-phenol-chloroform extraction. *Anal Biochem* 1987;162:156–9.
- Bieche I, Parfait B, Le Doussal V, Olivi M, Rio MC, Lidereau R, et al. Identification of CGA as a novel estrogen receptor-responsive gene in breast cancer: an outstanding candidate marker to predict the response to endocrine therapy. *Cancer Res* 2001;61:1652–8.
- Didier G, Brezellec P, Remy E, Henaut A. GeneANOVA-gene expression analysis of variance. *Bioinformatics* 2002;18:490–1.
- Bieche I, Nogués C, Lidereau R. Overexpression of BRCA2 gene in sporadic breast tumours. *Oncogene* 1999;18:5232–8.
- Cao Q, Mani RS, Ateeq B, Dhanasekaran SM, Asangani IA, Prensner JR, et al. Coordinated regulation of polycomb group complexes through microRNAs in Cancer. *Cancer Cell* 2011;20:187–99.
- Ye S, Yang L, Zhao X, Song W, Wang W, Zheng S. Bioinformatics method to predict two regulation mechanism: TF-miRNA-mRNA and lncRNA-miRNA-mRNA in pancreatic cancer. *Cell Biochem Biophys* 2014;5:41–9.
- Penna E, Orso F, Taverna D. miR-214 as a key hub that controls cancer networks: small player, multiple functions. *J Invest Dermatol* 2015;135:960–9.
- Zhang B, Liu XX, He JR, Zhou CX, Guo M, He M, et al. Pathologically decreased miR-26a antagonizes apoptosis and facilitates carcinogenesis by targeting MTDH and EZH2 in breast cancer. *Carcinogenesis* 2011;32:2–9.
- Sun YM, Lin KY, Chen YQ. Diverse functions of miR-125 family in different cell contexts. *J Hematol Oncol* 2013;6:6–14.
- Sepe R, Formisano U, Federico A, Forzati F, Bastos AU, D'Angelo D, et al. CBX7 and HMG1b proteins act in opposite way on the regulation of the SPP1 gene expression. *Oncotarget* 2015;6:2680–92.
- O'Loughlen A, Muñoz-Cabello AM, Gaspar-Maia A, Wu HA, Banito A, Kunowska N, et al. MicroRNA regulation of Cbx7 mediates a switch of Polycomb orthologs during ESC differentiation. *Cell Stem Cell* 2012;10:33–46.
- Congrains A, Kamide K, Ohishi M, Rakugi H. ANRIL: molecular mechanisms and implications in human health. *Int J Mol Sci* 2013;14:1278–92.
- Chen D, Zhang Z, Mao C, Zhou Y, Yu L, Yin Y, et al. ANRIL inhibits p15 (INK4b) through the TGFβ1 signaling pathway in human esophageal squamous cell carcinoma. *Cell Immunol* 2014;289:91–6.
- Hua L, Wang CY, Yao KH, Chen JT, Zhang JJ, Ma W. High expression of long non-coding RNA ANRIL is associated with poor prognosis in hepatocellular carcinoma. *Int J Clin Exp Pathol* 2015;8:3076–82.
- Lin L, Gu ZT, Chen WH, Cao KJ. Increased expression of the long non-coding RNA ANRIL promotes lung cancer cell metastasis and correlates with poor prognosis. *Diagn Pathol* 2015;10:14–21.
- Qiu JJ, Lin YY, Ding JX, Feng WW, Jin HY, Hua KQ. Long non-coding RNA ANRIL predicts poor prognosis and promotes invasion/metastasis in serous ovarian cancer. *Int J Oncol* 2015;46:2497–505.
- Shinjo K, Yamashita Y, Yamamoto E, Akatsuka S, Uno N, Kamiya A, et al. Expression of chromobox homolog 7 (CBX7) is associated with poor prognosis in ovarian clear cell adenocarcinoma via TRAIL-induced apoptotic pathway regulation. *Int J Cancer* 2014;135:308–18.
- Zheng X, Zhou J, Zhang B, Zhang J, Wilson J, Gu L, et al. Critical evaluation of Cbx7 downregulation in primary colon carcinomas and its clinical significance in Chinese patients. *BMC Cancer* 2015;15:145–55.
- Karamitopoulou E, Pallante P, Zlobec I, Tornillo L, Carafa V, Schaffner T, et al. Loss of the CBX7 protein expression correlates with a more aggressive phenotype in pancreatic cancer. *Eur J Cancer* 2010;46:1438–44.
- Pallante P, Terracciano L, Carafa V, Schneider S, Zlobec I, Lugli A, et al. The loss of the CBX7 gene expression represents an adverse prognostic marker for survival of colon carcinoma patients. *Eur J Cancer* 2010;46:2304–13.
- Pallante P, Forzati F, Federico A, Arra C, Fusco A. Polycombprotein family member CBX7 plays a critical role in cancer progression. *Am J Cancer Res* 2015;5:1594–601.
- Reijm EA, Timmermans AM, Look MP, Meijer-van Gelder ME, Stobbe CK, et al. High protein expression of EZH2 is related to unfavorable outcome to tamoxifen in metastatic breast cancer. *Ann Oncol* 2014;25:2185–90.
- Wu J, Crowe DL. The histone methyltransferase EZH2 promotes mammary stem and luminal progenitor cell expansion, metastasis and inhibits estrogen receptor-positive cellular differentiation in a model of basal breast cancer. *Oncol Rep* 2015;34:455–60.

39. Chen S, Huang L, Sun K, Wu D, Li M, Li M, et al. Enhancer of zeste homolog 2 as an independent prognostic marker for cancer: a meta-analysis. *PLoS One* 2015;10:e0125480.
40. Morin RD, Johnson NA, Severson TM, Mungall AJ, An J, Goya R, et al. Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. *Nat Genet* 2010;42:181–5.
41. Sudo T, Utsunomiya T, Mimori K, Nagahara H, Ogawa K, Inoue H, et al. Clinicopathological significance of EZH2 mRNA expression in patients with hepatocellular carcinoma. *Br J Cancer* 2005;92:1754–8.
42. Vélkel P, Dupret B, Le Bourhis X, Angrand PO. Diverse involvement of EZH2 in cancer epigenetics. *Am J Transl Res* 2015;7:175–93.
43. Koboldt DC, Fulton RS, McLellan MD, Schmidt H, Kalicki-Veizer J, McMichael JF, et al. Comprehensive molecular portraits of human breast tumours. *Nature* 2012;490:61–70.
44. Derfoul A, Juan AH, Difilippantonio MJ, Palanisamy N, Ried T, Sartorelli V. Decreased microRNA-214 levels in breast cancer cells coincides with increased cell proliferation, invasion and accumulation of the Polycomb Ezh2 methyltransferase. *Carcinogenesis* 2011;32:1607–14.
45. Xiangxiang W, Xiaoyun D, Shengcan C, Haojun S, Haizhong J, Ying F, et al. The functional sites of miRNAs and lncRNAs in gastric carcinogenesis. *Tumor Biol* 2015;36:521–32.
46. Mansueto G, Forzati F, Ferraro A, Pallante P, Bianco M, Esposito F, et al. Identification of a new pathway for tumor progression: microRNA-181b up-regulation and CBX7 down-regulation by HMGA1 protein. *Gene Cancer* 2010;1:210–24.
47. Lee W, Teckie S, Wiesner T, Ran L, Prieto Granada CN, Lin M, et al. PRC2 is recurrently inactivated through EED or SUZ12 loss in malignant peripheral nerve sheath tumors. *Nat Genet* 2014;46:1227–32.
48. Fu LL, Tian M, Li X, Li JJ, Huang J, Ouyang L, et al. Inhibition of BET bromodomains as a therapeutic strategy for cancer drug discovery. *Oncotarget* 2015;6:5501–16.
49. Baude A, Lindroth AM, Plass C. PRC2 loss amplifies Ras signaling in cancer. *Nat Genet* 2014;46:1154–5.
50. De Raedt T, Beert E, Pasmant E, Luscan A, Brems H, Ortonne N, et al. PRC2 loss amplifies Ras-driven transcription and confers sensitivity to BRD4-based therapies. *Nature* 2014;514:247–51.

SUPPLEMENTAL FIGURES LEGENDS

Supplemental figure 1: The *ANRIL* pathway.

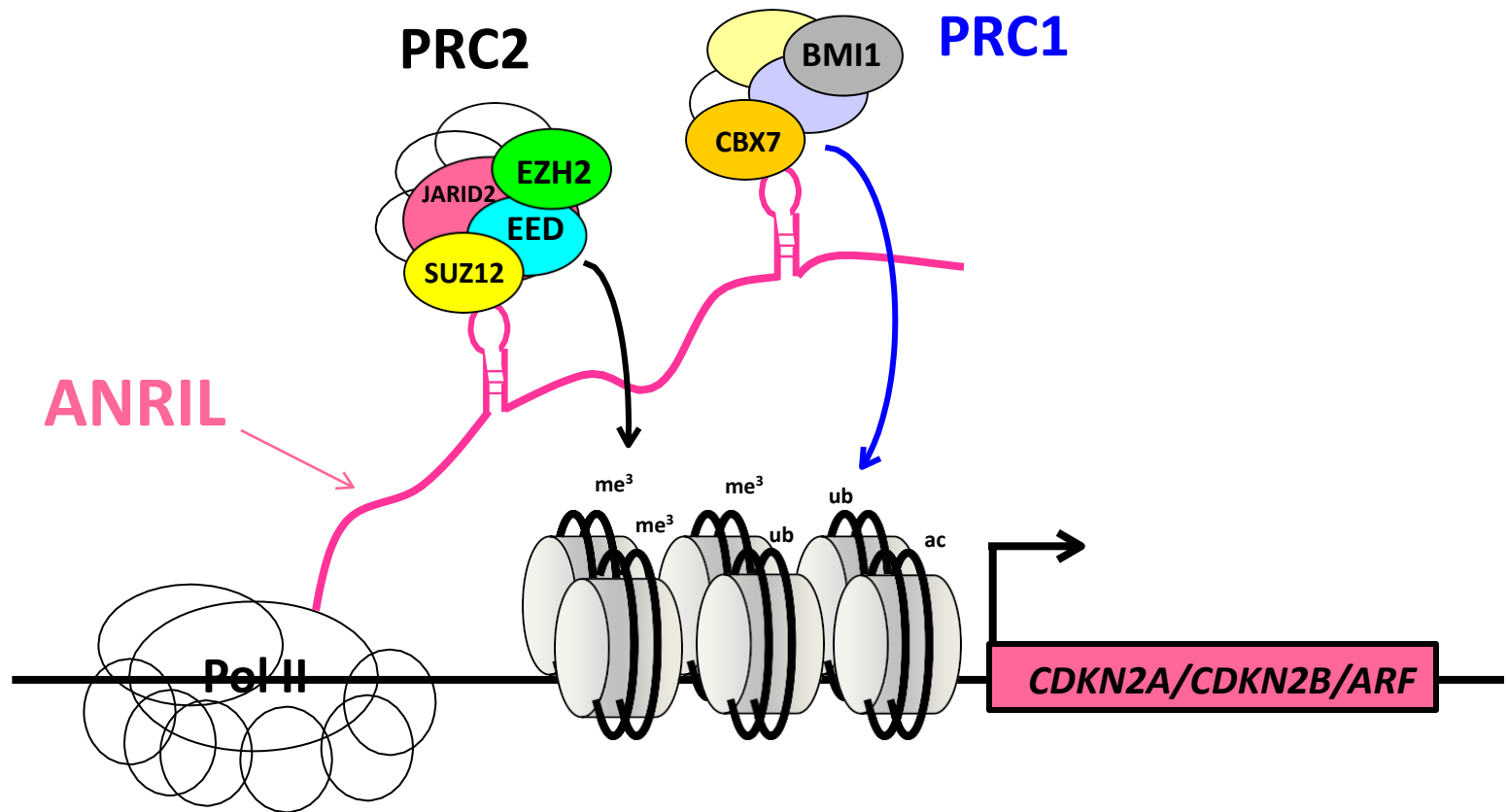
Supplemental figure 2: Immunohistochemical staining for p15, p16, and p14.

Overexpression (H score 2 and 3) of p15, p16 and p14 in cancer cells was observed respectively in 37.5%, 51.3%, and 41.2% of IBCs. Score 1: weak cytoplasmic staining for p14, p15 and p16 is seen in a subset of tumor cells. Score 2: moderate cytoplasmic staining for p14, p15 and p16 is seen in most of tumor cells. Score 3: strong and diffuse cytoplasmic staining for p14, p15 and p16 is seen in most of tumor cells.

Supplemental figure 3: Immunohistochemical staining for H3K27me3 and H3K27ac.

Nuclear under expression of H3K27me3 (H score 0 and 1) was observed in 76.3% of IBCs. Nuclear overexpression of H3K27ac (H score 2 and 3) in cancer cells was observed in 73.8% of IBCs. Score 0: no nuclear staining for H3K27ac and H3K27me3 is seen in tumor cells. Score 1: weak nuclear staining for H3K27ac and H3K27me3 is seen in a subset of tumor cells. Score 2: moderate nuclear staining for H3K27ac and H3K27me3 is seen in a subset of tumor cells. Score 3: strong and diffuse nuclear staining for H3K27ac and H3K27me3 is seen in most of tumor cells.

Supplemental figure 4: Dendrogram of the 10 *ANRIL* pathway genes constructed using hierarchical clustering, according to the gene profiling of the 456 breast tumors.

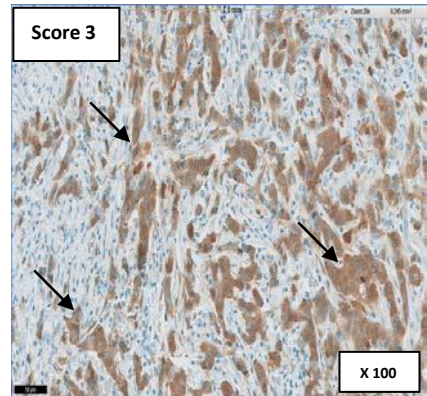
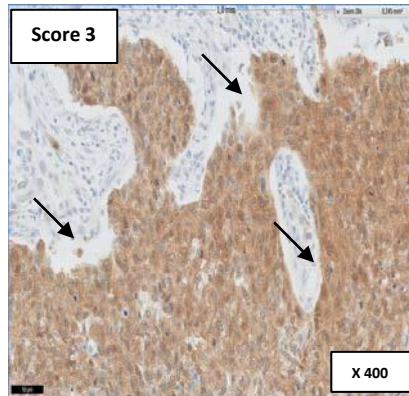
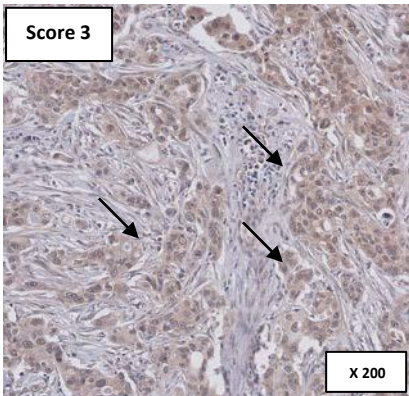
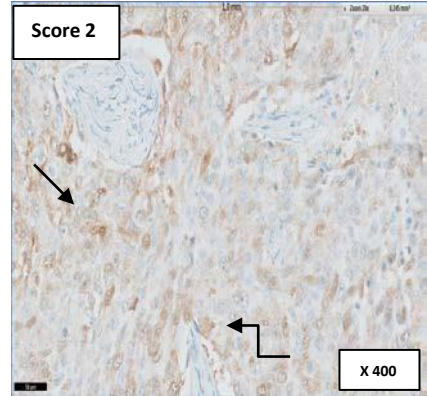
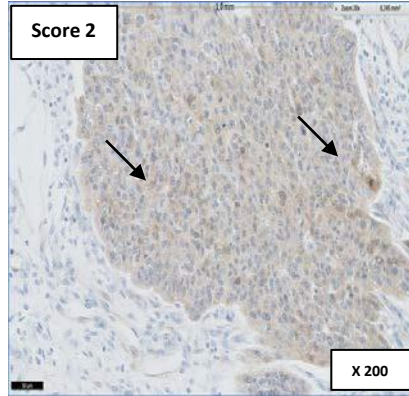
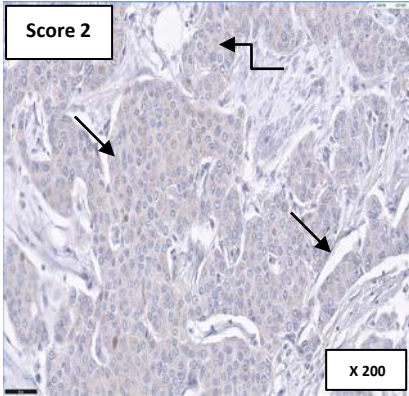
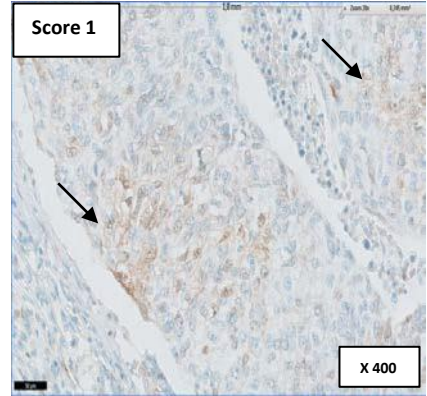
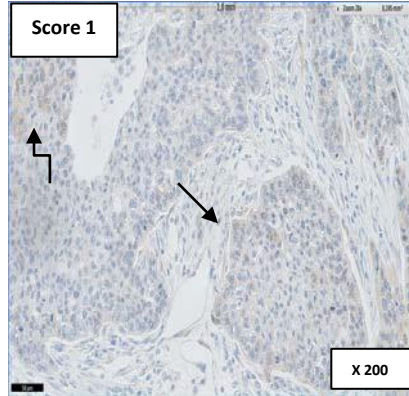
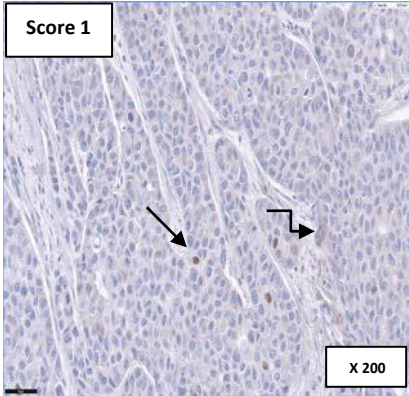


Supplemental Figure 1: Meseure et al.

p14

p15

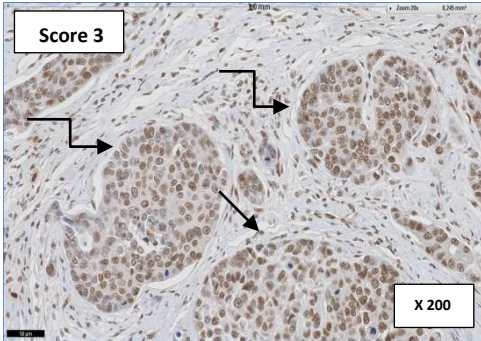
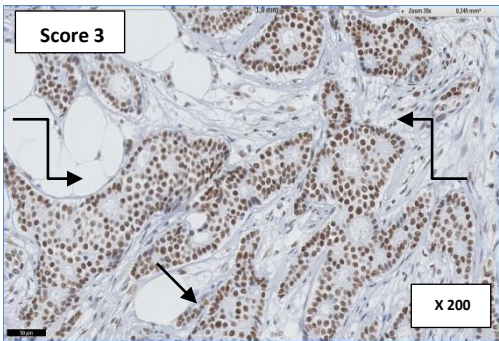
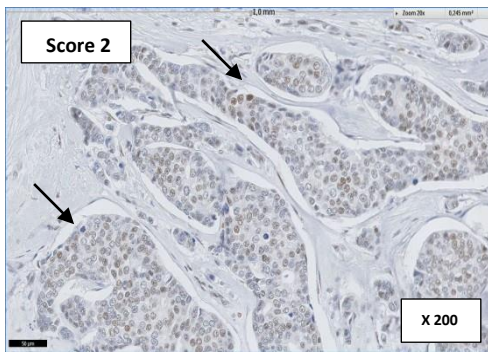
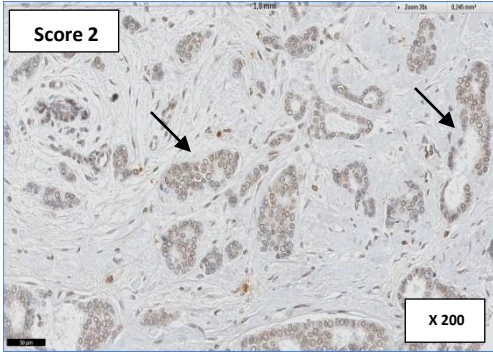
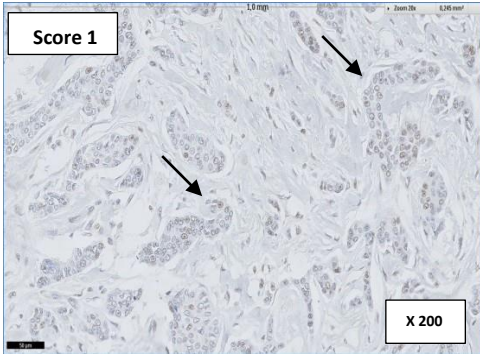
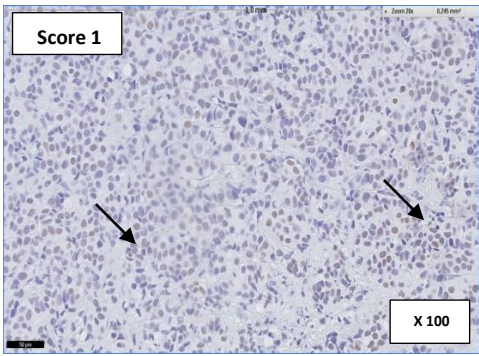
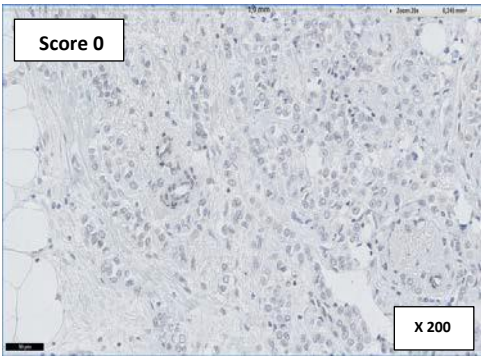
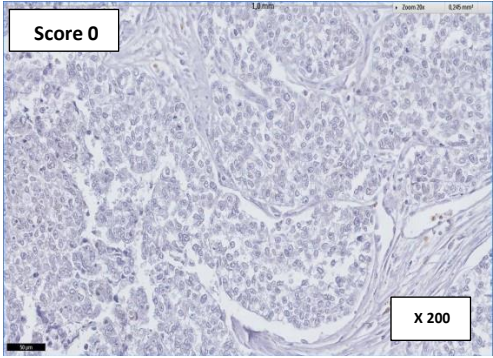
p16



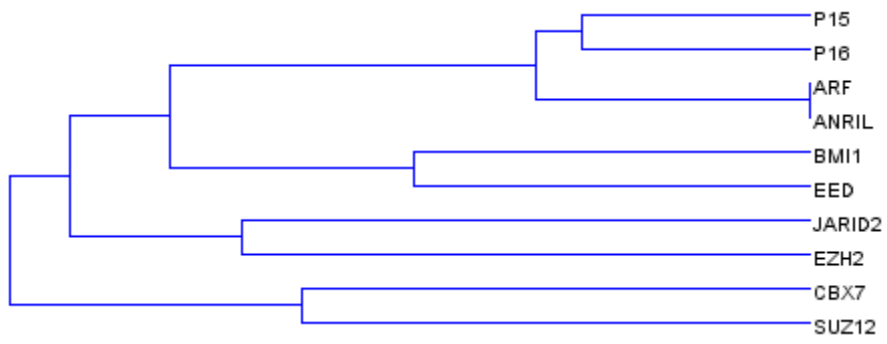
Supplemental Figure 2: Meseure et al.

H3K27me3

H3K27ac



Supplemental figure 3, Meseure et al



Supplemental figure 4: Meseure et al.

Supplementary Table 1: Pathological and clinical characteristics of the 456 patients in relation to metastasis free survival (MFS)

	Number of patients (%)	Number with metastases (%)	MFS p-value ^a
Total	456 (100.0)	169 (37.1)	
Age			
≤50	98 (21.5)	37 (37.8)	0.55 (NS)
>50	358 (78.5)	132 (36.9)	
SBR histological grade ^{b, c}			
I	58 (13)	8 (13.8)	0.000090
II	229 (51.2)	84 (36.7)	
III	160 (35.8)	73 (45.6)	
Lymph node status ^d			
0	119 (26.1)	35 (29.4)	0.00000087
1-3	237 (52.1)	77 (32.5)	
>3	99 (21.8)	57 (57.6)	
Macroscopic tumor size ^e			
≤25mm	223 (49.8)	63 (28.3)	0.000018
>25mm	225 (50.2)	105 (46.7)	
ER α status			
Negative	118 (25.9)	50 (42.4)	0.010
Positive	338 (74.1)	119 (35.2)	
PR status			
Negative	194 (42.5)	85 (43.8)	0.0014
Positive	262 (57.5)	84 (32.1)	
ERBB2 status			
Negative	357 (78.3)	127 (35.6)	0.098 (NS)
Positive	99 (21.7)	42 (42.4)	
Molecular subtypes			
HR- ERBB2-	68 (14.9)	26 (38.2)	0.0096
HR- ERBB2+	45 (9.9)	23 (51.1)	
HR+ ERBB2-	289 (63.4)	101 (34.9)	
HR+ ERBB2+	54 (11.8)	19 (35.2)	
PIK3CA mutation status			
wild type	307 (67.3)	124 (40.4)	0.014
mutated	149 (32.7)	45 (30.2)	
Treatment			
No adjuvant	87 (19.1)	32 (36.8)	0.063 (NS)
Chemotherapy	91 (20.0)	41 (45.1)	
Hormonotherapy	176 (38.6)	67 (38.1)	
Chemo and hormonotherapy	102 (22.4)	29 (28.4)	

Abbreviations: ER α : oestrogen receptor alpha; PR: progesterone receptor; ERBB2: human epidermal growth factor receptor 2; HR: hormone receptor.

The bold values are statistically significant (p-value<0.05). NS: not significant. ^a Log-rank test. ^b Scarff Bloom Richardson classification. ^c Information available for 447 patients. ^d Information available for 455 patients. ^e Information available for 448 patients.

Supplementary Table 2: Pathological and clinical characteristics of the 80 patients tested using immunohistochemistry in relation to metastasis free survival (MFS)

	Number of patients (%)	Number with metastases (%)	MFS p-value ^a
Total	80 (100)	36 (45.0)	
Age			
≤50	35 (43.8)	12 (34.3)	0.095 (NS)
>50	45 (56.3)	24 (53.3)	
SBR histological grade ^{b, c}			
I	14 (17.7)	4 (28.6)	0.089 (NS)
II	35 (44.3)	14 (40.0)	
III	30 (38.0)	17 (56.7)	
Lymph node status			
0	17 (21.3)	4 (23.5)	0.00020
1-3	50 (62.5)	21 (42.0)	
>3	13 (16.3)	11 (84.6)	
Macroscopic tumor size ^c			
≤25mm	40 (50.6)	16 (40.0)	0.12 (NS)
>25mm	39 (49.4)	20 (51.3)	
ER α status			
Negative	17 (21.3)	10 (58.8)	0.025
Positive	63 (78.8)	26 (41.3)	
PR status			
Negative	24 (30.0)	13 (54.2)	0.029
Positive	56 (70.0)	23 (41.1)	
ERBB2 status			
Negative	66 (82.5)	27 (40.9)	0.020
Positive	14 (17.5)	9 (64.3)	
Molecular subtypes			
HR- ERBB2-	9 (11.3)	5 (55.6)	0.0062
HR- ERBB2+	7 (8.8)	5 (71.4)	
HR+ ERBB2-	57 (71.3)	22 (38.6)	
HR+ ERBB2+	7 (8.8)	4 (57.1)	
PIK3CA mutation status ^d			
wild type	35 (58.3)	20 (57.1)	0.083 (NS)
mutated	25 (41.7)	10 (40.0)	
Treatment ^e			
No adjuvant	15 (19.5)	3 (20.0)	0.11 (NS)
Chemotherapy	32 (41.6)	17 (53.1)	
Hormonotherapy	7 (9.1)	5 (71.4)	
Chemo and hormonotherapy	23 (29.9)	9 (39.1)	

Abbreviations: ER α : oestrogen receptor alpha; PR: progesterone receptor; ERBB2: human epidermal growth factor receptor 2; HR: hormone receptor.

The bold values are statistically significant (p-value<0.05). NS: not significant. ^a Log-rank test. ^b Scarff Bloom Richardson classification. ^c Information available for 79 patients. ^d Information available for 60 patients. ^e Information available for 77 patients

Supplementary table 3: Relationship between ANRIL transcripts levels and classical clinical biological parameters in a series of 456 breast cancers.

	Total population (%)	Number of patients (%)		p-value ^a
		ANRIL mRNA expression ≤ 3 relative to normals	ANRIL mRNA expression > 3 relative to normals	
Total	456 (100.0)	366 (80.3)	90 (19.7)	
Age				
≤ 50	98 (21.5)	79 (21.6)	19 (21.1)	0.92 (NS)
> 50	358 (78.5)	287 (78.4)	71 (78.9)	
SBR histological grade ^{b, c}				
I	58 (13)	50 (13.8)	8 (9.3)	0.16 (NS)
II	229 (51.2)	189 (52.4)	40 (46.5)	
III	160 (35.8)	122 (33.8)	38 (44.2)	
Lymph node status ^d				
0	119 (26.1)	92 (25.1)	27 (30.3)	0.37 (NS)
1-3	237 (52.1)	190 (51.9)	47 (52.8)	
> 3	99 (21.8)	84 (23)	15 (16.9)	
Macroscopic tumor size ^e				
≤ 25 mm	223 (49.8)	185 (51.4)	38 (43.2)	0.17 (NS)
> 25 mm	225 (50.2)	175 (48.6)	50 (56.8)	
ER α status				
Negative	118 (25.9)	87 (23.8)	31 (34.4)	0.038
Positive	338 (74.1)	279 (76.2)	59 (65.6)	
PR status				
Negative	194 (42.5)	145 (39.6)	49 (54.4)	0.011
Positive	262 (57.5)	221 (60.4)	41 (45.6)	
ERBB2 status				
Negative	357 (78.3)	286 (78.1)	71 (78.9)	0.88 (NS)
Positive	99 (21.7)	80 (21.9)	19 (21.1)	
Treatment				
No adjuvant	87 (19.1)	73 (83.9)	14 (16.1)	0.68 (NS)
Chemotherapy	91 (20.0)	70 (76.9)	21 (23.1)	
Hormonotherapy	176 (38.6)	140 (79.5)	36 (20.5)	
Chemo and hormonotherapy	102 (22.4)	83 (81.4)	19 (18.6)	

Abbreviations: ER α : oestrogen receptor alpha; PR: progesterone receptor; ERBB2: human epidermal growth factor receptor 2; HR: hormone receptor. The bold values are statistically significant (p-value < 0.05).

^a χ^2 Test; ^b Scarff Bloom Richardson classification; ^c Information available for 447 patients; ^d Information available for 455 patients; ^e Information available for 448 patients.

Supplementary table 4: Relationship between EZH2 transcripts levels and classical clinical biological parameters in a series of 456 breast cancers.

	Total population (%)	Number of patients (%)		p-value ^a
		EZH2 mRNA expression ≤ 3 relative to normals	EZH2 mRNA expression > 3 relative to normals	
Total	456 (100.0)	105 (23)	351 (77)	
Age				
≤ 50	98 (21.5)	23 (21.9)	75 (21.4)	0.91 (NS)
> 50	358 (78.5)	82 (78.1)	276 (78.6)	
SBR histological grade ^{b, c}				
I	58 (13)	28 (26.7)	30 (87.7)	<0.0000001
II	229 (51.2)	65 (61.9)	164 (48)	
III	160 (35.8)	12 (11.4)	148 (43.3)	
Lymph node status ^d				
0	119 (26.1)	21 (20)	98 (28)	0.067(NS)
1-3	237 (52.1)	65 (61.9)	172 (49.1)	
> 3	99 (21.8)	19 (18.1)	80 (22.9)	
Macroscopic tumor size ^e				
≤ 25 mm	223 (49.8)	64 (61.5)	159 (46.2)	0.0062
> 25 mm	225 (50.2)	40 (38.5)	185 (53.8)	
ER α status				
Negative	118 (25.9)	7 (6.7)	111 (31.6)	0.00000030
Positive	338 (74.1)	98 (93.3)	240 (68.4)	
PR status				
Negative	194 (42.5)	22 (21)	172 (49)	0.00000034
Positive	262 (57.5)	83 (79)	179 (51)	
ERBB2 status				
Negative	357 (78.3)	96 (91.4)	261 (74.4)	0.00020
Positive	99 (21.7)	9 (8.6)	90 (25.6)	
Treatment				
No adjuvant	87 (19.1)	14 (16.1)	73 (83.9)	0.0069
Chemotherapy	91 (20.0)	17 (18.7)	74 (81.3)	
Hormonotherapy	176 (38.6)	38 (21.6)	138 (78.4)	
Chemo and hormonotherapy	102 (22.4)	36 (35.3)	66 (64.7)	

Abbreviations: ER α : oestrogen receptor alpha; PR: progesterone receptor; ERBB2: human epidermal growth factor receptor 2; HR: hormone receptor. The bold values are statistically significant (p-value <0.05).

^a χ^2 Test; ^b Scarff Bloom Richardson classification; ^c Information available for 447 patients; ^d Information available for 455 patients; ^e Information available for 448 patients.

Supplementary table 5: Relationship between SUZ12 transcripts levels and classical clinical biological parameters in a series of 456 breast cancers.

	Total population (%)	Number of patients (%)		p-value ^a
		SUZ12 mRNA expression ≤ 3 relative to normals	SUZ12 mRNA expression > 3 relative to normals	
Total	456 (100.0)	444 (97.4)	12 (2.6)	
Age				
≤ 50	98 (21.5)	96 (21.6)	2 (16.7)	0.96 (NS)
> 50	358 (78.5)	348 (78.4)	10 (83.3)	
SBR histological grade ^{b, c}				
I	58 (13)	57 (13.1)	1 (8.3)	0.56 (NS)
II	229 (51.2)	221 (50.8)	8 (66.7)	
III	160 (35.8)	157 (36.1)	3 (25)	
Lymph node status ^d				
0	119 (26.1)	112 (25.3)	7 (58.3)	<0.0000001
1-3	237 (52.1)	234 (52.8)	3 (25)	
> 3	99 (21.8)	97 (21.9)	2 (16.7)	
Macroscopic tumor size ^e				
≤ 25 mm	223 (49.8)	214 (49.1)	9 (75)	0.076 (NS)
> 25 mm	225 (50.2)	222 (50.9)	3 (25)	
ER α status				
Negative	118 (25.9)	117 (26.4)	1 (8.3)	0.28 (NS)
Positive	338 (74.1)	327 (73.6)	11 (91.7)	
PR status				
Negative	194 (42.5)	190 (42.8)	4 (33.3)	0.51 (NS)
Positive	262 (57.5)	254 (57.2)	8 (66.7)	
ERBB2 status				
Negative	357 (78.3)	350 (78.8)	7 (58.3)	0.18 (NS)
Positive	99 (21.7)	94 (21.2)	5 (41.7)	
Treatment				
No adjuvant	87 (19.1)	84 (96.6)	3 (3.4)	0.097 (NS)
Chemotherapy	91 (20.0)	91 (100)	0 (0)	
Hormonotherapy	176 (38.6)	168 (95.5)	8 (4.5)	
Chemo and hormonotherapy	102 (22.4)	101 (99.0)	1 (1.0)	

Abbreviations: ER α : oestrogen receptor alpha; PR: progesterone receptor; ERBB2: human epidermal growth factor receptor 2; HR: hormone receptor. The bold values are statistically significant (p-value<0.05).

^a χ^2 Test; ^b Scarff Bloom Richardson classification; ^c Information available for 447 patients; ^d Information available for 455 patients; ^e Information available for 448 patients.

Supplementary table 6: Relationship between EED transcripts levels and classical clinical biological parameters in a series of 446 breast cancers.

	Number of patients (%)				p-value ^a
	Total population (%)	EED mRNA expression <0.33 relative to normals	EED mRNA expression ≥ 0.33 and ≤ 3 relative to normals	EED mRNA expression >3 relative to normals	
Total	446 (100)	14 (3.1)	424 (95.1)	8 (1.8)	
Age					
≤ 50	94 (21.1)	1 (1.1)	90 (95.7)	3 (3.2)	0.23 (NS)
>50	352 (78.9)	13 (3.7)	334 (94.9)	5 (1.4)	
SBR histological grade ^{b, c}					
I	57 (13.0)	7 (12.3)	50 (87.7)	0 (0.0)	0.0011
II	223 (51.0)	5 (2.2)	214 (96.0)	4 (1.8)	
III	157 (35.9)	2 (1.3)	152 (96.8)	3 (1.9)	
Lymph node status ^d					
0	118 (26.5)	5 (4.2)	111 (94.1)	2 (1.7)	0.86 (NS)
1-3	232 (52.1)	7 (3.0)	220 (94.8)	5 (2.2)	
>3	95 (21.3)	2 (2.1)	92 (96.8)	1 (1.1)	
Macroscopic tumor size ^e					
≤ 25 mm	220 (50.2)	6 (2.7)	210 (95.5)	4 (1.8)	0.86 (NS)
>25mm	218 (49.8)	8 (3.7)	206 (94.5)	4 (1.8)	
ER α status					
Negative	115 (25.8)	0 (0.0)	112 (97.4)	3 (2.6)	0.061 (NS)
Positive	331 (74.2)	14 (4.2)	312 (94.3)	5 (1.5)	
PR status					
Negative	190 (42.6)	0 (0.0)	185 (97.4)	5 (2.6)	0.0028
Positive	256 (57.4)	14 (5.5)	239 (93.4)	3 (1.2)	
ERBB2 status					
Negative	353 (79.1)	14 (4.0)	332 (94.1)	7 (2.0)	0.12 (NS)
Positive	93 (20.9)	0 (0.0)	92 (98.9)	1 (1.1)	
Treatment					
No adjuvant	86 (19.3)	1 (1.2)	84 (97.7)	1 (1.2)	0.46 (NS)
Chemotherapy	87 (19.5)	1 (1.1)	83 (95.4)	3 (3.4)	
Hormonotherapy	173 (38.8)	7 (4.1)	163 (94.2)	3 (1.7)	
Chemo and hormonotherapy	100 (22.4)	5 (5.0)	94 (94.0)	1 (1.0)	

Abbreviations: ER α : oestrogen receptor alpha; PR: progesterone receptor; ERBB2: human epidermal growth factor receptor 2; HR: hormone receptor. The bold values are statistically significant (p-value<0.05).

^a χ^2 Test; ^b Scarff Bloom Richardson classification; ^c Information available for 437 patients; ^d Information available for 445 patients; ^e Information available for 438 patients.

Supplementary table 7: Relationship between JARID2 transcripts levels and classical clinical biological parameters in a series of 454 breast cancers.

	Total population (%)	Number of patients (%)		p-value ^a
		JARID2 mRNA expression ≤3 relative to normals	JARID2 mRNA expression >3 relative to normals	
Total	454 (100.0)	431 (94.9)	23 (5.1)	
Age				
≤50	97 (21.4)	93 (21.6)	4 (17.4)	0.83 (NS)
>50	357 (78.6)	338 (78.4)	19 (82.6)	
SBR histological grade ^{b, c}				
I	58 (13.0)	56 (13.2)	2 (9.5)	0.26 (NS)
II	228 (51.2)	220 (51.9)	8 (38.1)	
III	159 (35.7)	148 (34.9)	11 (52.4)	
	9	7	2	
Lymph node status ^d				
0	119 (26.3)	112 (26.0)	7 (30.4)	0.88 (NS)
1-3	237 (52.3)	226 (52.6)	11 (47.8)	
>3	97 (21.4)	92 (21.4)	5 (21.8)	
	1	1		
Macroscopic tumor size ^e				
≤25mm	222 (49.8)	210 (49.6)	12 (52.2)	0.81 (NS)
>25mm	224 (50.2)	213 (50.4)	11 (47.8)	
	8	8		
ERα status				
Negative	116 (25.6)	104 (24.1)	12 (52.2)	0.0027
Positive	338 (74.4)	327 (75.9)	11 (47.8)	
PR status				
Negative	192 (42.3)	179 (41.5)	13 (56.5)	0.16 (NS)
Positive	262 (57.7)	252 (58.5)	10 (43.5)	
ERBB2 status				
Negative	357 (78.6)	345 (80.0)	12 (52.2)	0.0035
Positive	97 (21.4)	86 (19.9)	11 (47.8)	
Treatment				
No adjuvant	87 (19.2)	82 (94.3)	5 (5.7)	0.44 (NS)
Chemotherapy	89 (19.6)	84 (94.4)	5 (5.6)	
Hormonotherapy	176 (38.8)	165 (93.8)	11 (6.3)	
Chemo and hormonotherapy	102 (22.5)	100 (98.0)	2 (2.0)	

Abbreviations: ERα: oestrogen receptor alpha; PR: progesterone receptor; ERBB2: human epidermal growth factor receptor 2; HR: hormone receptor. The bold values are statistically significant (p-value<0.05).

^a χ^2 Test; ^b Scarff Bloom Richardson classification; ^c Information available for 445 patients; ^d Information available for 453 patients; ^e Information available for 446 patients.

Supplementary table 8: Relationship between CBX7 transcripts levels and classical clinical biological parameters in a series of 456 breast cancers.

	Total population (%)	Number of patients (%)		p-value ^a
		CBX7 mRNA expression <0.33 relative to normals	CBX7 mRNA expression ≥0.33 relative to normals	
Total	456 (100.0)	181 (39.7)	275 (60.3)	
Age				
≤50	98 (21.5)	48 (26.5)	50 (18.2)	0.034
>50	358 (78.5)	133 (73.5)	225 (81.8)	
SBR histological grade ^{b, c}				
I	58 (13)	13 (7.3)	45 (16.8)	0.000025
II	229 (51.2)	80 (44.7)	149 (55.6)	
III	160 (35.8)	86 (48)	74 (27.6)	
Lymph node status ^d				
0	119 (26.1)	46 (25.6)	73 (26.55)	0.18 (NS)
1-3	237 (52.1)	87 (48.3)	150 (54.55)	
>3	99 (21.8)	47 (26.1)	52 (18.9)	
Macroscopic tumor size ^e				
≤25mm	223 (49.8)	77 (43)	146 (54.3)	0.020
>25mm	225 (50.2)	102 (57)	123 (45.7)	
ER α status				
Negative	118 (25.9)	67 (37)	51 (18.5)	0.000011
Positive	338 (74.1)	114 (63)	224 (81.5)	
PR status				
Negative	194 (42.5)	102 (56.4)	92 (33.5)	0.0000013
Positive	262 (57.5)	79 (43.6)	183 (66.5)	
ERBB2 status				
Negative	357 (78.3)	128 (70.7)	229 (83.3)	0.0015
Positive	99 (21.7)	53 (29.3)	46 (16.7)	
Treatment				
No adjuvant	87 (19.1)	33 (37.9)	54 (62.1)	0.011
Chemotherapy	91 (20.0)	49 (53.8)	42 (46.2)	
Hormonotherapy	176 (38.6)	58 (33.0)	118 (67.0)	
Chemo and hormonotherapy	102 (22.4)	41 (40.2)	61 (59.8)	

Abbreviations: ER α : oestrogen receptor alpha; PR: progesterone receptor; ERBB2: human epidermal growth factor receptor 2; HR: hormone receptor. The bold values are statistically significant (p-value<0.05).

^a χ^2 Test; ^b Scarff Bloom Richardson classification; ^c Information available for 447 patients; ^d Information available for 455 patients; ^e Information available for 448 patients.

Supplementary table 9: Relationship between BMI1 transcripts levels and classical clinical biological parameters in a series of 456 breast cancers.

	Total population (%)	Number of patients (%)		p-value ^a
		BMI mRNA expression ≤ 3 relative to normals	BMI mRNA expression > 3 relative to normals	
Total	456 (100.0)	410 (89.9)	46 (10.1)	
Age				
≤ 50	98 (21.5)	93 (22.7)	5 (10.9)	0.064 (NS)
> 50	358 (78.5)	317 (77.3)	41 (89.1)	
SBR histological grade ^{b, c}				
I	58 (13)	51 (12.7)	7 (15.2)	0.20 (NS)
II	229 (51.2)	201 (50.1)	28 (60.9)	
III	160 (35.8)	149 (37.2)	11 (23.9)	
Lymph node status ^d				
0	119 (26.1)	104 (25.4)	15 (32.6)	0.55 (NS)
1-3	237 (52.1)	216 (52.8)	21 (45.7)	
> 3	99 (21.8)	89 (21.8)	10 (21.7)	
Macroscopic tumor size ^e				
≤ 25 mm	223 (49.8)	206 (51)	17 (38.6)	0.12 (NS)
> 25 mm	225 (50.2)	198 (49)	27 (61.4)	
ER α status				
Negative	118 (25.9)	112 (27.3)	6 (13)	0.036
Positive	338 (74.1)	298 (72.7)	40 (87)	
PR status				
Negative	194 (42.5)	179 (43.7)	15 (32.6)	0.15 (NS)
Positive	262 (57.5)	231 (56.3)	31 (67.4)	
ERBB2 status				
Negative	357 (78.3)	319 (77.8)	38 (82.6)	0.45 (NS)
Positive	99 (21.7)	91 (22.2)	8 (17.4)	
Treatment				
No adjuvant	87 (19.1)	76 (87.4)	11 (12.6)	0.38 (NS)
Chemotherapy	91 (20.0)	86 (94.5)	5 (5.5)	
Hormonotherapy	176 (38.6)	156 (88.6)	20 (11.4)	
Chemo and hormonotherapy	102 (22.4)	92 (90.2)	10 (9.8)	

Abbreviations: ER α : oestrogen receptor alpha; PR: progesterone receptor; ERBB2: human epidermal growth factor receptor 2; HR: hormone receptor. The bold values are statistically significant (p-value <0.05).

^a χ^2 Test; ^b Scarff Bloom Richardson classification; ^c Information available for 447 patients; ^d Information available for 455 patients; ^e Information available for 448 patients.

Supplementary table 10: Relationship between ARF transcripts levels and classical clinical biological parameters in a series of 456 breast cancers.

	Total population (%)	Number of patients (%)		p-value ^a
		ARF mRNA expression ≤ 3 relative to normals	ARF mRNA expression > 3 relative to normals	
Total	456 (100.0)	274 (60.1)	182 (39.9)	
Age				
≤ 50	98 (21.5)	60 (21.9)	38 (20.9)	0.80 (NS)
> 50	358 (78.5)	214 (78.1)	144 (79.1)	
SBR histological grade ^{b, c}				
I	58 (13)	45 (16.7)	13 (7.3)	<0.0000001
II	229 (51.2)	157 (58.4)	72 (40.5)	
III	160 (35.8)	67 (24.9)	93 (52.2)	
Lymph node status ^d				
0	119 (26.1)	67 (24.4)	52 (28.7)	0.53 (NS)
1-3	237 (52.1)	144 (52.6)	93 (51.4)	
> 3	99 (21.8)	63 (23)	36 (19.9)	
Macroscopic tumor size ^e				
≤ 25 mm	223 (49.8)	146 (54.5)	77 (42.8)	0.015
> 25 mm	225 (50.2)	122 (45.5)	103 (57.2)	
ER α status				
Negative	118 (25.9)	47 (17.2)	71 (39)	0.00000018
Positive	338 (74.1)	227 (82.8)	111 (61)	
PR status				
Negative	194 (42.5)	94 (34.3)	100 (54.9)	0.000013
Positive	262 (57.5)	180 (65.7)	82 (45.1)	
ERBB2 status				
Negative	357 (78.3)	215 (78.5)	142 (78)	0.91 (NS)
Positive	99 (21.7)	59 (21.5)	40 (22)	
Treatment				
No adjuvant	87 (19.1)	47 (54.0)	40 (46.0)	0.27 (NS)
Chemotherapy	91 (20.0)	50 (54.9)	41 (45.1)	
Hormonotherapy	176 (38.6)	113 (64.2)	63 (35.8)	
Chemo and hormonotherapy	102 (22.4)	64 (62.7)	38 (37.3)	

Abbreviations: ER α : oestrogen receptor alpha; PR: progesterone receptor; ERBB2: human epidermal growth factor receptor 2; HR: hormone receptor. The bold values are statistically significant (p-value<0.05).

^a χ^2 Test; ^b Scarff Bloom Richardson classification; ^c Information available for 447 patients; ^d Information available for 455 patients; ^e Information available for 448 patients.

Supplementary table 11: Relationship between P16 transcripts levels and classical clinical biological parameters in a series of 456 breast cancers.

	Number of patients (%)			p-value ^a
	Total population (%)	P16 mRNA expression ≤ 3 versus a Ct of 35 equal to 1	P16 mRNA expression > 3 versus a Ct of 35 equal to 1	
Total	456 (100.0)	384 (84.2)	72 (15.8)	
Age				
≤ 50	98 (21.5)	79 (80.6)	19 (19.4)	0.27 (NS)
> 50	358 (78.5)	305 (85.2)	53 (14.8)	
SBR histological grade ^{b, c}				
I	58 (13)	53 (91.4)	5 (8.6)	0.00019
II	229 (51.2)	204 (89.1)	25 (10.9)	
III	160 (35.8)	119 (74.4)	41 (25.6)	
Lymph node status ^d				
0	119 (26.1)	95 (79.8)	24 (20.2)	0.30 (NS)
1-3	237 (52.1)	202 (85.2)	35 (14.8)	
> 3	99 (21.8)	86 (86.9)	13 (13.1)	
Macroscopic tumor size ^e				
≤ 25 mm	223 (49.8)	193 (86.5)	30 (13.5)	0.13 (NS)
> 25 mm	225 (50.2)	183 (81.3)	42 (18.7)	
ER α status				
Negative	118 (25.9)	79 (66.9)	39 (33.1)	<0.0000001
Positive	338 (74.1)	305 (90.2)	33 (9.8)	
PR status				
Negative	194 (42.5)	145 (74.7)	49 (25.3)	0.0000018
Positive	262 (57.5)	239 (91.2)	23 (8.8)	
ERBB2 status				
Negative	357 (78.3)	296 (82.9)	61 (17.1)	0.15 (NS)
Positive	99 (21.7)	88 (88.9)	11 (11.1)	
Treatment				
No adjuvant	87 (19.1)	70 (80.5)	17 (19.5)	0.34 (NS)
Chemotherapy	91 (20.0)	73 (80.2)	18 (19.8)	
Hormonotherapy	176 (38.6)	152 (86.4)	24 (13.6)	
Chemo and hormonotherapy	102 (22.4)	89 (87.3)	13 (12.7)	

Abbreviations: ER α : oestrogen receptor alpha; PR: progesterone receptor; ERBB2: human epidermal growth factor receptor 2; HR: hormone receptor. The bold values are statistically significant (p-value <0.05).

^a χ^2 Test; ^b Scarff Bloom Richardson classification; ^c Information available for 447 patients; ^d Information available for 455 patients; ^e Information available for 448 patients.

Supplementary table 12: Relationship between P15 transcripts levels and classical clinical biological parameters in a series of 456 breast cancers.

	Number of patients (%)				p-value ^a
	Total population (%)	P15 mRNA expression <0.33 relative to normals	P15 mRNA expression \geq 0.33 and \leq 3 relative to normals	P15 mRNA expression >3 relative to normals	
Total	456 (100.0)	32 (7)	325 (71.3)	99 (21.7)	
Age					
\leq 50	98 (21.5)	7 (21.9)	71 (21.8)	20 (20.2)	0.94 (NS)
>50	358 (78.5)	25 (78.1)	254 (78.2)	79 (79.8)	
SBR histological grade ^{b, c}					
I II	58 (13)	8 (25)	40 (12.6)	10 (10.2)	0.071 (NS)
III	229 (51.2)	15 (46.9)	170 (53.6)	44 (44.9)	
Lymph node status ^d	160 (35.8)	9 (28.1)	107 (33.8)	44 (44.9)	
0					0.86 (NS)
1-3	119 (26.1)	9 (28.1)	88 (27.1)	22 (22.4)	
>3	237 (52.1)	15 (46.9)	169 (52)	53 (54.1)	
Macroscopic tumor size ^e	99 (21.8)	8 (25)	68 (20.9)	23 (23.5)	
\leq 25mm					0.84 (NS)
>25mm	223 (49.8)	17 (54.8)	158 (49.4)	48 (49.5)	
ER α status					
Negative	118 (25.9)	3 (9.4)	80 (24.6)	35 (35.4)	0.0090
Positive	338 (74.1)	29 (90.6)	245 (75.4)	64 (64.6)	
PR status					
Negative	194 (42.5)	6 (18.75)	133 (40.9)	55 (55.6)	0.00080
Positive	262 (57.5)	26 (81.25)	192 (59.1)	44 (44.4)	
ERBB2 status					
Negative	357 (78.3)	30 (93.75)	249 (76.6)	78 (78.8)	0.078 (NS)
Positive	99 (21.7)	2 (6.25)	76 (23.4)	21 (21.2)	
Treatment					
No adjuvant	87 (19.1)	3 (3.5)	69 (79.3)	15 (17.2)	0.46 (NS)
Chemotherapy	91 (20.0)	5 (5.5)	66 (72.5)	20 (22.0)	
Hormonotherapy	176 (38.6)	15 (8.5)	123 (69.9)	38 (21.6)	
Chemo and hormonotherapy	102 (22.4)	9 (8.8)	67 (65.7)	26 (25.5)	

Abbreviations: ER α : oestrogen receptor alpha; PR: progesterone receptor; ERBB2: human epidermal growth factor receptor 2; HR: hormone receptor. The bold values are statistically significant (p-value<0.05).

^a χ^2 Test; ^b Scarff Bloom Richardson classification; ^c Information available for 447 patients; ^d Information available for 455 patients; ^e Information available for 448 patients.

Supplemental Table 13: ANRIL pathway genes mRNA levels in the 4 breast tumor subtypes in the series of 456 breast tumors.

	Normal breast tissues n=10	All tumors n=456	Tumor subtypes				p-value ^a
			HR-ERBB2- (Group 1) n=68	HR-ERBB2+ (Group 2) n=45	HR+ERBB2- (Group 3) n=289	HR+ERBB2+ (Group 4) n=54	
ANRIL	1.0 (0.25-2.45) ^b	1.58 (0.05-13.5)	1.92 (0.05-12.5)	1.35 (0.22-7.29)	1.50 (0.23-13.5)	1.91 (0.35-10.1)	
Non overexpressed tumors (%)		366 (80.3)	42 (61.8)	40 (88.9)	244 (84.4)	40 (74.1)	0.00013
Overexpressed tumors (%)		90 (19.7)	26 (38.2)	5 (11.1)	45 (15.6)	14 (25.9)	
EZH2	1.0 (0.48-2.32)	5.09 (0.76-48.3)	9.90 (0.95-45.8)	7.86 (2.68-35.6)	3.92 (0.76-48.3)	6.14 (1.00-17.6)	
Non overexpressed tumors (%)		105 (23.0)	3 (4.4)	2 (4.4)	93 (32.2)	7 (13.0)	<0.0000001
Overexpressed tumors (%)		351 (77.0)	65 (95.6)	43 (95.6)	196 (67.8)	47 (87.0)	
SUZ12	1.0 (0.56-1.25)	1.23 (0.43-59.2)	1.16 (0.43-6.74)	1.0 (0.49-2.98)	1.23 (0.46-59.2)	1.53 (0.48-5.00)	
Non overexpressed tumors (%)		444 (97.4)	67 (98.5)	45 (100)	283 (97.9)	49 (90.7)	0.011
Overexpressed tumors (%)		12 (2.6)	1 (1.5)	0 (0)	6 (2.1)	5 (9.3)	
EED^c	1.0 (0.58-1.84)	1.21 (0.03-7.74)	1.48 (0.36-4.01)	1.31 (0.37-2.81)	1.13 (0.03-7.74)	1.22 (0.35-3.33)	
Underexpressed tumors (%)		14 (3.1)	0 (0.0)	0 (0.0)	14 (4.9)	0 (0.0)	0.042
Non underexpressed tumors (%)		432 (96.9)	68 (100)	42 (100.0)	271 (95.1)	51 (100)	
JARID2^d	1.0 (0.83-1.46)	1.37 (0.34-8.53)	1.47 (0.64-8.53)	1.76 (0.79-7.73)	1.29 (0.34-8.39)	1.62 (0.53-8.21)	
Non overexpressed tumors (%)		431 (94.9)	63 (92.6)	36 (83.7)	282 (97.6)	50 (92.6)	0.00090
Overexpressed tumors (%)		23 (5.1)	5 (7.4)	7 (16.3)	7 (2.4)	4 (7.4)	
BMI	1.0 (0.59-1.47)	1.23 (0.00-14.0)	1.10 (0.38-6.56)	1.07 (0.49-4.49)	1.33 (0.00-14.0)	1.15 (0.15-4.85)	
Non overexpressed tumors (%)		410 (89.9)	65 (95.6)	42 (93.3)	254 (87.9)	49 (90.7)	0.23 (NS)
Overexpressed tumors (%)		46 (10.1)	3 (4.4)	3 (6.7)	35 (12.1)	5 (9.3)	
CBX7	1.0 (0.39-1.62)	0.38 (0.00-2.59)	0.31 (0.06-2.43)	0.25 (0.07-1.21)	0.42 (0.05-2.59)	0.35 (0.00-1.19)	
Underexpressed tumors (%)		181 (39.7)	38 (55.9)	27 (60.0)	90 (31.1)	26 (48.1)	0.000020
Non underexpressed tumors (%)		275 (60.3)	30 (44.1)	18 (40.0)	199 (68.9)	28 (51.9)	
ARF	1.0 (0.21-2.41)	2.28 (0.00-105)	10.7 (0.22-105)	2.80 (0.28-70.8)	1.76 (0.08-47.3)	2.29 (0.0-13.4)	
Non overexpressed tumors (%)		274 (60.1)	18 (26.5)	25 (55.6)	197 (68.2)	34 (63.0)	<0.0000001
Overexpressed tumors (%)		182 (39.9)	50 (73.5)	20 (44.4)	92 (31.8)	20 (37.0)	
P16^e	0.0 (0.0-6.95)	0.65 (0.0-63.4)	2.47 (0.0-63.4)	0.75 (0.0-8.18)	0.56 (0.0-15.2)	0.52 (0.0-18.5)	
Non overexpressed tumors (%)		384 (84.2)	35 (51.5)	39 (86.7)	261 (90.3)	49 (90.7)	<0.0000001
Overexpressed tumors (%)		72 (15.8)	33 (48.5)	6 (13.3)	28 (9.7)	5 (9.3)	
P15	1.0 (0.11-6.03)	1.62 (0.00-40.5)	1.65 (0.21-40.5)	1.67 (0.36-8.73)	1.60 (0.00-13.7)	1.66 (0.11-9.71)	
Non overexpressed tumors (%)		357 (78.3)	44 (64.7)	35 (77.8)	235 (81.3)	43 (79.6)	0.029
Overexpressed tumors (%)		99 (21.7)	24 (35.3)	10 (22.2)	54 (18.7)	11 (20.4)	

^a Chi2 Test (comparison of percentage of under or overexpression in group 1 versus group 2 versus group 3 versus group 4).

^b Median (range) of gene mRNA levels. The mRNA values of the samples were normalized such that the median of the 10 normal breast tissues mRNA values was 1.

^c Data available in 446 samples.

^d Data available in 454 samples.

^e The mRNA values of the samples were normalized such that a Ct value of 35 was 1.

Supplemental Table 14: Multivariate COX analysis of MFS for EZH2 and CBX7 combined expression levels in the series of 456 breast tumors.

Characteristics		HR ^a	95% CI ^b	p-value ^c
SBR histological grade	I	1		NS
	II	1.23	0.97-1.56	
	III	1.51	0.94-2.42	
Lymph node status	0	1		0.00021
	1-3	1.54	1.22-1.93	
	>3	2.36	1.50-3.72	
Macroscopic tumor size	<=25mm	1		0.014
	>25mm	1.50	1.09-2.08	
EZH2 and CBX7 combined expression levels	Normal EZH2 and CBX7 expression	1		0.001
	Normal EZH2 and under-CBX7	1.34	1.13-1.60	
	Over-EZH2 and normal CBX7	1.80	1.27-2.56	
	Over-EZH2 and under-CBX7	2.42	1.43-4.10	

^a Hazard ratio.

^b 95% Confidential Interval.

^c Multivariate COX analysis. 0

Supplemental Table 15: Statistical analysis and correlation between ANRIL pathway genes in the series of 456 breast tumors.

	ANRIL	EZH2	SUZ12	EED ^b	JARID2 ^c	BMI1	CBX7	ARF	P16	P15
ANRIL	1 ^a <0.0000001									
EZH2	+0.169 0.00039	1 <0.0000001								
SUZ12	+0.220 0.0000067	+0.193 0.000064	1 <0.0000001							
EED ^b	+0.184 0.00015	+0.295 <0.0000001	+0.162 0.00075	1 <0.0000001						
JARID2 ^c	+0.091 0.050 (NS)	+0.278 <0.0000001	+0.293 <0.0000001	+0.189 0.00010	1 <0.0000001					
BMI1	+0.110 0.018	+0.229 0.0000031	+0.195 0.000056	+0.386 <0.0000001	+0.110 0.018	1 <0.0000001				
CBX7	+0.233 0.0000022	-0.254 0.00000032	+0.317 <0.0000001	+0.188 0.00011	+0.155 0.0011	+0.034 0.47 (NS)	1 <0.0000001			
ARF	+0.594 <0.0000001	+0.408 <0.0000001	+0.064 0.17 (NS)	+0.302 <0.0000001	+0.061 0.19 (NS)	+0.214 0.000012	+0.032 0.50 (NS)	1 <0.0000001		
P16	+0.323 <0.0000001	+0.209 0.000017	-0.092 0.047	+0.294 <0.0000001	-0.011 0.81 (NS)	+0.147 0.0018	+0.048 0.31 (NS)	+0.597 <0.0000001	1 <0.0000001	
P15	+0.381 <0.0000001	+0.119 0.011	+0.046 0.33 (NS)	+0.323 <0.0000001	+0.065 0.16 (NS)	+0.256 0.00000025	+0.278 <0.0000001	+0.490 <0.0000001	+0.473 <0.0000001	1 <0.0000001

^a Spearman rank correlation Test.

^b Data available in 446 samples.

Supplemental Table 16: Statistical analysis of EMT associated gene expressions relative to ANRIL expression level.

MicroRNAs	Breast cancer with low expression of ANRIL (n=30)	Breast cancer with high expression of ANRIL (n=30)	p -value ^a	ROC-AUC ^b
ANRIL	1.0 (0.22-3.56) ^c	11.5 (5.64-62.1) ^c	<0.000001	1.000
EMT associated genes (n=9)				
TWIST1	1,0 (0,24-6,34)	0,99 (0,13-7,97)	0,99 (NS)	0,501
ZEB1	1,0 (0,24-3,53)	1,19 (0,30-2,96)	0,55 (NS)	0,545
ZEB2	1,0 (0,39-3,89)	1,19 (0,44-3,53)	0,30 (NS)	0,578
SNAI1	1,0 (0,30-6,80)	1,17 (0,31-6,57)	0,38 (NS)	0,566
SNAI2	1,0 (0,39-3,16)	1,39 (0,23-4,97)	0,24 (NS)	0,588
VIM	1,0 (0,45-2,72)	1,09 (0,26-3,87)	0,50 (NS)	0,551
CDH2	1,0 (0,17-159)	1,24 (0,16-55,4)	0,47 (NS)	0,554
CDH1	1,0 (0,22-4,96)	0,99 (0,01-7,41)	0,63 (NS)	0,463
ZO-1	1,0 (0,40-3,80)	0,86 (0,23-4,07)	0,59 (NS)	0,460

^a Kruskal et Wallis H Test.

^b ROC (Receiver Operating Characteristics) - AUC (Area Under Curve) analysis.

^c Median (range) of gene mRNA levels ; the mRNA values of the samples were normalized such that the

ARTICLE N°2 : Valeur pronostique d'un variant d'épissage alternatif du lncRNA *MALATI* dans les carcinomes mammaires infiltrants.

8.1 Objectif

MALATI est un lncRNA présentant des propriétés inattendues par rapport à la majorité des lncRNAs. Il est un lncRNA extrêmement abondant, exprimé de façon ubiquitaire et apparaît très conservé parmi les mammifères. De plus, *MALATI* est un lncRNA doté à la fois de propriétés structurales (*speckles*) et de propriétés fonctionnelles, telles que la régulation de l'épissage alternatif et le contrôle de la transcription de gènes impliqués dans le cycle cellulaire, la motilité cellulaire et l'EMT. Toutefois, les mécanismes moléculaires liés à sa dérégulation ne sont pas encore élucidés et peu d'études ont été consacrées à *MALATI* dans les cancers du sein. Pour analyser l'implication de *MALATI* dans les carcinomes mammaires, nous avons étudié son expression par RT-PCR en temps réel dans une large série de 446 patientes atteintes d'une tumeur mammaire unilatérale au suivi clinico-biologique documenté. Nous avons recherché des corrélations entre *MALATI*, les différents paramètres cliniques, pathologiques et évolutifs et les gènes et protéines impliqués dans les mécanismes moléculaires auxquels *MALATI* participe dans d'autres types de cancers humains.

Ce travail a fait l'objet de l'article suivant :

Article 2

Prognostic value of a new identified *MALATI* alternatively spliced transcript in breast cancer.

Didier Meseure, Sophie Vacher, François Lallemand, Andre Nicolas, Rana Hatem, Walid Chemlali, Kinan Drak Alsibai, Leanne De Koning, Eric Pasmant, Celine Callens, Rosette Lidereau, Antonin Morillon and Ivan Bieche.

Article publié dans la revue *British Journal of Cancer* 2016;114:1395-404.

8.2 Résumé

Le lncRNA *MALAT1* est impliqué dans divers cancers mais son rôle dans le cancer du sein est encore peu connu. Nous avons utilisé les techniques de RT-PCR, hybridation *in situ*, et RPPA afin de quantifier (i) la forme sauvage (*WT*) et un variant d'épissage alternatif de *MALAT1* et (ii) un panel de transcrits et de protéines impliqués dans la voie de signalisation interagissant avec *MALAT1* sur une large série de tumeurs du sein provenant de patientes ayant un statut clinico-pathologique et une évolution à long terme documentés. La forme *WT* de *MALAT1* est surexprimée dans 14.0% des tumeurs du sein et les cellules épithéliales tumorales surexprimant *MALAT1* présentent un marquage nucléaire diffus et de nombreux et volumineux *speckles*. Cependant, la forme *WT* de *MALAT1* n'apparaît ni corrélée aux caractères d'agressivité des tumeurs de notre série, ni à une évolutivité péjorative. Toutefois, le criblage de la base de données dbEST a permis l'identification d'un transcrit d'épissage dont le profil d'expression est très différent de celui de la forme sauvage *WT-MALAT1*. Ce transcrit alternatif de *MALAT1* (Δ -*MALAT1*) est en effet principalement sous exprimé (18.8%) et est un facteur pronostique indépendant en analyse multivariée. En conclusion, cette étude révèle un mode d'expression complexe du lncRNA *MALAT1* dans les tumeurs du sein, caractérisé par des variants d'épissage de significations biologique et pronostique différents et suggère que cette expression complexe doit être prise en compte pour évaluer *MALAT1* comme biomarqueur prédictif et comme cible thérapeutique potentielle dans les cancers du sein.

8.3 Résultats

La surexpression du transcrit *WT-MALAT1* a été détectée dans 14.0% des tumeurs du sein, suggérant un rôle oncogénique. En utilisant une technique d'hybridation *in situ* (RNA-FISH), nous avons montré que les transcrits de *MALAT1* sont localisés principalement dans les noyaux et se traduisent par des signaux nucléaires diffus et de nombreux et très volumineux *speckles*. Aucune corrélation significative n'a été observée entre la surexpression de *WT-MALAT1* et les marqueurs d'agressivité tumorale, dont le grade histopathologique, le statut ganglionnaire et la taille macroscopique. Nous avons cependant identifié une corrélation significative entre la surexpression de *WT-MALAT1* et le statut RH+ des tumeurs mammaires (un marqueur de bon pronostic). En outre, l'analyse des courbes de survie n'a pas révélé une MFS différente entre les patientes présentant une surexpression de *WT-MALAT1* et celles

pourvues d'une expression normale. Plus de 25% des gènes codant pour des lncRNAs subissent un épissage alternatif conduisant à la synthèse d'au moins deux isoformes différents. En criblant la base de données dbEST à l'aide de l'ADNc de *MALAT1*, nous avons ainsi pu identifier un important transcrit résultant d'un épissage alternatif de *MALAT1* (nommé Δ -*MALAT1*) qui comporte deux régions délétées concomitantes de 119 pb et 243 pb. Dans notre cohorte, Δ -*MALAT1* présente un profil d'expression très différent par rapport à celui de *WT-MALAT1*. En effet, son expression varie considérablement dans les tissus mammaires tumoraux, Δ -*MALAT1* étant à la fois sous-exprimé (18.8%) et surexprimé (5.4%). De façon inattendue, une corrélation positive significative a été observée entre la sous-expression de Δ -*MALAT1* et les tumeurs mammaires présentant une grande taille macroscopique, un statut RH- et des niveaux élevés en mRNA de MKI67. En outre, l'analyse de la survie a révélé que les patientes présentant un faible niveau d'expression de Δ -*MALAT1* ont une MFS plus courte. Enfin, l'analyse multivariée a confirmé que l'expression de Δ -*MALAT1* est un facteur pronostique indépendant. L'expression de Δ -*MALAT1* est associée à des altérations de la machinerie d'épissage alternatif des pré-mRNAs et du complexe Drosha-DGCR8 de biogénèse des ncRNAs. Nos résultats suggèrent que l'expression du variant d'épissage alternatif est dérégulé par la protéine YAP et associé à une activation de la voie PI3K/Akt/mTOR. Nous avons également analysé les niveaux d'expression de gènes potentiellement impliqués dans divers phénotypes cellulaires et moléculaires associés à la carcinogénèse *via* la dérégulation de *MALAT1*. Les gènes significativement associés à *MALAT1* sont essentiellement impliqués dans la migration cellulaire (*RHOB*, *PLAU/UPA* et *MMP14*), l'édification du complexe répressif Polycomb PRC2 (*EED*, *SUZ12*, *JARID2*, *TUG*), l'apoptose (*BIRC6*), la réparation de l'ADN (*ATM*, *MSH2*, *XRCC1*), la biogénèse des ncRNAs (*DGCR8*) et les interacteurs de l'épissage alternatif (*SRSF1*, *SRSF3*, *UHMK1*). Les gènes impliqués dans le contrôle du cycle cellulaire et l'EMT ne sont pas liés à *MALAT1* dans notre série de cancers du sein

8.4 Conclusion

Cette étude confirme que le lncRNA *MALAT1* est impliqué dans le cancer du sein et révèle de plus un mode d'expression complexe caractérisé par différents variants d'épissage de signification biologique très différente. Cette expression particulière doit être prise en compte lors de l'évaluation des thérapies antitumorales ciblant ce lncRNA.

8.5 Article N°2: Prognostic value of a new identified *MALATI* alternatively spliced transcript in breast cancer.

Keywords: MALAT1; alternatively spliced transcript; breast cancer; prognostic value; signalling pathways

Prognostic value of a newly identified MALAT1 alternatively spliced transcript in breast cancer

Didier Meseure^{1,2}, Sophie Vacher¹, François Lallemand¹, Kinan Drak Alsibai², Rana Hatem¹, Walid Chemlali¹, Andre Nicolas², Leanne De Koning³, Eric Pasmant⁴, Celine Callens¹, Rosette Lidereau¹, Antonin Morillon⁵ and Ivan Bieche^{*1,4}

¹Department of Genetics, Unit of Pharmacogenetics, Institut Curie, 26 rue d'Ulm, Paris Cedex F-75248, France; ²Department of Pathology, Platform of Investigative Pathology, Institut Curie, 26 rue d'Ulm, Paris Cedex F-75248, France; ³Department of Translational Research, Institut Curie, 26 rue d'Ulm, Paris Cedex F-75248, France; ⁴EA7331, Faculty of Pharmaceutical and Biological Sciences, Paris Descartes University, Sorbonne Paris Cité, Paris Cedex F-75006, France and ⁵CNRS UMR 3244, Institut Curie, 26 rue d'Ulm, Paris Cedex F-75248, France

Background: Epigenetic deregulation is considered as a new hallmark of cancer. The long non-coding RNA MALAT1 has been implicated in several cancers; however, its role in breast cancer is still little known.

Methods: We used RT-PCR, in situ hybridisation, and RPPA methods to quantify (i) the full-length (FL) and an alternatively spliced variant (Dsv) of MALAT1, and (ii) a panel of transcripts and proteins involved in MALAT1 pathways, in a large series of breast tumours from patients with known clinical/pathological status and long-term outcome.

Results: MALAT1 was overexpressed in 14% (63/446) of the breast tumours. MALAT1-overexpressed tumour epithelial cells showed marked diffuse nuclear signals and numerous huge nuclear speckles. Screening of the dbEST database led to the identification of Dsv-MALAT1, a major alternatively spliced MALAT1 transcript, with a very different expression pattern compared with FL-MALAT1. This alternative Dsv-MALAT1 transcript was mainly underexpressed (18.8%) in our breast tumour series. Multivariate analysis showed that alternative Dsv-MALAT1 transcript is an independent prognostic factor. Dsv-MALAT1 expression was associated with alterations of the pre-mRNAs alternative splicing machinery, and of the Drosha-DGCR8 complex required for non-coding RNA biogenesis. Alternative Dsv-MALAT1 transcript expression was associated to YAP protein status and with an activation of the PI3K-AKT pathway.

Conclusions: Our results reveal a complex expression pattern of various MALAT1 transcript variants in breast tumours, and suggest that this pattern of expressions should be taken into account to evaluate MALAT1 as predictive biomarker and therapeutic target.

Several studies have recently shown that expression of long non-coding RNAs (lncRNAs) are dysregulated in various cancers and that these lncRNAs have important roles in tumorigenesis and tumour progression (Spizzo et al, 2012). One example of such oncogenic lncRNA is HOTAIR, which is highly expressed in breast cancer and is a predictor for metastasis formation and associated with a poor prognosis (Gupta et al, 2010). Among these

lncRNAs, MALAT1 (metastasis-associated lung adenocarcinoma transcript 1), also referred as NEAT2 (nuclear-enriched abundant transcript 2) was discovered 10 years ago by using a subtractive hybridisation approach. MALAT1 was originally identified as a transcript showing significant expression in non-small cell lung tumours at high risk for metastasis (Ji et al, 2003). MALAT1 gene has a length of 8708 bp (NR_002819.2) and is localised in

*Correspondence: Dr I Bieche; E-mail: ivan.bieche@curie.fr

Received 20 November 2015; revised 16 February 2016; accepted 31 March 2016; published online 12 May 2016

© 2016 Cancer Research UK. All rights reserved 0007–0920/16



chromosome 11q13.1. Unlike most of lncRNAs, MALAT1 is extremely abundant, ubiquitously expressed and highly conserved among mammals, with potentially major functional roles in mammalian cells. MALAT1 is a nuclear-retained lncRNA, suggesting both structural and functional properties, for example, nuclear architecture and organisation, splicing, or gene-expression regulation (Gutschner et al, 2013a). MALAT1 has been implicated in alternative splicing regulation, showing interactions with several splicing factors, such as SRSF1 (Tripathi et al, 2010). MALAT1 has also been linked to transcriptional control of genes involved in cell cycle, cell motility and EMT (Gutschner et al, 2013a). MALAT1 could act as a transcription activator by mediating assembly of Polycomb repressive complexes (Yang et al, 2011).

MALAT1 upregulation has been reported in several tumour types and is also a negative prognostic factor in lung, pancreas, colorectal and bladder cancers (Zhang et al, 2015). Molecular mechanisms involved in MALAT1 dysregulation are still unclear. Activation of MALAT1 by gene amplification seems unlikely because MALAT1 is located in a chromosomal region (11q13.1) not recurrently amplified in human cancers (Curtis et al, 2012). Mutations in the MALAT1 gene were recently discovered in human cancers (The Cancer Genome Atlas studies; Kandath et al, 2013). MALAT1 seems the most frequently lncRNA mutated in human cancers. Rare cases of chromosomal translocations involving MALAT1 have also been reported in mesenchymal hamartomas and renal cell carcinomas (Davis et al, 2003; Mathews et al, 2013). MALAT1 epigenetic dysregulation mediated by CpG island methylation was not reported. A post-transcriptional MALAT1 regulation mechanism mediated by one microRNA (Hsa-miR-125b) has been only reported in bladder cancer (Han et al, 2013).

Few studies concerning MALAT1 in breast cancer are available. Guffanti et al (2009) identified MALAT1 as an abundantly expressed lncRNA in breast tumours. Rare mutations were recently described in luminal breast cancer (Ellis et al, 2012). Clinical prognostic value of MALAT1 dysregulation in breast cancer is little known at this time (Xu et al, 2015).

To obtain further insight concerning involvement of MALAT1 in molecular pathogenesis of breast cancer, we used quantitative real-time reverse-transcriptase–polymerase chain reaction (qRT–PCR) assay, to quantify the full-length (FL) and an alternatively spliced variant (Dsv) of MALAT1 mRNA expression in a series of 446 patients with unilateral invasive breast tumours and known long-term outcome. We sought links between MALAT1 mRNA expression pattern and classical clinical and pathological parameters, including patient outcome. We also sought relationships between MALAT1 and genes and proteins expression known to be involved in different steps of MALAT1 pathway dysregulation observed in others types of human cancers.

MATERIALS AND METHODS

Patients and samples. Samples of 446 unilateral invasive primary breast tumours excised from women managed at Institut Curie–Hôpital René Huguénin (Saint-Cloud, France) from 1978 to 2008 have been analysed. All patients cared in our institution before 2007 were informed that their tumour samples might be used for scientific purposes and had the opportunity to decline. Since 2007, patients treated in our institution have given their approval by signed inform consent. This study was approved by the local ethics committee (Breast Group of René Huguénin Hospital). Samples were immediately stored in liquid nitrogen until RNA extraction. A tumour sample was considered suitable for our study if the proportion of tumour cells exceeded 70%.

All patients (mean age 61.8 years, range 31–91 years) met the following criteria: primary unilateral nonmetastatic breast carcinoma for which complete clinical, histological and biological

data were available; no radiotherapy or chemotherapy before surgery; and full follow-up at Institut Curie–Hospital René Huguénin.

Treatment (information available for 438 patients) consisted of modified radical mastectomy in 278 cases (63.9%) and breast-conserving surgery plus locoregional radiotherapy in 160 cases (36.1%). The patients had a physical examination and routine chest radiotherapy every 3 months for 2 years, then annually. Mammograms were done annually. Adjuvant therapy was administered to 360 patients, consisting of chemotherapy alone in 87 cases, hormone therapy alone in 172 cases and both treatments in 101 cases. The histological type and the number of positive axillary nodes were established at the time of surgery. The malignancy of infiltrating carcinomas was scored according to Scarff Bloom Richardson's histoprognostic system.

Hormone receptor (HR; i.e., oestrogen receptor-alpha (ERα), progesterone receptor (PR) and human epidermal growth factor receptor 2 (ERBB2) statuses were determined at the protein level by using biochemical methods (dextran-coated charcoal method, enzyme immunoassay or immunohistochemistry) and confirmed by qRT–PCR assays (Bieche et al, 1999, 2001).

The population was divided into four groups according to HRs (ERα and PR) and ERBB2 status, as follows: two luminal subtypes (HR^P/ERBB2^P (n/45)) and (HR^P/ERBB2⁻ (n/195)); an ERBB2^P subtype (HR⁻/ERBB2^P (n/46)) and a triple-negative subtype (HR⁻/ERBB2⁻ (n/64)). Standard prognostic factors of this tumour set are presented in (Supplementary Table 1). During a median follow-up of 9.1 years (range 4.3 months to 33.2 years), 176 patients developed metastasis.

Ten specimens of adjacent normal breast tissue from breast cancer patients and normal breast tissue from women undergoing cosmetic breast surgery were used as sources of normal RNA.

RNA extraction. Total RNA was extracted from breast tissue samples by using acid–phenol guanidium. RNA quality was determined by electrophoresis through agarose gels, staining with ethidium bromide and visualisation of the 18S- and 28S-RNA bands under ultraviolet light.

qRT–PCR. Quantitative values were obtained from the cycle number (Ct value) at which the increase in the fluorescence signal associated with exponential growth of PCR products started to be detected by the laser detector of the ABI Prism 7900 sequence detection system (Perkin-Elmer Applied Biosystems, Foster City, CA, USA), using PE biosystems analysis software according to the manufacturer's manuals.

The precise amount of total RNA added to each reaction mix (based on optical density) and its quality (i.e., lack of extensive degradation) are both difficult to assess. Therefore, transcripts of the TBP gene (Genbank accession NM_003194) encoding the TATA box-binding protein (a component of the DNA-binding protein complex TFIID) were also quantified as an endogenous RNA control. Each sample was normalised on the basis of its TBP content. TBP was selected as an endogenous control due to the absence of known TBP retropseudogenes (retropseudogenes lead to co-amplification of contaminating genomic DNA and thus interfere with qRT–PCR, despite the use of primers in separate exons; Bieche et al, 1999).

Results, expressed as N-fold differences in target-gene expression relative to the TBP gene and termed 'Ntarget', were determined as $N_{target} = 2^{DCT_{sample}}$, where the DCt value of the sample was determined by subtracting the average Ct value of target gene from the average Ct value of TBP gene.

The target-gene values of the breast tumour samples were subsequently normalised such that the median of the target-gene values for the 10 normal breast tissues was 1.

The primers for TBP, MALAT1 and others target genes were chosen with the assistance of the Oligo 6.0 program (National

Biosciences, Plymouth, MN, USA; Supplementary Table 2). dbEST and nr databases were scanned to confirm the total gene specificity of the nucleotide sequences chosen for the primers and the absence of single-nucleotide polymorphisms. To avoid amplification of contaminating genomic DNA, one of the two primers was placed at the junction between two exons or on two different exons. Agarose gel electrophoresis was used to verify the specificity of PCR amplicons. The conditions of cDNA synthesis and PCR were as previously described (Bieche et al, 1999).

In situ hybridisation. We used Stellaris FISH Probes, Human MALAT1 with Quasar 570 Dye (Biosearch Technologies, Petaluma, CA, USA). First, paraffin-embedded tissue sliced at 4–5 mm thickness were obtained from normal and tumour tissues by using a microtome (Thermo scientific Sandom HE 340 E, Walldorf, Germany). Formalin-fixed paraffin-embedded breast tissue sections were deparaffinized by using 100% xylene, 100% ethanol, 95% ethanol, 70% ethanol and RNase-free PBS. Then, slides were incubated for 20 min at 37°C, and washed twice with PBS. We created a working probe solution at 125 nM (probe diluted in hybridisation buffer). We immersed tissue sections in a wash buffer for 2–5 min, while assembling the humidified chamber. Then, we dispensed 100 ml of working probe solution onto tissue sections, placed them in the humidified chamber, and covered them with parafilm. We incubated tissue sections in the dark at 37°C for at least 4 h. After decanting wash buffer, we added DAPI nuclear stain and immersed tissue sections in SSC. Finally, we added a small drop of antifade onto tissue sections and covered with a cover glass and proceeded to imaging.

RPPA. Samples were disrupted in Laemmli buffer (50 mM Tris pH ¼ 6.8, 2% SDS, 5% glycerol, 2 mM DTT, 2.5 mM EDTA, 2.5 mM EGTA, 1× HALT phosphatase inhibitor (Perbio, Villebon-sur-Yvette, France; 78420), protease inhibitor cocktail complete MINI EDTA-free (Roche, Basel, Switzerland; 1836170, 1 tablet per 10 ml), 4 mM Na₃VO₄ and 20 mM NaF) qsp 5 ml H₂O, using a Tissue Lyser (Qiagen, Venlo, Netherlands) and two 5 mm stainless beads per sample. Extracts were then boiled for 10 min at 100°C, passed through a fine needle to reduce viscosity and centrifuged for 15 min at 13 000 r.p.m. The supernatant was collected and stored at –80°C. Protein concentration was determined (Pierce BCA reducing agent compatible kit, Pierce, Waltham, MA, USA; ref 23252). Samples were deposited onto nitrocellulose covered slides (Fast slides, Maine Manufacturing, Sanford, ME, USA) using a dedicated arrayer (2470 arrayer, Aushon Biosystems, Billerica, MA, USA). Five serial dilutions, ranging from 1000 to 62.5 mg ml^{–1}, and two technical replicates per dilution were printed for each sample. Arrays were labelled with specific antibodies or without primary antibody (negative control), using an Autostainer Plus (Dako, Glostrup, Denmark). Briefly, slides were incubated with avidin, biotin and peroxides blocking reagents (Dako) before saturation with TBS containing 0.1% Tween-20 and 5% BSA (TBST-BSA). Slides were then probed overnight at 4°C with primary antibodies diluted in TBST-BSA. After washes with TBST, arrays were probed with horseradish peroxidase-coupled secondary antibodies (Jackson ImmunoResearch Laboratories, New market, UK) diluted in TBST-BSA for 1 h at room temperature (RT). To amplify the signal, slides were incubated with Bio-Rad Amplification Reagent (Bio-Rad, Hercules, CA, USA) for 15 min at RT. The arrays were washed with TBST, probed with Alexa647-Streptavidin (Molecular Probes, Eugene, OR, USA) diluted in TBST-BSA for 1 h at RT and washed again in TBST. For staining of total protein, arrays were incubated 15 min in 7% acetic acid and 10% methanol, rinsed twice in water, incubated 10 min in Sypro Ruby (Invitrogen, Carlsbad, CA, USA) and rinsed again. The processed slides were dried by centrifugation and scanned using a GenePix 4000B microarray scanner (Molecular Devices, Sunnyvale, CA, USA). Spot intensity was determined with MicroVigene software (VigeneTech Inc.,

Carlisle, MA, USA). All primary antibodies used in RPPA have been previously tested by Western Blotting to assess their specificity for the protein of interest.

Raw data were normalised using Normacurve (Troncale et al, 2012), which normalises for fluorescent background per spot, a total protein stain and potential spatial bias on the slide. Next, each RPPA slide was median centred and scaled (divided by median absolute deviation). We then corrected for remaining sample loadings effects individually for each array by correcting the dependency of the data for individual arrays on the median value of each sample over all arrays using a linear regression.

Statistical analysis. The distributions of target mRNA levels were characterised by their median values and ranges. Relationships between mRNA levels of the different target genes, and between mRNA levels and clinical parameters, were identified using nonparametric tests, namely the w^2 -test (relation between two qualitative parameters), the Mann–Whitney's U-test (relation between one qualitative parameter and one quantitative parameter) and the Spearman's rank correlation test (relation between two quantitative parameters). Differences were considered significant at confidence levels <95% ($P < 0.05$).

To visualise the efficacy of a molecular marker (MALAT1 level) to discriminate two populations (patients that developed/did not develop metastases) in the absence of an arbitrary cut-off value, data were summarised in an receiver operating characteristic curve (Hanley and McNeil, 1982). The AUC (area under curve) was calculated as a single measure for discriminate efficacy. Metastasis-free survival (MFS) was determined as the interval between initial diagnosis and detection of the first metastasis. Survival distributions were estimated by the Kaplan–Meier method, and the significance of differences between survival rates were ascertained with the log-rank test. The Cox-proportional hazards regression model was used to assess prognostic significance and the results are presented as hazard ratios and 95% confidence intervals.

RESULTS

MALAT1 expression in breast tumours and relationship with classical clinico-pathological parameters and patient outcome. In order to determine the prognostic significance of MALAT1 expression pattern in human breast tumours, we analysed MALAT1 mRNA levels in a large series of 446 primary breast tumours from patients with known clinical/pathological status and long-term outcome (Supplementary Table 1).

Among the 446 breast tumour RNA samples tested, 63 (14.1%) tumours showed MALAT1 mRNA overexpression (NMALAT1 from 3.02 to 13.4), and only 13 (2.9%) tumours showed MALAT1 mRNA underexpression (NMALAT1 from 0.15 to 0.32), as compared with normal breast tissues. We sought links between MALAT1 mRNA level status and standard clinico-pathological and biological factors in breast cancer (Supplementary Table 3). Significant positive associations were observed between the tumour group showing MALAT1 overexpression and ER $^+$ -positive ($P = 0.000015$), PR-positive ($P = 0.00079$) and molecular subtypes ($P = 0.00000075$). It is noteworthy that majority (9/13) of the MALAT1 underexpressed tumours was of triple-negative (HR $^-$ /ERBB2 $^-$) subtype.

We also examined PIK3CA mutation status, and expressions of EGFR and MKI67 (which encodes the proliferation-related antigen Ki-67). None of these three markers showed significant link with MALAT1 expression.

To further investigate whether MALAT1 mRNA expression could be of prognostic relevance, the log-rank test was used to identify relations between MFS and MALAT1 mRNA expression.

Results showed that MFS was not significantly influenced by MALAT1 overexpression status ($P = 0.23$; data not shown).

Results of MALAT1 mRNA levels shown in Supplementary Table 3 were obtained by using a primer pair (U13/L13) that encompass region 4891–4975 of published MALAT1 cDNA sequence (GenBank #NR_002819.2; Figure 1). Similar results (frequency of MALAT1 overexpression links with classical clinicopathological parameters and patient outcome) were obtained with a second primer pair (U9/L9) localised in MALAT1 gene at region 6565–6658 (Figure 1).

Relationship between MALAT1 mRNA level and HOTAIR and ANRIL mRNA levels. We tested possible relation between MALAT1 and HOTAIR and ANRIL (the two most documented lncRNAs that also interact with Polycomb repressive complexes) mRNA levels. We did not observe any association between MALAT1 and these two Polycomb complexes associated lncRNAs ($r = 0.069$, $P = 0.14$ for HOTAIR; $r = 0.014$, $P = 0.76$ for ANRIL; Spearman's rank correlation test).

Localisation of MALAT1 transcript in epithelial tumour cells. We detected specific MALAT1 RNA in epithelial and stromal cells of all ten tumour samples studied by in situ hybridisation (ISH). MALAT1 mRNA was found exclusively in the nucleus of both stromal and tumour epithelial cells. We detected strong specific MALAT1 RNA level in epithelial cells of the five tumours, which overexpressed MALAT1 mRNA (using qRT-PCR analysis) and low specific MALAT1 RNA level in the five tumours which did not overexpress MALAT1 mRNA. We thus obtained a perfect match between MALAT1 mRNA expression by using qRT-PCR and ISH analysis. MALAT1-overexpressed tumour epithelial cells showed marked diffuse nuclear signals and numerous nuclear speckles of variable size and shape as compared with MALAT1 normal-expressed tumour epithelial cells (Figure 2).

Identification of a major alternatively spliced MALAT1 transcript in breast tumours, and relationships with classical clinicopathological parameters and patient outcome. Screening of the dbEST database with the MALAT1 cDNA led to identification of two major groups of alternatively spliced MALAT1 ESTs. The first major alternatively spliced MALAT1 transcript (named D1sv-MALAT1) had a 119-bp deletion (from 6446 to 6564, from the NR_002819.2 sequence), resulting from alternative splicing of MALAT1 mRNA, whereas the second alternatively spliced MALAT1 transcript (named D2sv-MALAT1) had a 243-bp deletion (from 4633 to 4875, from the NR_002819.2 sequence). The deleted nucleotide sequences show consensus sequences of donor/acceptor splice sites.

To verify presence and quantify mRNA levels of these alternative splicing variants in our breast cancer series, we carried

out non-quantitative (classical) RT-PCR using primer pairs with one of the two primers placed at the junction of the two spliced regions: U2/L2 for the 243-bp alternatively spliced MALAT1 transcript (D2sv-MALAT1) and U18/L18 for the 119-bp alternatively spliced MALAT1 transcript (D1sv-MALAT1; Figure 1). However, additional qualitative analyses using primer couples U1/L20 or U2/L18 (Figure 1 and Supplementary Figure 1) showed that these two splices were always associated together. This unique transcript, showing both 119-bp and 243-bp deletions, is named Dsv-MALAT1 for the remaining part of the manuscript, and FL-MALAT1 for the FL transcript.

All the 446 breast tumour RNA samples tested showed a marked presence of Dsv-MALAT1 transcript. We observed a highly positive correlation between Dsv-MALAT1 and FL-MALAT1 expressions (Supplementary Figure 2A). The relative expression of Dsv-MALAT1 and FL-MALAT1 for each individual sample listed as dot/box plots is indicated in Supplementary Figure 2B. As compared with normal breast tissues, 24 (5.4%) tumours showed Dsv-MALAT1 mRNA overexpression (NDsv-MALAT1 from 3.16 to 8.4), and surprisingly 84 (18.8%) tumours showed Dsv-MALAT1 mRNA underexpression (NDsv-MALAT1 from 0.05 to 0.32). Marked significant positive associations were observed between the tumour group showing Dsv-MALAT1 underexpression and large macroscopic tumour size ($P = 0.0023$), ER α -negative ($P = 0.000062$), PR-negative ($P = 0.0000051$) and molecular subtypes ($P = 0.00074$; Table 1). We observed the same associations between Dsv-MALAT1 mRNA levels and molecular subtypes in a series of 21 breast cell lines, including 4 non-cancerous cell lines, 8 triple-negative cell lines, 4 ERBB2 cell lines and 5 RE^P cell lines (Supplementary Table 4). In particular, we observed a marked Dsv-MALAT1 underexpression in the triple-negative subtype. Dsv-MALAT1 underexpression was also highly associated with MKI67 mRNA levels ($P = 0.00033$). MFS was significantly influenced by Dsv-MALAT1 (overexpression vs normal expression vs underexpression) expression status ($P = 0.0099$; Figure 3A). AUC analyses was then performed to identify a putative cut-point to divide the cohort into two relevant Dsv-MALAT1 expression subgroups. Results confirmed that MFS of patients with low Dsv-MALAT1-expressing tumours (5-year RFS $70.3 \pm 2.5\%$; 10-year RFS $58.5 \pm 2.8\%$; 15-year RFS $51.4 \pm 3.0\%$) was shorter than that of patients whose tumours highly expressed Dsv-MALAT1 (5-year RFS $85.9 \pm 3.5\%$; 10-year RFS $82.3 \pm 3.9\%$; 15-year RFS $76.0 \pm 5.1\%$; $P = 0.000015$; Figure 3B).

Results of Dsv-MALAT1 mRNA levels shown in Table 1 were obtained by using a primer pair (U18/L18) that encompasses the 119 bp deleted region. Similar results (frequency of Dsv-MALAT1 overexpression, links with classical clinicopathological parameters and patient outcome) were obtained with a second primer

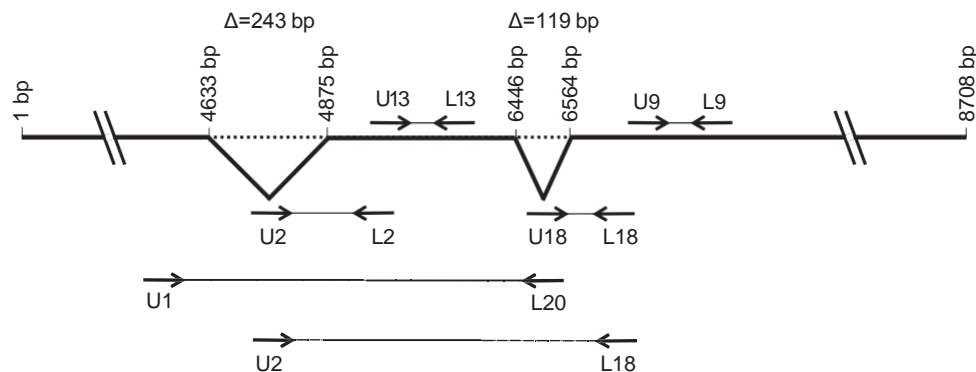


Figure 1. Location of primers used for MALAT1 mRNA expression analysis. Alternative splicing variants were amplified using primer pairs with one of the two primers placed at the junction of the two spliced regions: U2/L2 for the 243 bp alternatively spliced MALAT1 transcript (D2sv-MALAT1) and U18/L18 for the 119 bp alternatively spliced MALAT1 transcript (D1sv-MALAT1).

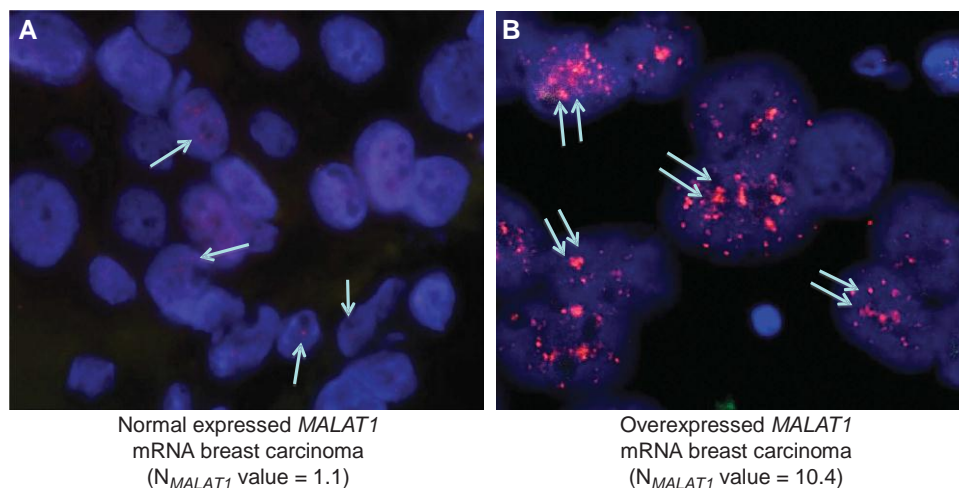


Figure 2. ISH of MALAT1 RNA in breast tumours. (A) Example of normal-expressed MALAT1 mRNA breast tumour (N_{MALAT1} value \approx 1.1, as determined by qRT-PCR analysis). Weak signals represented by small speckles (speckles in red, arrow) of equivalent size and shape, regularly distributed within nuclei of tumour cells (nuclei in blue) \times 400. (B) Example of overexpressed MALAT1 mRNA breast tumour (N_{MALAT1} value \approx 10.4). Marked diffuse signals and numerous and frequently huge nuclear speckles (speckles in red, two arrows) of variable size and shape within nuclei of tumour cells (nuclei in blue) \times 600.

pair (U2/L2) that encompass the 243 bp deleted region (Figure 1; Supplementary Table 5).

Finally, the prognostic significance of the five parameters identified in univariate analysis, including histopathological grade, lymph node status, macroscopic tumour size, PR status (Supplementary Table 1) and Dsv-MALAT1 expression status (Figure 3B) persisted (except for lymph node and PR status) in Cox multivariate regression analysis of MFS (Supplementary Table 6).

Relationship between Dsv-MALAT1 mRNA levels and Hsa-miR-125b expression status. As Hsa-miR-125b suppresses bladder cancer development by downregulating MALAT1 (Han et al, 2013), we tested the possible negative correlation between Dsv-MALAT1 and Hsa-miR-125b mRNA level in breast cancer. Hsa-miR-125b levels were analysed in 20 low-Dsv-MALAT1-expressing (median mRNA value: 0.36) and 20 high-Dsv-MALAT1-expressing breast tumours (median mRNA value: 2.13). We found no link between Dsv-MALAT1 and Hsa-miR-125b expression status; the median Hsa-miR-125b value was 0.13 in low-Dsv-MALAT1-expressing breast tumours and 0.11 in high-Dsv-MALAT1-expressing breast tumours. Similar results were obtained with 20 low-FL-MALAT1-expressing and 20 high-FL-MALAT1-expressing breast tumours.

Relationship between Dsv-MALAT1 mRNA levels and YAP protein level. As YAP protein regulates transcription of MALAT1 gene in liver cancer (Wang et al, 2014), we tested the possible positive correlation between YAP protein and Dsv-MALAT1 mRNA levels in breast cancer. YAP protein levels were analysed by using RPPA assay in 143 samples from our series of 446 breast tumours. We found a significant positive link with Dsv-MALAT1 mRNA level ($r = 0.303$, $P = 0.00032$; Spearman's rank correlation test) but no link between YAP protein and FL-MALAT1 mRNA levels of expression. As Dsv-MALAT1 low level is associated with a poor outcome and with low-YAP-protein level, we tested if YAP protein level could be also of prognostic relevance. We did not observe in our smaller series of 143 samples, any statistical correlation between low-YAP-protein expression and poor outcome.

Relationship between Dsv-MALAT1 mRNA levels and a large panel of selected genes involving in various signalling pathways. To obtain further insight into MALAT1 dysregulated pathways in breast cancer, we evaluated by qRT-PCR mRNA

expression of a large number of selected genes in 20 low-Dsv-MALAT1-expressing and 20 high-Dsv-MALAT1-expressing breast tumours. We assessed expression level of 48 genes involved in various cellular and molecular phenomena associated with carcinogenesis. These genes encode proteins involved in cell cycle control ($n = 7$), cell migration ($n = 5$), polycomb repressive complexes (PRC1) ($n = 3$) and PRC2 ($n = 5$), EMT ($n = 7$), apoptosis ($n = 6$) and DNA repair (Yang et al, 2011). We also focused on expression of well-known regulators (DGCR8; AGO2) and interactors (SRSF1, SRSF2, SRSF3, UHMK1) of MALAT1 (Gutschner et al, 2013a), as well as transcriptional dysregulated genes after MALAT1 depletion in A549 lung adenoma cell line (ROBO1, MCAM; Gutschner et al, 2013b) and HeLa cells (IFI44; Miyagawa et al, 2012).

Expression of 19 (39.5%) of these 48 genes was significantly positively associated with Dsv-MALAT1 expression (Table 2). Genes significantly associated to Dsv-MALAT1 were mainly involved in cell migration (RHOB, PLAU/UPA and MMP14), Polycomb repressive complex PRC2 (EED, SUZ12, JARID2, TUG1), apoptosis (BIRC6), DNA repair (ATM, MSH2, XRCC1) and regulators (DGCR8) and interactors (SRSF1, SRSF3, UHMK1) of MALAT1. Genes involved in cell cycle control and EMT, as well as putative MALAT1-inducible genes (identified by MALAT1 depletion in cell lines) were not linked to the MALAT1 in breast cancer.

Relationship between levels of Dsv-MALAT1 mRNA and RTK/MAPK/PI3K proteins. As several studies recently suggested that MALAT1 promotes proliferation and metastasis of various cancers by activating the RTK/MAPK/PI3K pathways (Wu et al, 2014; Dong et al, 2015; Xu et al, 2015), we tested possible correlation between Dsv-MALAT1 and various proteins involved in these signalling pathways.

Twenty-eight protein (non-phosphorylated or/and phosphorylated) levels were analysed using RPPA assays in 143 samples from our series of 446 breast tumours. These selected proteins are involved in TKR ($n = 9$), MAPK ($n = 4$) and PI3K/AKT ($n = 15$) pathways (Table 3). Low-Dsv-MALAT1 mRNA level were associated to high levels of 4 among the 15 proteins involved in the PI3K/AKT pathway (i.e., FOXO1, p70 S6 Kinase total protein and phosphorylated in Threonine 389, S6 ribosomal protein phosphorylated in Ser240/Ser244), but to none of the two others signalling pathways. Low-FL-MALAT1 mRNA level was exclusively associated to high level of FOXO1 (Table 3).

Table 1. Relationship between Dsv-MALAT1-spliced transcript levels and classical clinical biological parameters in a series of 446 breast cancer

	Number of patients (%)				P-value ^a
	Total population (%)	Dsv-MALAT1 underexpression	Dsv-MALAT1 normal expression	Dsv-MALAT1 overexpression	
Total	446 (100.0)	84 (18.8)	338 (75.8)	24 (5.4)	
Age (years)					
≤50	94 (21.1)	25 (26.6)	67 (71.3)	2 (2.1)	0.038
>50	352 (78.9)	59 (16.8)	271 (77.0)	22 (6.3)	
SBR histological grade^{b,c}					
I	57 (13.0)	4 (7.0)	49 (86.0)	4 (7.0)	0.043
II-III	380 (87.0)	79 (20.8)	282 (74.2)	19 (5.0)	
Lymph node status^d					
0	117 (26.3)	17 (14.5)	90 (76.9)	10 (8.5)	0.043
1-3	231 (51.9)	39 (16.9)	182 (78.8)	10 (4.3)	
≥3	97 (21.8)	27 (27.8)	66 (68.0)	4 (4.1)	
Macroscopic tumour size^e					
≤25 mm	218 (49.8)	28 (12.8)	174 (79.8)	16 (7.3)	0.0023
>25 mm	220 (50.2)	55 (25.0)	157 (71.4)	8 (3.6)	
ERα status					
Negative	115 (25.8)	37 (32.2)	76 (66.1)	2 (1.7)	0.000062
Positive	331 (74.2)	47 (14.2)	262 (79.2)	22 (6.6)	
PR status					
Negative	191 (42.8)	55 (28.8)	132 (69.1)	4 (2.1)	0.0000051
Positive	255 (57.2)	29 (11.4)	206 (80.8)	20 (7.8)	
ERBB2 status					
Negative	353 (79.1)	60 (17.0)	273 (77.3)	20 (5.7)	0.15 (NS)
Positive	93 (20.9)	24 (25.8)	65 (70.0)	4 (4.3)	
Molecular subtypes					
RH ⁻ ERBB2 ⁻	68 (15.2)	22 (32.4)	46 (67.6)	0 (0)	0.00074
RH ⁻ ERBB2 ^P	42 (9.4)	14 (33.3)	26 (61.9)	2 (4.8)	
RH ^P ERBB2 ⁻	285 (63.9)	38 (13.3)	227 (79.6)	20 (7.0)	
RH ^P ERBB2 ^P	51 (11.4)	10 (19.6)	39 (76.5)	2 (3.9)	
PIK3CA mutation status					
Wild type	299 (67.0)	69 (23.1)	215 (71.9)	15 (5.0)	0.0049
Mutated	147 (33.0)	15 (10.2)	123 (83.7)	9 (6.1)	
MKI67 mRNA expression					
Median	12.5 (0.80-117)	17.6 (0.86-117)	11.8 (0.80-94.5)	11.2 (0.93-47.1)	0.00033 ^f
EGFR mRNA expression					
Median	0.22 (0.00-106)	0.13 (0.02-3.20)	0.23 (0.00-106)	0.20 (0.03-1.51)	0.019 ^f

Abbreviations: ERα ¼ oestrogen receptor-α; ERBB2 ¼ human epidermal growth factor receptor 2; HR ¼ hormone receptor; NS ¼ not significant; PR ¼ progesterone receptor. The bold values are statistically significant (P-value < 0.05).

^aχ²-test.

^bScarff Bloom Richardson classification.

^cInformation available for 437 patients.

^dInformation available for 445 patients.

^eInformation available for 438 patients.

^fKruskal-Wallis's H test.

DISCUSSION

Recent studies have demonstrated the importance of non-protein-coding part of human genome in carcinogenesis. Among numerous kinds of non-protein-coding RNAs, lncRNAs have a key regulatory role in cancer biology. lncRNAs are dysregulated in different types of cancer and the expression levels of certain lncRNAs are associated with metastasis and prognosis of cancer. Overexpression of certain lncRNAs, behaving like oncogenes, can promote tumour growth and cancer cell invasion (Cheetham et al, 2013).

In this study, we focused on the lncRNA MALAT1 that has been shown dysregulated in various cancer types (Zhang et al, 2015), but poorly studied in breast cancer. One study, using deep-sequencing

technology, identified MALAT1 as one of the highly expressed lncRNAs in breast tumours (Guffanti et al, 2009).

We tested 10 normal breast tissue RNAs and 446 unilateral invasive primary breast tumour RNAs, using qRT-PCR method. MALAT1 mRNA was detected in all breast tumour samples and also in all normal breast tissues.

Overexpression of MALAT1 mRNA was detected in 14% (63/446) of breast tumours, confirming the oncogenic role of MALAT1. Indeed, MALAT1 is overexpressed in several cancer types, including lung, colon and hepatocarcinoma, and overexpression of MALAT1 in various cell lines enhanced cell proliferation, whereas in nude mice, increased levels of MALAT1 promoted tumour formation (Ji et al, 2003; Guo et al, 2010; Gupta et al, 2010; Gibb et al, 2011; Schmidt et al, 2011; Lai et al, 2012). Additional studies have also demonstrated that depletion of

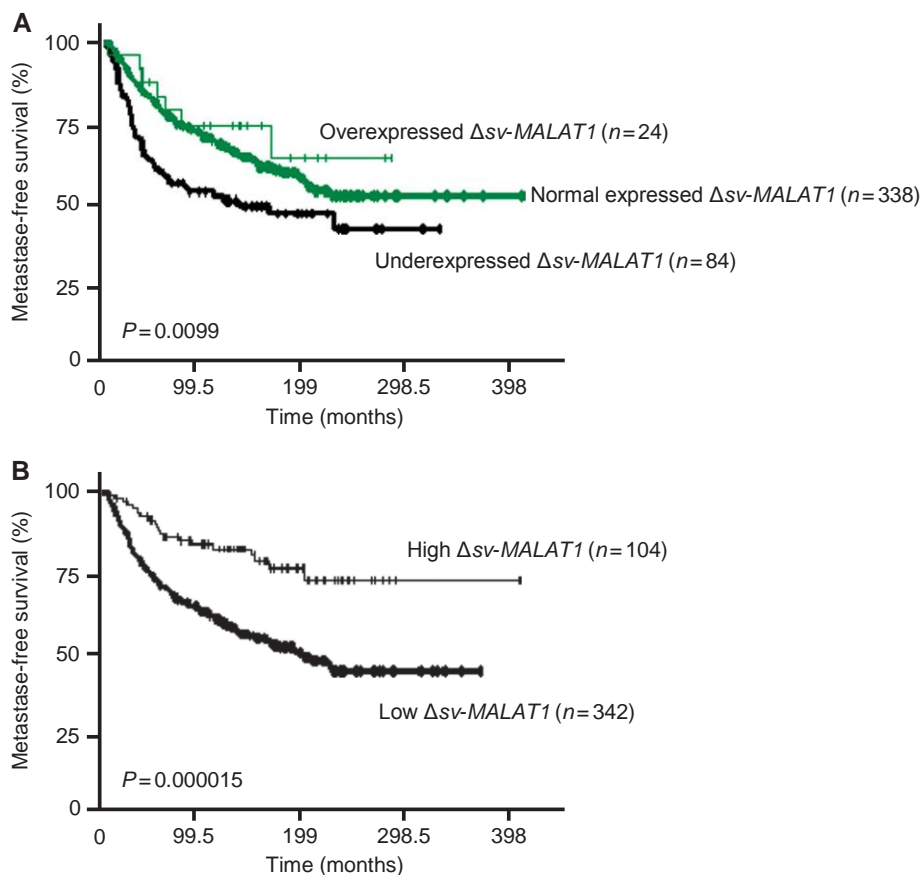


Figure 3. MFS curves of patient groups according to Dsv-MALAT1 mRNA expression level in the series of 446 breast tumours. (A) MFS curves of three patients groups with under, normal and overexpressed Dsv-MALAT1 tumours, as compared with normal breast tissues. (B) MFS curves for patients with high-Dsv-MALAT1-expressing and low-MALAT1-expressing tumours, using an optimal cut-off value.

MALAT1 impaired proliferative and invasive properties of cancer cells (Guo et al, 2010, Schmidt et al, 2011, Gutschner et al, 2013b).

By using ISH, we showed that MALAT1 transcripts were predominantly localised in nuclear speckles. Nucleus of the MALAT1-overexpressed tumour epithelial cells showed marked diffuse nuclear signals and numerous huge nuclear speckles.

No significant links were observed between MALAT1 mRNA overexpression and markers of aggressiveness, including histopathological grade, lymph node status and macroscopic tumour size, suggesting that overexpression of MALAT1 does not have a major role in aggressiveness of breast carcinomas. Moreover, we observed a link between MALAT1 mRNA overexpression and HR-positive tumours (a marker of good prognostic), suggesting that MALAT1 could be an ER-induced gene in breast cancer. Finally, survival analysis did not reveal that patients with MALAT1-overexpressed tumour had shorter MFS.

Alternative mRNA splicing is a common mechanism for regulating gene expression in higher eukaryotes, and there are many examples of development-specific, tissue-specific and tumour-specific differences in splicing events. In the GENCODE v7 catalogue of human lncRNAs, 425% of lncRNA genes show evidence of alternative splicing with at least two different transcript isoforms per gene locus (Derrien et al, 2012). The vast majority of alternatively spliced lncRNA introns are flanked by canonical splice sites (GT/AG), with no differences in splicing signal compared with the protein-coding genes (Derrien et al, 2012). In the present study, by screening the dbEST database with the FL-MALAT1 cDNA (named FL-MALAT1), we identified a major alternatively spliced MALAT1 transcript (named Dsv-MALAT1) with two concomitant deleted regions of 119 bp and 243 bp. As

expected, these alternatively spliced sequences showed consensus sequences of donor/acceptor splice sites. In our cohort, Dsv-MALAT1 showed a very different expression pattern, as compared with FL-MALAT1. Indeed, Dsv-MALAT1 expression varied widely in tumour tissues, being both underexpressed (18.8%) and overexpressed (5.4%). Surprisingly, a significant link was observed between Dsv-MALAT1 underexpression and tumours with large macroscopic size, negative for HRs and expressing high MKI67 mRNA levels, suggesting that underexpression of Dsv-MALAT1 has a role in aggressiveness of breast tumours. In this regard, in contrast to the FL-MALAT1 expression, survival analysis revealed that patients with low-Dsv-MALAT1-expressed tumours had shorter MFS. Moreover, multivariate analysis showed that Dsv-MALAT1 expression status was an independent prognostic marker for MFS. This alternatively spliced MALAT1 transcript isoform could act as decoys, sequestering biomolecules that fixed on the FL-MALAT1 transcript and thus dysregulating its function. Taken together, these results suggest that this alternatively spliced Dsv-MALAT1 transcript isoform has a significant contribution to overall MALAT1 function and breast carcinogenesis.

Further studies are necessary to elucidate the genetic (or epigenetic) mechanisms responsible for the observed under-expression of Dsv-MALAT1, in breast cancer. It is unlikely that gene amplification is one of the mechanisms for MALAT1 overexpression because MALAT1 is located in a chromosomal region (11q13.1) non-recurrently amplified in breast cancer (Curtis et al, 2012; Zack et al, 2013). Mutations in the MALAT1 gene, recently discovered in human cancers, are rare in breast (1.1%) as compared with other cancer types such as bladder cancer (15.3%) (Kandath et al, 2013). MALAT1 epigenetic dysregulation mediated

Table 2. Relationship between Dsv-MALAT1 mRNA and target-gene expression					
Genes	Normal breast tissues (n ¼ 10)	Breast tumours with low level of Dsv-MALAT1 (n ¼ 20)	Breast tumours with high level of Dsv-MALAT1 (n ¼ 20)	P-value ^a	ROC-AUC
Cell cycle control (n ¼ 7)					
MKI67	1.0 (0.00–4.74) ^b	12.11 (1.88–25.62) ^b	13.01 (6.15–81.52) ^b	0.42 (NS)	0.575
AURKA	1.0 (0.17–2.73)	8.07 (2.79–27.11)	7.0 (4.28–123.17)	0.45 (NS)	0.570
FOXM1	1.0 (0.00–12.25)	10.91 (2.87–30.02)	13.74 (4.1–101.38)	0.40 (NS)	0.577
PCNA	1.0 (0.46–3.22)	2.72 (0.65–8.31)	3.59 (0.45–6.88)	0.24 (NS)	0.626
CCNE1	1.0 (0.00–6.82)	3.34 (0.90–26.50)	3.18 (1.14–52.07)	0.96 (NS)	0.505
E2F1	1.0 (0.44–3.08)	4.50 (1.00–8.60)	5.05 (1.76–51.10)	0.73 (NS)	0.532
AURKB	1.0 (0.00–16.86)	23.59 (2.90–37.43)	15.38 (6.77–160.48)	0.75 (NS)	0.47
Cell migration (n ¼ 5)					
RHOB	1.0 (0.48–5.75)	0.88 (0.19–1.36)	1.13 (0.36–65.33)	0.015	0.724
PLAU	1.0 (0.42–2.29)	1.76 (0.31–12.59)	2.86 (0.84–33.28)	0.017	0.72
MMP14	1.0 (0.69–1.52)	1.31 (0.05–6.21)	2.52 (0.30–39.17)	0.0087	0.743
RHOA	1.0 (0.42–3.34)	1.36 (0.22–3.85)	1.59 (0.22–6.39)	0.38 (NS)	0.581
MMP13	1.0 (0.00–4.69)	45.38 (1.33–688.61)	38.70 (2.62–250.99)	0.83 (NS)	0.48
Polycomb repressive complex 1 (n ¼ 3)					
CBX7	1.0 (0.39–1.62)	0.36 (0.05–0.95)	0.43 (0.15–2.43)	0.045	0.685
CBX4	1.0 (0.41–3.80)	1.46 (0.79–2.98)	1.55 (0.00–3.10)	0.91 (NS)	0.501
BMI1	1.0 (0.59–1.47)	1.42 (0.17–4.91)	1.98 (0.66–6.43)	0.014	0.728
Polycomb repressive complex 2 (n ¼ 5)					
SUZ12	1.0 (0.56–1.25)	1.19 (0.53–1.93)	1.63 (0.92–2.99)	0.00084	0.793
JARID2	1.0 (0.83–1.46)	1.19 (0.71–1.83)	1.89 (0.51–4.13)	0.0015	0.794
EED	1.0 (0.58–1.84)	1.02 (0.18–2.61)	1.36 (0.19–3.13)	0.0036	0.769
TUG1	1.0 (0.79–3.06)	0.73 (0.35–2.01)	1.16 (0.34–1.98)	0.044	0.686
EZH2	1.0 (0.48–2.32)	4.58 (1.36–12.70)	5.99 (2.76–14.64)	0.15 (NS)	0.632
EMT (n ¼ 7)					
CDH1	1.0 (0.33–1.79)	0.91 (0.06–2.74)	1.28 (0.23–5.55)	0.051 (NS)	0.68
VIM	1.0 (0.42–3.11)	0.22 (0.09–0.70)	0.31 (0.10–2.16)	0.22 (NS)	0.614
ZEB2	1.0 (0.44–5.58)	0.30 (0.12–0.64)	0.32 (0.14–2.35)	0.60 (NS)	0.549
ZEB1	1.0 (0.43–4.08)	0.31 (0.19–1.47)	0.44 (0.16–1.09)	0.34 (NS)	0.587
SNAI2	1.0 (0.43–1.50)	0.23 (0.05–0.82)	0.24 (0.11–1.41)	0.52 (NS)	0.56
TWIST1	1.0 (0.53–3.62)	0.30 (0.08–1.16)	0.32 (0.07–3.83)	0.63 (NS)	0.545
SNAI1	1.0 (0.34–7.88)	0.83 (0.06–2.66)	0.89 (0.20–6.03)	0.80 (NS)	0.524
Apoptosis (n ¼ 6)					
BIRC6	1.0 (0.74–3.10)	0.85 (0.47–1.23)	1.13 (0.52–1.75)	0.00017	0.841
BAX	1.0 (0.41–8.88)	1.38 (0.62–3.68)	1.88 (0.80–8.26)	0.02	0.715
BIRC2	1.0 (0.73–3.61)	0.62 (0.30–1.18)	0.82 (0.26–1.99)	0.044	0.686
BIRC4 ^c	0.0 (0.00–1.16)	1.09 (0.00–6.88)	2.16 (0.00–8.59)	0.12 (NS)	0.645
BCL2L1	1.0 (0.48–3.76)	1.19 (0.65–2.19)	1.47 (0.52–4.69)	0.22 (NS)	0.612
BCL2	1.0 (0.34–6.13)	0.85 (0.08–3.75)	0.78 (0.24–2.43)	0.82 (NS)	0.521
DNA repair (n ¼ 6)					
ATM	1.0 (0.67–1.66)	0.60 (0.40–1.87)	1.36 (0.55–2.24)	0.00007	0.868
MSH2	1.0 (0.66–1.55)	0.98 (0.60–1.57)	1.40 (0.83–2.95)	0.0003	0.834
BRCA1	1.0 (0.00–2.56)	2.23 (0.05–13.56)	3.33 (0.00–8.85)	0.11 (NS)	0.649
BRCA2	1.0 (0.27–2.49)	3.12 (0.86–7.82)	3.87 (0.00–9.57)	0.19 (NS)	0.631
RAD51	1.0 (0.0–2.00)	7.55 (1.62–18.9)	5.29 (1.35–21.6)	0.37 (NS)	0.417
XRCC1	1.0 (0.70–1.30)	0.78 (0.53–1.34)	1.41 (0.78–2.84)	0.00001	0.908
Regulators of MALAT1 (n ¼ 6)					
DGCR8	1.0 (0.70–2.59)	0.80 (0.40–1.71)	1.12 (0.80–5.05)	0.000083	0.864
AGO2	1.0 (0.57–4.50)	0.8 (0.36–2.64)	0.89 (0.33–5.87)	0.26 (NS)	0.604
SFRS1	1.0 (0.66–4.32)	1.13 (0.87–1.68)	1.51 (0.98–2.80)	0.00097	0.805
SFRS3	1.0 (0.65–3.93)	1.06 (0.69–1.87)	1.30 (0.78–3.05)	0.036	0.694
SFRS2	1.0 (0.60–2.76)	1.52 (0.88–2.24)	1.89 (0.81–4.04)	0.18 (NS)	0.624
UHMK1	1.0 (0.59–5.09)	1.48 (0.93–2.74)	2.04 (0.95–3.88)	0.0094	0.74
MALAT1-inducible genes (n ¼ 3)					
ROBO1	1.0 (0.54–1.54)	0.36 (0.13–4.14)	0.52 (0.14–1.33)	0.14 (NS)	0.635
MCAM	1.0 (0.57–33.8)	0.52 (0.22–0.81)	0.61 (0.22–10.9)	0.21 (NS)	0.616
IFI44	1.0 (0.34–9.38)	1.92 (0.06–29.23)	2.63 (0.11–30.15)	0.43 (NS)	0.572

Abbreviations: AUC ¼ area under the curve; NS ¼ not significant; ROC ¼ receiver operating curves. Values in bold correspond to significant P-values (P<0.05).

^aKruskal–Wallis H-test.

^bMedian (range) of gene mRNA levels; the mRNA values of the samples were normalised such that the median of the 10 normal breast tissues mRNA values was 1.

^cThe mRNA values of the samples were normalised such that a Ct value of 35 was 1.

Table 3. Relationship between levels of Dsv-MALAT1 and FL-MALAT1 and a panel of RTK/MAPK/PI3K proteins in a panel of 143 breast tumours

	Dsv-MALAT1		FL-MALAT1	
	r ^a	P-value ^a	r ^a	P-value ^a
RTK proteins (n/4 9)				
EGFR	0.068	NS	0.033	NS
p.EGFR.Thr669	0.008	NS	-0.047	NS
p.EGFR.Tyr1173	0.036	NS	0.052	NS
Her2/ErbB2	0.091	NS	-0.032	NS
Her3/ErbB3	-0.128	NS	0.000	NS
p.Her3/ErbB3.Tyr1289	-0.060	NS	-0.020	NS
p.Her4.Tyr1162	0.108	NS	0.018	NS
Met	-0.025	NS	-0.029	NS
p.Met.Tyr1234.1235	-0.064	NS	-0.026	NS
MAPK pathway proteins (n/4 4)				
MEK1/2 p.MEK1/2.Ser217/221	-0.029	NS	-0.073	NS
p44.42.MAPK	0.113	NS	-0.047	NS
p.p44.42.MAPK.Thr202.Tyr204	0.116	NS	0.150	NS
	0.127	NS	0.030	NS
PI3K/AKT pathway proteins (n/4 15)				
PTEN	0.101	NS	0.050	NS
p.PTEN.Ser380.Thr382.383	-0.018	NS	-0.026	NS
INPP4b Akt1	0.001	NS	0.042	NS
p.Akt1.Ser473	0.030	NS	-0.017	NS
Akt2 mTor	0.119	NS	-0.032	NS
p.mTor.Ser2448	-0.102	NS	0.028	NS
FOXO1	-0.155	NS	-0.066	NS
TSC2 p70.S6.Kinase	-0.077	NS	0.008	NS
p.p70.S6.Kinase.Thr389	-0.194	0.0210	-0.196	0.0180
S6.Ribosomal.protein	-0.075	NS	0.003	NS
p.S6.Ribosomal.protein.	-0.172	0.0400	0.099	NS
Ser235.236	-0.225	0.0070	0.073	NS
p.S6.Ribosomal.protein.	-0.034	NS	0.075	NS
Ser240.244	-0.095	NS	0.006	NS
	-0.201	0.0160	-0.024	NS

Abbreviations: Akt ¼ protein kinase B, also known as Akt; EGFR ¼ epidermal growth factor receptor; MAPK ¼ mitogen-activated protein kinase; mTor ¼ mechanistic target of rapamycin; NS ¼ not significant; PI3K ¼ phosphoinositide 3-kinase; PTEN ¼ phosphatase and tensin homologue.
^aSpearman's rank correlation test.

et al, 2009). We observed a positive link between Dsv-MALAT1 overexpression and several interactors of MALAT1, in particular SFRS1 and SFRS3, confirming the involvement of MALAT1 in the regulation of alternative splicing of pre-mRNA in nuclear speckle domains (Tripathi et al, 2010). We also identified a link between Dsv-MALAT1 overexpression and several genes involved in DNA repair (ATM, MSH2, XRCC1). Only one study has recently suggested that MALAT1 depletion could dysregulate ATM-CHK2 pathway in oesophageal squamous cell carcinoma (Hu et al, 2015). More interesting, we observed a link between Dsv-MALAT1 expression and the major members (except EZH2) of Polycomb repressive complex PRC2, including SUZ12, EED and JARID2, but no (or little) link with the members of the Polycomb repressive complex PRC1 (i.e., CBX4, CBX7 and BMI1), as well as the genes regulated by this complex: PCNA and CCNE1 (Yang et al, 2011). In this regard, Yang et al (2011) showed a major role for MALAT1 in the relocation of transcription units by the PRC2 complex in the three-dimensional space of the nucleus, to coordinate the gene-expression programs.

In conclusion, this study suggests that the lncRNA MALAT1 (as the well-known lncRNA HOTAIR) is involved in breast cancer. These data revealed a complex expression pattern of various MALAT1 transcript variants, and suggest that this pattern of expression should be taken into account when evaluating antitumoural drugs designed to target this lncRNA. Further studies are also necessary to elucidate roles of these different MALAT1 transcript variants in breast tumourigenesis and their genetic (or epigenetic) dysregulation molecular mechanisms in this cancer.

ACKNOWLEDGEMENTS

We thank the staff of Institut Curie–Hospital René Huguenin for their assistance in specimen collection and patient care. We thank Aurélie Barbet, Floriane Bard and Caroline Lecerf for performing RPPA experiments. The RPPA platform is supported by Cancéro-pôle Ile-de-France. This work was supported by the Comité départemental des Hauts-de-Seine de la Ligue Nationale Contre le Cancer, the Association pour la recherche en cancérologie de Saint-Cloud (ARCS), and by Grant INCa-DGOS-4654.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Bieche I, Onody P, Laurendeau I, Olivi M, Vidaud D, Lidereau R, Vidaud M (1999) Real-time reverse transcription-PCR assay for future management of ERBB2-based clinical applications. *Clin Chem* 45: 1148–1156.
 Bieche I, Parfait B, Laurendeau I, Girault I, Vidaud M, Lidereau R (2001) Quantification of estrogen receptor alpha and beta expression in sporadic breast cancer. *Oncogene* 20: 8109–8115.
 Cheetham SW, Gruhl F, Mattick JS, Dinger ME (2013) Long noncoding RNAs and the genetics of cancer. *Br J Cancer* 108: 2419–2425.
 Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, Gräf S, Ha G, Haffari G, Bashashati A, Russell R, McKinney S, METABRIC Group, Langerød A, Green A, Provenzano E, Wishart G, Pinder S, Watson P, Markowitz F, Murphy L, Ellis I, Purushotham A, Børresen-Dale AL, Brenton JD, Tavaré S, Caldas C, Aparicio S (2012) The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 486: 346–352.
 Davis JJ, Hsi BL, Arroyo JD, Vargas SO, Yeh YA, Motyckova G, Valencia P, Perez-Atayde AR, Argani P, Ladanyi M, Fletcher JA, Fisher DE (2003) Cloning of an Alpha-TFEB fusion in renal tumors harboring the

by CpG island methylation is not reported till date. In the present study, post-transcriptional regulation of MALAT1 by Hsa-miR-125b as described in bladder cancer (Han et al, 2013) was not observed in our breast tumour series. Conversely, our data suggested a transcriptional regulation of the Dsv-MALAT1 (but not of the FL-MALAT1) by the transcriptional co-activator YAP, as described in liver cancer (Wang et al, 2014). Further studies are necessary, using functional assays, to confirm this association. In the downstream MALAT1 pathway, our results suggested that Dsv-MALAT1 (but not FL-MALAT1) could activate the PI3K/AKT pathway, in partial agreement with previous data (Wu et al, 2014; Dong et al, 2015; Xu et al, 2015).

We also assessed the expression levels of gene panel putatively involved in various cellular and molecular phenomena associated with carcinogenesis via dysregulation of Dsv-MALAT1. These genes encode proteins involved in cell cycle control, cell migration, polycomb repressive complexes (PRC1 and 2), EMT, apoptosis and DNA repair, as well as regulators and interactors of Dsv-MALAT1, or known MALAT1-induced genes (Miyagawa et al, 2012; Gutschner et al, 2013b). We identified a strong positive link between Dsv-MALAT1 overexpression and DGCR8 expression, suggesting that Drosha-DGCR8 complex (Microprocessor) controlled the abundance of Dsv-MALAT1 in breast cancer as in HEK 293T cells (Macias et al, 2012). No such link was observed with Ago2, a second putative major regulator of MALAT1 (Weinmann

- (6;11)(p21;q13) chromosome translocation. *Proc Natl Acad Sci USA* 100: 6051–6056.
- Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, Lagarde J, Veeravalli L, Ruan X, Ruan Y, Lassmann T, Carninci P, Brown JB, Lipovich L, Gonzalez JM, Thomas M, Davis CA, Shiekhatah R, Gingeras TR, Hubbard TJ, Notredame C, Harrow J, Guigó R (2012) The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res* 22: 1775–1789.
- Dong Y, Liang G, Yuan B, Yang C, Gao R, Zhou X (2015) MALAT1 promotes the proliferation and metastasis of osteosarcoma cells by activating the PI3K/Akt pathway. *Tumour Biol* 36: 1477–1486.
- Ellis MJ, Ding L, Shen D, Luo J, Suman VJ, Wallis JW, Van Tine BA, Hoog J, Goiffon RJ, Goldstein TC, Ng S, Lin L, Crowder R, Snider J, Ballman K, Weber J, Chen K, Koboldt DC, Kandoth C, Schierding WS, McMichael JF, Miller CA, Lu C, Harris CC, McLellan MD, Wendl MC, DeSchryver K, Allred DC, Esserman L, Unzeitig G, Margenthaler J, Babiera GV, Marcom PK, Guenther JM, Leitch M, Hunt K, Olson J, Tao Y, Maher CA, Fulton LL, Fulton RS, Harrison M, Oberkfell B, Du F, Demeter R, Vickery TL, Elhammali A, Piwnica-Worms H, McDonald S, Watson M, Dooling DJ, Ota D, Chang LW, Bose R, Ley TJ, Piwnica-Worms D, Stuart JM, Wilson RK, Mardis ER (2012) Whole-genome analysis informs breast cancer response to aromatase inhibition. *Nature* 486: 353–360.
- Gibb EA, Vucic EA, Enfield KS, Stewart GL, Lonergan KM, Kennett JY, Becker-Santos DD, MacAulay CE, Lam S, Brown CJ, Lam WL (2011) Human cancer long non-coding RNA transcriptomes. *PLoS One* 6: e25915.
- Guffanti A, Iacono M, Pelucchi P, Kim N, Soldà G, Croft LJ, Taft RJ, Rizzi E, Askarian-Amiri M, Bonnal RJ, Callari M, Mignone F, Pesole G, Bertalot G, Bernardi LR, Albertini A, Lee C, Mattick JS, Zucchi I, De Bellis G (2009) A transcriptional sketch of a primary human breast cancer by 454 deep sequencing. *BMC Genomics* 10: 163.
- Guo F, Li Y, Liu Y, Wang J, Li Y, 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* 42: 224–229.
- Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, Tsai MC, Hung T, Argani P, Rinn JL, Wang Y, Brzoska P, Kong B, Li R, West RB, van de Vijver MJ, Sukumar S, Chang HY (2010) Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 464: 1071–1076.
- Gutschner T, Hämmerle M, Diederichs S (2013a) MALAT1—a paradigm for long noncoding RNA function in cancer. *J Mol Med* 91: 791–801.
- Gutschner T, Hammerle M, Eissmann M, Hsu J, Kim Y, Hung G, Revenko A, Arun G, Stenrup M, Gross M, Zörnig M, MacLeod AR, Spector DL, Diederichs S (2013b) The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. *Cancer Res* 73: 1180–1189.
- Han Y, Liu Y, Zhang H, Wang T, Diao R, Jiang Z, Gui Y, Cai Z (2013) Hsa-miR-125b suppresses bladder cancer development by down-regulating oncogene SIRT7 and oncogenic long noncoding RNA MALAT1. *FEBS Lett* 587: 3875–3882.
- Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143: 29–36.
- Hu L, Wu Y, Tan D, Meng H, Wang K, Bai Y, Yang K (2015) Up-regulation of long noncoding RNA MALAT1 contributes to proliferation and metastasis in esophageal squamous cell carcinoma. *J Exp Clin Cancer Res* 34: 7–18.
- Ji P, Diederichs S, Wang W, Boing S, Metzger R, Schneider PM, Tidow N, Brandt B, Buerger H, Bulk E, Thomas M, Berdel WE, Serve H, Müller-Tidow C (2003) 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.
- Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, Xie M, Zhang Q, McMichael JF, Wyczalkowsky MA, Leiserson MD, Miller CA, Welch JS, Walter MJ, Wendl MC, Ley TJ, Wilson RK, Raphael BJ, Ding L (2013) Mutational landscape and significance across 12 major cancer types. *Nature* 502: 333–339.
- Lai MC, Yang Z, Zhou L, Zhu QQ, Xie HY, Zhang F, Wu LM, Chen LM, Zheng SS (2012) Long non-coding RNA MALAT-1 overexpression predicts tumor recurrence of hepatocellular carcinoma after liver transplantation. *Med Oncol* 29: 1810–1816.
- Macias S, Plass M, Stajuda A, Michlewski G, Eyraes E, Cáceres JF (2012) DGCR8 HITS-CLIP reveals novel functions for the Microprocessor. *Nat Struct Mol Biol* 19: 760–766.
- Mathews J, Duncavage EJ, Pfeifer JD (2013) Characterization of translocations in mesenchymal hamartoma and undifferentiated embryonal sarcoma of the liver. *Exp Mol Pathol* 95: 319–324.
- Miyagawa R, Tano K, Mizuno R, Nakamura Y, Ijiri K, Rakwal R, Shibato J, Masuo Y, Mayeda A, Hirose T, Akimitsu N (2012) Identification of cis- and trans-acting factors involved in the localization of MALAT-1 noncoding RNA to nuclear speckles. *RNA* 18: 738–751.
- Schmidt LH, Spieker T, Koschmieder S, Schäffers S, Humberg J, Jungen D, Bulk E, Hascher A, Wittmer D, Marra A, Hillejan L, Wiebe K, Berdel WE, Wiewrodt R, Müller-Tidow C (2011) The long noncoding MALAT-1 RNA indicates a poor prognosis in non-small cell lung cancer and induces migration and tumor growth. *J Thorac Oncol* 6: 1984–1992.
- Spizzo R, Almeida MI, Colombatti A, Calin GA (2012) Long non-coding RNAs and cancer: a new frontier of translational research? *Oncogene* 31: 4577–4587.
- Tripathi V, Ellis JD, Shen Z, Song DY, Pan Q, Watt AT, Freier SM, Bennett CF, Sharma A, Bubulya PA, Blencowe BJ, Prasanth SG, Prasanth KV (2010) The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol Cell* 39: 925–938.
- Troncale S, Barbet A, Coulibaly L, Henry E, He B, Barillot E, Dubois T, Hupé P, de Koning L (2012) NormaCurve: a SuperCurve-based method that simultaneously quantifies and normalizes reverse phase protein array data. *PLoS One* 7: e36886.
- Wang J, Wang H, Zhang Y, Zhen N, Zhang L, Qiao Y, Weng W, Liu X, Ma L, Xiao W, Yu W, Chu Q, Pan Q, Sun F (2014) Mutual inhibition between YAP and SRSF1 maintains long non-coding RNA, Malat1-induced tumorigenesis in liver cancer. *Cell Signal* 26: 1048–1059.
- Weinmann L, Höck J, Ivacevic T, Ohrt T, Mütze J, Schwille P, Kremmer E, Benes V, Urlaub H, Meister G (2009) Importin 8 is a gene silencing factor that targets argonaute proteins to distinct mRNAs. *Cell* 136: 496–507.
- Wu XS, Wang XA, Wu WG, Hu YP, Li ML, Ding Q, Weng H, Shu YJ, Liu TY, Jiang L, Cao Y, Bao RF, Mu JS, Tan ZJ, Tao F, Liu YB (2014) MALAT1 promotes the proliferation and metastasis of gallbladder cancer cells by activating the ERK/MAPK pathway. *Cancer Biol Ther* 15: 806–814.
- Xu S, Sui S, Zhang J, Bai N, Shi Q, Zhang G, Gao S, You Z, Zhan C, Liu F, Pang D (2015) Downregulation of long noncoding RNA MALAT1 induces epithelial-to-mesenchymal transition via the PI3K-AKT pathway in breast cancer. *Int J Clin Exp Pathol* 8: 4881–4891.
- Yang L, Lin C, Liu W, Zhang J, Ohgi KA, Grinstein JD, Dorrestein PC, Rosenfeld MG (2011) ncRNA- and Pc2 methylation-dependent gene relocation between nuclear structures mediates gene activation programs. *Cell* 147: 773–78838.
- Zack TI, Schumacher SE, Carter SL, Cherniack AD, Saksena G, Tabak B, Lawrence MS, Zhsng CZ, Wala J, Mermel CH, Sougnez C, Gabriel SB, Hernandez B, Shen H, Laird PW, Getz G, Meyerson M, Beroukhi R (2013) Pan-cancer patterns of somatic copy number alteration. *Nat Genet* 45: 1134–1140.
- Zhang J, Zhang B, Wang T, Wang H (2015) LncRNA MALAT1 overexpression is an unfavorable prognostic factor in human cancer: evidence from a meta-analysis. *Int J Clin Exp Med* 8: 5499–5505.

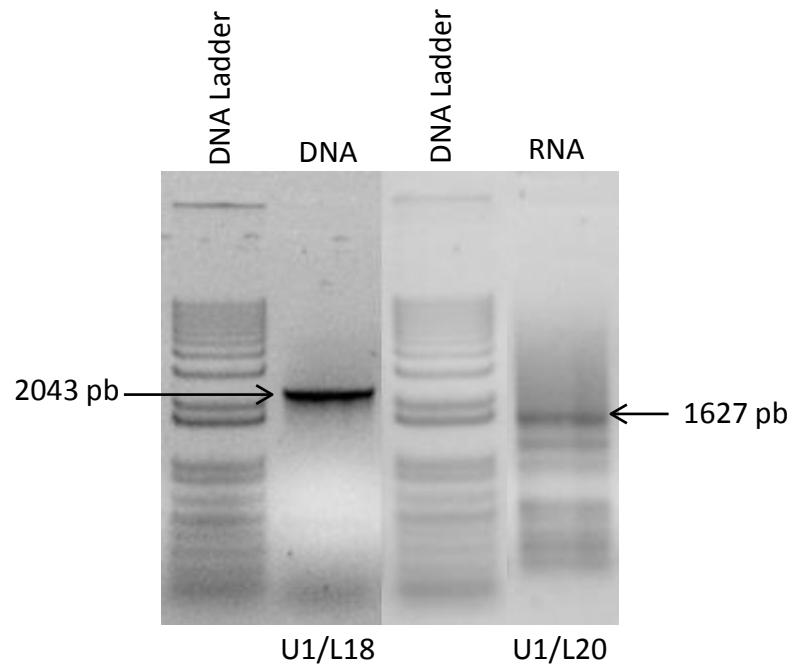


This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

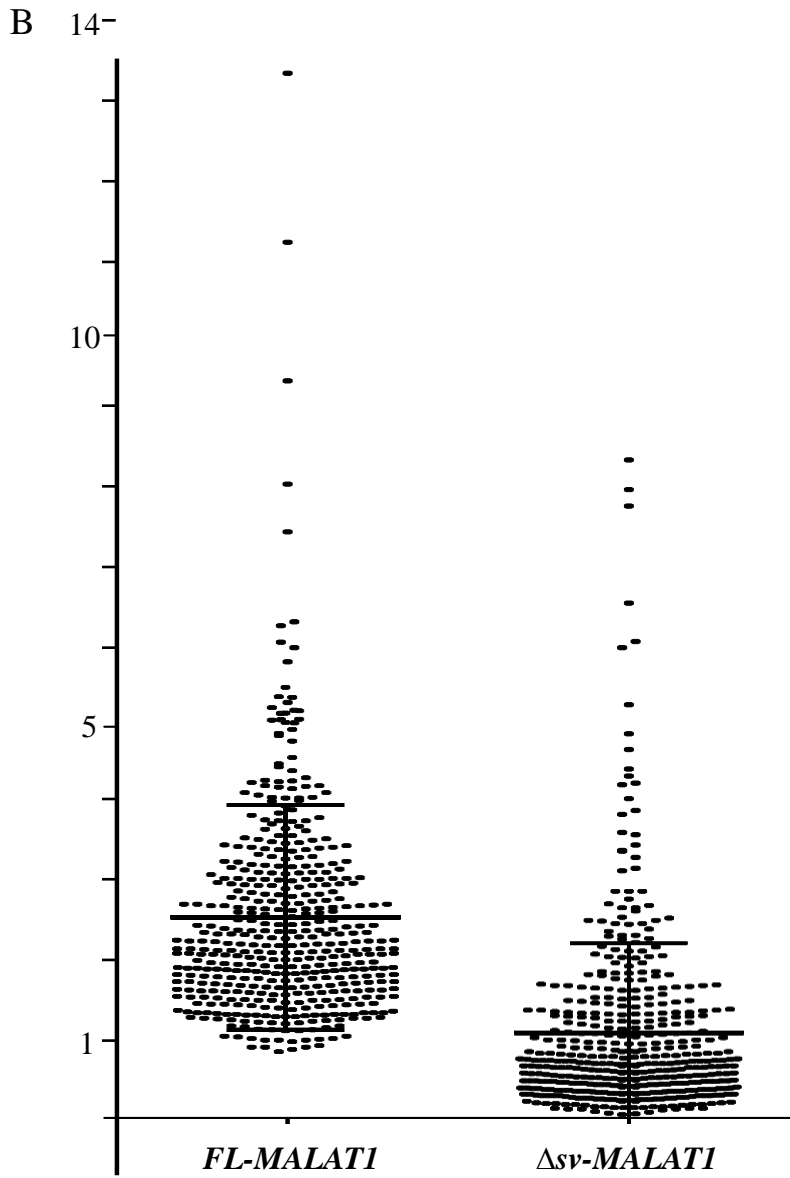
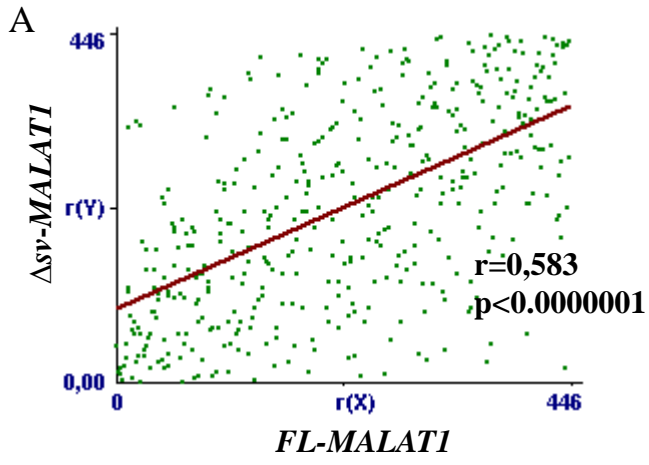
Supplementary Information accompanies this paper on British Journal of Cancer website (<http://www.nature.com/bjc>)

Supplemental figure 1: Visualization of *FL-MALAT1* (on genomic DNA) and Δ sv-*MALAT1* (on total RNA) RT-PCR products by agarose gel electrophoresis. Using primer pair U1/L20 with one of the two primers (U20) placed at the junction of the 119-pb spliced region, we observed only a unique RT-PCR product showing both 119-bp and 243-bp deletions (1627-pb), but not the RT-PCR product only showing the 119-bp deletion (expected size of 1870-pb =1627-pb + 243-pb).

Supplemental Figure 2: (A) Spearman rank correlation test between *FL-MALAT1* and Δ sv-*MALAT1*. (B) Scatter dot plot of qRT-PCR data for *FL-MALAT1* and Δ sv-*MALAT1* in the series of 446 breast cancer.



Supplemental figure 1 : Meseure et al.



Supplemental Table 1: Pathological and clinical characteristics of patients in relation to metastasis free survival (MFS)

	Number of patients (%)	Number with metastases (%)	MFS p-value ^a
Total	446 (100.0)	176 (39.5)	
Age			
≤50	94 (21.1)	37 (39.4)	0.96 (NS)
>50	352 (78.9)	139 (39.5)	
SBR histological grade ^{b, c}			
I	57 (13.0)	11 (19.3)	0.00073
II + III	380 (87.0)	161 (42.4)	
Lymph node status ^d			
0	117 (26.3)	36 (30.8)	0.00000036
1-3	231 (51.9)	80 (34.6)	
>3	97 (21.8)	60 (61.9)	
Macroscopic tumor size ^e			
≤25mm	218 (49.8)	66 (30.3)	0.0000095
>25mm	220 (50.2)	109 (49.5)	
ER α status			
Negative	115 (25.8)	47 (40.9)	0.18 (NS)
Positive	331 (74.2)	129 (39.0)	
PR status			
Negative	191 (42.8)	83 (43.5)	0.028
Positive	255 (57.2)	93 (36.5)	
ERBB2 status			
Negative	353 (79.1)	137 (38.8)	0.45 (NS)
Positive	93 (20.9)	39 (41.9)	
Molecular subtypes			
HR- ERBB2-	68 (15.2)	26 (38.2)	0.20 (NS)
HR- ERBB2+	42 (9.4)	20 (47.6)	
HR+ ERBB2-	285 (63.9)	111 (38.9)	
HR+ ERBB2+	51 (11.4)	19 (37.3)	
PIK3CA mutation status			
wild type	299 (67.0)	124 (41.5)	0.12 (NS)
mutated	147 (33.0)	52 (35.4)	

Abbreviations: ER α : oestrogen receptor alpha; PR: progesterone receptor; ERBB2: human epidermal growth factor receptor 2; HR: hormone receptor.

^a Log-rank test : the bold values are statistically significant (p-value<0.05). NS: not significant. ^b Scarff Bloom Richardson classification. ^c Information available for 437 patients. ^d Information available for 445 patients. ^e Information available for 438 patients.

Supplemental Table 2 : Primer sequences used for RT-PCR.

Genes	Primer sequences (5' to 3')	
	Upper	Lower
fl-MALAT1 U13/L13	CATCTTCTGAGTCATAACCAGCCT	TGCCTTTACTTATCAATTCACCAA
fl-MALAT1 U9/L9	ACTTCACAGAGAATGCAGTTGTCTTG	AATAGCAGCGGGATCAGAACAGT
Δsv-MALAT1 U18/L18	GGCTCTTCTTCTGTTCTAACTTCA	GAACAGTACTGCATTTACTTGCCAA
Δsv-MALAT1 U2/L2	TGGTAGCTTTTGTATTATCAAACCTT	CTGCCAGGCTGGTTATGACT
Δsv-MALAT1 U1/L20	TGAAGGGATTTCTTCTGATGGT	AACTGCATTCTCTGTGAAGTTAGAACA
HOTAIR	GCTCTCAATCAGAAAGGTCTGCT	CAGGCTTCTAAATCCGTTCCATTC
ANRIL	CCGGGGAGATCTATTTGGAATGTAT	TGGTGAAACCCCGTCTCTACTGTTA
Cell cycle control (n=7)		
MKI67	ATTGAACCTGCGGAAGAGCTGA	GGAGCGCAGGGATATTCCCTTA
AURKA	GCATTTTCAGGACCTGTTAAGGCTA	TGCTGAGTCACGAGAACACGTTT
FOXM1	GGGAGACCTGTGCAGATGGTGA	TCGAAGCCACTGGATGTTGGAT
PCNA	TCGATAAAGAGGAGGAAAGCTGTTAC	GCAGACATACTGAGTGTACCCGTT
CCNE1	GGGCGTCGCTGATGAAGATG	CTTTGCTCGGGCTTTGTCCA
E2F1	CCGTGGACTCTTCGGAGAACTT	GATCCACCTACGGTCTCCTCA
AURKB	TCCAATGTCCAGCCCACAGC	ACGCCAATCTCAAAGTCATCAAT
Cell migration (n=5)		
RHOB	CTACCCGGACACCGACGTCA	CGGGGATGTTCTCCAGCGA
PLAU	AAAATGCTGTGTGCTGCTGACC	GCCTTGGAGGGAACAGACGA
MMP14	CAACATTGGAGGAGACCCACT	CCAGGAAGATGTCATTTCCATTCA
RHOA	ATCGACAGCCCTGATAGTTTAGAAA	CTTCTTATCCCAACCAGGATGAT
MMP13	AGGAAACCAGGTCTGGAGATATGAT	CCAATTCCTGGGAAGTCTTCTTCT
Polycomb repressive complex 1 (n=3)		
CBX7	CCCCCTGAAGAGGAGGCAGAT	GTGATGGAGTTGGCGGTGATGT
CBX4	TCTGGTGAAATGGAGAGGCTGGT	TGCTCCTGCCGTTCCCTGTT
BMI1	GCAGCAATGACTGTGATGCACTT	TCCTCATACATGACATCAATCTGGAA
Polycomb repressive complex 2 (n=5)		
SUZ12	TCGCAACGGACCAGTTAAGAGAA	TGTTCTGTTTTGGCCTGCACA
JARID2	GTACCGCTACGATGAGGAACAGA	GCTGCCGTTTTTACCAGACACT
EED	TTATGGAATATCCAGACGGACACTC	CACCCAAAAGATCATAATCAGCACT
TUG1	TTCCAGAAAAGGCTTCTCTTGATACA	ACAGTTTTCACTCAAAGGGCTTCAT
EZH2	AACTTCGAGCTCCTCTGAAGCAA	AGCACCCTCCACTCCACATTCT
EMT (n=7)		
CDH1	CGCATTGCCACATACACTCTCTT	TCGGGCTTGTGTCATTCTGAT
VIM	CTCCCTCTGGTTGATACCCACTC	AGAAGTTTCGTTGATAACCTGTCCA
ZEB2	GCCTCTGTAGATGGTCCAGAAGAA	TGTTAATTGCCGTCTGGATCGT
ZEB1	ACAGTGTACCAGGGAGGAGCAGT	TTTCTTCCCTTCTTTCTGTCAAT
SNAI2	GCGAACTGGACACACATACAGTGA	GCAGCGGTAGTCCACACAGTGA
TWIST1	CGCCCCGCTCTTCTCCTCT	TGGACACGTCTGTCATCATCTCT
SNAI1	GCTGCAGGACTCTAATCCAGAGTT	GACAGAGTCCCAGATGAGCATTG
Apoptosis (n=6)		
BIRC6	GACAAACATGCAGCTTAATTCTGGT	TGAGCAGTACTCTCCTGAGTTCCAA
BAX	GTGCCTCAGGATGCGTCCA	ACGGCGGCAATCATCTCTCT
BIRC2	CCAACAGAAGATGTTTCAGGTCTGT	CGTTCTTCTTGCAACCTCCTCA
BIRC4	CTCATTCACTTGAGGAGTGTCTGGT	GGATTTTGGAAAGATGGTATCATCAA
BCL2L1	TGAATGACCACCTAGAGCCTTGGAA	GGAAACCAGCGGTTGAAGCGT
BCL2	ATTTTCATATTTTGTGCGGAGTTCA	CTTCTTTACAAAGCCATTTTCCCA
DNA repair (n=6)		
ATM	CCAGCTGTGCAGCGAACAAT	TCTAAGCACGTTTCTGCTAACCAGT
MSH2	GCCCTTAACCTTTTTTCAGGGTTCT	TTGAGGGGTTTTTACACTTATTCAGCA
BRCA1	AAGGGTCAACAAAAGAAATGTCCA	GTGATGTGGTGTGTTTCTGGCAA
BRCA2	GGCAGCTATGGAATGTGCCT	AGCCGATCTTCTGCTTCTATCAAT
RAD51	TGCTGGCAGTGGCTGAGAGGTAT	AGGAGCTGGGTCTGGTGGTCTGT
XRCC1	AGCAGCCAGGACTCGACTCACT	TATCTGCTCCTCTTCTCCAAGTGA
Regulators of MALAT1 (n=6)		
DGCR8	CATGTCAAGAAGTGGGGTCTTTACT	CACTCTTGTCCGATGTCTCCTGCT
AGO2	AAGGCCAGTTCAGCAGGTTCT	CCCGGGCTGGTAGTCTTTTTTCT
SFRS1	GCACTGGTGTGCTGGAGTTTGTGA	GGATGTAGGCAGTTTCTCCCTCA
SFRS3	GGCTACTATGGACCACTCCGAA	CAGCCACATAGTGTCTTCCATCT
SFRS2	AGTCTCCCCAGTGTCCAA	CACTGCTTGGCAGTACATCATTT
UHMK1	AATCCTGGCAGAGGACAAGTCTT	GTAGAATGTAGCCACAACAACTTCC
MALAT1 inducible genes (n=3)		
ROBO1	AGCTGCCCAAATCCAGATATGAA	GTATGAACCCATGTCACCAGCTGT
MCAM	GCCTGGACTTGGACACCATGATA	ACTCGGACGTCAGACACATAGTTCCAC
IFI44	AAAATCTTGGACTTGCTCAAAATTGTA	CTTTCATCCAGTGAATCTTCGCA

Supplemental Table 3: Relationship between MALAT1 transcript levels and classical clinical biological parameters in a series of 446 breast cancer

	Number of patients (%)				p-value ^a
	Total population (%)	MALAT1 under expression	MALAT1 normal expression	MALAT1 over expression	
Total	446 (100.0)	13 (2.9)	370 (83.0)	63 (14.1)	
Age					
≤50	94 (21.1)	4 (4.3)	79 (84.0)	11 (11.7)	0.54 (NS)
>50	352 (78.9)	9 (2.6)	291 (82.7)	52 (14.8)	
SBR histological grade ^{b, c}					
I	57 (13.0)	2 (3.5)	49 (86.0)	6 (10.5)	0.74 (NS)
II + III	380 (87.0)	11 (2.9)	315 (82.9)	54 (14.2)	
Lymph node status ^d					
0	117 (26.3)	4 (3.4)	92 (78.6)	21 (17.9)	0.58 (NS)
1-3	231 (51.9)	5 (2.2)	198 (85.7)	28 (12.1)	
>3	97 (21.8)	3 (3.1)	80 (82.5)	14 (14.4)	
Macroscopic tumor size ^e					
≤25mm	218 (49.8)	4 (1.8)	179 (82.1)	35 (16.1)	0.25 (NS)
>25mm	220 (50.2)	8 (3.6)	186 (84.5)	26 (11.8)	
ER α status					
Negative	115 (25.8)	9 (7.8)	101 (87.8)	5 (4.3)	0.000015
Positive	331 (74.2)	4 (1.2)	269 (81.3)	58 (17.5)	
PR status					
Negative	191 (42.8)	11 (5.8)	162 (84.8)	18 (9.4)	0.00079
Positive	255 (57.2)	2 (0.8)	208 (81.6)	45 (17.6)	
ERBB2 status					
Negative	353 (79.1)	13 (3.7)	290 (82.2)	50 (14.2)	0.17 (NS)
Positive	93 (20.9)	0 (0)	80 (86.0)	13 (14.0)	
Molecular subtypes					
HR- ERBB2-	68 (15.2)	9 (13.2)	57 (83.8)	2 (2.9)	0.00000075
HR- ERBB2+	42 (9.4)	0 (0)	39 (92.9)	3 (7.1)	
HR+ ERBB2-	285 (63.9)	4 (1.4)	233 (81.8)	48 (16.8)	
HR+ ERBB2+	51 (11.4)	0 (0)	41 (80.4)	10 (19.6)	
PIK3CA mutation status					
wild type	299 (67.0)	9 (3.0)	250 (83.6)	40 (13.4)	0.81 (NS)
mutated	147 (33.0)	4 (2.7)	120 (81.6)	23 (15.6)	
MKI67 mRNA expression					
median	12.5 (0.80-117)	13.7 (6.90-40.6)	12.4 (0.80-117)	12.2 (0.98-81.5)	0.66 (NS) ^f
EGFR mRNA expression					
median	0.22 (0.00-106)	0.52 (0.07-3.20)	0.22 (0.00-106)	0.20 (0.03-7.56)	0.27 (NS) ^f

Abbreviations: ER α : oestrogen receptor alpha; PR: progesterone receptor; ERBB2: human epidermal growth factor receptor 2; HR: hormone receptor.

The bold values are statistically significant (p-value<0.05). NS: not significant. ^a Chi-squared test. ^b Scarff Bloom Richardson classification. ^c Information available for 437 patients. ^d Information available for 445 patients. ^e Information available for 438 patients. ^f Kruskal Wallis's H Test.

Supplemental Table 4 : Δ sv-MALAT1 mRNA expression level in a series of 21 mammary cell lines.

Cell line	Status	Molecular subtype	Δ sv-MALAT1 mRNA expression level ^a
HMEC	N	-	2906
hTert HME1	N	-	1722
MCF-12A	N	-	1207
MCF-12F	N	-	2551
BT20	T	HR-ERBB2-	383
BT549	T	HR-ERBB2-	1032
HBL100	T	HR-ERBB2-	852
MDA-MB-157	T	HR-ERBB2-	114
MDA-MB-231	T	HR-ERBB2-	1265
MDA-MB-436	T	HR-ERBB2-	1015
MDA-MB-453	T	HR-ERBB2-	1076
MDA-MB-468	T	HR-ERBB2-	301
HCC-1569	T	ERBB2+	1637
HCC-1954	T	ERBB2+	1005
HCC-202	T	ERBB2+	1385
SKBR3	T	ERBB2+	4039
BT474	T	HR+	2004
BT483	T	HR+	1433
MCF7	T	HR+	1933
T47D	T	HR+	874
ZR75-1	T	HR+	136
Non-cancerous breast epithelial cell lines (n=4)			2136 (1207-2906) ^b
Total breast cancer cell lines (n=17)			1023 (114-4039) ^b
HR-ERBB2- breast cancer cell lines (n=8)			933 (114-1265) ^b
ERBB2+ breast cancer cell lines (n=4)			1511 (1005-4039) ^b
HR+ breast cancer cell lines (n=5)			1433 (136-2004) ^b

Abbreviations: HR: hormone receptor; ERBB2: human epidermal growth factor receptor 2;

N: non-cancerous cell line (n=4); T: breast cancer cell line (n=17).

^a mRNA expression relative to Ct35=1

^b Median (range) of genes mRNA levels

Supplemental Table 5: Relationship between Δ sv-MALAT1 spliced transcripts levels (using U2/L2 primers) and classical clinical biological parameters in a series of 446 breast cancer.

	Number of patients (%)				p-value ^a
	Total population (%)	Δ sv-MALAT1 under-expression	Δ sv-MALAT1 normal expression	Δ sv-MALAT1 over expression	
Total	446 (100.0)	99 (22.2)	306 (68.6)	41 (9.2)	
Age					
≤ 50	94 (21.1)	30 (31.9)	63 (67.0)	1 (1.1)	0.0012
> 50	352 (78.9)	69 (19.6)	243 (69.0)	40 (11.4)	
SBR histological grade ^{b, c}					
I	57 (13.0)	5 (8.8)	46 (80.7)	6 (10.5)	0.029
II + III	380 (87.0)	93 (24.5)	254 (66.8)	33 (8.7)	
Lymph node status ^d					
0	117 (26.3)	21 (17.9)	81 (69.2)	15 (12.8)	0.20 (NS)
1-3	231 (51.9)	49 (21.2)	164 (71.0)	18 (7.8)	
> 3	97 (21.8)	28 (28.9)	61 (62.9)	8 (8.2)	
Macroscopic tumor size ^e					
≤ 25 mm	218 (49.8)	38 (17.4)	159 (72.9)	21 (9.6)	0.044
> 25 mm	220 (50.2)	60 (27.3)	143 (65.0)	17 (7.7)	
ER α status					
Negative	115 (25.8)	40 (34.8)	68 (59.1)	7 (6.1)	0.00077
Positive	331 (74.2)	59 (17.8)	238 (71.9)	34 (10.3)	
PR status					
Negative	191 (42.8)	59 (30.9)	121 (63.4)	11 (5.8)	0.00027
Positive	255 (57.2)	40 (15.7)	185 (72.5)	30 (11.8)	
ERBB2 status					
Negative	353 (79.1)	75 (21.2)	244 (69.1)	34 (9.6)	0.58 (NS)
Positive	93 (20.9)	24 (25.8)	62 (66.7)	7 (7.5)	
Molecular subtypes					
RH- ERBB2-	68 (15.2)	25 (36.8)	40 (58.8)	3 (4.4)	0.0082
RH- ERBB2+	42 (9.4)	14 (33.3)	24 (57.1)	4 (9.5)	
RH+ ERBB2-	285 (63.9)	50 (17.5)	204 (71.6)	31 (10.9)	
RH+ ERBB2+	51 (11.4)	10 (19.6)	38 (74.5)	3 (5.9)	
PIK3CA mutation status					
wild type	299 (67.0)	76 (25.4)	198 (66.2)	25 (8.4)	0.057 (NS)
mutated	147 (33.0)	23 (15.6)	108 (73.5)	16 (10.9)	
MKI67 mRNA expression					
median	12.5 (0.80-117)	16.6 (2.09-75.1)	11.6 (0.80-117)	11.3 (0.85-62.6)	0.00018 ^f
EGFR mRNA expression					
median	0.22 (0.00-106)	0.15 (0.02-3.20)	0.24 (0.00-106)	0.22 (0.04-2.56)	0.0044 ^f

Abbreviations: ER α : oestrogen receptor alpha; PR: progesterone receptor; ERBB2: human epidermal growth factor receptor 2; HR: hormone receptor.

The bold values are statistically significant (p-value<0.05). NS: not significant. ^a Chi-squared test. ^b Scarff Bloom Richardson classification. ^c Information available for 437 patients. ^d Information available for 445 patients. ^e Information available for 438 patients. ^f Kruskal Wallis's H Test.

Supplemental Table 6 : Multivariable COX analysis of MFS in the series of 446 breast tumors.

Characteristics		HR ^a	95% CI ^b	p-value ^c
Lymph node status	0	1		0.00018
	1-3	1.54	1.23-1.93	
	>3	2.37	1.51-3.73	
Macroscopic tumor size	<=25mm	1		0.00078
	>25mm	1.67	1.24-2.25	
Δ sv-MALAT1 expression level	Low	1		0.0011
	High	0.64	0.47-0.87	

^a Hazard ratio.

^b 95% Confidential Interval.

^c Multivariate COX analysis. 0

RESULTATS COMPLEMENTAIRES : Etude préliminaire de l'épitranscriptome non codant des carcinomes mammaires infiltrants.

Cette étude préliminaire est le résultat d'un travail collaboratif au sein de l'Institut Curie entre l'Unité CNRS UMR 3244 Epigénétique, ARN non codant et Fluidité du génome (Dr Antonin Morillon), l'Unité de Pharmacogénomique du Service de Génétique de l'Ensemble Hospitalier (Dr Ivan Bièche) et l'Unité de Pathologie Expérimentale (Dr Didier Meseure) du Pôle Diagnostique et Théranostique de l'Ensemble Hospitalier (Dr Anne Vincent-Salomon).

9.1 Objectif

Cette étude préliminaire a pour but d'identifier par une stratégie transcriptomique dédiée aux lncRNAs, des signatures diagnostiques caractéristiques des tumeurs malignes par rapport aux tissus sains et des différents sous-types moléculaires ainsi que des signatures pronostiques indépendantes de ces sous types.

9.2 Résumé

La stratégie transcriptomique a consisté en un *screening* par RNA-seq d'une cohorte initiale de tumeurs malignes appartenant aux différents sous-types de carcinomes mammaires infiltrants et de tissus mammaires sains. Une sélection d'une liste restreinte de lncRNAs a également été réalisée. La combinaison de ces deux stratégies a conduit à l'identification d'un *set* de gènes d'intérêt.

9.3 Résultats

Cette stratégie transcriptomique a permis d'identifier 3 signatures constituées de lncRNAs antisens (asRNAs), de lincRNAs et de mRNAs à partir de 8 tumeurs et de 4 tissus sains.

- ***Une signature diagnostique globale constituée de 833 asRNAs, lncRNAs et mRNAs dérégulés.*** Les RNAs sous-exprimés comportent 82 asRNAs (62 nouveaux antisens et 20 antisens annotés), 124 lncRNAs (74 lncRNAs annotés et 50 nouveaux lincRNAs)

et 453 mRNAs qui sont sous exprimés dans les tumeurs du sein par rapport aux 4 tissus mammaires normaux. Les RNAs surexprimés comportent 12 asRNAs (8 nouveaux antisens et 4 antisens annotés), 18 lincRNAs (13 annotés et 5 nouveaux lincRNAs) et 144 mRNAs. La signature permet de séparer les échantillons normaux des tumeurs et également de différencier les échantillons tumoraux selon leur sous-type moléculaire. Les mécanismes moléculaires plus particulièrement associés à ces signatures concernent la division cellulaire, la mitose et la ségrégation chromosomique, l'homéostasie chimique, l'adhésion et la migration, l'angiogénèse et la cicatrisation.

- ***Une signature diagnostique antisens comportant 94 asRNAs dérégulés*** permet également de séparer les échantillons tumoraux des tissus mammaires normaux et de distinguer les différents sous-types moléculaires des cancers du sein. Parmi ces 94 asRNAs, nous avons identifié trois transcrits, RP11-414K1.3, as_S1 et as_S2, fortement dérégulés entre chaque sous-type tumoral et les tissus mammaires normaux. L'analyse des différents paramètres tumoraux (expression des récepteurs hormonaux, statut HER2 et survie sans récurrence) a révélé une différence dans le nombre d'asRNAs dérégulés, suggérant un rôle spécifique des RNAs antisens dans la progression tumorale et le devenir de ces tumeurs.
- ***Une signature pronostique préliminaire*** a été déduite de l'analyse différentielle entre les tumeurs d'évolution métastatique (n=2) et les tumeurs d'évolution non métastatique (n=6). Parmi les 6 gènes de cette signature, 4 gènes sont des lincRNAs et 2 des mRNAs.

9.4 Conclusion

Cette analyse préliminaire du transcriptome non codant des carcinomes mammaires a été effectuée à partir d'une cohorte initiale de 8 carcinomes canaux infiltrants et de 4 tissus mammaires normaux. Cette analyse a consisté avant tout en un test de faisabilité qui a permis la mise en place de l'étude principale du transcriptome non codant de 40 tumeurs mammaires. Cette étude, actuellement en cours de réalisation, nous permettra d'identifier de nouveaux lincRNAs, dont des nouveaux lincRNAs antisens et des lincRNAs circulaires. Elle renforcera la

robustesse des signatures identifiées sur la cohorte initiale et validera nos premiers résultats. L'intérêt de certains de ces gènes sera confirmé par l'analyse de leur niveau d'expression par RT-PCR ou technique *Nanostring* dans une large série de 456 carcinomes mammaires dans laquelle *ANRIL* et *MALAT1* ont été étudiés.

9.5 Etude préliminaire du transcriptome non codant des carcinomes mammaires

9.5.1 Matériel

Pour cette analyse concernant les carcinomes du sein, un séquençage de 12 échantillons mammaires a été effectué, dont 4 tissus normaux et 8 tumeurs mammaires de sous-types morpho-moléculaires différents (2 de type luminal, 2 de type basal et 4 de type HER2) (**Tableau 1**).

<i>Id séquençage</i>	<i>Id de l'échantillon</i>	IHC	IHC	IHC	IHC	<i>Sous-types</i>	<i>Stade</i>	<i>Métastase</i>	<i>Survie</i>
		RE	RP	HER2	Ki67 %		TNM		
A249N1	409076(S076)					normal			
A249N2	441113(S113)					normal			
A249N3	493754(S754)					normal			
A249N4	469625(S625)					normal			
A249T5	409076(T076)	0	0	0	50	TN	T2N0	non	oui
A249T6	476048(T048)	0	0	0	75	TN	T3N0	oui	dcd
A249T7	441113(T113)	0	0	3	60	HER2	T2N0	non	oui
A249T8	448962(T962)	0	0	3	55	HER2	T2N3	oui	dcd
A249T9	493754(T754)	3	3	0	5	LUM	T1N0	non	oui
A249T10	499395(T395)	3	3	0	5	LUM	T1N0	non	oui
A249T11	469625(T625)	3	3	3	50	HER2	T1N0	non	oui
A249T12	488894(T894)	0	0	3	60	HER2	T2N1	non	oui

Tableau 1: Description des échantillons de tissus mammaires normaux et tumoraux séquencés

Les banques ont été préparées avec le kit *TrueSeq Total* directionnel d'Illumina.

Les ARNs tumoraux ont été séquencés par la plateforme ICGex de l'institut Curie sur une machine Illumina *HiSeq 2500*.

9.5.2 Protocole d'analyse bioinformatique entre conditions normales et tumorales

Nous avons utilisé le *pipeline* bioinformatique créée pour l'analyse transcriptomique non codante des carcinomes prostatiques. Nous avons tenté de définir une signature transcriptomique qui permet de séparer les tumeurs mammaires des échantillons normaux et les différentes tumeurs selon leur sous-type et leur évolutivité. Pour définir cette signature, nous avons utilisé le *package R* d'analyse différentielle des données RNA-seq DEseq (Anders and Huber, 2010).

Le logiciel DEseq a permis de comparer et de normaliser les comptages bruts de chaque gène par rapport au nombre total et de sortir les gènes dérégulés entre les deux conditions avec une p-value ajustée. Cette analyse a été réalisée sur tous les RNAs qui possèdent au moins 10 reads dans une des banques en considérant tous les échantillons normaux comme des répliquats de la condition normale et tous les échantillons tumoraux comme des répliquats de la condition tumorale. Nous avons fixé le seuil de surexpression à 2, le seuil de sous-expression à 0.5 et la p-value ≤ 0.05 .

L'analyse en composante principale (ACP) est une méthode d'étude de données multivariées qui consiste à rechercher les directions de l'espace qui représentent le mieux les corrélations entre les échantillons analysés selon l'expression des gènes dérégulés et de projeter les échantillons sur un plan. Pour l'analyse en composante principale de notre signature définie par DEseq, nous avons choisi d'ajouter des données (déjà obtenues) d'échantillons normaux et tumoraux de prostate pour éprouver la robustesse de notre signature et observer si elle est spécifique des tumeurs du sein. L'ACP sur tous les RNAs dérégulés a permis de séparer au mieux les tumeurs des tissus sains en plus d'une bonne séparation avec des échantillons de prostate, suggérant que cette signature est spécifique du cancer du sein (Saci et al., Annexe 1, *supplemental figure 10*).

9.5.3 Signatures transcriptomiques « cancer du sein »

9.5.3.1 Signature diagnostique globale

- Elle a été identifiée en comparant 4 tissus mammaires normaux et 8 cancers du sein. Il a été réalisé une analyse différentielle avec le logiciel Deseq.

- Le seuil de dérégulation a été fixé à 2 (facteur 2 de surexpression ou de sous-expression) et la p-value ajustée à 0.05 (**Fig. 1**).

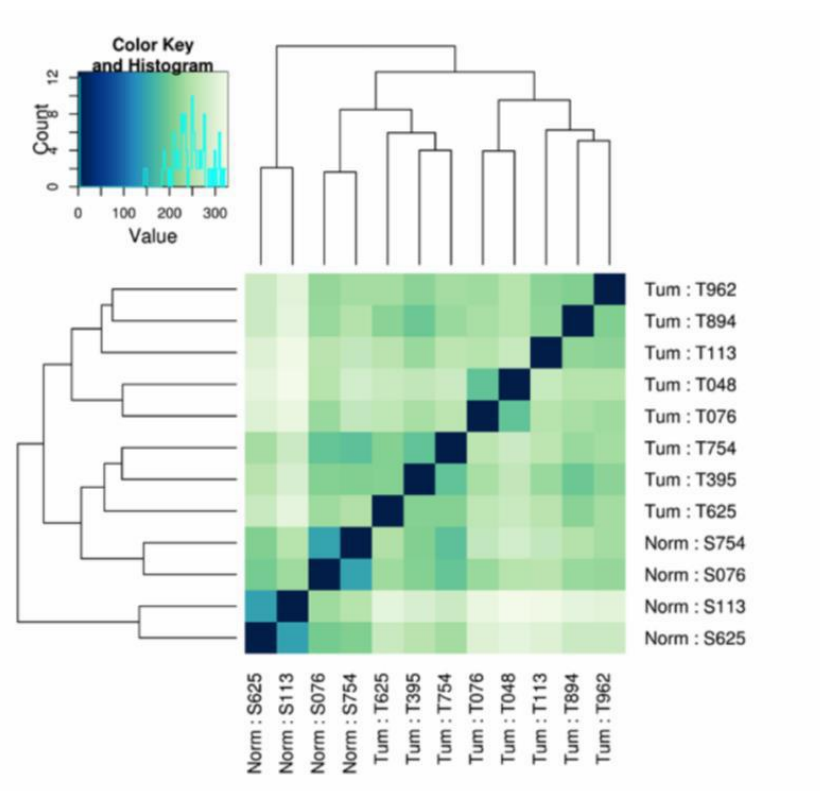


Figure 1 : Clustering non supervisé : distance euclidienne entre les différents échantillons du sein en fonction de l'expression des mRNAs, lncRNAs et asRNAs

- Les 4 tissus normaux et les 8 tumeurs *clusterisent* globalement ensemble par analyse non supervisée.
- La signature est constituée de 833 asRNAs, lncRNAs et mRNAs dont les niveaux d'expression varient significativement par rapport aux tissus mammaires normaux.
 - ✓ Les 659 RNAs sous-exprimés comportent 82 asRNAs (62 nouveaux antisens et 20 antisens annotés), 124 lncRNAs (74 lncRNAs annotés et 50 nouveaux lincRNAs) et 453 mRNAs (**Tableau 2**).
 - ✓ Les 174 RNAs surexprimés comprennent 12 asRNAs (8 nouveaux antisens et 4 antisens annotés), 18 lncRNAs (13 annotés et 5 nouveaux lincRNAs) et 144 mRNAs (**Tableau 2**).

id	nom	baseMean	baseMean_Norm	baseMean_Tum	foldChange	log10 FC	pval	padj
ARNm	FIGF	1028,45	2960,30	62,54	0,02	-1,68	2,42E-15	7,30E-11
ARNm	SCARA5	1139,93	3251,40	84,18	0,03	-1,59	2,18E-14	3,28E-10
ARNm	C2orf40	330,09	939,40	25,46	0,03	-1,57	6,53E-13	6,58E-09
ARNm	INMT	1533,64	4231,50	184,72	0,04	-1,36	1,17E-11	8,86E-08
ARNm	VIT	773,55	2153,00	83,83	0,04	-1,41	1,75E-11	1,05E-07
ARNm	E2F8	850,20	44,03	1253,20	28,46	1,45	3,52E-07	1,77E-04
ARNm	PYDC1	84,30	0,77	126,10	163,45	2,21	4,30E-07	2,09E-04
ARNm	INHBA	9112,20	709,46	13313,50	18,77	1,27	4,47E-07	2,11E-04
ARNm	S100A14	10732,60	834,60	15681,60	18,79	1,27	5,15E-07	2,32E-04
ARNm	OLR1	2347,10	174,57	3433,40	19,67	1,29	6,41E-07	2,75E-04
lncRNA	RP11-286B14.1	410,47	1144,80	43,31	0,04	-1,42	3,18E-11	1,60E-07
lncRNA	XLOC_69328	147,93	417,00	13,37	0,03	-1,49	2,11E-10	7,95E-07
lncRNA	AC132217.4	1328,37	3514,60	235,26	0,07	-1,17	2,62E-09	7,19E-06
lncRNA	RP11-483C6.1	483,92	1289,90	80,96	0,06	-1,20	3,88E-09	8,22E-06
lncRNA	CTD-2527121.4	897,78	2506,40	93,45	0,04	-1,43	4,96E-09	9,36E-06
lncRNA	XLOC_15125	75,90	1,66	113,00	68,16	1,83	3,47E-06	1,14E-03
lncRNA	PLAC2	652,60	52,12	952,90	18,28	1,26	1,29E-05	2,64E-03
lncRNA	UHRF1	910,10	81,62	1324,40	16,23	1,21	2,32E-05	3,72E-03
lncRNA	RP11-429J17.6	5622,30	945,04	7960,90	8,42	0,93	9,12E-05	8,77E-03
lncRNA	RECQL4	1192,40	194,06	1691,60	8,72	0,94	9,72E-05	9,19E-03
antisens	TMEM252	261,36	704,70	39,68	0,06	-1,25	6,36E-09	1,13E-05
antisens	VAC14	43,21	128,70	0,48	0,00	-2,43	2,41E-08	2,79E-05
antisens	NUDT6	86,53	242,80	8,39	0,03	-1,46	6,40E-08	5,86E-05
antisens	SKIV2L2	970,79	2362,30	275,02	0,12	-0,93	1,53E-06	5,94E-04
antisens	ERCC1	126,18	323,00	27,77	0,09	-1,07	4,41E-06	1,34E-03
antisens	KIF7	79,60	6,65	116,10	17,46	1,24	2,23E-04	1,62E-02
antisens	OIP5-AS1	407,50	76,97	572,80	7,44	0,87	4,64E-04	2,60E-02
antisens	PPDPF	99,80	12,31	143,50	11,66	1,07	4,80E-04	2,65E-02
antisens	STXBPSL	227,30	33,24	324,30	9,76	0,99	5,31E-04	2,79E-02
antisens	LRR41	235,50	41,17	332,70	8,08	0,91	5,83E-04	2,94E-02

Tableau 2 : « Top 5 » up et down de chaque classe de mRNAs, lncRNAs et asRNAs les plus différentiellement exprimés dans les cancers du sein classés par p-value ajustée

- Les représentations des niveaux d'expression des 833 RNAs de la signature (Signature 833) en *heatmap* et en ACP confirment que la signature permet de séparer les échantillons sains des échantillons tumoraux (**Fig. 2A et 2B**) et de distinguer les prélèvements tumoraux selon leur sous-type moléculaire : luminal, HER2 et basal (**Fig. 3A et 3B**).
- L'analyse des *GO-term* des mRNAs de la signature (453 mRNAs sous-exprimés et 144 mRNAs surexprimés) montre qu'ils sont particulièrement impliqués dans la division nucléaire, l'angiogénèse, la cicatrisation et l'adhésion cellulaire. Les différents processus biologiques, les fonctions moléculaires ainsi que les voies de signalisation associés aux mRNAs de la signature sont regroupés au sein du **Tableau 3**.

term ID	description	pvalue
GO:0000280	nuclear division	2,96E-12
GO:0000087	mitotic M phase	4,80E-12
GO:0048285	organelle fission	8,65E-12
GO:0000278	mitotic cell cycle	1,76E-10
GO:0051301	cell division	1,33E-09
GO:0007155	cell adhesion	1,71E-07
GO:0022610	biological adhesion	1,73E-07
GO:0007059	chromosome segregation	2,34E-07
GO:0007049	cell cycle	6,39E-07
GO:0065004	protein-DNA complex assembly	1,02E-06
GO:0001568	blood vessel development	1,06E-06
GO:0009611	response to wounding	2,49E-06
GO:0042060	wound healing	1,13E-05
GO:0006323	DNA packaging	2,07E-05
GO:0048878	chemical homeostasis	4,32E-05
GO:0050878	regulation of body fluid levels	4,32E-05
GO:0050817	coagulation	9,47E-05
GO:0030334	regulation of cell migration	9,99E-05
GO:0007610	behavior	1,01E-04
GO:0048015	phosphatidylinositol-mediated signaling	1,06E-04
GO:0000075	cell cycle checkpoint	1,44E-04
GO:0007017	microtubule-based process	1,77E-04
GO:0009725	response to hormone	3,04E-04
GO:0042127	regulation of cell proliferation	3,76E-04
GO:0007051	spindle organization	5,05E-04
GO:0042592	homeostatic process	5,21E-04
GO:0009719	response to endogenous stimulus	5,65E-04
GO:0006333	chromatin assembly or disassembly	7,38E-04
GO:0051046	regulation of secretion	7,53E-04
GO:0030808	regulation of nucleotide biosynthetic process	7,57E-04
GO:0044057	regulation of system process	9,27E-04
GO:0051303	establishment of chromosome localization	1,05E-03
GO:0050000	chromosome localization	1,05E-03
GO:0006140	regulation of nucleotide metabolic process	1,18E-03
GO:0031279	regulation of cyclase activity	1,21E-03
GO:0051339	regulation of lyase activity	1,40E-03
GO:0030817	regulation of cAMP biosynthetic process	1,40E-03
GO:0019932	second-messenger-mediated signaling	1,43E-03
GO:0006111	regulation of gluconeogenesis	1,60E-03
GO:0033043	regulation of organelle organization	1,61E-03
GO:0007267	cell-cell signaling	1,63E-03
GO:0019226	transmission of nerve impulse	1,94E-03
GO:0007626	locomotory behavior	3,03E-03
GO:0019216	regulation of lipid metabolic process	3,04E-03
GO:0006260	DNA replication	3,09E-03
GO:0007188	adenylate cyclase-modulating G-protein coupled receptor signaling pathway	5,62E-03
GO:0050873	brown fat cell differentiation	6,51E-03
GO:0000079	regulation of cyclin-dependent protein serine/threonine kinase activity	7,16E-03
GO:0007268	synaptic transmission	7,19E-03

Tableau 3 : Les 50 premiers processus biologiques enrichis dans les mRNAs dérégulés de la signature globale

Z
en

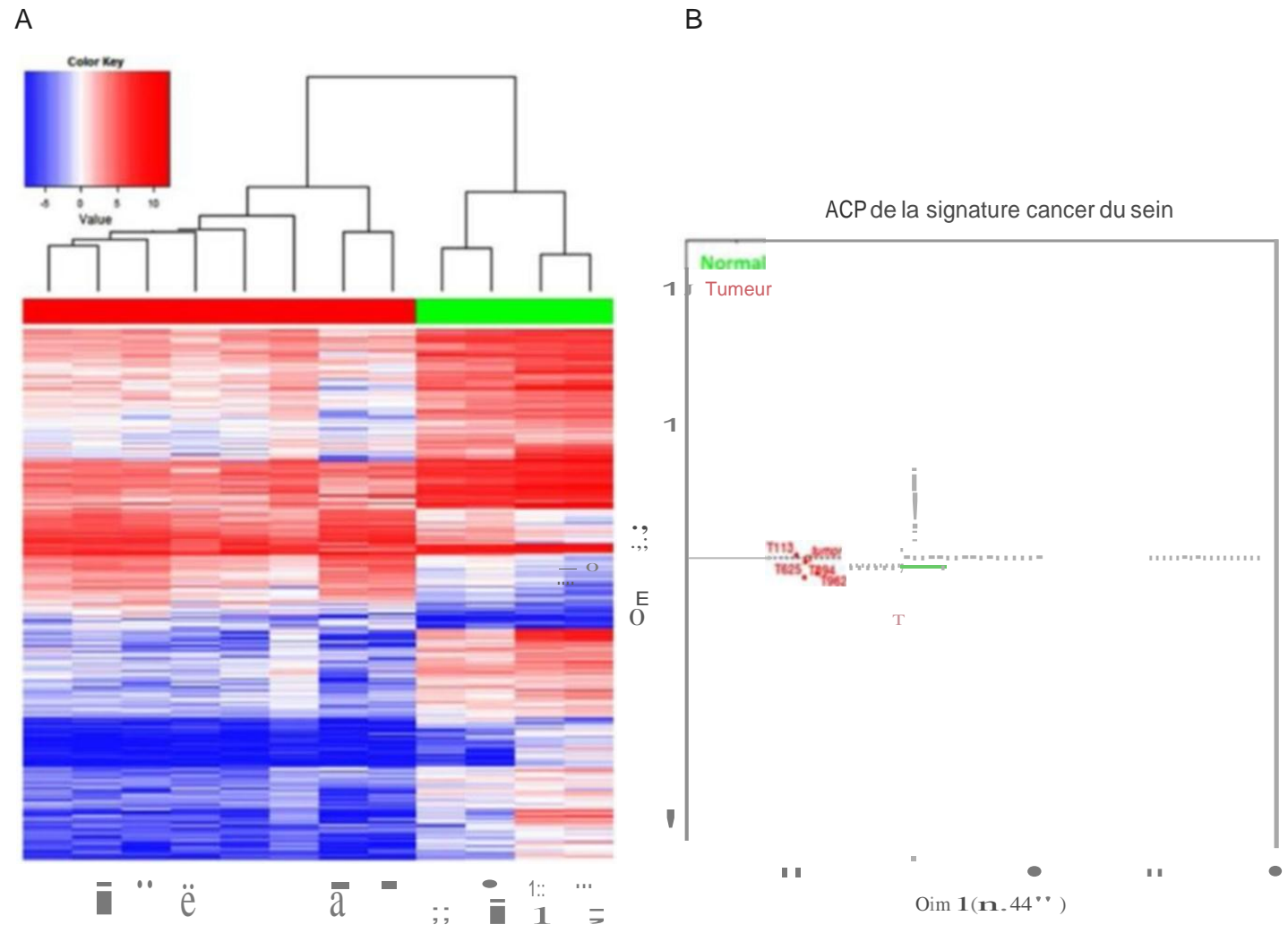


Figure 2 : Clustering des 12 échantillons avec la signature de 833 gènes
2A) Représentation en *heat map* du log2 de l'expression des RNAs de la signature.
2B) ACP des échantillons basée sur la signature des RNAs dérégulés.

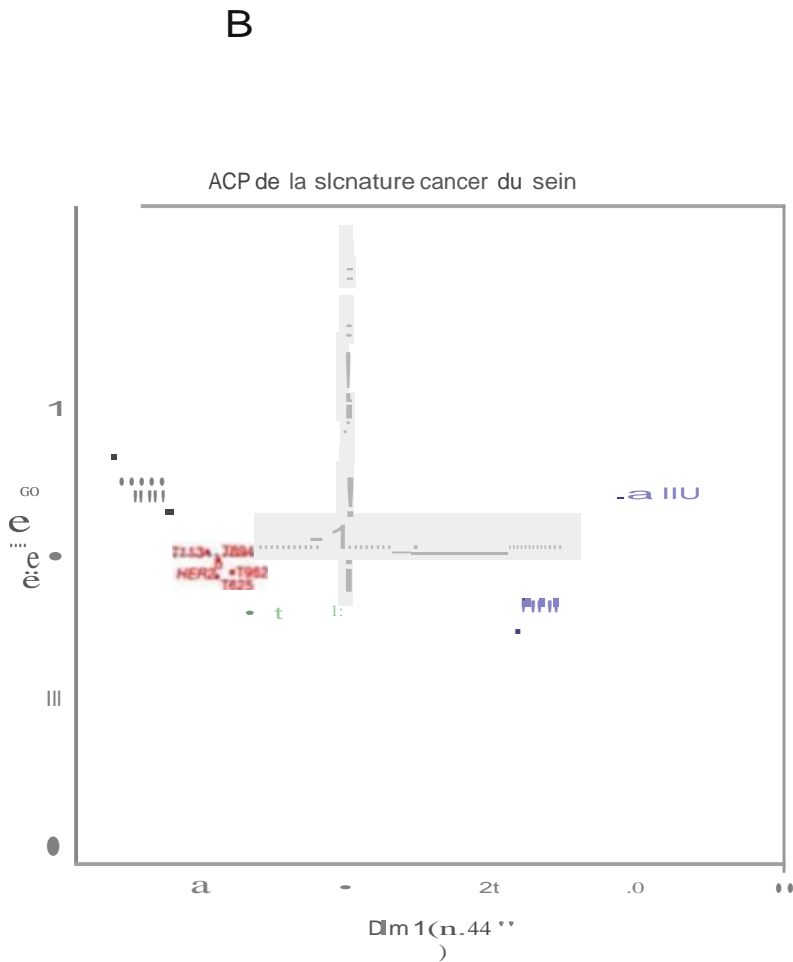
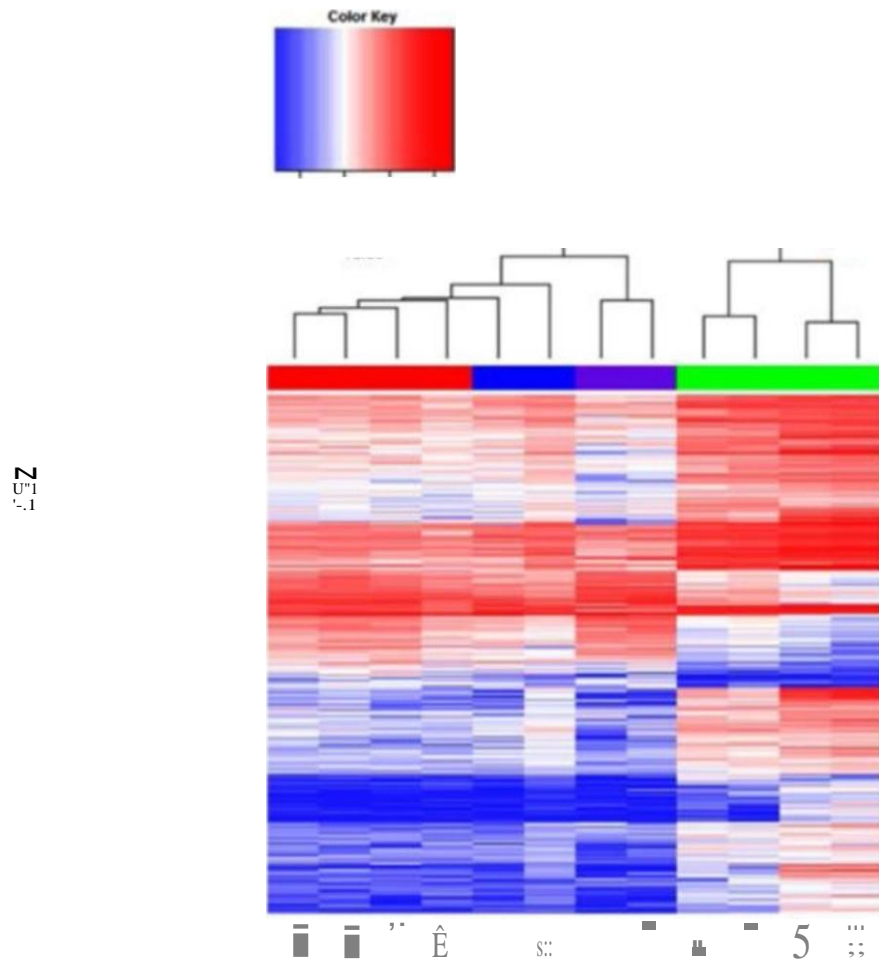


Figure 3 : *Clustering* des 12 échantillons avec la signature de 833 gènes par sous-types moléculaires
 3A) Représentation en *heatmap* du log₂ de l'expression des différents RNAs.
 3B) ACP des échantillons basée sur la signature.

9.5.3.2 Signature diagnostique antisens

- Dans la signature globale, 94 asRNAs apparaissent dérégulés entre la condition tumorale et la condition normale.
- Cette sous-signature (signature 94as) comporte 82 asRNAs sous-exprimés et 12 asRNAs surexprimés.
- La signature 94as permet aussi de séparer les 4 échantillons normaux des 8 échantillons tumoraux (**Fig. 4 A et 4B**).
- L'analyse des asRNAs de la sous-signature 94as montre qu'il existe une variation dans le nombre de candidats dérégulés par sous-type tumoral. Trois transcrits fortement dérégulés entre chaque sous-type de tumeur et les tissus mammaires normaux ont été identifiés (**Fig. 5**).
- ✓ L'ARN antisens annoté RP11-414K1.3 est surexprimé dans les tumeurs de type basal et de type HER2+ avec une expression très hétérogène entre les différents échantillons mais ne l'est pas dans les tumeurs de type luminal (**Fig. 5B**).
- ✓ Le nouvel asRNA as_S1 est surexprimé dans les tumeurs de type luminal et HER2 avec moins de variabilité que le précédent (**Fig. 5C**).
- ✓ Enfin, le nouvel antisens as_S2 est surexprimé dans les tumeurs de type luminal (**Fig. 5D**) mais il est également surexprimé dans un des 4 échantillons du type HER2 (**Fig. 5A**).

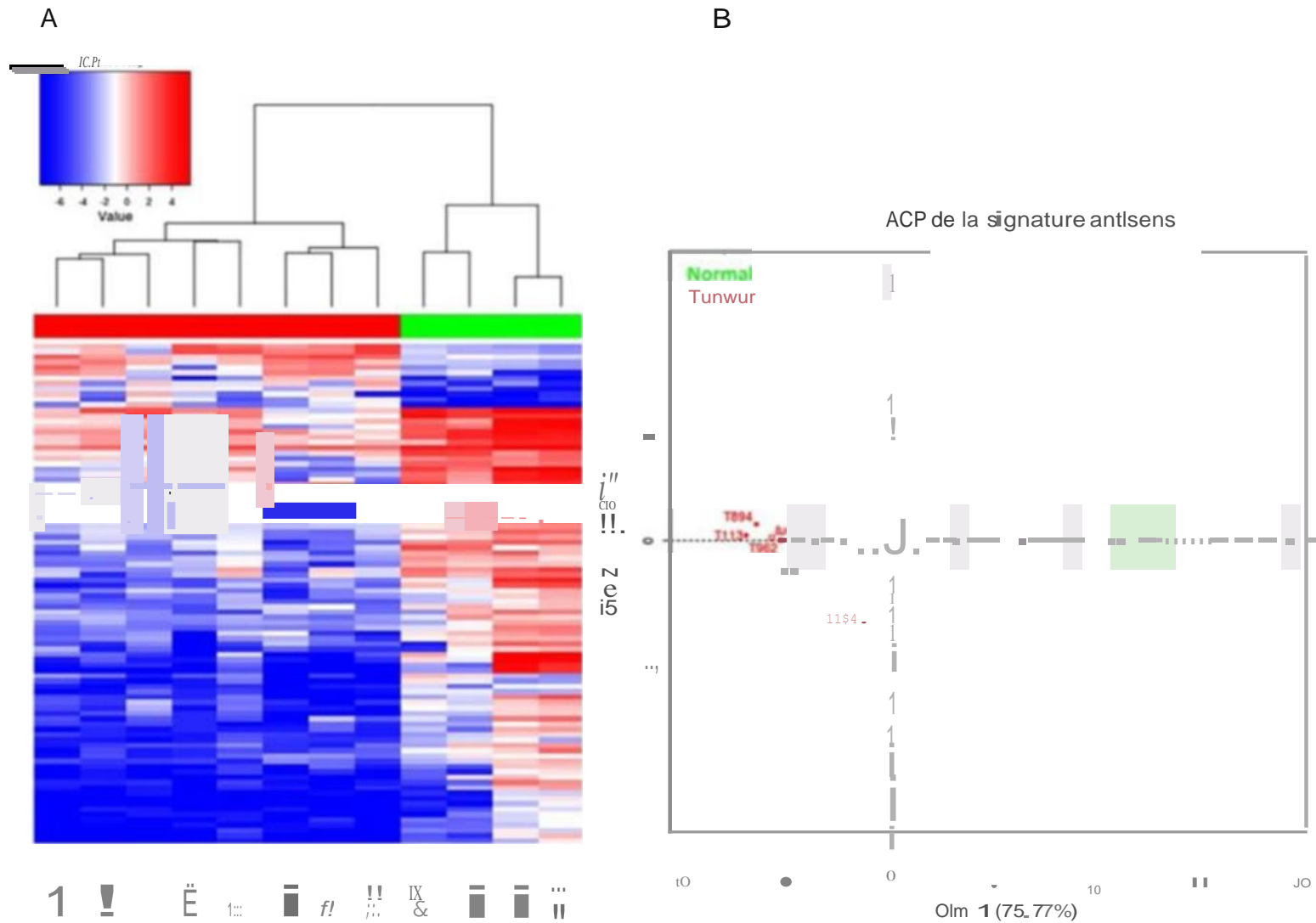


Figure 4: *Clustering* des 12 échantillons avec la signature 94as dans les carcinomes mammaires
 4A) Représentation en *heatmap* du log2 de l'expression des asRNAs.
 4B) ACP des échantillons basée sur la signature antisens.

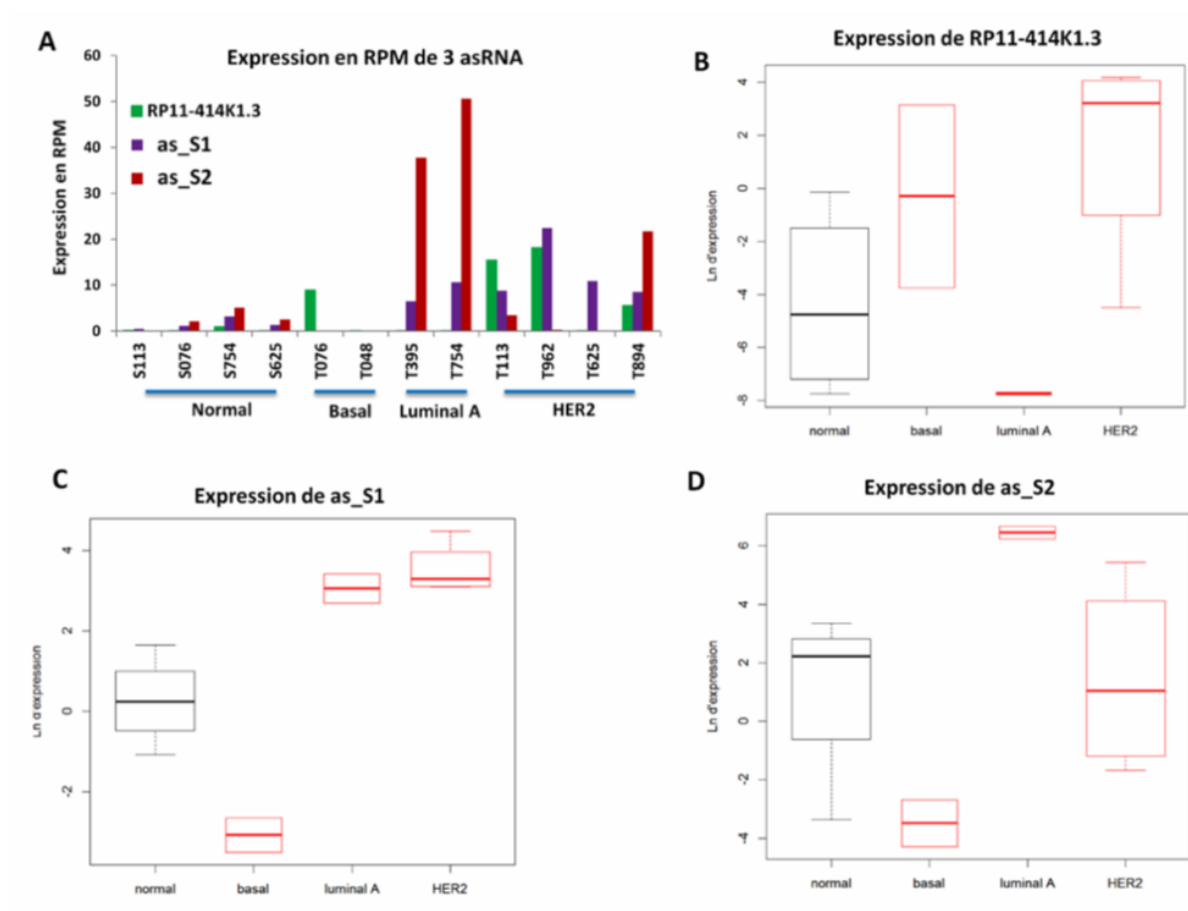


Figure 5 : Expression de trois lincRNAs antisens dans les divers sous-types moléculaires de cancers mammaires. A : Distribution de l'expression de RP11-414K1.3, as_S1 et as_S2 dans chaque échantillon; B : \log_2 de l'expression de l'antisens annoté RP11-414K1.3 par type tumoral ; C : \log_2 de l'expression du nouveau antisens as_S1; D: \log_2 de l'expression du nouvel antisens as_S2.

9.5.3.3 Signature pronostique préliminaire

- L'analyse différentielle entre les 2 tumeurs d'évolution métastatique (448962-T962 et 476048-T048) (**Tableau 1**) et les 6 tumeurs d'évolution non métastatique (quiescentes ou récidivantes) a permis d'identifier 6 gènes d'intérêt. Parmi ces 6 gènes, 4 gènes sont des lincRNAs (685235_ENSG00000138336.8_TET1, RP11-142120, RP11-638F5.1 et RP11-25H12.1) et 2 des mRNAs (*CLCA2* et *GABRB3*).

9.5.4 Validation du *set* d'intérêt dans une large série de tumeurs mammaires

Pour confirmer les modifications du niveau d'expression du *set* de gènes et la pertinence des signatures identifiées lors de cette première analyse transcriptomique préliminaire, une étude additionnelle transcriptomique est actuellement en cours et une validation des résultats sera réalisée dans la série de 456 CCIIs utilisée pour l'analyse d'*ANRIL* et *MALAT1* par RT-PCR ou par la technique *Nanostring*.

9.5.5 Conclusion

L'analyse des données bioinformatiques issues de l'étude par RNA-seq de notre panel initial de carcinomes mammaires a permis de définir des signatures transcriptomiques globale, antisens et pronostique, composées de mRNAs, de lncRNAs et d'asRNAs. Ces signatures séparent les tissus sains des tissus tumoraux, les différents sous-types morpho-moléculaires et les tumeurs présentant ou non une évolution métastatique. Cette première étude très préliminaire a servi de preuve de concept et fera place à une deuxième analyse par RNA-seq d'une cohorte plus importante de 40 carcinomes mammaires incluant les 4 sous-types moléculaires (20 cas avec évolution métastatique survenue en 1 à 3 ans *versus* 20 cas sans évolution métastatique durant 10 ans), de manière à renforcer la robustesse de nos signatures et de valider nos premiers résultats.

CONCLUSION ET PERSPECTIVES

Le cancer du sein demeure un problème majeur de santé publique. A l'échelle mondiale, il représente le deuxième cancer humain le plus fréquent (12 %) et occupe la première place des tumeurs malignes de la femme. Le cancer du sein est une maladie très hétérogène caractérisée par des sous-types morphologiques et moléculaires pourvus de profils génomiques, épigénomiques et transcriptomiques distincts. Les classifications cliniques, morphologiques et immunophénotypiques ne permettant pas d'explorer la totalité de l'hétérogénéité tumorale propre à chaque patiente. De nouvelles classifications intégrées comportant des sous-types moléculaires des carcinomes mammaires ont été récemment proposées en analysant et en intégrant les données issues des diverses techniques, telles la CGH array, les SNP arrays, le séquençage d'exome et de *whole genome*, la méthylation de l'ADN, les cDNA arrays, le RNAseq et la RPPA. Actuellement, seuls certains biomarqueurs morphologiques (récepteurs des estrogènes et de la progestérone, récepteur HER2, index de prolifération Ki67) et la nouvelle signature PAM50 sont pris en compte dans la classification et l'orientation thérapeutique des cancers du sein. La recherche de nouveaux biomarqueurs et de nouvelles cibles thérapeutiques s'avère donc impérative.

Les analyses transcriptomiques à haut débit ont révélé que moins de 2 % du génome humain code pour des protéines, alors que plus de 80% de l'ADN est transcrit en ARN. Les lncRNAs, dont la longueur est supérieure à 200 nt, constituent une famille émergente de molécules caractérisées par leur localisation intranucléaire prédominante, leur plus grande spécificité cellulaire et tissulaire et des fonctions biologiques très variées s'exerçant préférentiellement sous la forme de complexes ribonucléoprotidiques. Les lncRNAs sont responsables de l'intégrité de l'organisation nucléaire et présentent des propriétés régulatrices très diversifiées jouant des rôles décisifs dans de multiples processus physiologiques participant au maintien de l'homéostasie cellulaire et tissulaire. Certains lncRNAs comportent des profils d'expression anormaux dans diverses pathologies, dont le cancer.

L'objectif principal du projet de thèse a consisté en l'analyse de l'expression, de la fonctionnalité et du rôle des lncRNAs dans l'oncogénèse mammaire, en recherchant plus particulièrement l'intérêt potentiel de ces molécules comme biomarqueurs à visée diagnostique, pronostique ou thérapeutique. La première partie du projet a été consacrée à l'étude de deux lncRNAs majeurs, *ANRIL* et *MALAT1*, impliqués dans de nombreux types de cancers, mais dont les mécanismes d'action biologiques et la signification clinique au cours de la cancérogénèse mammaire sont encore controversés et/ou méconnus. La deuxième partie du projet a eu pour objectif l'identification de nouveaux types de lncRNAs, en analysant le

transcriptome non codant des carcinomes mammaires infiltrants par une stratégie épitranscriptomique appliquée initialement à une petite série (n=8), qui sera suivie par une cohorte élargie (n=40) de cancers du sein.

***ANRIL* et *MALATI* sont particulièrement impliqués dans la cancérogénèse mammaire.**

ANRIL et *MALATI* sont deux lncRNAs très étudiés en raison de leur implication principale dans la prolifération tumorale, caractéristique la plus importante de tous les *hallmarks* du cancer. Les analyses de gain et de perte de fonction réalisées *in vitro* et *in vivo* suggèrent qu'*ANRIL* et *MALATI* sont des lncRNAs oncogéniques dans une majorité de cancers. La déplétion d'*ANRIL* induite expérimentalement dans différents types de tumeurs malignes, s'accompagne d'une réduction de la prolifération tumorale et de la lymphangiogénèse et d'une majoration de l'apoptose et de la sénescence, tandis que sa surexpression favorise la progression du cycle cellulaire, la prolifération tumorale et l'adhésion cellulaire. De même, la déplétion de *MALATI* inhibe la prolifération cellulaire, augmente l'apoptose et diminue la migration des cellules cancéreuses dans divers types de cancers alors que sa surexpression favorise la prolifération, l'EMT, la migration des cellules malignes et la formation de métastases.

ANRIL, *MALATI* et *NEATI* présentent le plus d'altérations structurales parmi les lncRNAs au cours de la cancérogénèse. *ANRIL* apparaît ainsi comme l'un des lncRNAs les plus fréquemment altérés dans divers types de cancers. Les délétions et les translocations avec transcrits de fusion impliquant *ANRIL* peuvent conduire à la dérégulation de l'expression des gènes cibles en *cis* et en *trans* et sont associées à une augmentation de la prolifération et de l'adhésion cellulaire. Des altérations structurales ont de même été identifiées au niveau du locus de *MALATI*, dont des mutations ponctuelles, des translocations chromosomiques et des amplifications. Une série récente de 566 cancers du sein par *whole-genome sequencing* montre que *MALATI* et *NEATI* sont les 2 plus fréquents lncRNAs mutés dans les cancers du sein (Nik-Zainal et al., 2016). Toutefois, bien qu'*ANRIL* et *MALATI* soient dérégulés dans de nombreux types de tumeurs malignes, leurs rôles et leurs mécanismes d'action demeurent actuellement largement méconnus dans la tumorigénèse mammaire, ainsi que leur intérêt clinique dans le diagnostic, le pronostic et la classification moléculaire des cancers du sein.

Une surexpression d'*ANRIL* et de *MALATI* a été observée dans notre série de carcinomes mammaires infiltrants, confirmant le rôle pro-oncogénique des 2 lncRNAs. *ANRIL* apparaît ainsi surexprimé dans 20% des CCI et sa surexpression est présente plus particulièrement dans le sous-type agressif triple-négatif. Dans notre étude, une surexpression de *MALATI* a été également observée dans 14% des tumeurs malignes du sein. En utilisant la technique de RNA-FISH, nous avons localisé les transcrits de *MALATI* principalement dans les *speckles* intra-nucléaires et les noyaux des cellules épithéliales tumorales surexprimant *MALATI* présentent une fluorescence intense située au niveau de nombreux et volumineux *speckles*.

Des résultats expérimentaux observés à partir de modèles murins après inactivation de *MALATI*, tels l'absence de létalité et de phénotype anormal, suggèrent que l'activité des lncRNAs n'est pas seulement liée à la présence de séquences conservées entre les espèces, mais semble aussi résulter de facteurs supplémentaires, dont la structure secondotertiaire, les domaines fonctionnels, l'épissage alternatif et la transmission synténique. La connaissance précise des fonctions biologiques des lncRNAs et de leurs dérégulations au cours de la cancérogénèse constitue actuellement une étape indispensable à leur ciblage thérapeutique. La localisation cellulaire des lncRNAs étant liées à leurs fonctions, nous utiliserons la technique de RNA-FISH, afin d'inclure les transcrits d'intérêt dans les 3 sous-groupes topographiques permettant d'orienter leurs fonctions moléculaires (localisation intranucléaire au niveau du site de transcription, localisation intranucléaire avec relocalisation nucléoplasmique ou dans des corps intra-nucléaires et localisation intracytoplasmique). Nous tenterons de réaliser des tests fonctionnels par invalidation *in vitro* et *ex vivo* sur des lignées cellulaires et des xénogreffes de tumeurs humaines, en privilégiant la technologie CRISPr/Cas9. Cette technique de *genome editing* maîtrisée à l'Institut Curie, offre en effet d'importantes possibilités de ciblage thérapeutique des lncRNAs.

Des profils transcriptomiques complexes liés à la dérégulation d'*ANRIL* et de *MALATI* ont été identifiés dans les carcinomes du sein par rapport à leurs profils d'expression dans le parenchyme mammaire normal.

La complexité de leur dérégulation est liée à la présence d'isoformes et de réseaux d'interactions avec les mRNAs, les miRNAs et d'autres lncRNAs. Cette dérégulation de leurs profils transcriptomiques est en grande partie liée à la diversité potentielle de leurs liaisons

avec des oncogènes et des suppresseurs de tumeurs appartenant aux principales voies de signalisation, des complexes multiprotéiques intranucléaires structuraux ou fonctionnels modifiant la chromatine, des miRNAs et des mRNAs, des facteurs de transcription ainsi qu'avec d'autres lncRNAs. Si les fonctions biologiques principales d'*ANRIL* et de *MALAT1* semblent globalement conservées au cours de la cancérogénèse mammaire, les profils transcriptomiques dérégulés d'*ANRIL* et de *MALAT1* observés dans notre série de tumeurs résultent d'altérations de leurs niveaux d'expression et de leurs processus de maturation, mais aussi de modifications de leurs mécanismes d'action et de leurs réseaux d'interactions.

ANRIL se comporte dans les carcinomes mammaires comme un transcrit naturel antisens régulant l'expression des gènes au niveau épigénétique en recrutant les deux complexes répresseurs polycomb PRC2 et PRC1, dont la fonction est de réorganiser la chromatine en inactivant le locus *p16-CDKN2A/p15-CDKN2B/p14-ARF*. Toutefois, l'expression des sous-unités appartenant aux deux complexes polycomb PRC2/PRC1 est altérée dans notre série et est caractérisée par une importante surexpression d'*EZH2* dans 77% des CCI et par une sous-expression inattendue de *CBX7* dans 40% des tumeurs. A ce jour, il n'a pas été observé de mutation inactivatrice de *CBX7* et les données de la littérature suggèrent l'existence d'interactions multiples entre les lncRNAs, les mRNAs, les complexes polycomb et les miRNAs au cours de la progression tumorale. Ainsi, *ANRIL*, en se liant à *EZH2*, induit l'inactivation épigénétique des deux miRNAs suppresseurs miR-49a et miR-449, favorisant la prolifération tumorale et inhibant l'apoptose et la sénescence des cellules cancéreuses gastriques. Nous suggérons que dans les cancers du sein, le lncRNA *ANRIL* pourrait également réguler en *cis* et en *trans* l'expression de gènes par des mécanismes épigénétiques *via* des interactions entre les sous-unités des complexes polycomb et des miRNAs. Dans les cancers gastriques, la surexpression d'*ANRIL* et de *PVT1* inactive *p16-CDKN2A* et *p15-CDKN2B* par des mécanismes épigénétiques en *cis* (*ANRIL*) et en *trans* (*PVT1*) en favorisant tous deux un accroissement de l'apport en *EZH2* au niveau du *cluster p16-CDKN2A* et *p15-CDKN2B*. Des réseaux d'interactions complexes analogues pourraient également expliquer la surexpression d'*EZH2* et la sous-expression de *CBX7* dans des cancers du sein. Notre étude a ainsi révélé que les 3 miRNAs onco-suppresseurs (miR-26A1, miR-125B et miR-214) qui ciblent normalement *EZH2*, sont sous-exprimés dans notre série de CCI. Les 2 oncomiRs miR-181B1 and miR-181A2 qui ciblent et inactivent *CBX7*, apparaissent surexprimés dans notre série, en relation avec l'activation de l'oncogène *HMGAI*.

MALATI est principalement localisé dans les *speckles* des cellules tumorales mammaires étudiées et est impliqué dans la régulation transcriptionnelle de l'expression génique et dans l'épissage alternatif en interagissant avec de multiples réseaux protéiques variant selon le type cellulaire et tissulaire de manière à contrôler directement ou indirectement l'expression de multiples gènes. Nous avons évalué les niveaux d'expression d'un panel de gènes potentiellement impliqués dans divers processus moléculaires et cellulaires associés à la carcinogénèse *via* la dérégulation de *MALATI*. Les gènes significativement associés à *MALATI* sont essentiellement impliqués dans la migration cellulaire (*RHOB*, *PLAU/UPA* et *MMP14*), l'édification du complexe répressif Polycomb PRC2 (*EED*, *SUZ12*, *JARID2*, *TUG*), l'apoptose (*BIRC6*), la réparation de l'ADN (*ATM*, *MSH2*, *XRCC1*), la biogénèse des ncRNAs (*DGCR8*) et les interacteurs de l'épissage alternatif (*SRSF1*, *SRSF3*, *UHMK1*). Le criblage de la base de données dbEST avec le cDNA de la forme complète de *MALATI* nous a permis d'identifier un important variant d'épissage alternatif, Δ -*MALATI*, comportant deux régions concomitantes délétées de 119 pb et 243 pb. Dans notre série de carcinomes mammaires infiltrants, le profil transcriptomique dérégulé de *MALATI* associe une surexpression de *WT-MALATI* dans 14% des tumeurs et une sous-expression du transcrit alternatif Δ -*MALATI* dans 20% des cancers du sein. Les causes de la dérégulation de *MALATI* demeurent encore largement méconnues et d'autres études seront nécessaires afin d'identifier les mécanismes génétiques et/ou épigénétiques responsables de la surexpression de *WT-MALATI* et de la sous-expression du transcrit alternatif Δ -*MALATI* dans les cancers du sein. Des mutations du gène *MALATI* sont plus rarement observées dans les carcinomes mammaires que dans d'autres cancers, tels le cancer de la vessie, mais constituent avec celles de *NEATI* les plus fréquentes mutations des lncRNAs dans les cancers du sein. Une amplification du gène *MALATI* est improbable car ce gène n'est pas situé dans une région fréquemment amplifiée (11q13.1). De plus, une dérégulation épigénétique de *MALATI* médiée par la méthylation des îlots CpG n'a pas été rapportée dans la littérature. En outre, la régulation post-transcriptionnelle de *MALATI* par HSA-miR-125b identifiée dans les cancers de la vessie n'a pas été confirmée dans notre série de tumeurs du sein. Nos données suggèrent cependant l'existence d'une régulation transcriptionnelle de Δ -*MALATI* (mais non de *WT-MALATI*) par le coactivateur de transcription YAP, également observée dans le cancer du foie.

Les profils transcriptomiques d'*ANRIL* et de *MALATI* dans les cancers du sein semblent caractéristiques de la localisation mammaire lorsqu'ils sont comparés à leurs profils transcriptomiques dans des tumeurs malignes d'autres sites.

Le profil transcriptomique dérégulé d'*ANRIL* est caractérisé dans notre série de carcinomes mammaires par une importante sous-expression de *CBX7* et une corrélation positive inattendue entre la surexpression d'*ANRIL* et celle du locus *p16-CDKN2A/p15-CDKN2B/p14-ARF*, et plus particulièrement de *p14/ARF*, alors que ce profil apparait inverse (association négative) dans les carcinomes de prostate où ce même *cluster* est inactivé par *ANRIL*. De même, *ANRIL* inactive épigénétiquement les miRNAs onco-suppresseurs miR-99a/miR-449a dans les carcinomes gastriques et non dans notre série de CCI. Bien que ces résultats semblent contradictoires, nous pouvons supposer que *CBX7* pourrait influencer la corrélation positive ou négative entre *ANRIL* et le locus *p16-CDKN2A/p15-CDKN2B/p14-ARF* et se comporter à la fois comme un suppresseur de tumeur ou un oncogène selon le type de tumeur, la nature du microenvironnement tumoral et les réseaux d'interaction spécifiques composés de mRNAs, de miRNAs, de lncRNAs et de facteurs de transcription interagissant avec les complexes lncRNA-PRC2/PRC1.

Concernant *MALATI*, aucune variation du niveau d'expression du lncRNA inhibiteur transcriptionnel *TUG1* n'a pu être observée dans notre série de cancers du sein, alors que *TUG1* apparait co-surexprimé avec *MALATI* dans les carcinomes de vessie. De même, il n'a pas été identifié dans notre série de corrélations entre *MALATI* et les gènes codant pour des protéines impliquées dans la régulation du cycle cellulaire, la migration, l'apoptose (à l'exception de *BIRC6*) et l'EMT, comme décrites dans divers types de cancers. En outre, l'analyse de *MALATI* via la technique de RNA-FISH révèle une surexpression diffuse et non seulement limitée aux seuls *speckles*, évoquant une relocalisation nucléoplasmique et suggérant l'acquisition d'autres fonctions moléculaires.

***ANRIL* et *MALATI* sont des biomarqueurs pronostiques et des cibles thérapeutiques potentiels.**

Les lncRNAs sont actuellement considérés comme des biomarqueurs diagnostiques et pronostiques potentiels particulièrement intéressants en raison d'une propriété caractéristique leur conférant une plus grande spécificité tissulaire, cellulaire et mécanistique.

- **Biomarqueurs pronostiques**

En utilisant l'analyse par log-rank afin d'établir des corrélations entre la survie sans métastase (MFS) et les niveaux d'expression des mRNAs des 10 gènes de la voie *ANRIL*, il n'a pas été observé de corrélation significative entre l'expression d'*ANRIL* et la survie sans métastase (MFS). La surexpression d'*EZH2* et la sous-expression de *CBX7* sont par contre significativement associées à une survie plus courte. La surexpression d'*EZH2* est fortement associée à un grade SBR élevé et un statut RE- et RP-. La sous-expression de *CBX7* est fortement corrélée avec un haut grade histologique SBR, un statut RE- et RP- et l'expression de marqueurs d'EMT et de cellules souches cancéreuses mammaires. Mais, à la différence de *CBX7*, la surexpression d'*EZH2* est étroitement liée au marqueur de prolifération *MKI67*, suggérant un rôle différent de ces deux protéines dans la tumorigénèse mammaire. L'analyse multivariée a révélé que le rapport *EZH2/CBX7* est un facteur pronostique indépendant (P=0.001).

Une corrélation significative a été observée dans notre série de tumeurs mammaires entre la sous-expression d'un transcrit d'épissage Δ -*MALATI* et les tumeurs mammaires présentant une grande taille macroscopique, un statut RE- et RP- et des niveaux d'expression élevés de *MKI67*, suggérant que la sous-expression de Δ -*MALATI* joue un rôle dans l'agressivité des tumeurs malignes du sein. A l'inverse, aucune corrélation significative n'a été observée entre la surexpression du mRNA de *WT-MALATI* et les marqueurs d'agressivité. L'analyse de la survie a révélé que les patientes présentant un faible niveau d'expression de Δ -*MALATI* ont une survie plus courte. En outre, l'analyse multivariée a montré que le statut d'expression de Δ -*MALATI* est un facteur pronostique indépendant pour la MFS. De façon analogue aux isoformes du lncRNA *PTENP1*, ce variant d'épissage alternatif de *MALATI* pourrait agir comme un leurre, séquestrant des biomolécules normalement liées au transcrit *WT-MALATI* et dérégulant ainsi ses fonctions. De plus, en raison de ses modifications structurales, Δ -*MALATI* pourrait acquérir de nouvelles fonctions ou perdre des sites de fixation aux miRNAs. Globalement, ces résultats suggèrent que l'isoforme d'épissage alternatif Δ -*MALATI* participe activement aux fonctions biologiques de *MALATI* et à la carcinogénèse mammaire.

- **Cibles thérapeutiques potentielles**

Les lncRNAs et les molécules avec lesquelles ils interagissent pourraient constituer des cibles thérapeutiques novatrices. Toutefois, la présence de variants d'épissage alternatifs aux fonctions variées, d'interactions multiples et d'une spécificité d'organe doit être prise en compte lors de l'évaluation des thérapies dirigées contre ces lncRNAs.

En raison de la surexpression d'*ANRIL* dans de nombreux cancers, dont les carcinomes mammaires, la déplétion d'*ANRIL* pourrait se révéler une option thérapeutique pertinente de manière à réduire la prolifération tumorale et la lymphangiogénèse et augmenter la sensibilité des cellules cancéreuses aux agents endommageant l'ADN (Wan et al., 2013), à l'apoptose et à la sénescence. Des données récentes ont révélé que les complexes polycomb PRC2 et PRC1 jouent un rôle bivalent dans la cancérogénèse, présentant à la fois des propriétés oncogéniques et oncosuppressives dépendant du type cellulaire. Le déficit en composants de ces complexes pourrait diminuer les niveaux d'expression des marques épigénétiques répressives H3K27me3, contribuant à l'acétylation de H3K27, à l'augmentation de la transcription et à l'activation de voies oncogéniques. Ainsi, la définition d'un statut d'activité PRC2/PRC1 dans les carcinomes mammaires pourrait se révéler pertinente lors du choix des options thérapeutiques. Nous avons observé dans notre série de tumeurs mammaires une prédominance de la marque épigénétique activatrice H3K27ac dans 73.8% des CCI et une forte réduction de la marque épigénétique répressive H3K27me3 dans 76.3% des CCI. Ces données suggèrent que la perte des fonctions répressives polycomb dans les CCI pourrait être due à des altérations structurales ou à des modifications épigénétique d'*ANRIL* et de PRC1/PRC2, avec pour conséquence la perte de la marque répressive H3K27me3 et l'augmentation de la marque activatrice H3K27ac. Les thérapies ciblant les processus épigénétiques dans le cancer connaissent un développement croissant avec l'utilisation des inhibiteurs des HDAC et de l'ADN méthyltransférase, et plus récemment les inhibiteurs des bromodomaines BET (Baude et al., 2014). En effet, des études concernant les tumeurs malignes des gaines nerveuses périphériques (MPNSTs) ont montré que les altérations de sous-unités PRC2 conduisent à l'amplification de la transcription induite par la voie Ras et confèreraient une sensibilité aux inhibiteurs des bromodomaines (De Raedt et al., 2014).

MALAT1 est impliqué dans la régulation transcriptionnelle de l'expression génique et dans l'épissage alternatif. Sa dérégulation s'accompagne d'importantes conséquences dans le contrôle du cycle cellulaire et la prolifération tumorale, l'apoptose, la plasticité et l'angiogénèse. L'injection d'ASO ciblant *MALAT1* inhibe efficacement la croissance des métastases de diverses tumeurs malignes. En raison de son implication dans de nombreux

cancers, *MALATI* pourrait constituer une cible thérapeutique pertinente dans les tumeurs malignes métastatiques et en particulier dans le cancer du sein (Arun et al., 2015).

L'analyse du transcriptome non codant des carcinomes mammaires par une stratégie pangénomique a permis d'identifier de nouveaux types de lncRNAs, tels les nouveaux lncRNAs antisens.

L'étude du transcriptome non codant des carcinomes mammaires a permis de définir des signatures comportant des lncRNAs. Cette étude préliminaire a eu pour but d'identifier des signatures diagnostiques caractéristiques des différents sous-types tumoraux et des signatures pronostiques indépendantes des sous types moléculaires classiques. Cette stratégie pangénomique a consisté en un *screening* par RNAseq d'une cohorte initiale de 8 tumeurs malignes appartenant aux différents sous-types de carcinomes mammaires infiltrants et de 4 tissus mammaires sains. Elle a permis d'identifier une signature diagnostique globale constituée d'asRNAs, de lincRNAs et de mRNAs et une signature constituée exclusivement d'asRNAs. Une signature pronostique préliminaire a été déduite de l'analyse différentielle entre les tumeurs d'évolution métastatique (n=2) et les tumeurs d'évolution non métastatique (n=6). Cette première analyse ayant surtout servi de preuve de concept, a permis la mise en place de l'étude principale à partir d'une cohorte élargie. L'étude du transcriptome non codant réalisée à partir d'une série de 40 carcinomes mammaires, actuellement en cours de réalisation, nous permettra d'identifier de nouveaux types de lncRNAs parmi les nouveaux asRNAs, les circRNAs, les sno-lncRNAs, les T-UCRs, les lncRNAs inductibles, les lncRNAs traduits en peptides et les lncRNAs interagissant avec les régulateurs chromatiniens. Cette analyse renforcera la robustesse des signatures déjà identifiées sur la cohorte préliminaire et validera nos premiers résultats. Nous tenterons de confirmer si nos signatures constituées de lncRNAs dérégulés sont associées avec le statut RE+ et RE- des CCI et avec les différents sous-types moléculaires des tumeurs mammaires. Nous rechercherons si les sous-types tumoraux identifiés par des lncRNAs sont associés à des altérations des différentes fonctions biologiques et des voies de signalisation et nous comparerons ces résultats avec ceux obtenus pour des signatures constituées de mRNAs. En effet, une étude récente a révélé que si les sous-types tumoraux mammaires identifiés par des signatures basées sur les mRNAs et les lncRNAs affectent globalement les mêmes voies signalétiques, des différences significatives peuvent être observées dans certains sous-types (le sous-type de type triple-négatif est lié à la

voie EGFR dans la signature lncRNA) et pourraient justifier le recours à des signatures mixtes mRNAs-lncRNAs (van Grembergen et al., 2016). L'intérêt de certains de ces gènes ou signatures sera confirmé par l'analyse de leur niveau d'expression dans une large série de 456 carcinomes mammaires par RT-PCR ou technique *Nanostring*, dans laquelle *ANRIL* et *MALATI* ont été étudiés. Les lncRNAs sont actuellement considérés comme des biomarqueurs diagnostiques, pronostiques et prédictifs potentiels pertinents, en raison de leurs fonctions homéostatiques, de leur implication dans la plupart des cancers et de leur plus grande spécificité cellulaire et tissulaire. A partir de ces données, nous tenterons d'identifier des biomarqueurs tissulaires spécifiques et des cibles thérapeutiques potentielles.

Parmi les domaines d'investigation faisant l'objet de recherches intensives, celui des lncRNAs compétiteurs endogènes régulant les mRNAs via les miRNAs, présente actuellement le plus d'implications cliniques potentielles.

Les lncRNAs appartenant aux réseaux de compétiteurs endogènes sont représentés par des lincRNAs, des lncRNAs antisens, des pseudogènes et des lncRNAs circulaires. Ces réseaux d'interactions assurent le maintien de l'homéostasie cellulaire et tissulaire, varient considérablement entre les tissus normaux et tumoraux, présentent des profils transcriptomiques comparables dans les types tumoraux de même origine tissulaire et contrôlent la plupart des *hallmarks* des cancers. A partir d'un panel multitumoral de 12 types différents de tissus normaux et tumoraux, nous étudierons les niveaux d'expression des lncRNAs d'intérêt et vérifierons l'hypothèse concernant leur « spécificité d'organe ». Nous tenterons d'identifier des isoformes alternatifs des principaux lncRNAs impliqués dans les mécanismes de leurs dérégulations. Nous séparerons les réseaux de compétiteurs endogènes communs à l'ensemble des cancers de ceux qui sont propres à des types particuliers de tumeur et donc susceptibles de représenter des biomarqueurs pertinents et des cibles thérapeutiques innovantes.

BIBLIOGRAPHIE

Abdelhamid RF, Plessy C, Yamauchi Y, Taoka M, de Hoon M, Gingeras TR, Isobe T, Carninci P. Multiplicity of 5' cap structures present on short RNAs. *PLoS One* 2014 ;9:e102895. doi: 10.1371.

Abel Y, Clerget G, Bourguignon-Igel V, Salone V, Rederstorff M. Beyond usual functions of snoRNAs. *Med Sci (Paris)* 2014;30:297-302.

Adriaenssens E, Lottin S, Dugimont T, Fauquette W, Coll J, Dupouy JP, Boilly B, Cury JJ. Steroid hormones modulate H19 gene expression in both mammary gland and uterus. *Oncogene* 1999;18:4460-73.

Affymetrix ENCODE Transcriptome Project and Cold Spring Harbor Laboratory ENCODE Transcriptome Project. Post-transcriptional processing generates a diversity of 50-modified long and short RNAs. *Nature* 2009;457:1028–32.

Aguilo F, Di Cecilia S, Walsh MJ. Long Non-coding RNA ANRIL and Polycomb in Human Cancers and Cardiovascular Disease. *Curr Top Microbiol Immunol* 2016;394:29-39.

Allen TA, von Kaenel S, Goodrich JA, Kugel JF. The SINE-encoded mouse B2 RNA represses mRNA transcription in response to heat shock. *Nat Struct Mol Biol* 2004;11:816–21.

Amaral PP, Mattick JS. Noncoding RNA in development. *Mamm Genome* 2008;19:454–92.

Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol* 2010;11:R106. doi: 10.1186/gb-2010-11-10-r106.

Anderson SM, Rudolph MC, McManaman JL, Neville MC. Key stages in mammary gland development. Secretory activation in the mammary gland : it's not just about milk protein synthesis! *Breast Cancer Res* 2007;9:204. PubMed PMID: 17338830; PubMed Central PMCID: PMC1851396.

Anderson DM, Anderson KM, Chang CL, Makarewich CA, Nelson BR, McAnally JR, Kasaragod P, Shelton JM, Liou J, Bassel-Duby R, Olson EN. A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. *Cell* 2015;160:595-606.

Arase M, Horiguchi K, Ehata S, Morikawa M, Tsutsumi S, Aburatani H, Miyazono K, Koinuma D. Transforming growth factor- β -induced lncRNA-Smad7 inhibits apoptosis of mouse breast cancer JygMC(A) cells. *Cancer Sci* 2014;105:974-82.

Arun G, Diermeier S, Akerman M, Chang KC, Wilkinson JE, Hearn S, Kim Y, MacLeod AR, Krainer AR, Norton L, Brogi E, Egeblad M, Spector DL. Differentiation of mammary tumors and reduction in metastasis upon Malat1 lncRNA loss. *Genes Dev* 2016;30:34-51.

Askarian-Amiri ME, Crawford J, French JD, Smart CE, Smith MA, Clark MB, Ru K, Mercer TR, Thompson ER, Lakhani SR, Vargas AC, Campbell IG, Brown MA, Dinger ME, Mattick JS. SNORD-host RNA Zfas1 is a regulator of mammary development and a potential marker for breast cancer. *RNA* 2011;17:878-91.

Askarian-Amiri ME, Seyfoddin V, Smart CE, Wang J, Kim JE, Hansji H, Baguley BC, Finlay GJ, Leung EY. Emerging role of long non-coding RNA SOX2OT in SOX2 regulation in breast cancer. *PLoS One* 2014;9:e102140.

Atala A. The long noncoding RNA SchLAPI promotes aggressive prostate cancer and antagonizes the SWI/SNF complex. *J Urol* 2014;192:613. doi: 10.1016/j.juro.2014.05.031.

Atianand MK, Hu W, Satpathy AT, Shen Y, Ricci EP, Alvarez-Dominguez JR, Bhatta A, Schattgen SA, McGowan JD, Blin J, Braun JE, Gandhi P, Moore MJ, Chang HY, Lodish HF, Caffrey DR, Fitzgerald KA. A Long Noncoding RNA lincRNA-EPS Acts as a Transcriptional Brake to Restrain Inflammation. *Cell* 2016;165:1672-85.

Audas TE, Jacob MD, Lee S. Immobilization of proteins in the nucleolus by ribosomal intergenic spacer noncoding RNA. *Mol Cell* 2012;45:147-57.

Audas TE, Lee S. Stressing out over long noncoding RNA. *Biochim Biophys Acta* 2016;1859:184-91.

Augoff K, McCue B, Plow EF, Sossey-Alaoui K. miR-31 and its host gene lincRNA LOC554202 are regulated by promoter hypermethylation in triple-negative breast cancer. *Mol Cancer* 2012;11:5-17.

Ayers D. Long Non-Coding RNAs: Novel Emergent Biomarkers for Cancer Diagnostics. *J Cancer Res Treat* 2013;1:31-5.

Azzalin CM, Reichenbach P, Khoriauli L, Giulotto E, Lingner J. Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science* 2007;318:798-801.

Bak RO, Mikkelsen JG. miRNA sponges: soaking up miRNAs for regulation of gene expression. *Wiley Interdiscip Rev RNA* 2014;5:317-33.

Bardou F, Ariel F, Simpson CG, Romero-Barrios N, Laporte P, Balzergue S, Brown JW, Crespi M. Long noncoding RNA modulates alternative splicing regulators in Arabidopsis. *Dev Cell* 2014;30:166-76.

Bartolomei MS, Zemel S, Tilghman SM. Parental imprinting of the mouse H19 gene. *Nature* 1991;351:153-5.

Bartolomei MS. Genomic imprinting: employing and avoiding epigenetic processes. *Genes Dev* 2009;23:2124-33.

Barry G, Mattick JS. The role of regulatory RNA in cognitive evolution. *Trends Cogn Sci* 2012;16:497-503.

Barry G, Briggs JA, Vanichkina DP et al. The long noncoding RNA Gomafu is acutely regulated in response to neuronal activation and involved in schizophrenia-associated alternative splicing. *Mol Psychiatry* 2014;19:486-494.

Barsyte-Lovejoy D, Lau SK, Boutros PC, Khosravi F, Jurisica I, Andrulis IL, Tsao MS, Penn LZ. The c-Myc oncogene directly induces the H19 noncoding RNA by allele-specific binding to potentiate tumorigenesis. *Cancer Res* 2006;66:5330-7.

Baude A, Lindroth AM, Plass C. PRC2 loss amplifies RASS1 signaling in cancer. *Nat Genet* 2014;46:1154-5.

Beckedorff FC, Amaral MS, Deocesano-Pereira C, Verjovski-Almeida S. Long non-coding RNAs and their implications in cancer epigenetics. *Biosci Rep* 2013;33:e00061. doi: 10.1042/BSR20130054.

Bejerano G, Pheasant M, Makunin I, Stephen S, Kent WJ, Mattick JS, Haussler D. Ultraconserved elements in the human genome. *Science* 2004;304:1321-5.

Belancio VP, Deininger PL, Roy-Engel AM. LINE dancing in the human genome: transposable elements and disease. *Genome Med* 2009;1:97. doi:10.1186/gm97.

Beltran M, Puig I, Peña C, García JM, Alvarez AB, Peña R, Bonilla F, de Herreros AG. A natural antisense transcript regulates *Zeb2/Sip1* gene expression during Snail1-induced epithelial-mesenchymal transition. *Genes Dev* 2008;22:756-69.

Benjaminov A, Westhof E, Krol A. Distinctive structures between chimpanzee and human in a brain noncoding RNA. *RNA* 2008;14:1270-5.

Bennett CF, Swayze EE. RNA targeting therapeutics: molecular mechanisms of antisense oligonucleotides as a therapeutic platform. *Annu Rev Pharmacol Toxicol* 2010;50:259-93.

Berghoff EG, Clark MF, Chen S, Cajigas I, Leib DE, Kohtz JD. *Evf2 (Dlx6as)* lncRNA regulates ultraconserved enhancer methylation and the differential transcriptional control of adjacent genes. *Development* 2013;140:4407-16.

Bergmann JH, Spector DL. Long non-coding RNAs: modulators of nuclear structure and function. *Curr Opin Cell Biol* 2014;26:10-8.

Bernard D, Prasanth KV, Tripathi V, Colasse S, Nakamura T, Xuan Z, Zhang MQ, Sedel F, Jourdain L, Couplier F, Triller A, Spector DL, Bessis A. A long nuclear-retained non-coding RNA regulates synaptogenesis by modulating gene expression. *EMBO J* 2010;29:3082-93.

Berretta J, Morillon A. Pervasive transcription constitutes a new level of eukaryotic genome regulation. *EMBO Rep* 2009;10:973-82.

Bertani S, Sauer S, Bolotin E, Sauer F. The noncoding RNA *Mistral* activates *Hoxa6* and *Hoxa7* expression and stem cell differentiation by recruiting MLL1 to chromatin. *Mol Cell* 2011;43:1040-6.

Berteaux N, Lottin S, Monté D, Pinte S, Quatannens B, Coll J, Hondermarck H, Cury JJ, Dugimont T, Adriaenssens E. H19 mRNA-like noncoding RNA promotes breast cancer cell proliferation through positive control by E2F1. *J Biol Chem* 2005;280:29625-36.

Bertozzi D, Iurlaro R, Sordet O, Marinello J, Zaffaroni N, Capranico G. Characterization of novel antisense HIF-1 α transcripts in human cancers. *Cell Cycle* 2011;10:3189–97.

Bhan A, Hussain I, Ansari KI, Kasiri S, Bashyal A, Mandal SS. Antisense Transcript Long Noncoding RNA (lncRNA) HOTAIR is Transcriptionally Induced by Estradiol. *J Mol Biol* 2013;425:3707-22.

Bhartiya D, Pal K, Ghosh S, Kapoor S, Jalali S, Panwar B, Jain S, Sati S, Sengupta S, Sachidanandan C, Raghava GP, Sivasubbu S, Scaria V. lncRNome: a comprehensive knowledgebase of human long noncoding RNAs. *Database (Oxford)*. 2013;2013:bat034. doi: 10.1093/database/bat034.

Bida JP, Maher LJ 3rd. Improved prediction of RNA tertiary structure with insights into native state dynamics. *RNA* 2012;18:385-93.

Bonsang-Kitzis H, Sadacca B, Hamy-Petit AS, Moarii M, Pinheiro A, Laurent C, Reyat F. Biological network-driven gene selection identifies a stromal immune module as a key determinant of triple-negative breast carcinoma prognosis. *Oncoimmunology* 2015;5:e1061176.

Boon RA, Jaé N, Holdt L, Dimmeler S. Long Noncoding RNAs: From Clinical Genetics to Therapeutic Targets? *J Am Coll Cardiol* 2016;67:1214-26.

Braconi C, Valeri N, Kogure T, Gasparini P, Huang N, Nuovo GJ, et al. Expression and functional role of a transcribed noncoding RNA with an ultraconserved element in hepatocellular carcinoma. *Proc Natl Acad Sci U S A* 2011;108:786–91.

Braidotti G, Baubec T, Pauler F, Seidl C, Smrzka O, Stricker S, Yotova I, Barlow DP. The Air noncoding RNA: An imprinted cis-silencing transcript. *Cold Spring Harb Symp. Quant Biol* 2004;69:55–66.

Brinton LA, Schairer C, Hoover RN, Fraumeni JF Jr. Menstrual factors and risk of breast cancer. *Cancer Invest* 1988;6:245-54.

Briskin C, O'Malley B. Hormone action in the mammary gland. *Cold Spring Harb Perspect Biol* 2010;2:a003178. doi: 10.1101/cshperspect.a003178.

Briskin C. Progesterone signalling in breast cancer: a neglected hormone coming into the limelight. *Nat Rev Cancer* 2013;13:385-96.

Brown CJ, Hendrich BD, Rupert JL, Lafrenière RG, Xing Y, Lawrence J, Willard HF. 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* 1992;71:527-42.

Brown JA, Valenstein ML, Yario TA, Tycowski KT, Steitz JA. Formation of triple-helical structures by the 3'-end sequences of MALAT1 and MEN β noncoding RNAs. *Proc Natl Acad Sci U S A* 2012;109:19202–07.

Buechner J, Einvik C. N-myc and noncoding RNAs in neuroblastoma. *Mol Cancer Res* 2012;10:1243-53.

Bulger M, Groudine M. Functional and mechanistic diversity of distal transcription enhancers. *Cell* 2011;144:327–39.

Burd CE, Jeck WR, Liu Y, Sanoff HK, Wang Z, Sharpless NE. Expression of linear and novel circular forms of an INK4/ARF-associated non-coding RNA correlates with atherosclerosis risk. *PLoS Genet* 2010;6:e1001233. doi: 10.1371/journal.pgen.1001233.

Burnett JC, Rossi JJ, Tiemann K. Current progress of siRNA/shRNA therapeutics in clinical trials. *Biotechnol J* 2011;6:1130–46.

Burstein MD, Tsimelzon A, Poage GM, Covington KR, Contreras A, Fuqua SA, Savage MI, Osborne CK, Hilsenbeck SG, Chang JC, Mills GB, Lau CC, Brown PH. Comprehensive genomic analysis identifies novel subtypes and targets of triple-negative breast cancer. *Clin Cancer Res* 2015;21:1688–98.

Busch H, Reddy R, Rothblum L, Choi YC. SnRNAs, SnRNPs, and RNA processing. *Annu Rev Biochem* 1982;51:617–54.

Byun JS, Gardner K. Wounds that will not heal: pervasive cellular reprogramming in cancer *Am J Pathol* 2013;182:1055–64.

Cadoo KA, McArdle O, O'Shea AM, Power CP, Hennessy BT. Management of unusual histological types of breast cancer. *Oncologist* 2012;17:1135–45.

Cai X, Cullen BR. The imprinted H19 noncoding RNA is a primary microRNA precursor. *RNA* 2007;13:313–6.

Cai B, Wu Z, Liao K, Zhang S. Long noncoding RNA HOTAIR can serve as a common molecular marker for lymph node metastasis: a meta-analysis. *Tumour Biol* 2014;35:8445–50.

Cajigas I, Leib DE, Cochrane J et al Evf2 lncRNA/ RG1/DLX1 interactions reveal RNA-dependent chromatin Mechanisms of long noncoding RNA function in development and disease remodeling inhibition. *Development* 2015;142:2641–2652.

Calin GA, Liu CG, Ferracin M, Hyslop T, Spizzo R, Sevignani C, Fabbri M, Cimmino A, Lee EJ, Wojcik SE, Shimizu M, Tili E, Rossi S, Taccioli C, Pichiorri F, Liu X, Zupo S, Herlea V, Gramantieri L, Lanza G, Alder H, Rassenti L, Volinia S, Schmittgen TD, Kipps TJ, Negrini M, Croce CM. Ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas. *Cancer Cell* 2007;12:215–29.

Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* 2012;490:61–70.

Cancer Genome Atlas Research Network, Kandoth C, Schultz N, Cherniack AD, Akbani R, Liu Y, Shen H, Robertson AG, Pashtan I, Shen R, Benz CC, Yau C, Laird PW, Ding L, Zhang W, Mills GB, Kucherlapati R, Mardis ER, Levine DA. Integrated genomic characterization of endometrial carcinoma. *Nature* 2013;497:67–73.

Cao S, Liu W, Li F, Zhao W, Qin C. Decreased expression of lncRNA GAS5 predicts a poor prognosis in cervical cancer. *Int J Clin Exp Pathol* 2014;7:6776–83.

Capper R, Britt-Compton B, Tankimanova M, Rowson J, Letsolo B, Man S, Haughton M, Baird DM. The nature of telomere fusion and a definition of the critical telomere length in human cells. *Genes Dev* 2007;21:2495-508.

Caretti G, Schiltz RL, Dilworth FJ, Di Padova M, Zhao P, Ogryzko V, Fuller-Pace FV, Hoffman EP, Tapscott SJ, Sartorelli V. The RNA helicases p68/p72 and the noncoding RNA SRA are coregulators of MyoD and skeletal muscle differentiation. *Dev Cell* 2006;11:547-60.

Carnero E, Barriocanal M, Prior C, Pablo Unfried J, Segura V, Guruceaga E, Enguita M, Smerdou C, Gastaminza P, Fortes P. Long noncoding RNA EGOT negatively affects the antiviral response and favors HCV replication. *EMBO Rep* 2016;17:1013-28.

Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, Oyama R, Ravasi T, Lenhard B, Wells C, Kodzius R, Shimokawa K, Bajic VB, Brenner SE, Batalov S, Forrest AR, Zavolan M, Davis MJ, Wilming LG, Aidinis V, Allen JE, Ambesi-Impombato A, Apweiler R, Aturaliya RN, Bailey TL, Bansal M, Baxter L, Beisel KW, Bersano T, Bono H, Chalk AM, Chiu KP, Choudhary V, Christoffels A, Clutterbuck DR, Crowe ML, Dalla E, Dalrymple BP, de Bono B, Della Gatta G, di Bernardo D, Down T, Engstrom P, Fagiolini M, Faulkner G, Fletcher CF, Fukushima T, Furuno M, Futaki S, Gariboldi M, Georgii-Hemming P, Gingeras TR, Gojobori T, Green RE, Gustincich S, Harbers M, Hayashi Y, Hensch TK, Hirokawa N, Hill D, Huminiecki L, Iacono M, Ikeo K, Iwama A, Ishikawa T, Jakt M, Kanapin A, Katoh M, Kawasawa Y, Kelso J, Kitamura H, Kitano H, Kollias G, Krishnan SP, Kruger A, Kummerfeld SK, Kurochkin IV, Lareau LF, Lazarevic D, Lipovich L, Liu J, Liuni S, McWilliam S, Madan Babu M, Madera M, Marchionni L, Matsuda H, Matsuzawa S, Miki H, Mignone F, Miyake S, Morris K, Mottagui-Tabar S, Mulder N, Nakano N, Nakauchi H, Ng P, Nilsson R, Nishiguchi S, Nishikawa S, Nori F, Ohara O, Okazaki Y, Orlando V, Pang KC, Pavan WJ, Pavesi G, Pesole G, Petrovsky N, Piazza S, Reed J, Reid JF, Ring BZ, Ringwald M, Rost B, Ruan Y, Salzberg SL, Sandelin A, Schneider C, Schönbach C, Sekiguchi K, Semple CA, Seno S, Sessa L, Sheng Y, Shibata Y, Shimada H, Shimada K, Silva D, Sinclair B, Sperling S, Stupka E, Sugiura K, Sultana R, Takenaka Y, Taki K, Tammoja K, Tan SL, Tang S, Taylor MS, Tegner J, Teichmann SA, Ueda HR, van Nimwegen E, Verardo R, Wei CL, Yagi K, Yamanishi H, Zabarovsky E, Zhu S, Zimmer A, Hide W, Bult C, Grimmond SM, Teasdale RD, Liu ET, Brusica V, Quackenbush J, Wahlestedt C, Mattick JS, Hume DA, Kai C, Sasaki D, Tomaru Y, Fukuda S, Kanamori-Katayama M, Suzuki M, Aoki J, Arakawa T, Iida J, Imamura K, Itoh M, Kato T, Kawaji H, Kawagashira N, Kawashima T, Kojima M, Kondo S, Konno H, Nakano K, Ninomiya N, Nishio T, Okada M, Plessy C, Shibata K, Shiraki T, Suzuki S, Tagami M, Waki K, Watahiki A, Okamura-Oho Y, Suzuki H, Kawai J, Hayashizaki Y; FANTOM Consortium; RIKEN Genome Exploration Research Group and Genome Science Group (Genome Network Project Core Group). The transcriptional landscape of the mammalian genome. *Science* 2005;309:1559-63. Erratum in: *Science* 2006;311:1713.

Carpenter S. Long noncoding RNA: Novel links between gene expression and innate immunity. *Virus Res* 2016;212:137-45.

Cariaso M, Lennon G. SNPedia: a wiki supporting personal genome annotation, interpretation and analysis. *Nucleic Acids Res* 2012;40(Database issue):D1308-12. doi: 10.1093/nar/gkr798.

Carrieri C, Cimatti L, Biagioli M, Beugnet A, Zucchelli S, Fedele S, Pesce E, Ferrer I, Collavin L, Santoro C, Forrest AR, Carninci P, Biffo S, Stupka E, Gustincich S. Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. *Nature* 2012;491:454-7.

Caslini C. Transcriptional regulation of telomeric non-coding RNA: implications on telomere biology, replicative senescence and cancer. *RNA Biol* 2010;7:18–22.

Caudron-Herger M, Müller-Ott K, Mallm JP, Marth C, Schmidt U, Fejes-Tóth K, Rippe K. Coding RNAs with a non-coding function: maintenance of open chromatin structure. *Nucleus* 2011;2:410-24.

Cawley S, Bekiranov S, Ng HH, Kapranov P, Sekinger EA, Kampa D, Piccolboni A, Sementchenko V, Cheng J, Williams AJ et al. Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell* 2004;116:499–509.

Ceja-Rangel HA, Sánchez-Suárez P, Castellanos-Juárez E, Peñaroja-Flores R, Arenas-Aranda DJ, Gariglio P, Benítez-Bribiesca L. Shorter telomeres and high telomerase activity correlate with a highly aggressive phenotype in breast cancer cell lines. *Tumour Biol* 2016. [Epub ahead of print] PubMed PMID:27072825.

Cesana M, Cacchiarelli D, Legnini I, Santini T, Sthandier O, Chinappi M, Tramontano A, Bozzoni I. A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell* 2011;147:358-69. Erratum in: *Cell*. 2011;147(4):947.

Chaffer CL, Weinberg RA. A perspective on cancer cell metastasis. *Science* 2011;331:1559-64.

Chaumeil J, Le Baccon P, Wutz A, Heard E. A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. *Genes Dev* 2006;20:2223-37.

Cheetham SW, Gruhl F, Mattick JS, Dinger ME. Long noncoding RNAs and the genetics of cancer. *Br J Cancer* 2013;108:2419–25.

Chen LL, Carmichael GG. Altered nuclear retention of mRNAs containing inverted repeats in human embryonic stem cells: functional role of a nuclear noncoding RNA. *Mol Cell* 2009;35:467-78.

Chen J, Wang R, Zhang K, Chen LB. Long non-coding RNAs in non-small cell lung cancer as biomarkers and therapeutic targets. *J Cell Mol Med* 2014a;18:2425-36.

Chen Z, Li J, Tian L, Zhou C, Gao Y, Zhou F, Shi S, Feng X, Sun N, Yao R, Shao K, Li N, Qiu B, Tan F, He J. MiRNA expression profile reveals a prognostic signature for esophageal squamous cell carcinoma. *Cancer Lett* 2014b;350:34-42.

Chen ZX, Sturgill D, Qu J, Jiang H, Park S, Boley N, et al. Comparative validation of the *D. melanogaster* modENCODE transcriptome annotation. *Genome Res* 2014c;24:1209-23.

Chen Y, Li C, Tan C, Liu X. Circular RNAs: a new frontier in the study of human diseases. *J Med Genet* 2016a;53:359-65.

Chen LL. Linking Long Noncoding RNA Localization and Function. *Trends Biochem Sci* 2016b;41:761-72.

Chin K, de Solorzano CO, Knowles D, Jones A, Chou W, Rodriguez EG, Kuo WL, Ljung BM, Chew K, Myambo K, Miranda M, Krig S, Garbe J, Stampfer M, Yaswen P, Gray JW, Lockett SJ. In situ analyses of genome instability in breast cancer. *Nat Genet* 2004;36:984-8.

Chisholm KM, Wan Y, Li R, Montgomery KD, Chang HY, West RB. Detection of long non-coding RNA in archival tissue: correlation with polycomb protein expression in primary and metastatic breast carcinoma. *PLoS One* 2012;7:e47998.

Chooniedass-Kothari S, Hamedani MK, Troup S, Hubé F, Leygue E. The steroid receptor RNA activator protein is expressed in breast tumor tissues. *Int J Cancer* 2006;118:1054-9.

Choudhry H, Albukhari A, Morotti M, Haider S, Moralli D, Smythies J, Schödel J, Green CM, Camps C, Buffa F, Ratcliffe P, Ragoussis J, Harris AL, Mole DR. Tumor hypoxia induces nuclear paraspeckle formation through HIF-2 α dependent transcriptional activation of NEAT1 leading to cancer cell survival. *Oncogene* 2015;34:4482-90. Erratum in: *Oncogene* 2015;34:4546-58.

Chujo T, Yamazaki T, Hirose T. Architectural RNAs (arcRNAs): A class of long noncoding RNAs that function as the scaffold of nuclear bodies. *Biochim Biophys Acta* 2016;1859:139-46.

Chung S, Nakagawa H, Uemura M, Piao L, Ashikawa K, Hosono N, Takata R, Akamatsu S, Kawaguchi T, Morizono T, Tsunoda T, Daigo Y, Matsuda K, Kamatani N, Nakamura Y, Kubo M. Association of a novel long non-coding RNA in 8q24 with prostate cancer susceptibility. *Cancer Sci* 2011;102:245-52.

Ciccia A, Elledge SJ. The DNA damage response: making it safe to play with knives. *Mol Cell* 2010;40:179-204.

Cichon MA, Degnim AC, Visscher DW, Radisky DC. Microenvironmental influences that drive progression from benign breast disease to invasive breast cancer. *J Mammary Gland Biol Neoplasia* 2010;15:389-97.

Citti L, Rainaldi G. Synthetic hammerhead ribozymes as therapeutic tools to control disease genes. *Curr Gene Ther* 2005;5:11-24.

Clark MB, Mattick JS. Long noncoding RNAs in cell biology. *Semin Cell Dev Biol* 2011a;22:366-76.

Clark MB, Amaral PP, Schlesinger FS, Dinger ME, Taft RJ, Rinn JL, Ponting CP, Stadler PF, Morris KV, Morillon A, Rozowsky JS, Gerstein MB, Wahlestedt C, Hayashizaki Y, Carninci P, Gingeras TR, Mattick JS. The Reality of Pervasive Transcription. *PLoS Biol* 2011b;9:e1000625.

Clark MB, Johnston RL, Inostroza-Ponta M, Fox AH, Fortini E, Moscato P, Dinger ME, Mattick JS. Genome-wide analysis of long noncoding RNA stability. *Genome Res* 2012;22:885-98.

Clark BS, Blackshaw S. Long non-coding RNA-dependent transcriptional regulation in neuronal development and disease. *Front Genet* 2014;5:164-178.

Clemson CM, Hutchinson JN, Sara SA, Ensminger AW, Fox AH, Chess A, et al. An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. *Mol Cell* 2009;33:717-26.

Cohen SM. Everything old is new again: (linc)RNAs make proteins! *EMBO J* 2014;33:937-8.

Colley SM, Leedman PJ. SRA and its binding partners: an expanding role for RNA-binding coregulators in nuclear receptor-mediated gene regulation. *Crit Rev Biochem Mol Biol* 2009;44:25–33.

Colley SM, Leedman PJ. Steroid Receptor RNA Activator - A nuclear receptor coregulator with multiple partners: Insights and challenges. *Biochimie* 2011a;93:1966-72.

Colley HE, Hearnden V, Jones AV, Weinreb PH, Violette SM, Macneil S, Thornhill MH, Murdoch C. Development of tissue-engineered models of oral dysplasia and early invasive oral squamous cell carcinoma. *Br J Cancer* 2011b;105:1582-92.

Colombo T, Farina L, Macino G, Paci P. PVT1: a rising star among oncogenic long noncoding RNAs. *Biomed Res Int* 2015;2015:304208. doi: 10.1155/2015/304208.

Comings DE. Replicative heterogeneity of mammalian DNA. *Exp Cell Res* 1972;71:106-12.

Congrains Ada, Kei Kamide, Mitsuru Ohishi and Hiromi Rakugi. ANRIL: Molecular Mechanisms and Implications in Human Health. *J Mol Sci* 2013;14:1278-1292.

Cooper C, Guo J, Yan Y, Chooniedass-Kothari S, Hube F, Hamedani MK, Murphy LC, Myal Y, Leygue E. Increasing the relative expression of endogenous non-coding Steroid Receptor RNA Activator (SRA) in human breast cancer cells using modified oligonucleotides. *Nucleic Acids Res* 2009;37:4518-31.

Cooper C, Vincett D, Yan Y, Hamedani MK, Myal Y, Leygue E. Steroid Receptor RNA Activator bifaceted genetic system: Heads or Tails? *Biochimie* 2011;93:1973–80.

Core LJ, Waterfall JJ, Lis JT. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* 2008;322:1845-8.

Courjal F, Theillet C. Comparative genomic hybridization analysis of breast tumors with predetermined profiles of DNA amplification. *Cancer Res* 1997;57:4368-77.

Cowin P, Wysolmerski J. Molecular mechanisms guiding embryonic mammary gland development. *Cold Spring Harb Perspect Biol* 2010;2:a003251. doi: 10.1101/cshperspect.a003251.

Crea F, Clermont PL, Parolia A, Wang Y, Helgason CD. The non coding transcriptome as a dynamic regulator of cancer metastasis. *Cancer Metastasis Rev* 2014a;33:1–16.

Crea F, Watahiki A, Quagliata L, Xue H, Pikor L, Parolia A, et al. Identification of a long non-coding RNA as a novel biomarker and potential therapeutic target for metastatic prostate cancer. *Oncotarget* 2014b;5:764-74.

Crick FH. On protein synthesis. *Symp Soc Exp Biol* 1958;12:138-63.

Croce CM. LINCing chromatin remodeling to metastasis. *Nat Biotechnol* 2010;28:931-2.

Cruickshanks HA, Vafadar-Isfahani N, Dunican DS, Lee A, Sproul D, Lund JN, Meehan RR, Tufarelli C. Expression of a large LINE-1-driven antisense RNA is linked to epigenetic silencing of the metastasis suppressor gene TFPI-2 in cancer. *Nucleic Acids Res* 2013;41:6857-69.

Cui H, Onyango P, Brandenburg S, Wu Y, Hsieh CL, Feinberg AP. Loss of imprinting in colorectal cancer linked to hypomethylation of H19 and IGF2. *Cancer Res* 2002;62:6442-6.

Cui Z, Ren S, Lu J, Wang F, Xu W, Sun Y, et al. The prostate cancer-up-regulated long noncoding RNA PlncRNA-1 modulates apoptosis and proliferation through reciprocal regulation of androgen receptor. *Urologic oncology* 2013;1117-23.

Cui M, Xiao Z, Wang Y, Zheng M, Song T, Cai X, Sun B, Ye L, Zhang X. Long noncoding RNA HULC modulates abnormal lipid metabolism in hepatoma cells through an miR-9-mediated RXRA signaling pathway. *Cancer Res* 2015;75:846-57.

Cui M, You L, Ren X, Zhao W, Liao Q, Zhao Y. Long non-coding RNA PVT1 and cancer. *Biochem Biophys Res Commun* 2016;471:10-14.

Cunnington MS, Santibanez Koref M, Mayosi BM, Burn J, Keavney B. Chromosome 9p21 SNPs Associated with Multiple Disease Phenotypes Correlate with ANRIL Expression. *PLoS Genet* 2010;6:e1000899. doi: 10.1371/journal.pgen.1000899.

Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, Gräf S, Ha G, Haffari G, Bashashati A, Russell R, McKinney S; METABRIC Group, Langerød A, Green A, Provenzano E, Wishart G, Pinder S, Watson P, Markowitz F, Murphy L, Ellis I, Purushotham A, Børresen-Dale AL, Brenton JD, Tavaré S, Caldas C, Aparicio S. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 2012;486:346-52.

Cusanelli E, Romero CA, Chartrand P. Telomeric noncoding RNA TERRA is induced by telomere shortening to nucleate telomerase molecules at short telomeres. *Mol Cell* 2013;51:780-91.

Da Sacco L, Baldassarre A, Masotti A. Bioinformatics tools and novel challenges in long non-coding RNAs (lncRNAs) functional analysis. *Int J Mol Sci* 2012;13:97-114.

Darfeuille F, Reigadas S, Hansen JB, Orum H, Di Primo C, Toulmé JJ. Aptamers targeted to an RNA hairpin show improved specificity compared to that of complementary oligonucleotides. *Biochemistry* 2006;45:12076-82.

Das S, Ghosal S, Sen R, Chakrabarti J. InCeDB: database of human long noncoding RNA acting as competing endogenous RNA. *PLoS One* 2014;9:e98965. doi: 10.1371/journal.pone.0098965.

Das PP, Hendrix DA, Apostolou E, Buchner AH, Canver MC, Beyaz S, Ljuboja D, Kuintzle R, Kim W, Karnik R, Shao Z, Xie H, Xu J, De Los Angeles A, Zhang Y, Choe J, Jun DL, Shen X, Gregory RI, Daley GQ, Meissner A, Kellis M, Hochedlinger K, Kim J, Orkin SH. PRC2 Is Required to

Maintain Expression of the Maternal Gtl2-Rian-Mirg Locus by Preventing De Novo DNA Methylation in Mouse Embryonic Stem Cells. *Cell Rep* 2015;12:1456-70.

De Andres-Pablo A, Morillon A, Wery M. LncRNAs, lost in translation or licence to regulate? *Curr Genet* 2016. [Epub ahead of print] Review. PubMed PMID:27230909.

De los Campos G, Gianola D, Allison DB. Predicting genetic predisposition in humans: the promise of whole-genome markers. *Nat Rev Genet* 2010;11:880-6.

De Souza FS, Franchini LF, Rubinstein M. Exaptation of transposable elements into novel cis-regulatory elements: is the evidence always strong? *Mol Biol Evol* 2013;30:1239-51.

Del Castillo M, Chibon F, Arnould L, Croce S, Ribeiro A, Perot G, Hostein I, Geha S, Bozon C, Garnier A, Lae M, Vincent-Salomon A, MacGrogan G. Secretory Breast Carcinoma : A Histopathologic and Genomic Spectrum Characterized by a Joint Specific ETV6-NTRK3 Gene Fusion. *Am J Surg Pathol* 2015;39:1458-67.

Delouis C. Physiology of colostrum production. *Ann Rech Vet* 1978;9:193-203.

Demongeot J, Hazgui H, Bandiera S, Cohen O, Henrion-Caude A. MitomiRs, chloromiRs and modelling of the microRNA inhibition. *Acta Biotheor* 2013;61:367-83.

Denegri M, Moralli D, Rocchi M, Biggiogera M, Raimondi E, Cobianchi F, De Carli L, Riva S, Biamonti G. Human chromosomes 9, 12, and 15 contain the nucleation sites of stress-induced nuclear bodies. *Mol Biol Cell* 2002;13:2069-79.

Deng G, Sui G. Noncoding RNA in oncogenesis: a new era of identifying key players. *Int J Mol Sci* 2013;14:18319-49.

Deng L, Yang SB, Xu FF, Zhang JH. Long noncoding RNA CCAT1 promotes hepatocellular carcinoma progression by functioning as let-7 sponge. *J Exp Clin Cancer Res* 2015;34:18. doi: 10.1186/s13046-015-0136-7.

Deng M, Brägelmann J, Schultze JL, Perner S. Web-TCGA: an online platform for integrated analysis of molecular cancer data sets. *BMC Bioinformatics* 2016a;17:72. doi: 10.1186/s12859-016-0917-9.

Deng K, Wang H, Guo X, Xia J. The cross talk between long, non-coding RNAs and microRNAs in gastric cancer. *Acta Biochim Biophys Sin (Shanghai)*. 2016b;48:111-6.

Deng W, Wang J, Zhang J, Cai J, Bai Z, Zhang Z. TET2 regulates LncRNA-ANRIL expression and inhibits the growth of human gastric cancer cells. *IUBMB Life* 2016c;68:355-64.

Denoeud F, Kapranov P, Ucla C, Frankish A, Castelo R, Drenkow J, Lagarde J, Alioto T, Manzano C, Chrast J, Dike S, Wyss C, Henrichsen CN, Holroyd N, Dickson MC, Taylor R, Hance Z, Foissac S, Myers RM, Rogers J, Hubbard T, Harrow J, Guigó R, Gingeras TR, Antonarakis SE, Reymond A. Prominent use of distal 5' transcription start sites and discovery of a large number of additional exons in ENCODE regions. *Genome Res* 2007;17:746-59.

Derrien T, Guigó R, Johnson R. The Long Non-Coding RNAs: A New (P)layer in the "Dark Matter". *Front Genet* 2012a;2:107. doi: 10.3389/fgene.

Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, et al. The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Res* 2012b;22:1775–89.

Dey BK, Pfeifer K, Dutta A et al. The H19 long noncoding RNA gives rise to microRNAs miR-675-3p and miR-675-5p to promote skeletal muscle differentiation and regeneration. *Genes Dev* 2014; 28:491–501.

Dhamija S, Diederichs S. From junk to master regulators of invasion: lncRNA functions in migration, EMT and metastasis. *Int J Cancer* 2016;139:269-80.

Di Gesualdo F, Capaccioli S, Lulli M. A pathophysiological view of the long non-coding RNA world. *Oncotarget* 2014;5:10976–96.

Di Ruscio A, Ebralidze AK, Benoukraf T, Amabile G, Goff LA, Terragni J, Figueroa ME, de Figueiredo PL, Alberich-Jorda M, Zhang P et al. DNMT1-interacting RNAs block gene-specific DNA methylation. *Nature* 2013;503:371–76.

Dieci G, Fiorino G, Castelnuovo M, Teichmann M, Pagano A. The expanding RNA polymerase III transcriptome. *Trends Genet* 2007;23:614-22.

Dieci MV, Criscitiello C, Goubar A, Viale G, Conte P, Guarneri V, Ficarra G, Mathieu MC, Delalogue S, Curigliano G, Andre F. Prognostic value of tumor-infiltrating lymphocytes on residual disease after primary chemotherapy for triple-negative breast cancer: a retrospective multicenter study. *Ann Oncol* 2014a;25:611-8.

Dieci MV, Criscitiello C, Goubar De Raedt T, Beert E, Pasmant E, Luscan A, Brems H, Ortonne N, et al. PRC2 loss amplifies Ras-driven transcription and confers sensitivity to BRD4-based therapies. *Nature* 2014b;514:247-51.

Diederichs S. The four dimensions of noncoding RNA conservation. *Trends Genet* 2014;30:121-3.

Dimitrova N, Zamudio JR, Jong RM, Soukup D, Resnick R, Sarma K, Ward AJ, Raj A, Lee JT, Sharp PA et al. LincRNA-p21 activates p21 in cis to promote polycomb target gene expression and to enforce the G1/S checkpoint. *Mol Cell* 2014;54:777–90.

Dinger ME, Gascoigne DK, Mattick JS. The evolution of RNAs with multiple functions. *Biochimie* 2011;93:2013–18.

Djebali S, Kapranov P, Foissac S, Lagarde J, Reymond A, Ucla C, Wyss C, Drenkow J, Dumais E, Murray RR, Lin C, Szeto D, Denoeud F, Calvo M, Frankish A, Harrow J, Makrythanasis P, Vidal M, Salehi-Ashtiani K, Antonarakis SE, Gingeras TR, Guigó R. Efficient targeted transcript discovery via array-based normalization of RACE libraries. *Nat Methods* 2008;5:629-35.

Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A. Landscape of transcription in human cells. *Nature* 2012;489:101-8.

Dogini DB, Pascoal VD, Avansini SH, Vieira AS, Pereira TC, Lopes-Cendes I. The new world of RNAs. *Genet Mol Biol* 2014;37(1 Suppl):285-93.

Dorsett Y, Tuschl T. siRNAs: applications in functional genomics and potential as therapeutics. *Nat Rev Drug Discov* 2004;3:318–29.

Douc-Rasy S, Barrois M, Fogel S, Ahomadegbe JC, Stéhelin D, Coll J, Riou G. High incidence of loss of heterozygosity and abnormal imprinting of H19 and IGF2 genes in invasive cervical carcinomas. Uncoupling of H19 and IGF2 expression and biallelic hypomethylation of H19. *Oncogene* 1996;12:423-30.

Du W, Pogoriler J. Retinoblastoma family genes. *Oncogene* 2006;25:5190-200.

Du Y, Kong G, You X, Zhang S, Zhang T, Gao Y, Ye L, Zhang X. Elevation of highly up-regulated in liver cancer (HULC) by hepatitis B virus X protein promotes hepatoma cell proliferation via down-regulating p18. *J Biol Chem* 2012;287:26302-11.

Du Z, Fei T, Verhaak RG, Su Z, Zhang Y, Brown M, Chen Y, Liu S. Integrative genomic analyses reveal clinically relevant long noncoding RNA in human cancer. *Nat Struct Mol Biol* 2013;20:908–913.

Du Toit A. Non-coding RNA: RNA stability control by Pol II. *Nat Rev Mol Cell Biol* 2013;14:128–29.

Duret L, Chureau C, Samain S, Weissenbach J, Avner P. The Xist RNA gene evolved in eutherians by pseudogenization of a protein-coding gene. *Science* 2006;312:1653-5.

Eades G, Zhang YS, Li QL, Xia JX, Yao Y, Zhou Q. Long non-coding RNAs in stem cells and cancer. *World J Clin Oncol* 2014;5:134-41.

Easwaran H, Hsing-Chen Tsai H, Baylin SB; Cancer epigenetics: Tumor Heterogeneity, Plasticity of Stem-like States, and Drug Resistance. *Mol Cell* 2014; 54:716–727.

Ebbesen KK, Kjems J, Hansen TB. Circular RNAs: Identification, biogenesis and function. *Biochim Biophys Acta* 2016;1859:163-8.

Edwards CA, Ferguson-Smith AC. Mechanisms regulating imprinted genes in clusters. *Curr Opin Cell Biol* 2007;19:281-9.

Ellis MJ, Ding L, Shen D, Luo J, Suman VJ, Wallis JW, Van Tine BA, Hoog J, Goiffon RJ, Goldstein TC, Ng S, Lin L, Crowder R, Snider J, Ballman K, Weber J, Chen K, Koboldt DC, Kandoth C, Schierding WS, McMichael JF, Miller CA, Lu C, Harris CC, McLellan MD, Wendl MC, DeSchryver K, Allred DC, Esserman L, Unzeitig G, Margenthaler J, Babiera GV, Marcom PK, Guenther JM, Leitch M, Hunt K, Olson J, Tao Y, Maher CA, Fulton LL, Fulton RS, Harrison M, Oberkfell B, Du F, Demeter R, Vickery TL, Elhammali A, Piwnica-Worms H, McDonald S, Watson M, Dooling DJ, Ota

D, Chang LW, Bose R, Ley TJ, Piwnica-Worms D, Stuart JM, Wilson RK, Mardis ER. Whole-genome analysis informs breast cancer response to aromatase inhibition. *Nature* 2012;486:353-60.

Elshamy WM, Duhé RJ. Overview: cellular plasticity, cancer stem cells and metastasis. *Cancer Lett* 2013;341:2-8.

ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012;489:57-74.

Engreitz J.M, Pandya-Jones A, McDonel P, Shishkin A, Sirokman K, Surka C, Kadri S, Xing J, Goren A, Lander E.S et al. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* 2013;341:1237973.

Faghihi MA, Modarresi F, Khalil AM, Wood DE, Sahagan BG, Morgan TE, Finch CE, St Laurent G 3rd, Kenny PJ, Wahlestedt C. Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase. *Nat Med* 2008;14:723-30.

Falahi F, van Kruchten M, Martinet N, Hospers GA, Rots MG. Current and upcoming approaches to exploit the reversibility of epigenetic mutations in breast cancer. *Breast Cancer Res* 2014;16:412. doi: 10.1186/s13058-014-0412-2.

Fan M, Li X, Jiang W, Huang Y, Li J, Wang Z. A long non-coding RNA, PTCSC3, as a tumor suppressor and a target of miRNAs in thyroid cancer cells. *Exp Ther Med* 2013;5:1143–6.

Fan Y, Shen B, Tan M, Mu X, Qin Y, Zhang F, Liu Y. Long non-coding RNA UCA1 increases chemoresistance of bladder cancer cells by regulating Wnt signaling. *FEBS J* 2014;281:1750-8.

Fang K, Han B-W, Chen Z-H, Lin K-Y, Zeng C-W, Li X-J, et al. A distinct set of long non-coding RNAs in childhood MLL-rearranged acute lymphoblastic leukemia: biology and epigenetic target. *Hum Mol Genet* 2014;23:3278–88.

Fang XY, Pan HF, Leng RX, Ye DQ. Long noncoding RNAs: novel insights into gastric cancer. *Cancer Lett* 2015;356:357-66.

Fang Y, Fullwood MJ. Roles, Functions, and Mechanisms of Long Non-coding RNAs in Cancer. *Genomics Proteomics Bioinformatics* 2016;14:42-54.

Fang Z, Xu C, Li Y, Cai X, Ren S, Liu H, Wang Y, Wang F, Chen R, Qu M, Wang Y, Zhu Y, Zhang W, Shi X, Yao J, Gao X, Hou J, Xu C, Sun Y. A feed-forward regulatory loop between androgen receptor and PlncRNA-1 promotes prostate cancer progression. *Cancer Lett* 2016b;374:62-74.

FANTOM Consortium and the RIKEN PMI and CLST (DGT), Forrest AR, Kawaji H, Rehli M, Baillie JK, de Hoon MJ, Hablerle V, Lassmann T, Kulakovskiy IV, Lizio M, Itoh M, Andersson R, Mungall CJ, Meehan TF, Schmeier S, Bertin N, Jørgensen M, Dimont E, Arner E, Schmidl C, Schaefer U, Medvedeva YA, Plessy C, Vitezic M, Severin J, Semple C, Ishizu Y, Young RS, Francescato M, Alam I, Albanese D, Altschuler GM, Arakawa T, Archer JA, Arner P, Babina M, Rennie S, Balwierz PJ, Beckhouse AG, Pradhan-Bhatt S, Blake JA, Blumenthal A, Bodega B, Bonetti A, Briggs J, Brombacher F, Burroughs AM, Califano A, Cannistraci CV, Carbajo D, Chen Y, Chierici

M, Ciani Y, Clevers HC, Dalla E, Davis CA, Detmar M, Diehl AD, Dohi T, Drabløs F, Edge AS, Edinger M, Ekwall K, Endoh M, Enomoto H, Fagiolini M, Fairbairn L, Fang H, Farach-Carson MC, Faulkner GJ, Favorov AV, Fisher ME, Frith MC, Fujita R, Fukuda S, Furlanello C, Furino M, Furusawa J, Geijtenbeek TB, Gibson AP, Gingeras T, Goldowitz D, Gough J, Guhl S, Guler R, Gustincich S, Ha TJ, Hamaguchi M, Hara M, Harbers M, Harshbarger J, Hasegawa A, Hasegawa Y, Hashimoto T, Herlyn M, Hitchens KJ, Ho Sui SJ, Hofmann OM, Hoof I, Hori F, Huminiecki L, Iida K, Ikawa T, Jankovic BR, Jia H, Joshi A, Jurman G, Kaczkowski B, Kai C, Kaida K, Kaiho A, Kajiyama K, Kanamori-Katayama M, Kasianov AS, Kasukawa T, Katayama S, Kato S, Kawaguchi S, Kawamoto H, Kawamura YI, Kawashima T, Kempfle JS, Kenna TJ, Kere J, Khachigian LM, Kitamura T, Klinken SP, Knox AJ, Kojima M, Kojima S, Kondo N, Koseki H, Koyasu S, Krampitz S, Kubosaki A, Kwon AT, Laros JF, Lee W, Lennartsson A, Li K, Lilje B, Lipovich L, Mackay-Sim A, Manabe R, Mar JC, Marchand B, Mathelier A, Mejhert N, Meynert A, Mizuno Y, de Lima Morais DA, Morikawa H, Morimoto M, Moro K, Motakis E, Motohashi H, Mummery CL, Murata M, Nagao-Sato S, Nakachi Y, Nakahara F, Nakamura T, Nakamura Y, Nakazato K, van Nimwegen E, Ninomiya N, Nishiyori H, Noma S, Noma S, Nozaki T, Ogishima S, Ohkura N, Ohimiya H, Ohno H, Ohshima M, Okada-Hatakeyama M, Okazaki Y, Orlando V, Ovchinnikov DA, Pain A, Passier R, Patrikakis M, Persson H, Piazza S, Prendergast JG, Rackham OJ, Ramiłowski JA, Rashid M, Ravasi T, Rizzu P, Roncador M, Roy S, Rye MB, Saijyo E, Sajantila A, Saka A, Sakaguchi S, Sakai M, Sato H, Savvi S, Saxena A, Schneider C, Schultes EA, Schulze-Tanzil GG, Schwegmann A, Sengstag T, Sheng G, Shimoji H, Shimoni Y, Shin JW, Simon C, Sugiyama D, Sugiyama T, Suzuki M, Suzuki N, Swoboda RK, Hoen PA, Tagami M, Takahashi N, Takai J, Tanaka H, Tatsukawa H, Tatum Z, Thompson M, Toyodo H, Toyoda T, Valen E, van de Wetering M, van den Berg LM, Verado R, Vijayan D, Vorontsov IE, Wasserman WW, Watanabe S, Wells CA, Winteringham LN, Wolvetang E, Wood EJ, Yamaguchi Y, Yamamoto M, Yoneda M, Yonekura Y, Yoshida S, Zabierowski SE, Zhang PG, Zhao X, Zucchelli S, Summers KM, Suzuki H, Daub CO, Kawai J, Heutink P, Hide W, Freeman TC, Lenhard B, Bajic VB, Taylor MS, Makeev VJ, Sandelin A, Hume DA, Carninci P, Hayashizaki Y. A promoter-level mammalian expression atlas. *Nature* 2014;507:462-70.

Fanucchi S, Shibayama Y, Burd S, Weinberg MS, Mhlanga MM. Chromosomal contact permits transcription between coregulated genes. *Cell* 2013;155:606-20.

Farazi TA, Juranek SA, Tuschl T. The growing catalog of smallRNAs and their association with distinct Argonaute/Piwi family members. *Development* 2008;135:1201–14.

Fasolo A, Sessa C. Targeting mTOR pathways in human malignancies. *Curr Pharm Des* 2012;18:2766-77.

Fatemi RP, Velmeshev D, Faghihi MA. De-repressing LncRNA-Targeted Genes to Upregulate Gene Expression: Focus on Small Molecule Therapeutics. *Mol Ther Nucleic Acids* 2014;3:e196.

Fatima R, Akhade VS, Pal D, Rao SM. Long noncoding RNAs in development and cancer: potential biomarkers and therapeutic targets. *Mol Cell Ther* 2015;3:5.doi: 10.1186/s40591-015-0042-6.

Faulkner GJ, Kimura Y, Daub CO, Wani S, Plessy C, Irvine KM, Schroder K, Cloonan N, Steptoe AL, Lassmann T, Waki K, Hornig N, Arakawa T, Takahashi H, Kawai J, Forrest AR, Suzuki H, Hayashizaki Y, Hume DA, Orlando V, Grimmond SM, Carninci P. The regulated retrotransposon transcriptome of mammalian cells. *Nat Genet* 2009;41:563-71.

Feil R. Epigenetics: Ready for the marks. *Nature* 2009;461:359-60.

Fellig Y, Ariel I, Ohana P, Schachter P, Sinelnikov I, Birman T, Ayesh S, Schneider T, de Groot N, Czerniak A, Hochberg A. H19 expression in hepatic metastases from a range of human carcinomas. *J Clin Pathol* 2005;58:1064-8.

Feinberg AP, Vogelstein B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* 1983;301:89-92.

Ferdin J, Nishida N, Wu X, Nicoloso MS, Shah MY, Devlin C, Ling H, Shimizu M, Kumar K, Cortez MA, Ferracin M, Bi Y, Yang D, Czerniak B, Zhang W, Schmittgen TD, Voorhoeve MP, Reginato MJ, Negrini M, Davuluri RV, Kunej T, Ivan M, Calin GA. HINCUTs in cancer: hypoxia-induced noncoding ultraconserved transcripts. *Cell Death Differ* 2013;20:1675-87.

Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F: Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer* 2015;136:359-386.

Ferreira LB, Palumbo A, de Mello KD, Sternberg C, Caetano MS, de Oliveira FL, et al. PCA3 noncoding RNA is involved in the control of prostate-cancer cell survival and modulates androgen receptor signaling. *BMC Cancer* 2012;12:507.

Flippot R, Malouf GG, Su X, Mouawad R, Spano JP, Khayat D. Cancer subtypes classification using long non-coding RNA. *Oncotarget* 2016. doi:10.18632/oncotarget.10213.

Flynn RA and Chang HY. Long noncoding RNAs in cell-fate programming and reprogramming. *Cell Stem Cell* 2014;14:752-61.

Folkersen L, Kyriakou T, Goel A, Peden J, Mälarstig A, Paulsson-Berne G, et al. Relationship between CAD risk genotype in the chromosome 9p21 locus and gene expression. Identification of eight new ANRIL splice variants. *PLoS One* 2009;4:e7677.

Foulds CE, Tsimelzon A, Long W, Le A, Tsai SY, Tsai MJ, O'Malley BW. Research resource: expression profiling reveals unexpected targets and functions of the human steroid receptor RNA activator (SRA) gene. *Mol Endocrinol* 2010;24:1090-105.

Foulkes WD, Stefansson IM, Chappuis PO, Bégin LR, Goffin JR, Wong N, Trudel M, Akslen LA. Germline BRCA1 mutations and a basal epithelial phenotype in breast cancer. *J Natl Cancer Inst* 2003;95:1482-5.

Frankish A, Harrow J. GENCODE pseudogenes. *Methods Mol Biol* 2014;1167:129-55.

Frith MC, Bailey TL, Kasukawa T, Mignone F, Kummerfeld SK, Madera M, Sunkara S, Furuno M, Bult CJ, Quackenbush J, Kai C, Kawai J, Carninci P, Hayashizaki Y, Pesole G, Mattick JS. Discrimination of non-protein-coding transcripts from protein-coding mRNA. *RNA Biol* 2006;3:40-8.

Fujimoto A, Furuta M, Totoki Y, Tsunoda T, Kato M, Shiraishi Y, Tanaka H, Taniguchi H, Kawakami Y, Ueno M, Gotoh K, Ariizumi S, Wardell CP, Hayami S, Nakamura T, Aikata H, Arihiro K,

Boroevich KA, Abe T, Nakano K, Maejima K, Sasaki-Oku A, Ohsawa A, Shibuya T, Nakamura H, Hama N, Hosoda F, Arai Y, Ohashi S, Urushidate T, Nagae G, Yamamoto S, Ueda H, Tatsuno K, Ojima H, Hiraoka N, Okusaka T, Kubo M, Marubashi S, Yamada T, Hirano S, Yamamoto M, Ohdan H, Shimada K, Ishikawa O, Yamaue H, Chayama K, Miyano S, Aburatani H, Shibata T, Nakagawa H. Whole-genome mutational landscape and characterization of noncoding and structural mutations in liver cancer. *Nat Genet* 2016;48:500-9. Erratum: Whole-genome mutational landscape and characterization of noncoding and structural mutations in liver cancer. *Nat Genet* 2016;48:700. doi: 10.1038/ng0616-700a.

Gabory A, Jammes H, Dandolo L. The H19 locus: role of an imprinted non-coding RNA in growth and development. *Bioessays* 2010;32:473-80.

Ganesan S, Silver DP, Greenberg RA, Avni D, Drapkin R, Miron A, Mok SC, Randrianarison V, Brodie S, Salstrom J, Rasmussen TP, Klimke A, Marrese C, Marahrens Y, Deng CX, Feunteun J, Livingston DM. BRCA1 supports XIST RNA concentration on the inactive X chromosome. *Cell* 2002;111:393-405.

Gao W, Chan JY, Wong TS. Differential expression of long noncoding RNA in primary and recurrent nasopharyngeal carcinoma. *Biomed Res Int* 2014:404567.

Ge X, Chen Y, Liao X, Liu D, Li F, Ruan H, Jia W. Overexpression of long noncoding RNA PCAT-1 is a novel biomarker of poor prognosis in patients with colorectal cancer. *Med Oncol* 2013;30:588-599.

Geisler S, Collier J. RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. *Nat Rev Mol Cell Biol* 2013;14:699–712.

Geng YJ, Xie SL, Li Q, Ma J, Wang GY. Large intervening non-coding RNA HOTAIR is associated with hepatocellular carcinoma progression. *J Int Med Res* 2011;39:2119–28.

Gibb EA, Brown CJ, Lam WL. The functional role of long non-coding RNA in human carcinomas. *Mol Cancer* 2011;10:38. doi: 1186/1476-4598-10-38.

Ginger MR, Shore AN, Contreras A, Rijnkels M, Miller J, Gonzalez-Rimbau MF, Rosen JM. A noncoding RNA is a potential marker of cell fate during mammary gland development. *Proc Natl Acad Sci U S A* 2006;103:5781-6.

Gloss BS and Dinger ME. The specificity of long noncoding RNA expression. *Biochim Biophys Acta* 2016;1859:16-22.

Gnant M, Filipits M, Greil R, Stoeger H, Rudas M, Bago-Horvath Z, Mlineritsch B, Kwasny W, Knauer M, Singer C, Jakesz R, Dubsky P, Fitzal F, Bartsch R, Steger G, Balic M, Ressler S, Cowens JW, Strohoff J, Ferree S, Schaper C, Liu S, Fesl C, Nielsen TO; Austrian Breast and Colorectal Cancer Study Group. Predicting distant recurrence in receptor-positive breast cancer patients with limited clinicopathological risk : using the PAM50 Risk of Recurrence score in 1478 postmenopausal patients of the ABCSG-8 trial treated with adjuvant endocrine therapy alone. *Ann Oncol* 2014;25:339-45.

- Godlewski J, Nowicki MO, Bronisz A, Williams S, Otsuki A, Nuovo G, Raychaudhury A, Newton HB, Chiocca EA, Lawler S. Targeting of the Bmi-1 oncogene/stem cell renewal factor by microRNA-128 inhibits glioma proliferation and self-renewal. *Cancer Res* 2008;68:9125-30.
- Gómez-Maldonado L, Tiana M, Roche O, Prado-Cabrero A, Jensen L, Fernandez-Barral A, Guijarro-Muñoz I, Favaro E, Moreno-Bueno G, Sanz L, Aragonés J, Harris A, Volpert O, Jimenez B, del Peso L. EFNA3 long noncoding RNAs induced by hypoxia promote metastatic dissemination. *Oncogene* 2015;34:2609-20.
- Goncalves R, Warner WA, Luo J, Ellis MJ. New concepts in breast cancer genomics and genetics. *Breast Cancer Res* 2014;16:460.
- Gong C, Maquat LE. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. *Nature* 2011;470:284-8.
- Graham LD, Pedersen SK, Brown GS, Ho T, Kassir Z, Moynihan AT, et al. Colorectal Neoplasia Differentially Expressed (CRNDE), a Novel Gene with Elevated Expression in Colorectal Adenomas and Adenocarcinomas. *Genes Cancer* 2011;2:829-40.
- Gredell JA, Berger AK, Walton SP. Impact of target mRNA structure on siRNA silencing efficiency: A large-scale study. *Biotechnol Bioeng* 2008;100:744-55.
- Groen JN, Capraro D, Morris KV. The emerging role of pseudogene expressed non-coding RNAs in cellular functions. *Int J Biochem Cell Biol* 2014;54:350-5.
- Grote P, Wittler L, Hendrix D et al. The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. *Dev Cell* 2013;24:206-214.
- Guarnerio J, Bezzi M, Jeong JC, Paffenholz SV, Berry K, Naldini MM, Lo-Coco F, Tay Y, Beck AH, Pandolfi PP. Oncogenic Role of Fusion-circRNAs Derived from Cancer-Associated Chromosomal Translocations. *Cell* 2016;165:289-302.
- Guffanti A, Iacono M, Pelucchi P, Kim N, Soldà G, Croft LJ, Taft RJ, Rizzi E, Askarian-Amiri M, Bonnal RJ, Callari M, Mignone F, Pesole G, Bertalot G, Bernardi LR, Albertini A, Lee C, Mattick JS, Zucchi I, De Bellis G. A transcriptional sketch of a primary human breast cancer by 454 deep sequencing. *BMC Genomics* 2009;10:163. doi: 10.1186/1471-2164-10-163.
- Gumireddy K, Li A, Yan J, Setoyama T, Johannes GJ, Orom UA, Tchou J, Liu Q, Zhang L, Speicher DW, Calin GA, Huang Q. Identification of a long non-coding RNA-associated RNP complex regulating metastasis at the translational step. *EMBO J* 2013;32:2672-84.
- Guo F, Li Y, Liu Y, Wang J, Li Y, Li G. Inhibition of metastasis-associated lung adenocarcinoma transcript 1 in CaSki human cervical cancer cells suppresses cell proliferation and invasion. *Acta Biochim Biophys Sin (Shanghai)* 2010;42:224-9.
- Guo X, Tang Y. OCT4 pseudogenes present in human leukemia cells. *Clin Exp Med* 2012;12:207-16.
- Guo Q, Qian Z, Yan D, Li L, Huang L. LncRNA-MEG3 inhibits cell proliferation of endometrial carcinoma by repressing Notch signaling. *Biomed Pharmacother* 2016;82:589-594.

Guohui Wan, Rohit Mathur, Xiaoxiao Hu, Yunhua Liu, Xinna Zhang, Guang Peng, Xiongbin Lu. Long non-coding RNA ANRIL (CDKN2B-AS) is induced by the ATM-E2F1 signaling pathway. *Cellular Signalling* 2013;25:1086–1095.

Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 2010;464:1071–6.

Gusterson BA, Stein T: Human breast development. *Semin Cell Dev Biol* 2012;23:567-573.

Gutschner T, Diederichs S. The hallmarks of cancer: a long non-coding RNA point of view. *RNA Biol* 2012;9:703–19.

Gutschner T, Hämmerle M, Eissmann M, Hsu J, Kim Y, Hung G, et al. The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. *Cancer Res* 2013a;73:1180–9.

Gutschner T, Hämmerle M and Diederichs S. MALAT1, a paradigm for long noncoding RNA function in cancer *J Mol Med* 2013b;DOI 10.1007/s00109-013-1028-y.

Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* 2009;458:223–7.

Guttman M, Donaghey J, Carey BW, Garber M, Grenier JK, Munson G, Young G, Lucas AB, Ach R, Bruhn L, Yang X, Amit I, Meissner A, Regev A, Rinn JL, Root DE, Lander ES. lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* 2011;477:295-300.

Guttman M, Rinn JL. Modular regulatory principles of large non-coding RNAs. *Nature* 2012;482:339-46. doi: 10.1038/nature10887.

Ha H, Song J, Wang S, Kapusta A, Feschotte C, Chen KC, Xing J. A comprehensive analysis of piRNAs from adult human testis and their relationship with genes and mobile elements. *BMC Genomics* 2014b;15:545.

Haemmerle M, Gutschner T. Long non-coding RNAs in cancer and development : where do we go from here? *Int J Mol Sci* 2015;16:1395-405.

Hah N, Danko CG, Core L, Waterfall JJ, Siepel A, Lis JT, Kraus WL. A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. *Cell* 2011;145:622-34.

Hah N, Murakami S, Nagari A, Danko CG, Kraus WL. Enhancer transcripts mark active estrogen receptor binding sites. *Genome Res* 2013;23:1210-23.

Hah N, Kraus WL. Hormone-regulated transcriptomes: lessons learned from estrogen signaling pathways in breast cancer cells. *Mol Cell Endocrinol* 2014a;382:652-64.

Hacisuleyman E, Goff LA, Trapnell C, Williams A, Henao-Mejia J, Sun L, et al. Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre. *Nat Struct Mol Biol* 2014;21:198–206.

Hackermüller J, Reiche K, Otto C, Hösler N, Blumert C, Brocke-Heidrich K, Böhlig L, Nitsche A, Kasack K, Ahnert P, Krupp W, Engeland K, Stadler PF, Horn F. Cell cycle, oncogenic and tumor suppressor pathways regulate numerous long and macro non-protein-coding RNAs. *Genome Biol* 2014;15:R48. doi: 10.1186/gb-2014-15-3-r48.

Haemmerle M, Gutschner T. Long non-coding RNAs in cancer and development: where do we go from here? *Int J Mol Sci* 2015;16:1395–405.

Haffari G et al: The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* 2012;486:395-399.

Hajjari M, Behmanesh M, Sadeghizadeh M, Zeinoddini M. Up-regulation of HOTAIR long non-coding RNA in human gastric adenocarcinoma tissues. *Med Oncol Northwood Lond Engl* 2013a;30:670.

Hajjari M, Khoshnevisan A, Shin YK. Long non-coding RNAs in hematologic malignancies: road to translational research. *Front Genet* 2013b;4:250.

Halley P, Khorkova O, Wahlestedt C. Natural antisense transcripts as therapeutic targets. *Drug Discov Today Ther Strateg* 2013;10:e119-e125.

Halley P, Kadakkuzha BM, Faghihi MA, Magistri M, Zeier Z, Khorkova O, Coito C, Hsiao J, Lawrence M, Wahlestedt C. Regulation of the apolipoprotein gene cluster by a long noncoding RNA. *Cell Rep* 2014;6:222-30.

Han Y, Liu Y, Gui Y, Cai Z. Long intergenic non-coding RNA TUG1 is overexpressed in urothelial carcinoma of the bladder. *J Surg Oncol* 2013;107:555–9.

Han X, Yang F, Cao H, Liang Z. Malat1 regulates serum response factor through miR-133 as a competing endogenous RNA in myogenesis. *FASEB J* 2015;29:3054-64.

Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell* 2000;100:57–70.

Hanahan D, Weinberg RA: Hallmarks of cancer: the next generation. *Cell* 2011;144:646-674.

Hanahan D, Coussens LM. Accessories to the crime: functions of cells recruited to the tumor microenvironment. *Cancer Cell* 2012;21:309-22.

Hansji H, Leung EY, Baguley BC, Finlay GJ, Askarian-Amiri ME. Keeping abreast with long non-coding RNAs in mammary gland development and breast cancer. *Front Genet* 2014;5:37.

Harismendy O, Notani D, Song X, Rahim NG, Tanasa B, Heintzman N, Ren B, Fu XD, Topol EJ, Rosenfeld MG, Frazer KA. 9p21 DNA variants associated with coronary artery disease impair interferon- γ signalling response. *Nature* 2011;470:264-8.

Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S, Barnes I, Bignell A, Boychenko V, Hunt T, Kay M, Mukherjee G, Rajan J, Despacio-Reyes G, Saunders G, Steward C, Harte R, Lin M, Howald C, Tanzer A, Derrien T, Chrast J, Walters N, Balasubramanian S, Pei B, Tress M, Rodriguez JM, Ezkurdia I, van Baren J, Brent M, Haussler D, Kellis M, Valencia A, Reymond A, Gerstein M, Guigó R, Hubbard TJ. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* 2012;22:1760-74.

He S, Gu W, Li Y, Zhu H. ANRIL/CDKN2B-AS shows two-stage clade-specific evolution and becomes conserved after transposon insertions in simians. *BMC Evol Biol* 2013;13:247. doi: 10.1186/1471-2148-13-247.

Hendriks WT, Jiang X, Daheron L, Cowan CA. TALEN- and CRISPR/Cas9-Mediated Gene Editing in Human Pluripotent Stem Cells Using Lipid-Based Transfection. *Curr Protoc Stem Cell Biol* 2015;34:5B.3.1-5B.3.25.

Herriges MJ, Swarr DT, Morley MP, Rathi KS, Peng T, Stewart KM, Morrissey EE. Long noncoding RNAs are spatially correlated with transcription factors and regulate lung development. *Genes Dev* 2014;28:1363-79.

Hessels D, Schalken JA. The use of PCA3 in the diagnosis of prostate cancer. *Nat Rev Urol* 2009;6:255–61.

Hesson LB, Packham D, Pontzer E, Funchain P, Eng C, Ward RL. A reinvestigation of somatic hypermethylation at the PTEN CpG island in cancer cell lines. *Biol Proced Online* 2012;14:5. doi: 10.1186/1480-9222-14-5.

Heuston EF, Lemon KT, Arceci RJ. The Beginning of the Road for Non-Coding RNAs in Normal Hematopoiesis and Hematologic Malignancies. *Front Genet* 2011;2:94.

Hezroni H, Koppstein D, Schwartz MG, Avrutin A, Bartel DP, Ulitsky I. Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. *Cell Rep* 2015;11:1110-22.

Hibi K, Nakamura H, Hirai A, Fujikake Y, Kasai Y, Akiyama S, Ito K, Takagi H. Loss of H19 imprinting in esophageal cancer. *Cancer Res* 1996;56:480-2.

Hirose T, Mishima Y, Tomari Y. Elements and machinery of non-coding RNAs: toward their taxonomy. *EMBO Rep* 2014;15:489-507.

Holley RW, Apgar J, Everett GA, Madison JT, Marquisie M, Merrill SH, Penswick JR, Zamir A. Structure of a ribonucleic acid. *Science* 1965;147:1462-5.

Hu G, Tang Q, Sharma S, Yu F, Escobar TM, Muljo SA, Zhu J, Zhao K. Expression and regulation of intergenic long noncoding RNAs during T cell development and differentiation. *Nat Immunol* 2013;14:1190-8.

Hu L, Di C, Kai M, Yang YC, Li Y, Qiu Y, Hu X, Yip KY, Zhang MQ et al. Coding RNAs: Critical Players in Hepatocellular Carcinoma. *Int J Mol Sci* 2014a;15:20434-48.

- Hu Y, Chen HY, Yu CY, Xu J, Wang JL, Qian J, Zhang X, Fang JY. A long non-coding RNA signature to improve prognosis prediction of colorectal cancer. *Oncotarget* 2014b;5:2230-42.
- Hu X, Feng Y, Zhang D, Zhao SD, Hu Z, Greshock J, Zhang Y, Yang L, Zhong X, Wang LP, Jean S, Li C, Huang Q, Katsaros D, Montone KT, Tanyi JL, Lu Y, Boyd J, Nathanson KL, Li H, Mills GB, Zhang L. A functional genomic approach identifies FAL1 as an oncogenic long noncoding RNA that associates with BMI1 and represses p21 expression in cancer. *Cancer Cell* 2014c;26:344-57.
- Hu L, Wu Y, Tan D, Meng H, Wang K, Bai Y, Yang K. Up-regulation of long noncoding RNA MALAT1 contributes to proliferation and metastasis in esophageal squamous cell carcinoma. *J Exp Clin Cancer Res* 2015;34:7.
- Hu G, Gong AY, Wang Y, Ma S, Chen X, Chen J, Su CJ, Shibata A, Strauss-Soukup JK, Drescher KM, Chen XM. LincRNA-Cox2 Promotes Late Inflammatory Gene Transcription in Macrophages through Modulating SWI/SNF-Mediated Chromatin Remodeling. *J Immunol* 2016;196:2799-808.
- Huang JF, Guo YJ, Zhao CX, Yuan SX, Wang Y, Tang GN, Zhou WP, Sun SH. Hepatitis B virus X protein (HBx)-related long noncoding RNA (lncRNA) down-regulated expression by HBx (Dreh) inhibits hepatocellular carcinoma metastasis by targeting the intermediate filament protein vimentin. *Hepatology* 2013;57:1882-92.
- Huang J, Zhou N, Watabe K, Lu Z, Wu F, Xu M, Mo YY. Long non-coding RNA UCA1 promotes breast tumor growth by suppression of p27 (Kip1). *Cell Death Dis* 2014;5: e1008.
- Huang MD, Chen WM, Qi FZ, Xia R, Sun M, Xu TP, Yin L, Zhang EB, De W, Shu YQ. Long non-coding RNA ANRIL is upregulated in hepatocellular carcinoma and regulates cell apoptosis by epigenetic silencing of KLF2. *J Hematol Oncol* 2015;8:50. doi: 10.1186/s13045-015-0146-0.
- Huarte M, Rinn JL. Large non-coding RNAs: missing links in cancer? *Hum Mol Genet* 2010;19:R152–61.
- Huarte M, Guttman M, Feldser D, Garber M, Koziol MJ, Kenzelmann-Broz D, et al. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* 2010;142:409–19.
- Hube F, Guo J, Chooniedass-Kothari S, Cooper C, Hamedani MK, Dibrov AA, et al. Alternative splicing of the first intron of the steroid receptor RNA activator (SRA) participates in the generation of coding and noncoding RNA isoforms in breast cancer cell lines. *DNA Cell Biol* 2006;25:418–28.
- Hung T, Wang Y, Lin MF, Koegel AK, Kotake Y, Grant GD, et al. Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nat Genet* 2011;43:621–29.
- Hung KH, Wang Y, Zhao JC. Regulation of Mammalian Gene Dosage by Long Noncoding RNAs. *Biomolecules* 2013;3:124-42.
- Hutchins AP, Pei D. Transposable elements at the center of the crossroads between embryogenesis, embryonic stem cells, reprogramming, and long non-coding RNAs. *Sci Bull (Beijing)* 2015;60:1722-1733.

Hutchinson JN, Ensminger AW, Clemson CM, Lynch CR, Lawrence JB, Chess A. A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. *BMC Genomics* 2007;8:39.

Iacoangeli A, Lin Y, Morley EJ, Muslimov IA, Bianchi R, Reilly J, Weedon J, Diallo R, Böcker W, Tiedge H. BC200 RNA in invasive and preinvasive breast cancer. *Carcinogenesis* 2004;25:2125-33.

Iacoangeli A, Tiedge H. Translational control at the synapse: role of RNA regulators. *Trends Biochem Sci* 2013;38:47-55.

Iacobucci I, Sazzini M, Garagnani P, Ferrari A, Boattini A, Lonetti A, et al. A polymorphism in the chromosome 9p21 ANRIL locus is associated to Philadelphia positive acute lymphoblastic leukemia. *Leuk Res* 2011;35:1052–9.

Iden M, Fye S, Li K, Chowdhury T, Ramchandran R, Rader JS. The lncRNA PVT1 Contributes to the Cervical Cancer Phenotype and Associates with Poor Patient Prognosis. *PLoS One* 2016;11:e0156274. doi:10.1371.

Ilboudo A, Chouhan J, McNeil BK, Osborne JR, Ogunwobi OO. PVT1 Exon 9: A Potential Biomarker of Aggressive Prostate Cancer? *Int J Environ Res Public Health* 2015;13:ijerph13010012. doi: 10.3390/ijerph13010012.

Imamura K, Imamachi N, Akizuki G, Kumakura M, Kawaguchi A, Nagata K, Kato A, Kawaguchi Y, Sato H, Yoneda M, Kai C, Yada T, Suzuki Y, Yamada T, Ozawa T, Kaneki K, Inoue T, Kobayashi M, Kodama T, Wada Y, Sekimizu K, Akimitsu N. Long noncoding RNA NEAT1-dependent SFPQ relocation from promoter region to paraspeckle mediates IL8 expression upon immune stimuli. *Mol Cell* 2014;53:393-406.

International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* 2004;431:931-45.

Ishibashi M, Kogo R, Shibata K, Sawada G, Takahashi Y, Kurashige J, et al. Clinical significance of the expression of long non-coding RNA HOTAIR in primary hepatocellular carcinoma. *Oncol Rep* 2013;29:946–50.

Iwakiri J, Hamada M, Asai K. Bioinformatics tools for lncRNA research. *Biochim Biophys Acta* 2016;1859:23-30.

Iyer MK, Niknafs YS, Malik R et al. The landscape of long noncoding RNAs in the human transcriptome. *Nature Genetics* 2015;47:199–208.

Jacob MD, Audas TE, Uniacke J, Trinkle-Mulcahy L, Lee S. Environmental cues induce a long noncoding RNA-dependent remodeling of the nucleolus. *Mol Biol Cell* 2013;24:2943-53.

Jafri MA, Ansari SA, Alqahtani MH, Shay JW. Roles of telomeres and telomerase in cancer, and advances in telomerase-targeted therapies. *Genome Med* 2016;8:69. doi: 10.1186/s13073-016-0324-x.

Jarinova O, Stewart AF, Roberts R, Wells G, Lau P, Naing T, Buerki C, McLean BW, Cook RC, Parker JS, McPherson R. Functional analysis of the chromosome 9p21.3 coronary artery disease risk locus. *Arterioscler Thromb Vasc Biol* 2009;29:1671-7.

Jeck WR, Sorrentino JA, Wang K, Slevin MK, Burd CE, Liu J, Marzluff WF, Sharpless NE. Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* 2013;19:141-57.

Jendrzewski J, He H, Radomska HS, Li W, Tomsic J, Liyanarachchi S, Davuluri RV, Nagy R, de la Chapelle A. The polymorphism rs944289 predisposes to papillary thyroid carcinoma through a large intergenic noncoding RNA gene of tumor suppressor type. *Proc Natl Acad Sci U S A* 2012;109:8646-51.

Jeon BG, Rho GJ, Betts DH, Petrik JJ, Favetta LA, King WA. Low levels of X-inactive specific transcript in somatic cell nuclear transfer embryos derived from female bovine freemartin donor cells. *Sex Dev* 2012;6:151-9.

Ji P, Diederichs S, Wang W, Böing S, Metzger R, Schneider PM, et al. MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene* 2003;22:8031-41.

Ji Q, Zhang L, Liu X, Zhou L, Wang W, Han Z, et al. Long non-coding RNA MALAT1 promotes tumour growth and metastasis in colorectal cancer through binding to SFPQ and releasing oncogene PTBP2 from SFPQ/PTBP2 complex. *Br J Cancer* 2014;111:736-48.

Jiang Q, Wang J, Wang Y, Ma R, Wu X, Li Y. TF. LncRNA: identifying common transcription factors for a list of lncRNA genes from ChIP-Seq data. *Biomed Res Int* 2014;2014:317642.

Jiang Y, Malouf GG, Zhang J, Zheng X, Chen Y, Thompson EJ, Weinstein JN, Yuan Y, Spano JP, Broaddus R, Tannir NM, Khayat D, Lu KH, Su X. Long non-coding RNA profiling links subgroup classification of endometrioid endometrial carcinomas with trithorax and polycomb complex aberrations. *Oncotarget* 2015;6:39865-76.

Jinno Y, Ikeda Y, Yun K, Maw M, Masuzaki H, Fukuda H, Inuzuka K, Fujishita A, Ohtani Y, Okimoto T, et al. Establishment of functional imprinting of the H19 gene in human developing placentae. *Nat Genet* 1995;10:318-24.

Johnson R, Richter N, Jauch R, Gaughwin PM, Zuccato C, Cattaneo E, Stanton LW. Human accelerated region 1 noncoding RNA is repressed by REST in Huntington's disease. *Physiol Genomics* 2010;41:269-74.

Johnson R, Guigó R. The RIDL hypothesis: transposable elements as functional domains of long noncoding RNAs. *RNA* 2014;20:959-76.

Johnsson P, Ackley A, Vidarsdottir L, Lui WO, Corcoran M, Grandér D, Morris KV. A pseudogene long-noncoding-RNA network regulates PTEN transcription and translation in human cells. *Nat Struct Mol Biol* 2013;20:440-6.

- Johnsson P, Lipovich L, Grandér D, Morris KV. Evolutionary conservation of long non-coding RNAs; sequence, structure, function. *Biochim Biophys Acta* 2014;1840:1063-71.
- Jolly S, Kestin LL, Goldstein NS, Vicini FA. The impact of lobular carcinoma in situ in association with invasive breast cancer on the rate of local recurrence in patients with early-stage breast cancer treated with breast-conserving therapy. *Int J Radiat Oncol Biol Phys* 2006;66:365-71.
- Jones PA, Baylin SB. The fundamental role of epigenetic events in cancer. *Nat Rev Genet* 2002;3:415-28.
- Kallen AN, Zhou XB, Xu J, Qiao C, Ma J, Yan L, Lu L, Liu C, Yi JS, Zhang H, Min W, Bennett AM, Gregory RI, Ding Y, Huang Y. The imprinted H19 lncRNA antagonizes let-7 microRNAs. *Mol Cell* 2013;52:101-12.
- Kalyana-Sundaram S, Kumar-Sinha C, Shankar S, Robinson DR, Wu YM, Cao X, Asangani IA, Kothari V, Prensner JR, Lonigro RJ, Iyer MK, Barrette T, Shanmugam A, Dhanasekaran SM, Palanisamy N, Chinnaiyan AM. Expressed pseudogenes in the transcriptional landscape of human cancers. *Cell*. 2012;149:1622-34.
- Kanduri C, Whitehead J, Mohammad F. The long and the short of it: RNA-directed chromatin asymmetry in mammalian X-chromosome inactivation. *FEBS Lett* 2009;583:857-64.
- Kanduri C. *Kcnq1ot1*: A chromatin regulatory RNA. *Sem Cell Dev Biol* 2011;22:343–50.
- Kaneko S, Li G, Son J, Xu CF, Margueron R, Neubert TA, Reinberg D. Phosphorylation of the PRC2 component *Ezh2* is cell cycle-regulated and up-regulates its binding to ncRNA. *Genes Dev* 2010;24:2615-20.
- Kapranov P, Cawley SE, Drenkow J, Bekiranov S, Strausberg RL, Fodor SP, Gingeras TR. Large-scale transcriptional activity in chromosomes 21 and 22. *Science* 2002;296:916-9.
- Kapranov P, Cheng J, Dike S, Nix D.A, Dutttagupta R, Willingham A.T, Stadler P.F, Hertel J, Hackermuller J, Hofacker I.L et al. RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* 2007;316:1484–88.
- Kapranov P, Ozsolak F, Kim SW, Foissac S, Lipson D, Hart C, Roels S, Borel C, Antonarakis SE, Monaghan AP, John B, Milos PM. New class of gene-termini-associated human RNAs suggests a novel RNA copying mechanism. *Nature* 2010;466:642-6.
- Kapranov P, St Laurent G, Raz T, Ozsolak F, Reynolds CP, Sorensen PH, Reaman G, Milos P, Arceci RJ, Thompson JF, Triche TJ. The majority of total nuclear-encoded non-ribosomal RNA in a human cell is 'dark matter' un-annotated RNA. *BMC Biol* 2010;8:149. doi: 10.1186/1741-7007-8-149. Erratum in: *BMC Biol* 2011;9:86.
- Kapusta A, Kronenberg Z, Lynch VJ, Zhuo X, Ramsay L, Bourque G, Yandell M, Feschotte C. Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. *PLoS Genet* 2013;9:e1003470.doi: 10.1371/journal.pgen.1003470.

Kawakami T, Okamoto K, Sugihara H, Hattori T, Reeve AE, Ogawa O, Okada Y. The roles of supernumerical X chromosomes and XIST expression in testicular germ cell tumors. *J Urol* 2003;169:1546-52.

Kawakami T, Okamoto K, Ogawa O, Okada Y. XIST unmethylated DNA fragments in male-derived plasma as a tumour marker for testicular cancer. *Lancet* 2004;363:40-2.

Ke H, Zhao L, Feng X, Xu H, Zou L, Yang Q, Su X, Peng L, Jiao B. NEAT1 is Required for Survival of Breast Cancer Cells Through FUS and miR-548. *Gene Regul Syst Bio* 2016;10:11-7.

Kelley D, Rinn J. Transposable elements reveal a stem cell-specific class of long noncoding RNAs. *Genome Biol* 2012;13:R107. doi: 10.1186/gb-2012-13-11-r107.

Keniry A, Oxley D, Monnier P, Kyba M, Dandolo L, Smits G, Reik W. The H19 lincRNA is a developmental reservoir of miR-675 that suppresses growth and Igf1r. *Nat Cell Biol* 2012;14:659–65.

Key TJ, Verkasalo PK, Banks E. Epidemiology of breast cancer. *Lancet Oncol* 2001;2:133-40.

Khaitovich P, Kelso J, Franz H, Visagie J, Giger T, Joerchel S, Petzold E, Green, R.E, Lachmann M, Paabo S. Functionality of intergenic transcription: An evolutionary comparison. *PLoS Genet* 2006;2:e171.

Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, et al. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci U S A* 2009;106:11667–72.

Khurana E, Fu Y, Chakravarty D, Demichelis F, Rubin MA, Gerstein M. Role of non-coding sequence variants in cancer. *Nat Rev Genet* 2016;17:93-108.

Kim NH, Lee CH, Lee AY. H19 RNA downregulation stimulated melanogenesis in melasma. *Pigment Cell Melanoma Res* 2010;23:84-92.

Kim ED, Sung S. Long noncoding RNA: unveiling hidden layer of gene regulatory networks. *Trends Plant Sci* 2012;17:16-21.

Kim T, Reitmair A. Non-Coding RNAs: Functional Aspects and Diagnostic Utility in Oncology. *Int J Mol Sci* 2013a;14:4934-68.

Kim K, Jutooru I, Chadalapaka G, Johnson G, Frank J, Burghardt R, Kim S, Safe S. HOTAIR is a negative prognostic factor and exhibits pro-oncogenic activity in pancreatic cancer. *Oncogene* 2013b;32:1616-25.

Kino T, Hurt DE, Ichijo T, Nader N, Chrousos GP. Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. *Sci Signal* 2010;3:ra8. doi:10.1126/scisignal.2000568.

Kitagawa M, Kotake Y, Ohhata T. Long non-coding RNAs involved in cancer development and cell fate determination. *Curr Drug Targets* 2012;13:1616–21.

Klattenhoff CA, Scheuermann JC, Surface LE, Bradley RK, Fields PA, Steinhauser ML, Ding H, Butty VL, Torrey L, Haas S, Abo R, Tabebordbar M, Lee RT, Burge CB, Boyer LA. Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. *Cell* 2013;152:570-83.

Knoll M, Lodish HF, Sun L. Long non-coding RNAs as regulators of the endocrine system. *Nat Rev Endocrinol* 2015;11:151-60.

Knott GJ, Bond CS, Fox AH. The DBHS proteins SFPQ, NONO and PSPC1: a multipurpose molecular scaffold. *Nucleic Acids Res* 2016;44:3989-4004.

Koerner MV, Pauler FM, Huang R, Barlow DP. The function of non-coding RNAs in genomic imprinting. *Development* 2009;136:1771-83.

Kogo R, Shimamura T, Mimori K, Kawahara K, Imoto S, Sudo T, et al. Long noncoding RNA HOTAIR regulates polycomb-dependent chromatin modification and is associated with poor prognosis in colorectal cancers. *Cancer Res* 2011;71:6320–6.

Kondo Y, Shen L, Ahmed S, Bumber Y, Sekido Y, Haddad BR, Issa JP. Downregulation of histone H3 lysine 9 methyltransferase G9a induces centrosome disruption and chromosome instability in cancer cells. *PLoS One* 2008;3:e2037. doi: 10.1371/journal.pone.0002037.

Kong R, Zhang EB, Yin DD, You LH, Xu TP, Chen WM, Xia R, Wan L, Sun M, Wang ZX, De W, Zhang ZH. Long noncoding RNA PVT1 indicates a poor prognosis of gastric cancer and promotes cell proliferation through epigenetically regulating p15 and p16. *Mol Cancer* 2015;14:82. doi: 10.1186/s12943-015-0355-8.

Korneev SA, Korneeva EI, Lagarkova MA, Kiselev SL, Critchley G, O'Shea M. Novel noncoding antisense RNA transcribed from human anti-NOS2A locus is differentially regulated during neuronal differentiation of embryonic stem cells. *RNA* 2008;14:2030-7.

Kornfeld JW, Brüning JC. Regulation of metabolism by long, non-coding RNAs. *Front Genet* 2014;5:57.

Kotake Y, Nakagawa T, Kitagawa K, Suzuki S, Liu N, Kitagawa M, et al. Long non-coding RNA anril is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene. *Oncogene* 2011;30:1956–62.

Kotula JW, Pratico ED, Ming X, Nakagawa O, Juliano RL, Sullenger BA. Aptamer-mediated delivery of splice-switching oligonucleotides to the nuclei of cancer cells. *Nucleic Acid Ther* 2012;22:187-95.

Kour S, Rath PC. Long noncoding RNAs in aging and age-related diseases. *Ageing Res Rev* 2016;26:1-21.

Kouzarides T. Chromatin modifications and their function. *Cell* 2007;128:693-705.

Krawczyk M, Emerson BM. p50-associated COX-2 extragenic RNA (PACER) activates COX-2 gene expression by occluding repressive NF- κ B complexes. 2014. *elife* 3:e01776.

Kretz M, Webster DE, Flockhart RJ, Lee CS, Zehnder A, Lopez-Pajares V, et al. Suppression of progenitor differentiation requires the long noncoding RNA ANCR. *Genes Dev* 2012;26:338–43.

Kretz M, Saprashvili Z, Chu C et al. Control of somatic tissue differentiation by the long non-coding RNA TINCR. *Nature* 2013;493:231–235.

Kunej T, Obsteter J, Pogacar Z, Horvat S, Calin GA. The decalog of long non-coding RNA involvement in cancer diagnosis and monitoring. *Crit Rev Clin Lab Sci* 2014;51:344-57.

Kung JT, Colognori D, Lee JT. Long noncoding RNAs: past, present, and future. *Genetics* 2013;193:651-69.

Lai M, Yang Z, Zhou L, Zhu Q, Xie H, Zhang F, et al. Long non-coding RNA MALAT-1 overexpression predicts tumor recurrence of hepatocellular carcinoma after liver transplantation. *Med Oncol* 2012;29:1810–6.

Lai F, Orom UA, Cesaroni M, Beringer M, Taatjes DJ, Blobel GA, Shiekhattar R. Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature* 2013;494:497-501.

Lanctôt C, Cheutin T, Cremer M, Cavalli G, Cremer T. Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. *Nat Rev Genet* 2007;8:104-15.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann Y, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissole SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blöcker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A,

Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kasprzyk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowki J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ, Szustakowki J; International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* 2001;409:860-921. Erratum in: *Nature* 2001;411:720.

Lanz RB, McKenna NJ, Onate SA, Albrecht U, Wong J, Tsai SY, et al. A steroid receptor coactivator, SRA, functions as a RNA and is present in an SRC-1 complex. *Cell* 1999;97:17-27.

Lanz RB, Chua SS, Barron N, Söder BM, DeMayo F, O'Malley BW. Steroid receptor RNA activator stimulates proliferation as well as apoptosis in vivo. *Mol Cell Biol* 2003;23:7163-76.

Lauressergues D, Couzigou JM, Clemente HS, Martinez Y, Dunand C, Bécard G, Combier JP. Primary transcripts of microRNAs encode regulatory peptides. *Nature* 2015;520:90-3.

Lazorthes S, Vallot C, Briois S, Aguirrebengoa M, Thuret JY, St Laurent G, Rougeulle C, Kapranov P, Mann C, Trouche D, Nicolas E. Erratum: A vlincRNA participates in senescence maintenance by relieving H2AZ-mediated repression at the INK4 locus. *Nat Commun* 2015;6:6918. doi: 10.1038/ncomms7918.

Lee JT, Davidow LS, Warshawsky D. Tsix, a gene antisense to Xist at the X-inactivation centre. *Nat Genet* 1999;21:400-4.

Lee JT. Lessons from X-chromosome inactivation: long ncRNA as guides and tethers to the epigenome. *Genes Dev* 2009;23:1831-42.

Lee JT. Epigenetic regulation by long noncoding RNAs. *Science* 2012;338:1435-39.

Lee S, Kopp F, Chang TC, Sataluri A, Chen B, Sivakumar S, Yu H, Xie Y, Mendell JT. Noncoding RNA NORAD Regulates Genomic Stability by Sequestering PUMILIO Proteins. *Cell* 2016;164:69-80.

Leeb M, Steffen PA, Wutz A. X chromosome inactivation sparked by non-coding RNAs. *RNA Biol* 2009;6:94-9.

Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, Pietenpol JA. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J Clin Invest* 2011;121:2750-67.

Lehmann BD, Pietenpol JA, Tan AR. Triple-negative breast cancer: molecular subtypes and new targets for therapy. *Am Soc Clin Oncol Educ Book* 2015:e31-9. doi: 10.14694/EdBook_AM.

Lei B, Xu SP, Liang XS, Li YW, Zhang JF, Zhang GQ, Pang D. Long non-coding RNA MVIH is associated with poor prognosis and malignant biological behavior in breast cancer. *Tumour Biol* 2016;37:5257-64.

Lepoivre C, Belhocine M, Bergon A, Griffon A, Yammine M, Vanhille L, Zacarias-Cabeza J, Garibal MA, Koch F, Maqbool MA, Fenouil R, Loriol B, Holota H, Gut M, Gut I, Imbert J, Andrau JC, Puthier D, Spicuglia S. Divergent transcription is associated with promoters of transcriptional regulators. *BMC Genomics* 2013;14:914. doi: 10.1186/1471-2164-14-914.

Leygue E, Dotzlaw H, Watson PH, Murphy LC: Expression of the steroid receptor RNA activator in human breast tumors. *Cancer Res* 1999;59:4190–93.

Li CH, Chen Y. Targeting long non-coding RNAs in cancers: progress and prospects. *Int J Biochem Cell Biol* 2013a;45:1895-910.

Li L, Liu B, Wapinski OL, Tsai MC, Qu K, Zhang J, et al. Targeted disruption of Hotair leads to homeotic transformation and gene derepression. *Cell Reports* 2013b;5:3–12.

Li CH, Chen Y. Targeting long non-coding RNAs in cancers: progress and prospects. *Int J Biochem Cell Biol* 2013c;45:1895-910.

Li Z, Chao TC, Chang KY, Lin N, Patil VS, Shimizu C, Head SR, Burns JC, Rana TM. The long noncoding RNA THRIL regulates TNF α expression through its interaction with hnRNPL. *Proc Natl Acad Sci U S A* 2014a;111:1002-7.

Li H, Yu B, Li J, Su L, Yan M, Zhu Z, et al. Overexpression of lncRNA H19 enhances carcinogenesis and metastasis of gastric cancer. *Oncotarget* 2014b;5:2318–29.

Li R, Zhang L, Jia L, Duan Y, Li Y, Bao L, et al. Long non-coding RNA BANCR promotes proliferation in malignant melanoma by regulating MAPK pathway activation. *PLoS One* 2014c;9:e100893.

Li R, Qian J, Wang YY, Zhang JX, You YP. Long noncoding RNA profiles reveal three molecular subtypes in glioma. *CNS Neurosci Ther* 2014d;20:339-43.

Li F, Hu CP. Long Non-Coding RNA Urothelial Carcinoma Associated 1 (UCA1): Insight into Its Role in Human Diseases. *Crit Rev Eukaryot Gene Expr* 2015;25:191-7.

Li T, Mo X, Fu L, Xiao B, Guo J. Molecular mechanisms of long noncoding RNAs on gastric cancer. *Oncotarget* 2016;7:8601-12.

Lin N, Chang K-Y, Li Z et al. An evolutionarily conserved long noncoding RNA TUNA controls pluripotency and neural lineage commitment. *Mol Cell* 2014;53:1005–1019.

Ling H, Spizzo R, Atlasi Y, Nicoloso M, Shimizu M, Redis RS, et al. CCAT2, a novel noncoding RNA mapping to 8q24, underlies metastatic progression and chromosomal instability in colon cancer. *Genome Res* 2013;23:1446–61.

Little MP, Boice JD Jr. Comparison of breast cancer incidence in the Massachusetts tuberculosis fluoroscopy cohort and in the Japanese atomic bomb survivors. *Radiat Res* 1999;151:218-24.

Liu Q, Paroo Z. Molecular biology. Dicer's cut and switch. *Science* 2010;328:314-5.

Liu Z, Wang W, Jiang J, Bao E, Xu D, Zeng Y, et al. Downregulation of GAS5 promotes bladder cancer cell proliferation, partly by regulating CDK6. *PLoS One* 2013;8:e73991.

Liu J-H, Chen G, Dang Y-W, Li C-J, Luo D-Z. Expression and prognostic significance of lncRNA MALAT1 in pancreatic cancer tissues. *Asian Pac J Cancer Prev* 2014a;15:2971-7.

Liu W-T, Lu X, Tang G-H, Ren J-J, Liao W-J, Ge P-L, et al. LncRNAs expression signatures of hepatocellular carcinoma revealed by microarray. *World J Gastroenterol* 2014b;20:6314-21.

Liu Y, Sharma S, Watabe K. Roles of lncRNA in breast cancer. *Front Biosci* 2015a;7:94-108.

Liu B, Sun L, Liu Q, Gong C, Yao Y, Lv X, Lin L, Yao H, Su F, Li D, Zeng M, Song E. A cytoplasmic NF- κ B interacting long noncoding RNA blocks I κ B phosphorylation and suppresses breast cancer metastasis. *Cancer Cell* 2015b;27:370-81.

Liu Z, Yang T, Xu Z, Cao X. Upregulation of the long non-coding RNA BANCR correlates with tumor progression and poor prognosis in esophageal squamous cell carcinoma. *Biomed Pharmacother* 2016a;82:406-412.

Liu T, Zhang X, Yang YM, Du LT, Wang CX. Increased expression of the long noncoding RNA CRNDE-h indicates a poor prognosis in colorectal cancer, and is positively correlated with IRX5 mRNA expression. *Onco Targets Ther* 2016b;9:1437-48.

Liu FT, Zhu PQ, Ou YX, Lin QS, Qiu C, Luo HL. Long non-coding RNA-LET can indicate metastasis and a poor prognosis: a meta-analysis. *Minerva Med* 2016c;107:101-7.

Livasy CA, Karaca G, Nanda R, Tretiakova MS, Olopade OI, Moore DT, Perou CM. Phenotypic evaluation of the basal-like subtype of invasive breast carcinoma. *Mod Pathol* 2006;19:264-71.

Liz J, Portela A, Soler M, Gomez A, Ling H, Michlewski G, Calin GA, Guil S, Esteller M. Regulation of pri-miRNA processing by a long noncoding RNA transcribed from an ultraconserved region. *Mol Cell* 2014;55:138-47.

Liz J, Esteller M. lncRNAs and microRNAs with a role in cancer development. *Biochim Biophys Acta* 2016;1859:169-76.

Loewer S, Cabili MN, Guttman M, Loh YH, Thomas K, Park IH, et al. Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. *Nat Genet* 2010;12:1113-17.

Looijenga LH, Gillis AJ, van Gurp RJ, Verkerk AJ, Oosterhuis JW. X inactivation in human testicular tumors. XIST expression and androgen receptor methylation status. *Am J Pathol* 1997a;151:581-90.

Looijenga LH, Verkerk AJ, De Groot N, Hochberg AA, Oosterhuis JW. H19 in normal development and neoplasia. *Mol Reprod Dev* 1997b;46:419-39.

Lottin S, Adriaenssens E, Dupressoir T, Berteaux N, Montpellier C, Coll J, Dugimont T, Cury JJ. Overexpression of an ectopic H19 gene enhances the tumorigenic properties of breast cancer cells. *Carcinogenesis* 2002;23:1885-95.

Lopez-Pajares V. Long non-coding RNA regulation of gene expression during differentiation. *Pflugers Arch* 2016;468:971-81.

Louro R, Smirnova AS, Verjovski-Almeida S. Long intronic noncoding RNA transcription: expression noise or expression choice? *Genomics* 2009;93:291-8.

Lovén J, Hoke HA, Lin CY, Lau A, Orlando DA, Vakoc CR et al. Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* 2013;153:320–34.

Lu L, Zhu G, Zhang C, Deng Q, Katsaros D, Mayne ST, et al. Association of large noncoding RNA HOTAIR expression and its downstream intergenic CpG island methylation with survival in breast cancer. *Breast Cancer Res Treat* 2012;136:875-83.

Lukas J, Altmeyer M. A lncRNA to repair DNA. *EMBO Rep* 2015;e201541309.

Luo KQ, Chang DC. The gene-silencing efficiency of siRNA is strongly dependent on the local structure of mRNA at the targeted region. *Biochem Biophys Res Commun* 2004;318:303-10. Erratum in: *Biochem Biophys Res Commun* 2004;320:622.

Luo M, Li Z, Wang W, Zeng Y, Liu Z, Qiu J. Long non-coding RNA H19 increases bladder cancer metastasis by associating with EZH2 and inhibiting E-cadherin expression. *Cancer Lett* 2013;333:213-21.

Ma L, Bajic V.B, Zhang Z. On the classification of long noncoding RNAs. *RNA Biol* 2013;10:925–33.

Ma C, Nong K, Zhu H, Wang W, Huang X, Yuan Z, Ai K. H19 promotes pancreatic cancer metastasis by derepressing let-7's suppression on its target HMGA2-mediated EMT. *Tumour Biol* 2014;35:9163-9.

Macias S, Plass M, Stajuda A, Michlewski G, Eyraas E, Cáceres JF. DGCR8 HITS-CLIP reveals novel functions for the Microprocessor. *Nat Struct Mol Biol* 2012;19:760-6.

Magadum S, Banerjee U, Murugan P, Gangapur D, Ravikesavan R. Gene duplication as a major force in evolution. *J Genet* 2013;92:155-61.

Magny EG, Pueyo JI, Pearl FM, Cespedes MA, Niven JE, Bishop SA, Couso JP. Conserved regulation of cardiac calcium uptake by peptides encoded in small open reading frames. *Science* 2013;341:1116-20.

Makrythanasis P, Kapranov P, Bartoloni L, Reymond A, Deutsch S, Guigó R, Denoeud F, Drenkow J, Rossier C, Ariani F, Capra V, Excoffier L, Renieri A, Gingeras TR, Antonarakis SE. Variation in novel exons (RACEfrags) of the MECP2 gene in Rett syndrome patients and controls. *Hum Mutat* 2009;30:E866-79. doi: 10.1002/humu.21073.

Malek E, Jagannathan S, Driscoll JJ. Correlation of long non-coding RNA expression with metastasis, drug resistance and clinical outcome in cancer. *Oncotarget* 2014;5:8027-38.

Malik R, Patel L, Prensner JR, Shi Y, Iyer MK, Subramaniyan S, Carley A, Niknafs YS, Sahu A, Han S, Ma T, Liu M, Asangani IA, Jing X, Cao X, Dhanasekaran SM, Robinson DR, Feng FY, Chinnaiyan AM. The lncRNA PCAT29 inhibits oncogenic phenotypes in prostate cancer. *Mol Cancer Res* 2014;12:1081-7.

Mallon E, Osin P, Nasiri N, Blain I, Howard B, Gusterson B. The basic pathology of human breast cancer. *J Mammary Gland Biol Neoplasia* 2000;5:139-63.

Malouf GG, Zhang J, Yuan Y, Compérat E, Rouprêt M, Cussenot O, Chen Y, Thompson EJ, Tannir NM, Weinstein JN, Valero V, Khayat D, Spano JP, Su X. Characterization of long non-coding RNA transcriptome in clear-cell renal cell carcinoma by next-generation deep sequencing. *Mol Oncol* 2015;9:32-43.

Malumbres M, Pérez De Castro I, Hernández MI, Jiménez M, Corral T, Pellicer A. Cellular response to oncogenic ras involves induction of the Cdk4 and Cdk6 inhibitor p15(INK4b). *Mol Cell Biol* 2000;20:2915-25.

Mao YS, Zhang B, Spector DL. Biogenesis and function of nuclear bodies. *Trends Genet* 2011;27:295-306.

Marahrens Y, Panning B, Dausman J, Strauss W, Jaenisch R. Xist-deficient mice are defective in dosage compensation but not spermatogenesis. *Genes Dev* 1997;11:156-66.

Margueron R, Li G, Sarma K, Blais A, Zavadil J, Woodcock CL, Dynlacht BD, Reinberg D. Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. *Mol Cell* 2008;32:503-18.

Margueron R, Reinberg D. The Polycomb complex PRC2 and its mark in life. *Nature* 2011;469:343-49.

Mariner PD, Walters RD, Espinoza CA, Drullinger LF, Wagner SD, Kugel JF, Goodrich JA. Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock. *Mol Cell* 2008;29:499-509.

Martelotto LG, De Filippo MR, Ng CK, Natrajan R, Fuhrmann L, Cyrta J, Piscuoglio S, Wen HC, Lim RS, Shen R, Schultheis AM, Wen YH, Edelweiss M, Mariani O, Stenman G, Chan TA, Colombo PE, Norton L, Vincent-Salomon A, Reis-Filho JS, Weigelt B. Genomic landscape of adenoid cystic carcinoma of the breast. *J Pathol* 2015;237:179-89.

Martianov I, Ramadass A, Serra Barros A, Chow N, Akoulitchev A. Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. *Nature* 2007;445:666-70.

- Matouk I, Mezan S, Mizrahi A, Ohana P, Abu-Lail R, Fellig Y, et al. The oncofetal H19 RNA connection: hypoxia, p53 and cancer. *Biochim Biophys Acta* 2010;1803:443–51.
- Matouk I, Raveh E, Ohana P, Lail RA, Gershtain E, Gilon M, De Groot N, Czerniak A, Hochberg A. The increasing complexity of the oncofetal h19 gene locus: functional dissection and therapeutic intervention. *Int J Mol Sci* 2013;14:4298-316.
- Matouk IJ, Raveh E, Abu-lail R, Mezan S, Gilon M, Gershtain E, Birman T, Gallula J, Schneider T, Barkali M, Richler C, Fellig Y, Sorin V, Hubert A, Hochberg A, Czerniak A. Oncofetal H19 RNA promotes tumor metastasis. *Biochim Biophys Acta* 2014;1843:1414-26.
- Matos I, Dufloth R, Alvarenga M, Zeferino LC, Schmitt F. p63, cytokeratin 5, and P-cadherin: three molecular markers to distinguish basal phenotype in breast carcinomas. *Virchows Arch* 2005;447:688-94.
- Mattick JS, Gagen MJ. The evolution of controlled multitasked gene networks: the role of introns and other noncoding RNAs in the development of complex organisms. *Mol Biol Evol* 2001;18:1611-30.
- Mattick JS, Makunin IV. Non-coding RNA. *Hum Mol Genet* 2006;Spec No 1:R17-29.
- Mattick JS. The genetic signatures of noncoding RNAs. *PLoS Genet* 2009;5: e1000459.
- Mayer G, Famulok M. [Nucleic acids in the drug discovery process : aptamers as therapeutic and as active ingredients for drugs]. *Pharm Unserer Zeit* 2007;36:432-6.
- McClelland ML, Mesh K, Lorenzana E, Chopra VS, Segal E, Watanabe C, Haley B, Mayba O, Yaylaoglu M, Gnad F, Firestein R. CCAT1 is an enhancer-templated RNA that predicts BET sensitivity in colorectal cancer. *J Clin Invest* 2016;126:639-52.
- Mehlen P, Puisieux A. Metastasis: a question of life or death. *Nat Rev Cancer* 2006;6:449-58.
- Mei D, Song H, Wang K, Lou Y, Sun W, Liu Z, Ding X, Guo J. Up-regulation of SUMO1 pseudogene 3 (SUMO1P3) in gastric cancer and its clinical association. *Med Oncol* 2013; 30:709.
- Melissari MT, Grote P. Roles for long non-coding RNAs in physiology and disease. *Pflugers Arch* 2016;468:945-58.
- Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, Maier L, Mackowiak SD, Gregersen LH, Munschauer M, Loewer A, Ziebold U, Landthaler M, Kocks C, le Noble F, Rajewsky N. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* 2013;495:333-38.
- Mendell JT. Targeting a Long Noncoding RNA in Breast Cancer. *N Engl J Med* 2016;374:2287-9.
- Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. *Nat Rev Genet* 2009;10:155–59.

Mercer TR, Wilhelm D, Dinger ME, Soldà G, Korbie DJ, Glazov EA, Truong V, Schwenke M, Simons C, Matthaei KI, Saint R, Koopman P, Mattick JS. Expression of distinct RNAs from 3' untranslated regions. *Nucleic Acids Res* 2011;39:2393-403.

Mercer TR, Mattick JS. Structure and function of long noncoding RNAs in epigenetic regulation. *Nat Struct Mol Biol* 2013;20:300-7.

Mestdagh P, Fredlund E, Pattyn F, Rihani A, Van Maerken T, Vermeulen J, et al. An integrative genomics screen uncovers ncRNA T-UCR functions in neuroblastoma tumours. *Oncogene* 2010;29:3583-92.

Michalik KM, You X, Manavski Y, Doddaballapur A, Zörnig M, Braun T, John D, Ponomareva Y, Chen W, Uchida S, Boon RA, Dimmeler S. Long noncoding RNA MALAT1 regulates endothelial cell function and vessel growth. *Circ Res* 2014;114:1389-97.

Mizrahi A, Czerniak A, Levy T, Amiur S, Gallula J, Matouk I, Abu-lail R, Sorin V, Birman T, de Groot N, Hochberg A, Ohana P. Development of targeted therapy for ovarian cancer mediated by a plasmid expressing diphtheria toxin under the control of H19 regulatory sequences. *J Transl Med* 2009;7:69. doi: 10.1186/1479-5876-7-69.

Modarresi F, Faghihi MA, Lopez-Toledano MA, Fatemi RP, Magistri M, Brothers SP, van der Brug MP, Wahlestedt C. Inhibition of natural antisense transcripts in vivo results in gene-specific transcriptional upregulation. *Nat Biotechnol* 2012;30:453-9.

Mohammad F, Pandey RR, Nagano T, Chakalova L, Mondal T, Fraser P, et al. Kcnq1ot1/Lit1 noncoding RNA mediates transcriptional silencing by targeting to the perinucleolar region. *Mol Cell Biol* 2008;28:3713-28.

Mohammad F, Pandey GK, Mondal T, Enroth S, Redrup L, Gyllensten U, Kanduri C. Long noncoding RNA-mediated maintenance of DNA methylation and transcriptional gene silencing. *Development* 2012;139:2792-803.

Mondal T, Rasmussen M, Pandey GK, Isaksson A, Kanduri C. Characterization of the RNA content of chromatin. *Genome Res* 2010;20:899-907.

Monnier P, Martinet C, Pontis J, Stancheva I, Ait-Si-Ali S, Dandolo L. H19 lncRNA controls gene expression of the Imprinted Gene Network by recruiting MBD1. *PNAS* 2013;110:20693-98.

Morris KV. Long antisense non-coding RNAs function to direct epigenetic complexes that regulate transcription in human cells. *Epigenetics* 2009;4:296-301.

Morris KV, Mattick JS. The rise of regulatory RNA. *Nat Rev Genet* 2014;15:423-37.

Mourtada-Maarabouni M, Hedge VL, Kirkham L, Farzaneh F, Williams GT. Growth arrest in human T-cells is controlled by the non-coding RNA growth-arrest-specific transcript 5 (GAS5). *J Cell Sci* 2008;121:939-46.

Mourtada-Maarabouni M, Pickard MR, Hedge VL, Farzaneh F, Williams GT. GAS5, a non-protein-coding RNA, controls apoptosis and is downregulated in breast cancer. *Oncogene* 2009;28:195–208.

Mourtada-Maarabouni M, Hasan AM, Farzaneh F, Williams GT. Inhibition of human T-cell proliferation by mammalian target of rapamycin (mTOR) antagonists requires noncoding RNA growth-arrest-specific transcript 5 (GAS5). *Mol Pharmacol* 2010;78:19–28.

Mousavi K, Zare H, Dell'orso S, Grontved L, Gutierrez-Cruz G, Derfoul A, et al. eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci. *Mol Cell* 2013;51:606–17.

Müller S, Raulefs S, Bruns P, Afonso-Grunz F, Plötner A, Thermann R, Jäger C, Schlitter AM, Kong B, Regel I, Roth WK, Rotter B, Hoffmeier K, Kahl G, Koch I, Theis FJ, Kleeff J, Winter P, Michalski CW. Next-generation sequencing reveals novel differentially regulated mRNAs, lncRNAs, miRNAs, sdRNAs and a piRNA in pancreatic cancer. *Mol Cancer* 2015;14:94. doi: 10.1186/s12943-015-0358-5. Erratum in: *Mol Cancer* 2015;14:144.

Nagano T, Mitchell JA, Sanz LA, Pauler FM, Ferguson-Smith AC, Feil R, et al. The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science Reports* 2008;322:1717–20.

Nagano T, Fraser P. No-nonsense functions for long noncoding RNAs. *Cell* 2011;145:178-81.

Nakagawa S, Ip JY, Shioi G, Tripathi V, Zong X, Hirose T, Prasanth KV. Malat1 is not an essential component of nuclear speckles in mice. *RNA* 2012;18:1487-1499.

Nakagawa T, Endo H, Yokoyama M, Abe J, Tamai K, Tanaka N, et al. Large non-coding RNA HOTAIR enhances aggressive biological behavior and is associated with short disease-free survival in human non-small cell lung cancer. *Biochem Biophys Res Commun* 2013;436:319–24.

Nakano S, Murakami K, Meguro M, Soejima H, Higashimoto K, Urano T, et al. Expression profile of LIT1/KCNQ1OT1 and epigenetic status at the KvDMR1 in colorectal cancers. *Cancer Sci* 2006;97:1147–54.

Nakaya HI, Amaral PP, Louro R, Lopes A, Fachel AA, Moreira YB, El-Jundi TA, da Silva AM, Reis EM, Verjovski-Almeida S. Genome mapping and expression analyses of human intronic noncoding RNAs reveal tissue-specific patterns and enrichment in genes related to regulation of transcription. *Genome Biol* 2007;8:R43.

Nam JW, Bartel DP. Long noncoding RNAs in *C. elegans*. *Genome Res* 2012;22:2529-40.

Negishi M, Wongpalee SP, Sarkar S, Park J, Lee KY, Shibata Y, et al. A new lncRNA, APTR, associates with and represses the CDKN1A/p21 promoter by recruiting polycomb proteins. *PLoS One* 2014;9:e95216.

Negrini S, Gorgoulis VG, Halazonetis TD: Genomic instability--an evolving hallmark of cancer. *Nat Rev Mol Cell Biol* 2010;11:220-228.

Nelson BR, Makarewich CA, Anderson DM, Winders BR, Troupes CD, Wu F, Reese AL, McAnally JR, Chen X, Kavalali ET, Cannon SC, Houser SR, Bassel-Duby R, Olson EN. A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle. *Science* 2016;351:271-5.

Ng LJ, Cropley JE, Pickett HA, Reddel RR, Suter CM. Telomerase activity is associated with an increase in DNA methylation at the proximal subtelomere and a reduction in telomeric transcription. *Nucleic Acids Res* 2009;37:1152-9.

Ng SY, Johnson R, Stanton LW. Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors. *EMBO J* 2012;31:522-33.

Niazi F, Valadkhan S. Computational analysis of functional long noncoding RNAs reveals lack of peptide-coding capacity and parallels with 3' UTRs. *RNA* 2012;18:825-43.

Nie L, Wu H-J, Hsu J-M, Chang S-S, Labaff AM, Li C-W, et al. Long non-coding RNAs: versatile master regulators of gene expression and crucial players in cancer. *Am J Transl Res* 2012;4:127-50.

Nie Y, Liu X, Qu S, Song E, Zou H, Gong C. Long noncoding RNA HOTAIR is an independent prognostic marker for nasopharyngeal carcinoma progression and survival. *Cancer Science* 2013;4:458-64.

Nielsen TO, Hsu FD, Jensen K, Cheang M, Karaca G, Hu Z, Hernandez-Boussard T, Livasy C, Cowan D, Dressler L, Akslen LA, Ragaz J, Gown AM, Gilks CB, van de Rijn M, Perou CM. Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. *Clin Cancer Res* 2004;10:5367-74.

Niinuma T, Suzuki H, Nojima M, Noshio K, Yamamoto H, Takamaru H, et al. Upregulation of miR-196a and HOTAIR drive malignant character in gastrointestinal stromal tumors. *Cancer Res* 2012;72:1126-36.

Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, Jones D, Hinton J, Marshall J, Stebbings LA, Menzies A, Martin S, Leung K, Chen L, Leroy C, Ramakrishna M, Rance R, Lau KW, Mudie LJ, Varela I, McBride DJ, Bignell GR, Cooke SL, Shlien A, Gamble J, Whitmore I, Maddison M, Tarpey PS, Davies HR, Papaemmanuil E, Stephens PJ, McLaren S, Butler AP, Teague JW, Jönsson G, Garber JE, Silver D, Miron P, Fatima A, Boyault S, Langerød A, Tutt A, Martens JW, Aparicio SA, Borg Å, Salomon AV, Thomas G, Børresen-Dale AL, Richardson AL, Neuberger MS, Futreal PA, Campbell PJ, Stratton MR; Breast Cancer Working Group of the International Cancer Genome Consortium. Mutational processes molding the genomes of 21 breast cancers. *Cell* 2012;149:979-93.

Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, Martincorena I, Alexandrov LB, Martin S, Wedge DC, Van Loo P, Ju YS, Smid M, Brinkman AB, Morganella S, Aure MR, Lingjærde OC, Langerød A, Ringnér M, Ahn SM, Boyault S, Brock JE, Broeks A, Butler A, Desmedt C, Dirix L, Dronov S, Fatima A, Foekens JA, Gerstung M, Hooijer GK, Jang SJ, Jones DR, Kim HY, King TA, Krishnamurthy S, Lee HJ, Lee JY, Li Y, McLaren S, Menzies A, Mustonen V, O'Meara S, Pauporté I, Pivot X, Purdie CA, Raine K, Ramakrishnan K, Rodríguez-González FG, Romieu G, Sieuwerts AM,

Simpson PT, Shepherd R, Stebbings L, Stefansson OA, Teague J, Tommasi S, Treilleux I, Van den Eynden GG, Vermeulen P, Vincent-Salomon A, Yates L, Caldas C, van't Veer L, Tutt A, Knappskog S, Tan BK, Jonkers J, Borg Å, Ueno NT, Sotiriou C, Viari A, Futreal PA, Campbell PJ, Span PN, Van Laere S, Lakhani SR, Eyfjord JE, Thompson AM, Birney E, Stunnenberg HG, van de Vijver MJ, Martens JW, Børresen-Dale AL, Richardson AL, Kong G, Thomas G, Stratton MR. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* 2016;534:47-54.

Nkondjock A, Ghadirian P. [Risk factors and risk reduction of breast cancer]. *Med Sci (Paris)* 2005;21:175-80.

Novikova IV, Hennelly SP, Sanbonmatsu KY. Structural architecture of the human long non-coding RNA, steroid receptor RNA activator. *Nucleic Acids Res* 2012;40:5034-51.

Nowell PC. The clonal evolution of tumor cell populations. *Science* 1976;194:23-8.

O'Dea EL, Kearns JD, Hoffmann A. UV as an amplifier rather than inducer of NF-kappaB activity. *Mol Cell* 2008;30:632-41.

Okazaki Y, Furuno M, Kasukawa T, Adachi J, Bono H, Kondo S, et al. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* 2002;420:563-73.

Oliver PL, Chodroff RA, Gosal A, Edwards B, Cheung AF, Gomez-Rodriguez J, Elliot G, Garrett LJ, Lickiss T, Szele F, Green ED, Molnár Z, Ponting CP. Disruption of *Visc-2*, a Brain-Expressed Conserved Long Noncoding RNA, Does Not Elicit an Overt Anatomical or Behavioral Phenotype. *Cereb Cortex* 2015;25:3572-85.

Ørom UA, Derrien T, Beringer M, Gumireddy K, Gardini A, Bussotti G, et al. Long noncoding RNAs with enhancer-like function in human cells. *Cell* 2010;143:46-58.

Ørom UA, Shiekhatar R. Long noncoding RNAs usher in a new era in the biology of enhancers. *Cell* 2013;154:1190-3.

Paci P, Colombo T, Farina L. Computational analysis identifies a sponge interaction network between long non-coding RNAs and messenger RNAs in human breast cancer. *BMC Syst Biol* 2014;8:83. doi: 10.1186/1752-0509-8-83.

Pandey RR, Mondal T, Mohammad F, Enroth S, Redrup L, Komorowski J, Nagano T, Mancini-Dinardo D, Kanduri C. *Kcnq1ot1* antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol Cell* 2008;32:232-46.

Pang KC, Frith MC, Mattick JS. Rapid evolution of noncoding RNAs: Lack of conservation does not mean lack of function. *Trends Genet* 2006;22:1-5.

Panzitt K, Tschernatsch MMO, Guelly C, Moustafa T, Stradner M, Strohmaier HM, et al. Characterization of *HULC*, a novel gene with striking up-regulation in hepatocellular carcinoma, as noncoding RNA. *Gastroenterology* 2007;132:330-42.

Park C, Yu N, Choi I, Kim W, Lee S. IncRNAtor: a comprehensive resource for functional investigation of long noncoding RNAs. *Bioinformatics* 2014;30:2480-85.

Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, Davies S, Fauron C, He X, Hu Z, Quackenbush JF, Stijleman IJ, Palazzo J, Marron JS, Nobel AB, Mardis E, Nielsen TO, Ellis MJ, Perou CM, Bernard PS. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol* 2009;27:1160-7.

Pasmant E, Laurendeau I, Héron D, Vidaud M, Vidaud D, Bièche I. Characterization of a germ-line deletion, including the entire INK4/ARF locus, in a melanoma-neural system tumor family: identification of ANRIL, an antisense noncoding RNA whose expression coclusters with ARF. *Cancer Res* 2007;67:3963-69.

Pasmant E, Sabbagh A, Vidaud M, Bièche I. ANRIL, a long, noncoding RNA, is an unexpected major hotspot in GWAS. *FASEB J* 2011a;25:444-8.

Pasmant E, Sabbagh A, Masliah-Planchon J, Ortonne N, Laurendeau I, Melin L, Ferkal S, Hernandez L, Leroy K, Valeyrie-Allanore L, Parfait B, Vidaud D, Bièche I, Lantieri L, Wolkenstein P, Vidaud M, NF France Network. Role of noncoding RNA ANRIL in genesis of plexiform neurofibromas in neurofibromatosis type 1. *J Natl Cancer Inst* 2011b;103:1713-22.

Pasquale EB. Eph receptors and ephrins in cancer : bidirectional signalling and beyond. *Nat Rev Cancer* 2010;10:165-80.

Pauli A, Rinn JL, Schier AF. Non-coding RNAs as regulators of embryogenesis. *Nat Rev Genet* 2011;12:136-49.

Payer B, Lee JT. X chromosome dosage compensation: how mammals keep the balance. *Annu Rev Genet* 2008;42:733-72.

Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lønning PE, Børresen-Dale AL, Brown PO, Botstein D. Molecular portraits of human breast tumours. *Nature* 2000;406:747-52.

Polyak K, Kalluri R. The role of the microenvironment in mammary gland development and cancer. *Cold Spring Harb Perspect Biol* 2010;2:a003244. doi: 10.1101/cshperspect.a003244.

Peschansky V and Claes Wahlestedt C. Non-coding RNAs as direct and indirect modulators of epigenetic regulation. *Epigenetics* 2014;9:3-12.

Peschansky VJ, Pastori C, Zeier Z, Motti D, Wentzel K, Velmeshev D, Magistri M, Bixby JL, Lemmon VP, Silva JP, Wahlestedt C. Changes in expression of the long non-coding RNA FMR4 associate with altered gene expression during differentiation of human neural precursor cells. *Front Genet* 2015;6:263.

Petrovics G, Zhang W, Makarem M, Street JP, Connelly R, Sun L, et al. Elevated expression of PCGEM1, a prostate-specific gene with cell growth-promoting function, is associated with high-risk prostate cancer patients. *Oncogene* 2004;23:605-11.

Piao HL, Ma L. Non-coding RNAs as regulators of mammary development and breast cancer. *J Mammary Gland Biol Neoplasia* 2012;17:33-42.

Piao L, Li H, Teng L, Yung BC, Sugimoto Y, Brueggemeier RW, Lee RJ. Human serum albumin-coated lipid nanoparticles for delivery of siRNA to breast cancer. *Nanomedicine* 2013;9:122-9.

Pickard MR, Mourtada-Maarabouni M, Williams GT. Long non-coding RNA GAS5 regulates apoptosis in prostate cancer cell lines. *Biochim Biophys Acta* 1832;2013:1613–23.

Pickard MR, Williams GT. Regulation of apoptosis by long non-coding RNA GAS5 in breast cancer cells: implications for chemotherapy. *Breast Cancer Res Treat* 2014;145:359–70.

Pircher A, Gebetsberger J, Polacek N. Ribosome-associated ncRNAs: an emerging class of translation regulators. *RNA Biol* 2014;11:1335-9.

Plath K, Mlynarczyk-Evans S, Nusinow DA, Panning B. Xist RNA and the mechanism of X chromosome inactivation. *Annu Rev Genet* 2002;36:233-78.

Ploussard G, Durand X, Xylinas E, Moutereau S, Radulescu C, Forgue A, Nicolaiew N, Terry S, Allory Y, Loric S, Salomon L, Vacherot F, de la Taille A. Prostate cancer antigen 3 score accurately predicts tumour volume and might help in selecting prostate cancer patients for active surveillance. *European Urology* 2011;59:422–29.

Poliseno L, Salmena L, Zhang J, Carver B, Haveman WJ, Pandolfi PP. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* 2010;465:1033-8.

Pollard KS, Salama SR, King B, Kern AD, Dreszer T, Katzman S, Siepel A, Pedersen JS, Bejerano G, Baertsch R, Rosenbloom KR, Kent J, Haussler D. Forces shaping the fastest evolving regions in the human genome. *PLoS Genet* 2006;2:e168.

Polyak K: Heterogeneity in breast cancer. *J Clin Invest* 2011;121:3786-3788.

Ponjavic J, Ponting C.P, Lunter G. Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. *Genome Res* 2007;17:556–65.

Ponting C.P, Oliver P.L, Reik W. Evolution and functions of long noncoding RNAs. *Cell* 2009;136:629–641.

Popov N, Gil J. Epigenetic regulation of the INK4b-ARF-INK4a locus: in sickness and in health. *Epigenetics* 2010;5:685–90.

Powell WT, Coulson RL, Crary FK, Wong SS, Ach RA, Tsang P, Alice Yamada N, Yasui DH, Lasalle JM. A Prader-Willi locus lncRNA cloud modulates diurnal genes and energy expenditure. *Hum Mol Genet* 2013;22:4318-28.

Prabakaran S, Hemberg M, Chauhan R, Winter D, Tweedie-Cullen RY, Dittrich C, Hong E, Gunawardena J, Steen H, Kreiman G, Steen JA. Quantitative profiling of peptides from RNAs classified as noncoding. *Nat Commun* 2014;5:5429. doi: 10.1038/ncomms6429.

Prat A, Adamo B, Cheang MC, Anders CK, Carey LA, Perou CM. Molecular characterization of basal-like and non-basal-like triple-negative breast cancer. *Oncologist* 2013;18:123-33.

Prensner JR, Chinnaiyan AM. The emergence of lncRNAs in cancer biology. *Cancer Discov* 2011;1:391-407.

Prensner JR, Chen W, Iyer MK, Cao Q, Ma T, Han S, et al. PCAT-1, a long noncoding RNA, regulates BRCA2 and controls homologous recombination in cancer. *Cancer Res* 2014a;74:1651-60.

Prensner JR, Sahu A, Iyer MK, Malik R, Chandler B, Asangani IA, Poliakov A, Vergara IA, Alshalalfa M, Jenkins RB, Davicioni E, Feng FY, Chinnaiyan AM. The lncRNAs PCGEM1 and PRNCR1 are not implicated in castration resistant prostate cancer. *Oncotarget* 2014b;5:1434-8.

Przybytkowski E, Lenkiewicz E, Barrett MT, Klein K, Nabavi S, Greenwood CM, Basik M. Chromosome-breakage genomic instability and chromothripsis in breast cancer. *BMC Genomics* 2014;15:579. doi: 10.1186/1471-2164-15-579.

Pu H, Zheng Q, Li H, Wu M, An J, Gui X, Li T, Lu D. CUDR promotes liver cancer stem cell growth through upregulating TERT and C-Myc. *Oncotarget* 2015;6:40775-98.

Pueyo JI, Magny EG, Couso JP. New Peptides Under the s(ORF)ace of the Genome. *Trends Biochem Sci* 2016;41:665-78.

Qi P, Du X. The long non-coding RNAs, a new cancer diagnostic and therapeutic gold mine. *Mod Pathol* 2013;26:155-65.

Qi HL, Li CS, Qian CW, Xiao YS, Yuan YF, Liu QY, Liu ZS. The long noncoding RNA, EGFR-AS1, a target of GHR, increases the expression of EGFR in hepatocellular carcinoma. *Tumour Biol* 2016;37:1079-89.

Qian M, Yang X, Li Z, Jiang C, Song D, Yan W, Liu T, Wu Z, Kong J, Wei H, Xiao J. P50-associated COX-2 extragenic RNA (PACER) overexpression promotes proliferation and metastasis of osteosarcoma cells by activating COX-2 gene. *Tumour Biol* 2016;37:3879-86.

Qiao H-P, Gao W-S, Huo J-X, Yang Z-S. Long Non-coding RNA GAS5 Functions as a Tumor Suppressor in Renal Cell Carcinoma. *Asian Pacific J Cancer Prev* 2013;14:1077-82.

Qiu M-T, Hu J-W, Yin R, Xu L. Long noncoding RNA: an emerging paradigm of cancer research. *Tumour Biol* 2013;34:613-20.

Qiu JJ, Lin YY, Ye LC, Ding JX, Feng WW, Jin HY, Zhang Y, Li Q, Hua KQ. Overexpression of long non-coding RNA HOTAIR predicts poor patient prognosis and promotes tumor metastasis in epithelial ovarian cancer. *Gynecol Oncol* 2014a;134:121-28.

Qiu M, Xu Y, Yang X, Wang J, Hu J, Xu L et al. CCAT2 is a lung adenocarcinoma-specific long non-coding RNA and promotes invasion of non-small cell lung cancer. *Tumour Biol* 2014b;35:5375-8.

Qiu JJ, Wang Y, Liu YL, Zhang Y, Ding JX, Hua KQ. The long non-coding RNA ANRIL promotes proliferation and cell cycle progression and inhibits apoptosis and senescence in epithelial ovarian cancer. *Oncotarget* 2016. doi: 10.18632/oncotarget.8744.

Quagliata L, Terracciano LM. Liver Diseases and Long Non-Coding RNAs: New Insight and Perspective. *Front Med (Lausanne)* 2014a;1:35. doi: 10.3389/fmed.2014.00035.

Quagliata L, Matter MS, Piscuoglio S, Arabi L, Ruiz C, Procino A, Kovac M, Moretti F, Makowska Z, Boldanova T, Andersen JB, Hämmerle M, Tornillo L, Heim MH, Diederichs S, Cillo C, Terracciano LM. Long noncoding RNA HOTTIP/HOXA13 expression is associated with disease progression and predicts outcome in hepatocellular carcinoma patients. *Hepatology* 2014b;59:911-23.

Quan M, Chen J, Zhang D. Exploring the Secrets of Long Noncoding RNAs. *Int J Mol Sci* 2015;16:5467-5496.

Quinodoz S, Guttman M. Long noncoding RNAs: an emerging link between gene regulation and nuclear organization. *Trends Cell Biol* 2014;24:651-63.

Rachmilewitz J, Goshen R, Ariel I, Schneider T, de Groot N, Hochberg A. Parental imprinting of the human H19 gene. *FEBS Lett* 1992;309:25-8.

Rackham et al., 2011 Rackham O, Shearwood AM, Mercer TR, Davies SM, Mattick JS, Filipovska A. Long noncoding RNAs are generated from the mitochondrial genome and regulated by nuclear-encoded proteins. *RNA* 2011;17:2085-93.

Rakha EA, Tan DS, Foulkes WD, Ellis IO, Tutt A, Nielsen TO, Reis-Filho JS. Are triple-negative tumours and basal-like breast cancer synonymous? *Breast Cancer Res* 2007;9:404; author reply 405. doi: 10.1186/bcr1827.

Rando TA, Chang HY. Aging, rejuvenation, and epigenetic reprogramming: resetting the aging clock. *Cell* 2012;148:46-57.

Ranzani V, Rossetti G, Panzeri I, Arrigoni A, Bonnal RJ, Curti S, Gruarin P, Provasi E, Sugliano E, Marconi M, De Francesco R, Geginat J, Bodega B, Abrignani S, Pagani M. The long intergenic noncoding RNA landscape of human lymphocytes highlights the regulation of T cell differentiation by linc-MAF-4. *Nat Immunol* 2015;16:318-25.

Ravasi T, Suzuki H, Pang K.C, Katayama S, Furuno M, Okunishi R, Fukuda S, Ru K, Frith M.C, Gongora M.M et al. Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome. *Genome Res* 2006;16: 11–19.

Raveh E, Matouk IJ, Gilon M, Hochberg A. The H19 Long non-coding RNA in cancer initiation, progression and metastasis - a proposed unifying theory. *Mol Cancer* 2015;14:184. doi: 10.1186/s12943-015-0458-2.

Redis RS, Sieuwerts AM, Look MP, Tudoran O, Ivan C, Spizzo R, et al. CCAT2, a novel long non-coding RNA in breast cancer: expression study and clinical correlations. *Oncotarget* 2013;4:1748–62.

Redon S, Reichenbach P, Lingner J. The non-coding RNA TERRA is a natural ligand and direct inhibitor of human telomerase. *Nucleic Acids Res* 2010;38:5797–806.

Reiche K, Kasack K, Schreiber S, Lüders T, Due EU, Naume B, et al. Long non-coding RNAs differentially expressed between normal versus primary breast tumor tissues disclose converse changes to breast cancer-related protein-coding genes. *PLoS One* 2014;9:e106076.

Reis EM, Verjovski-Almeida S. Perspectives of Long Non-Coding RNAs in Cancer Diagnostics. *Front Genet* 2012;3:32. doi: 10.3389/fgene.

Ren S, Liu Y, Xu W, Sun Y, Lu J, Wang F, et al. Long noncoding RNA MALAT-1 is a new potential therapeutic target for castration resistant prostate cancer. *J Urol* 2013;190:2278–87.

Rennstam K, Ahlstedt-Soini M, Baldetorp B, Bendahl PO, Borg A, Karhu R, Tanner M, Tirkkonen M, Isola J. Patterns of chromosomal imbalances defines subgroups of breast cancer with distinct clinical features and prognosis. A study of 305 tumors by comparative genomic hybridization. *Cancer Res* 2003;63:8861-8.

Reon BJ, Dutta A. Biological Processes Discovered by High-Throughput Sequencing. *Am J Pathol* 2016;186:722-32.

Reuter JS, Mathews DH. RNA structure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* 2010;11:129. doi: 10.1186/1471-2105-11-129.

Richards EJ, Zhang G, Li ZP, Permeth-Wey J, Challa S, Li Y, Kong W, Dan S, Bui MM, Coppola D, Mao WM, Sellers TA, Cheng JQ. Long non-coding RNAs (LncRNA) regulated by transforming growth factor (TGF) β : LncRNA-hit-mediated TGF β -induced epithelial to mesenchymal transition in mammary epithelia. *J Biol Chem* 2015;290:6857-67.

Richardson AL, Wang ZC, De Nicolo A, Lu X, Brown M, Miron A, Liao X, Iglehart JD, Livingston DM, Ganesan S. X chromosomal abnormalities in basal-like human breast cancer. *Cancer Cell* 2006;9:121-32.

Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Bruggmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E, Chang HY: Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* 2007;129:1311–23.

Ripoche MA, Kress C, Poirier F, Dandolo L. Deletion of the H19 transcription unit reveals the existence of a putative imprinting control element. *Genes Dev* 1997;11:1596-604.

Rizzi N, Denegri M, Chiodi I, Corioni M, Valgardsdottir R, Cobiauchi F, Riva S, Biamonti G. Transcriptional activation of a constitutive heterochromatic domain of the human genome in response to heat shock. *Mol Biol Cell* 2004;15:543-51.

Rodriguez C, Borgel J, Court F, Cathala G, Forné T, Piette J. CTCF is a DNA methylation-sensitive positive regulator of the INK/ARF locus. *Biochem Biophys Res Commun* 2010;392:129-34.

Roger L, Jones RE, Heppel NH, Williams GT, Sampson JR, Baird DM. Extensive telomere erosion in the initiation of colorectal adenomas and its association with chromosomal instability. *J Natl Cancer Inst* 2013;105:1202-11.

Rong Shi, Jue-Yu Zhou, Hui Zhou, Zhen Zhao, Sang-Hua Liang, Wen-Ling Zheng, and Wen-Li Ma. The Role of PinX1 in Growth Control of Breast Cancer Cells and Its Potential Molecular Mechanism by mRNA and lncRNA Expression Profiles Screening. *BioMed Research International* 2014;1978984.

Roobol MJ, Schroder FH, van Leenders GL, Hessels D, van den Bergh RC, T. Wolters T, van Leeuwen PJ. Performance of prostate cancer antigen 3 (PCA3) and prostate-specific antigen in prescreened men: reproducibility and detection characteristics for prostate cancer patients with high PCA3 scores (≥ 100). *European Urology* 2010;58:893–99.

Ropero S, Fraga MF, Ballestar E, Hamelin R, Yamamoto H, Boix-Chornet M, Caballero R, Alaminos M, Setien F, Paz MF, Herranz M, Palacios J, Arango D, Orntoft TF, Aaltonen LA, Schwartz S Jr, Esteller M. A truncating mutation of HDAC2 in human cancers confers resistance to histone deacetylase inhibition. *Nat Genet* 2006;38:566-9.

Rosa A, Brivanlou AH. Regulatory non-coding RNAs in pluripotent stem cells. *Int J Mol Sci* 2013;14:14346-73.

Roshan F, Vijay S A, Debosree P, Satyanarayana MR. Long noncoding RNAs in development and cancer: potential biomarkers and therapeutic targets. *Mol Cell Ther* 2015;3:5. doi: 10.1186/s4059-015-0042.

Ross DT, Johnsen H, Akslen LA et al: Molecular portraits of human breast tumours. *Nature* 2000;406:747-752.

Rossi MN, Antonangeli F. LncRNAs: New Players in Apoptosis Control. *Int J Cell Biol* 2014;2014:473857. doi: 10.1155/2014/473857.

Rossouw JE, Anderson GL, Prentice RL, LaCroix AZ, Kooperberg C, Stefanick ML, Jackson RD, Beresford SA, Howard BV, Johnson KC, Kotchen JM, Ockene J; Writing Group for the Women's Health Initiative Investigators. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial. *JAMA* 2002;288:321-33.

Rother S, Meister G. Small RNAs derived from longer non-coding RNAs. *Biochimie* 2011;93:1905–15.

Rudas M, Schmidinger M, Wenzel C, Okamoto I, Budinsky A, Fazeny B, Marosi C. Karyotypic findings in two cases of male breast cancer. *Cancer Genet Cytogenet* 2000;121:190-3.

Ruffner DE, Stormo GD, Uhlenbeck OC. Sequence requirements of the hammerhead RNA self-cleavage reaction. *Biochemistry* 1990;29:10695-702.

Russo J, Lynch H, Russo IH. Mammary gland architecture as a determining factor in the susceptibility of the human breast to cancer. *Breast J* 2001;7:278-91.

St Laurent G, Shtokalo D, Tackett MR, Yang Z, Eremina T, Wahlestedt C, Urcuqui-Inchima S, Seilheimer B, McCaffrey TA, Kapranov P. Intronic RNAs constitute the major fraction of the non-coding RNA in mammalian cells. *BMC Genomics* 2012;13:504. doi: 10.1186/1471-2164-13-504.

St Laurent G, Shtokalo D, Dong B, Tackett MR, Fan X, Lazorthes S, Nicolas E, Sang N, Triche TJ, McCaffrey TA, Xiao W, Kapranov P. VlincRNAs controlled by retroviral elements are a hallmark of pluripotency and cancer. *Genome Biol* 2013;14(7):R73. doi: 10.1186/gb-2013-14-7-r73.

St Laurent G, Vyatkin Y, Kapranov P. Dark matter RNA illuminates the puzzle of genome-wide association studies. *BMC Med* 2014;12:97. doi: 10.1186/1741-7015-12-97.

St Laurent G, Wahlestedt C, Kapranov P. The landscape of long non coding RNA classification. *Trends Genet* 2015;31:239-51.

Sakurai K, Reon BJ, Anaya J, Dutta A. The lncRNA DRAIC/PCAT29 Locus Constitutes a Tumor-Suppressive Nexus. *Mol Cancer Res* 2015;13:828-38.

Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A ceRNA hypothesis : the Rosetta Stone of a hidden RNA language? *Cell* 2011;146:353-8.

Salzman J. Circular RNA Expression: Its Potential Regulation and Function. *Trends Genet* 2016;32:309-16.

Sana J, Faltejskova P, Svoboda M, Slaby O. Novel classes of non-coding RNAs and cancer. *J Transl Med* 2012;10:103. doi : 10.1186/1479-5876-10-103.

Sang Y, Zhou F, Wang D, Bi X, Liu X, Hao Z, Li Q, Zhang W. Up-regulation of long non-coding HOTTIP functions as an oncogene by regulating HOXA13 in non-small cell lung cancer. *Am J Transl Res* 2016;8:2022-32.

Sanger HL, Klotz G, Riesner D, Gross HJ, Kleinschmidt AK. Viroids are single-stranded covalently closed circular RNA molecules existing as highly base-paired rod-like structures. *Proc Natl Acad Sci U S A* 1976;73:3852-6.

Sánchez Y, Huarte M. Long non-coding RNAs: challenges for diagnosis and therapies. *Nucleic Acid Ther* 2013;23:15–20.

Sarver AL, Murray CD, Temiz NA, Tseng YY, Bagchi A. MYC and PVT1 synergize to regulate RSPO1 levels in breast cancer. *Cell Cycle* 2016;15:881-5.

Sas-Chen A, Aure MR, Leibovich L, Carvalho S, Eneka Y, Körner C, Polycarpou-Schwarz M, Lavi S, Nevo N, Kuznetsov Y, Yuan J, Azuaje F; Oslo Breast Cancer Research Consortium (OSBREAC). LIMT is a novel metastasis inhibiting lncRNA suppressed by EGF and downregulated in aggressive breast cancer. *EMBO Mol Med* 2016;8:1052-64.

Sato K, Nakagawa H, Tajima A, Yoshida K, Inoue I. ANRIL is implicated in the regulation of nucleus and potential transcriptional target of E2F1. *Oncol Rep* 2010;24:701-7.

Saxena A, Carninci P. Long non-coding RNA modifies chromatin: epigenetic silencing by long non-coding RNAs. *Bioessays* 2011;33:830-9.

Scaruffi P, Stigliani S, Moretti S, Coco S, De Vecchi C, Valdora F, Garaventa A, Bonassi S, Tonini GP. Transcribed-Ultra Conserved Region expression is associated with outcome in high-risk neuroblastoma. *BMC Cancer* 2009;9:441. doi: 10.1186/1471-2407-9-441.

Scaruffi P. The transcribed-ultraconserved regions: a novel class of long noncoding RNAs involved in cancer susceptibility. *Scientific World Journal* 2011;11:340-52.

Schmitt AM, Chang HY. Long Noncoding RNAs in Cancer Pathways. *Cancer Cell* 2016;29:452-63.

Schmitz SU, Grote P, Herrmann BG. Mechanisms of long noncoding RNA function in development and disease. *Cell Mol Life Sci* 2016;73:2491-509.

Schuettengruber B, Chourrout D, Vervoort M, Leblanc B, Cavalli G. Genome regulation by polycomb and trithorax proteins. *Cell* 2007;128:735-45.

Seila AC, Sharp PA. Small RNAs tell big stories in Whistler. *Nat Cell Biol* 2008;10:630-3. doi: 10.1038/ncb0608-630.

Semenza GL. Molecular mechanisms mediating metastasis of hypoxic breast cancer cells. *Trends Mol Med* 2012;18:534-43.

Serviss JT, Johnsson P, Grandér D. An emerging role for long non-coding RNAs in cancer metastasis. *Front Genet* 2014;5:234. doi: 10.3389/fgene.2014.00234.

Sestak I, Cuzick J, Dowsett M, Lopez-Knowles E, Filipits M, Dubsy P, Cowens JW, Ferree S, Schaper C, Fesl C, Gnant M. Prediction of late distant recurrence after 5 years of endocrine treatment: a combined analysis of patients from the Austrian breast and colorectal cancer study group 8 and arimidex, tamoxifen alone or in combination randomized trials using the PAM50 risk of recurrence score. *J Clin Oncol* 2015;33:916-22.

Shamovsky I, Ivannikov M, Kandel ES, Gershon D, Nudler E. RNA-mediated response to heat shock in mammalian cells. *Nature* 2006;440:556-60.

Sharif J, Muto M, Takebayashi S, Suetake I, Iwamatsu A, Endo TA, Shinga J, Mizutani-Koseki Y, Toyoda T, Okamura K, Tajima S, Mitsuya K, Okano M, Koseki H. The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* 2007;450:908-12.

Sharma S, Findlay GM, Bandukwala HS, Oberdoerffer S, Baust B, Li Z, Schmidt V, Hogan PG, Sacks DB, Rao A. 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* 2011;108:11381-6.

Sharma V, Khurana S, Kubben N, Abdelmohsen K, Oberdoerffer P, Gorospe M, Misteli T. A BRCA1-interacting lncRNA regulates homologous recombination. *EMBO Rep* 2015;16:1520-34.

Shen H, Laird PW. Interplay between the cancer genome and epigenome. *Cell* 2013;153:38-55.

Shen L, Chen L, Wang Y, Jiang X, Xia H, Zhuang Z. Long noncoding RNA MALAT1 promotes brain metastasis by inducing epithelial-mesenchymal transition in lung cancer. *J Neurooncol* 2014;121:101-8.

Shen Y, Katsaros D, Loo LW, Hernandez BY, Chong C, Canuto EM, Biglia N, Lu L, Risch H, Chu WM, Yu H. Prognostic and predictive values of long non-coding RNA LINC00472 in breast cancer. *Oncotarget* 2015;6:8579-92.

Shen F, Cai WS, Feng Z, Chen JW, Feng JH, Liu QC, Fang YP, Li KP, Xiao HQ, Cao J, Xu B. Long non-coding RNA SPRY4-IT1 promotes colorectal cancer metastasis by regulate epithelial-mesenchymal transition. *Oncotarget* 2016. doi: 10.18632/oncotarget.10407.

Shi X, Sun M, Liu H, Yao Y, Song Y. Long noncoding RNAs: Novel insights into hepatocellular carcinoma. *Cancer Lett* 2013;10:339:159-66.

Shi Y, Wang Y, Luan W, Wang P, Tao T, et al. Long Non-Coding RNA H19 Promotes Glioma Cell Invasion by Deriving miR-675. *PLoS one* 2014;9:e86295.

Shi X, Sun M, Wu Y, Yao Y, Liu H, Wu G, Yuan D, Song Y. Post-transcriptional regulation of long noncoding RNAs in cancer. *Tumour Biol* 2015a;36:503-13.

Shi SJ, Wang LJ, Yu B, Li YH, Jin Y, Bai XZ. LncRNA-ATB promotes trastuzumab resistance and invasion-metastasis cascade in breast cancer. *Oncotarget* 2015b;6:11652-63.

Shi X, Nie F, Wang Z, Sun M. Pseudogene-expressed RNAs: a new frontier in cancers. *Tumour Biol* 2016;37:1471-8.

Shibayama Y, Fanucchi S, Magagula L, Mhlanga MM. lncRNA and gene looping. *Transcription* 2014;5:e28658.

Shore AN, Kabotyanski EB, Roarty K, Smith MA, Zhang Y, Creighton CJ, et al. Pregnancy-Induced Noncoding RNA (PINC) Associates with Polycomb Repressive Complex 2 and Regulates Mammary Epithelial Differentiation. *PLoS Genet* 2012;8:1–20.

Shore AN, Rosen JM. Regulation of mammary epithelial cell homeostasis by lncRNAs. *Int J Biochem Cell Biol* 2014;54:318-30.

Silva JM, Boczek NJ, Berres MW, Ma X, Smith DI. LSINCT5 is over expressed in breast and ovarian cancer and affects cellular proliferation. *RNA Biol* 2011;8:496-505.

Sleutels F, Zwart R, Barlow DP. The non-coding RNA Air is required for silencing autosomal imprinted genes. *Nature* 2002;415:810–13.

Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS et al: Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 2001; 98:10869-10874.

Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S, Demeter J, Perou CM, Lønning PE, Brown PO, Børresen-Dale AL, Botstein D. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A* 2003;100:8418-23.

Span PN, Rao JU, Oude Ophuis SB, Lenders JW, Sweep FC, Wesseling P, Kusters B, van Nederveen FH, de Krijger RR, Hermus AR, Timmers HJ. Overexpression of the natural antisense hypoxia-inducible factor-1alpha transcript is associated with malignant pheochromocytoma/paraganglioma. *Endocr Relat Cancer* 2011;18:323-31.

Spano D, Heck C, De Antonellis P, Christofori G, Zollo M. Molecular networks that regulate cancer metastasis. *Semin Cancer Biol* 2012;22:234-49.

Spector DL, Lamond AI. Nuclear speckles. *Cold Spring Harb Perspect Biol* 2011;3. pii: a000646. doi: 10.1101/cshperspect.a000646.

Spicuglia S, Pekowska A, Zacarias-Cabeza J, Ferrier P. Epigenetic control of Tcrb gene rearrangement. *Semin Immunol* 2010;22:330-6.

Spicuglia S, Maqbool MA, Puthier D, Andrau JC. An update on recent methods applied for deciphering the diversity of the noncoding RNA genome structure and function. *Methods* 2013;63:3-17.

Spizzo R, Almeida MI, Colombatti A, Calin GA. Long non-coding RNAs and cancer: a new frontier of translational research? *Oncogene* 2012;31:4577-87.

Splinter E, de Wit E, Nora EP, Klous P, van de Werken HJ, Zhu Y, Kaaij LJ, van Ijcken W, Gribnau J, Heard E, de Laat W. The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on Xist RNA. *Genes Dev* 2011;25:1371-83.

Stehelin D, Varmus HE, Bishop JM, Vogt PK. DNA related to the transforming gene(s) of avian sarcoma viruses is present in normal avian DNA. *Nature* 1976;260:170-3.

Stephens PJ, Greenman CD, Fu B, Yang F, Bignell GR, Mudie LJ, Pleasance ED, Lau KW, Beare D, Stebbings LA, McLaren S, Lin ML, McBride DJ, Varela I, Nik-Zainal S, Leroy C, Jia M, Menzies A, Butler AP, Teague JW, Quail MA, Burton J, Swerdlow H, Carter NP, Morsberger LA, Iacobuzio-Donahue C, Follows GA, Green AR, Flanagan AM, Stratton MR, Futreal PA, Campbell PJ. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* 2011;144:27-40.

Stephens PJ, Tarpey PS, Davies H, Van Loo P, Greenman C, Wedge DC, Nik-Zainal S, Martin S, Varela I, Bignell GR, Yates LR, Papaemmanuil E, Beare D, Butler A, Cheverton A, Gamble J, Hinton J, Jia M, Jayakumar A, Jones D, Latimer C, Lau KW, McLaren S, McBride DJ, Menzies A, Mudie L, Raine K, Rad R, Chapman MS, Teague J, Easton D, Langerød A; Oslo Breast Cancer Consortium (OSBREAC), Lee MT, Shen CY, Tee BT, Huimin BW, Broeks A, Vargas AC, Turashvili G, Martens J, Fatima A, Miron P, Chin SF, Thomas G, Boyault S, Mariani O, Lakhani SR, van de Vijver M, van't Veer L, Foekens J, Desmedt C, Sotiriou C, Tutt A, Caldas C, Reis-Filho JS, Aparicio SA, Salomon AV, Børresen-Dale AL, Richardson AL, Campbell PJ, Futreal PA, Stratton MR. The landscape of cancer genes and mutational processes in breast cancer. *Nature* 2012;486:400-4.

Stransky N, Cerami E, Schalm S, Kim JL, Lengauer C. The landscape of kinase fusions in cancer. *Nat Commun* 2014;5:4846. doi: 10.1038/ncomms5846.

Su X, Malouf GG, Chen Y, Zhang J, Yao H, Valero V, Weinstein JN, Spano JP, Meric-Bernstam F, Khayat D, Esteva FJ. Comprehensive analysis of long non-coding RNAs in human breast cancer clinical subtypes. *Oncotarget* 2014;5:9864-76.

Sue SC, Alverdi V, Komives EA, Dyson HJ. Detection of a ternary complex of NF-kappaB and IkappaBalpha with DNA provides insights into how IkappaBalpha removes NF-kappaB from transcription sites. *Proc Natl Acad Sci U S A* 2011;108:1367-72.

Sültmann H, Diederichs S. Long noncoding RNA: "LNCs" to cancer. *Eur Urol* 2014;65:1152-3.

Sun L, Goff LA, Trapnell C, Alexander R, Lo KA, Hacisuleyman E, Sauvageau M, Tazon-Vega B, Kelley DR, Hendrickson DG, Yuan B, Kellis M, Lodish HF, Rinn JL. Long noncoding RNAs regulate adipogenesis. *Proc Natl Acad Sci U S A* 2013a;110:3387-92.

Sun W, Wu Y, Yu X, Liu Y, Song H, Xia T, Xiao B, Guo J. Decreased expression of long noncoding RNA AC096655.1-002 in gastric cancer and its clinical significance. *Tumour Biol* 2013b;34:2697-701.

Sun M, Jin F, Xia R, Kong R, Li J, Xu T, et al. Decreased expression of long noncoding RNA GAS5 indicates a poor prognosis and promotes cell proliferation in gastric cancer. *BMC Cancer* 2014a;14:319.

Sun M, Liu X-H, Wang K-M, Nie F, Kong R, Yang J, et al. Downregulation of BRAF activated non-coding RNA is associated with poor prognosis for non-small cell lung cancer and promotes metastasis by affecting epithelial-mesenchymal transition. *Mol Cancer* 2014b;13:68. Doi : 10.1186/1471-264-15-545.

Sun J, Bie B, Zhang S, Yang J and Li Z. Long Non-Coding RNAs: Critical Players in Hepatocellular Carcinoma. *Int J Mol Sci* 2014c;15:20434-20448.

Sun R. Analysis of Long Noncoding RNAs Controlling Cell-Cycle Gene Expression and Proliferation in Breast Cancer Cells. *Molecular Cell* 2015;59:698–711.

Sun Z, Ou C, Ren W, Xie X, Li X, Li G. Downregulation of long non-coding RNA ANRIL suppresses lymphangiogenesis and lymphatic metastasis in colorectal cancer. *Oncotarget* 2016. doi: 10.18632.

Sunwoo H, Dinger ME, Wilusz JE, Amaral PP, Mattick JS, Spector DL. MEN epsilon/beta nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. *Genome Res* 2009;19:347–59.

Sytņikova YA, Rahman R, Chirn GW, Clark JP, Lau NC. Transposable element dynamics and PIWI regulation impacts lncRNA and gene expression diversity in *Drosophila* ovarian cell cultures. *Genome Res* 2014;24:1977-90.

Taft RJ, Pheasant M, Mattick JS. The relationship between non-protein-coding DNA and eukaryotic complexity. *Bioessays* 2007;29:288-99.

Taft RJ, Kaplan CD, Simons C, Mattick JS: Evolution, biogenesis and function of promoter-associated RNAs. *Cell Cycle* 2009;8:2332–38.

Tahira AC, Kubrusly MS, Faria MF, Dazzani B, Fonseca RS, Maracaja-Coutinho V, Verjovski-Almeida S, Machado MC, Reis EM. Long noncoding intronic RNAs are differentially expressed in primary and metastatic pancreatic cancer. *Mol Cancer* 2011;10:141. doi: 10.1186/1476-4598-10-141.

Takahashi H, Carninci P. Widespread genome transcription: new possibilities for RNA therapies. *Biochem Biophys Res Commun* 2014a;452:294–301.

Takahashi Y, Sawada G, Kurashige J, Uchi R, Matsumura T, Ueo H, et al. Amplification of PVT-1 is involved in poor prognosis via apoptosis inhibition in colorectal cancers. *Br J Cancer* 2014b;110:164–71.

Tamborero D, Gonzalez-Perez A, Perez-Llamas C, Deu-Pons J, Kandath C, Reimand J, Lawrence MS, Getz G, Bader GD, Ding L, Lopez-Bigas N. Comprehensive identification of mutational cancer driver genes across 12 tumor types. *Sci Rep* 2013;3:2650.

Tang SS, Zheng BY, Xiong XD. LincRNA-p21: Implications in Human Diseases. *Int J Mol Sci* 2015;16:18732-40.

Tano K, Mizuno R, Okada T, Rakwal R, Shibato J, Masuo Y, et al. MALAT-1 enhances cell motility of lung adenocarcinoma cells by influencing the expression of motility-related genes. *FEBS Lett* 2010;584:4575–80.

Tano K, Akimitsu N. Long non-coding RNAs in cancer progression. *Front Genet* 2012;3:219. doi: 10.3389/fgene.2012.00219.

Tay Y, Tan SM, Karreth FA, Lieberman J, Pandolfi PP. Characterization of dual PTEN and p53-targeting microRNAs identifies microRNA-638/Dnm2 as a two-hit oncogenic locus. *Cell Rep* 2014a;8:714-22.

Tay Y, Karreth FA, Pandolfi PP. Aberrant ceRNA activity drives lung cancer. *Cell Res* 2014b;24:259-60.

Tedeschi L, Lande C, Cecchettini A, Citti L. Hammerhead ribozymes in therapeutic target discovery and validation. *Drug Discov Today* 2009;14:776-83.

Tee AE, Liu B, Song R, Li J, Pasquier E, Cheung BB, Jiang C, Marshall GM, Haber M, Norris MD, Fletcher JJ, Dinger ME, Liu T. The long noncoding RNA MALAT1 promotes tumor-driven angiogenesis by up-regulating pro-angiogenic gene expression. *Oncotarget* 2016;7:8663-75.

Teixeira MR, Pandis N, Dietrich CU, Reed W, Andersen J, Qvist H, Heim S. Chromosome banding analysis of gynecomastias and breast carcinomas in men. *Genes Chromosomes Cancer* 1998;23:16-20.

Teschendorff AE, Lee SH, Jones A, Fiegl H, Kalwa M, Wagner W, Chindera K, Evans I, Dubeau L, Orjalo A, Horlings HM, Niederreiter L, Kaser A, Yang W, Goode EL, Fridley BL, Jenner RG, Berns EM, Wik E, Salvesen HB, Wisman GB, van der Zee AG, Davidson B, Tropé CG, Lambrechts S,

Tollervey JR, Curk T, Rogelj B, Briese M, Cereda M, Kayikci M, König J, Hortobágyi T, Nishimura AL, Zupunski V, Patani R, Chandran S, Rot G, Zupan B, Shaw CE, Ule J. Characterizing the RNA targets and position-dependent splicing regulation by TDP-43. *Nat Neurosci* 2011;14:452-8.

Vergote I, Calvert H, Jacobs IJ, Widschwendter M. HOTAIR and its surrogate DNA methylation signature indicate carboplatin resistance in ovarian cancer. *Genome Med* 2015;7:108. PMC4619324.

The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012;489:57–74.

Tian D, Sun S, Lee JT. The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation. *Cell* 2010;143:390-403.

Timp W, Feinberg AP. Cancer as a dysregulated epigenome allowing cellular growth advantage at the expense of the host. *Nat Rev Cancer* 2013;13:497-510.

Tisseur M, Kwapisz M, Morillon A. Pervasive transcription - Lessons from yeast. *Biochimie* 2011;93:1889-96.

Tong YK, Lo YM. Plasma epigenetic markers for cancer detection and prenatal diagnosis. *Front Biosci* 2006;11:2647-56.

Torres-Ruiz R, Rodriguez-Perales S. CRISPR-Cas9: A Revolutionary Tool for Cancer Modelling. *Int J Mol Sci* 2015;16:22151-68.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 2010;28:511-5.

Tripathi V, Ellis JD, Shen Z, Song DY, Pan Q, Watt AT, Freier SM, Bennett CF, Sharma A, Bubulya PA et al. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol Cell* 2010;39:925–38.

Tripathi V, Shen Z, Chakraborty A, Giri S, Freier SM, Wu X, Zhang Y, Gorospe M, Prasanth SG, Lal A, Prasanth KV. Long noncoding RNA MALAT1 controls cell cycle progression by regulating the expression of oncogenic transcription factor B-MYB. *PLoS Genet* 2013;9:e1003368.

Tsai MC, Manor O, Wan Y, Mosammamarast N, Wang JK, Lan F, et al. Long noncoding RNA as modular scaffold of histone modification complexes. *Science* 2010;329:689–93.

Tsai MC, Spitale RC, Chang HY. Long intergenic noncoding RNAs: new links in cancer progression. *Cancer Res* 2011;71:3–7.

Tsang WP, Ng EK, Ng SS, Jin H, Yu J, Sung JJ, Kwok TT. Oncofetal H19-derived miR-675 regulates tumor suppressor RB in human colorectal cancer. *Carcinogenesis* 2010;31:350-8.

Tseng YY, Bagchi A. The PVT1-MYC duet in cancer. *Mol Cell Oncol* 2015;2:e974467. doi: 10.4161/23723556.2014.974467.

- Turner NC, Reis-Filho JS. Basal-like breast cancer and the BRCA1 phenotype. *Oncogene* 2006;25:5846-53.
- Turner NC, Reis-Filho JS, Russell AM, Springall RJ, Ryder K, Steele D, Savage K, Gillett CE, Schmitt FC, Ashworth A, Tutt AN. BRCA1 dysfunction in sporadic basal-like breast cancer. *Oncogene* 2007;26:2126-32.
- Tye CE, Gordon JA, Martin-Buley LA, Stein JL, Lian JB, Stein GS. Could lncRNAs be the missing links in control of mesenchymal stem cell differentiation? *J Cell Physiol* 2015;230:526-34.
- Ulitsky I, Shkumatava A, Jan CH, Sive H, Bartel DP. Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell* 2011;147:1537-50.
- Ulitsky I, Bartel DP. lincRNAs: genomics, evolution, and mechanisms. *Cell* 2013;154:26-46.
- Ulveling D, Francastel C, Hubé F: When one is better than two: RNA with dual functions. *Biochimie* 2011;93:633-44.
- Valen E, Preker P, Andersen PR, Zhao X, Chen Y, Ender C, Dueck A, Meister G, Sandelin A, Jensen TH. Biogenic mechanisms and utilization of small RNAs derived from human protein-coding genes. *Nat Struct Mol Biol* 2011;18:1075-82.
- Vallot C, Huret C, Lesecque Y, Resch A, Oudrhiri N, Bennaceur-Griscelli A, Duret L, Rougeulle C. XACT, a long noncoding transcript coating the active X chromosome in human pluripotent cells. *Nat Genet* 2013;45:239-41.
- Van Amerongen R, Bowman AN, Nusse R. Developmental stage and time dictate the fate of Wnt/ β -catenin-responsive stem cells in the mammary gland. *Cell Stem Cell* 2012;11:387-400.
- Van Dijk AD, van Mourik S, van Ham RC. Mutational robustness of gene regulatory networks. *PLoS One* 2012a;7:e30591. doi: 10.1371/journal.pone.0030591.
- Van Dijk M, Thulluru HK, Mulders J, Michel OJ, Poutsma A, Windhorst S, Kleiverda G, Sie D, Lachmeijer AM, Oudejans CB. HELLP babies link a novel lincRNA to the trophoblast cell cycle. *J Clin Invest* 2012b;122:4003-11.
- Van Grembergen 2016 (sciencedirect.com) Van Grembergen O, Bizet M, de Bony EJ, Calonne E, Putmans P, Brohée S, Olsen C, Guo M, Bontempi G, Sotiriou C, Defrance M, Fuks F. Portraying breast cancers with long noncoding RNAs. *Sci Adv* 2016;2:e1600220. doi:10.1126/sciadv.1600220.
- Van Heesch S, van Iterson M, Jacobi J, Boymans S, Essers PB, de Bruijn E, Hao W, MacInnes AW, Cuppen E, Simonis M. Extensive localization of long noncoding RNAs to the cytosol and mono- and polyribosomal complexes. *Genome Biol* 2014;15:R6. doi: 10.1186/gb-2014-15-1-r6.
- Van Keymeulen A, Rocha AS, Ousset M, Beck B, Bouvencourt G, Rock J, Sharma N, Dekoninck S, Blanpain C. Distinct stem cells contribute to mammary gland development and maintenance. *Nature* 2011;479:189-93.

Van Roozendaal CE, Gillis AJ, Klijn JG, van Ooijen B, Claassen CJ, Eggermont AM, Henzen-Logmans SC, Oosterhuis JW, Foekens JA, Looijenga LH. Loss of imprinting of IGF2 and not H19 in breast cancer, adjacent normal tissue and derived fibroblast cultures. *FEBS Lett* 1998;437:107-11.

Vance KW, Ponting CP. Transcriptional regulatory functions of nuclear long noncoding RNAs. *Trends Genet* 2014;30:348-55.

Veedu RN, Wengel J. Locked nucleic acids: promising nucleic acid analogs for therapeutic applications. *Chem Biodivers* 2010;7:536-42.

Venkatraman A, He XC, Thorvaldsen JL, Sugimura R, Perry JM, Tao F, Zhao M, Christenson MK, Sanchez R, Yu JY, Peng L, Haug JS, Paulson A, Li H, Zhong XB, Clemens TL, Bartolomei MS, Li L. Maternal imprinting at the H19-Igf2 locus maintains adult haematopoietic stem cell quiescence. *Nature* 2013;500:345-9.

Vihanto MM, Plock J, Erni D, Frey BM, Frey FJ, Huynh-Do U. Hypoxia up-regulates expression of Eph receptors and ephrins in mouse skin. *FASEB J* 2005;19:1689-91.

Vikram R, Ramachandran R, Abdul KS. Functional significance of long non-coding RNAs in breast cancer. *Breast Cancer* 2014;21:5:515-521.

Visvader JE, Stingl J: Mammary stem cells and the differentiation hierarchy: current status and perspectives. *Genes Dev* 2014;28:1143-1158.

Volders PJ, Verheggen K, Menschaert G, Vandepoele K, Martens L, Vandesompele J, Mestdagh P. An update on LNCipedia: a database for annotated human lncRNA sequences. *Nucleic Acids Res* 2015;43 (Database issue):D174-80. doi: 10.1093/nar/gku1060. Erratum in: *Nucleic Acids Res* 2015;43:4363-4.

Vogelstein B, Kinzler KW. Cancer genes and the pathways they control. *Nat Med* 2004;10:789-99.

Vollmers C, Schmitz RJ, Nathanson J, Yeo G, Ecker JR, Panda S. Circadian oscillations of protein-coding and regulatory RNAs in a highly dynamic mammalian liver epigenome. *Cell Metab* 2012;16:833-47.

Von Roretz C, Di Marco S, Mazroui R, Gallouzi IE. Turnover of AU-rich-containing mRNAs during stress: a matter of survival. *Wiley Interdiscip Rev RNA* 2011;2:336-47.

Wagschal A, Sutherland HG, Woodfine K, Henckel A, Chebli K, Schulz R, Oakey RJ, Bickmore WA, Feil R. G9a histone methyltransferase contributes to imprinting in the mouse placenta. *Mol Cell Biol* 2008;28:1104-13.

Wahlestedt C. Targeting long non-coding RNA to therapeutically upregulate gene expression. *Nat Rev Drug Discov* 2013;12:433-46.

Walsh AL, Tuzova AV, Bolton EM, Lynch TH, Perry AS. Long noncoding RNAs and prostate carcinogenesis: the missing 'linc'? *Trends Mol Med* 2014;20:428-36.

Wan LB, Bartolomei MS. Regulation of imprinting in clusters: noncoding RNAs versus insulators. *Adv Genet* 2008;61:207-23.

Wan G, Mathur R, Hu X, Liu Y, Zhang X, Peng G, Lu X. Long non-coding RNA ANRIL (CDKN2B-AS) is induced by the ATM-E2F1 signaling pathway. *Cell Signal* 2013a; 25:1086–95.

Wan J, Huang M, Zhao H, Wang C, Zhao X, Jiang X, et al. A novel tetranucleotide repeat polymorphism within KCNQ1OT1 confers risk for hepatocellular carcinoma. *DNA Cell Biol* 2013b;32:628–34.

Wan G, Hu X, Liu Y, Han C, Sood AK, Calin GA, Zhang X, Lu X. A novel non-coding RNA lncRNA-JADE connects DNA damage signalling to histone H4 acetylation. *EMBO J* 2013c;32:2833-47.

Wang Y, Leung FC. An evaluation of new criteria for CpG islands in the human genome as gene markers. *Bioinformatics* 2004;20:1170-7.

Wang X, Arai S, Song X, Reichart D, Du K, Pascual G, et al. Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. *Nature* 2008;454:126–30.

Wang KC, Chang HY: Molecular mechanisms of long noncoding RNAs. *Mol Cell* 2011a;43:904–14.

Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y, et al. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* 2011b;472:120–24.

Wang Y, Chen W, Yang C, Wu W, Wu S, Qin X, et al. Long non-coding RNA UCA1a (CUDR) promotes proliferation and tumorigenesis of bladder cancer. *Int J Oncol* 2012;41:276–84.

Wang Y, Xu Z, Jiang J, Xu C, Kang J, Xiao L, Wu M, Xiong J, Guo X, Liu H. Endogenous miRNA sponge lincRNA-RoR regulates Oct4, Nanog, and Sox2 in human embryonic stem cell self-renewal. *Dev Cell* 2013;25:69-80.

Wang F, Yuan J-H, Wang S-B, Yang F, Yuan S-X, Ye C, et al. Oncofetal long noncoding RNA PVT1 promotes proliferation and stem cell-like property of hepatocellular carcinoma cells by stabilizing NOP2. *Hepatology* 2014a;60:1278–90.

Wang P, Xue Y, Han Y et al. The STAT3-binding long noncoding RNA lnc-DC controls human dendritic cell differentiation. *Science* 2014b;344:310–313.

Wang G, Li Z, Zhao Q, Zhu Y, Zhao C, Li X, et al. LincRNA-p21 enhances the sensitivity of radiotherapy for human colorectal cancer by targeting the Wnt/ β -catenin signaling pathway. *Oncol Rep* 2014c;31:1839–45.

Wang L, Fu D, Qiu Y, Xing X, Xu F, Han C, Xu X1, Wei Z, Zhang Z, Ge J, Cheng W5, Xie HL. Genome-wide screening and identification of long noncoding RNAs and their interaction with protein coding RNAs in bladder urothelial cell carcinoma. *Cancer Lett* 2014d;349:77-86.

Wang J, Wang H, Zhang Y, Zhen N, Zhang L, Qiao Y, Weng W, Liu X, Ma L, Xiao W, Yu W, Chu Q, Pan Q, Sun F. Mutual inhibition between YAP and SRSF1 maintains long non-coding RNA, Malat1-induced tumorigenesis in liver cancer. *Cell Signal* 2014e;26:1048-59.

Wang L, Zhang N, Wang Z, Ai DM, Cao ZY, Pan HP. Pseudogene PTENP1 Functions as a Competing Endogenous RNA (ceRNA) to Regulate PTEN Expression by Sponging miR-499-5p. *Biochemistry (Mosc)* 2016;81:739-47.

Wapinski O, Chang HY. Long noncoding RNAs and human disease. *Trends Cell Biol* 2011;21:354-61.

Watanabe T, Totoki Y, Toyoda A, Kaneda M, Kuramochi-Miyagawa S, Obata Y, Chiba H, Kohara Y, Kono T, Nakano T, Surani MA, Sakaki Y, Sasaki H. Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* 2008;453:539-43.

Watrin M, Von Pelchrzim F, Dausse E, Schroeder R, Toulmé JJ. In vitro selection of RNA aptamers derived from a genomic human library against the TAR RNA element of HIV-1. *Biochemistry* 2009;48:6278-84.

Watters KM, Bryan K, Foley NH, Meehan M, Stallings RL. Expressional alterations in functional ultra-conserved non-coding RNAs in response to all-trans retinoic acid--induced differentiation in neuroblastoma cells. *BMC Cancer* 2013;13:184. doi: 10.1186/1471-2407-13-184.

Watts JK, Corey DR. Gene silencing by siRNAs and antisense oligonucleotides in the laboratory and the clinic. *J Pathol* 2012;226:365-79.

Weakley SM, Wang H, Yao Q, Chen C. Expression and function of a large non-coding RNA gene XIST in human cancer. *World J Surg* 2011;35:1751-6.

Wei G, Luo H, Sun Y, Li J, Tian L, Liu W, Liu L, Luo J, He J, Chen R. Transcriptome profiling of esophageal squamous cell carcinoma reveals a long noncoding RNA acting as a tumor suppressor. *Oncotarget* 2015;6:17065-80.

Weick and Miska, 2014 Weick EM, Miska EA. piRNAs: from biogenesis to function. *Development* 2014;141:3458-71.

Wen J, Xu J, Sun Q, Xing C, Yin W. Upregulation of long non coding RNA PCAT-1 contributes to cell proliferation, migration and apoptosis in hepatocellular carcinoma. *Mol Med Rep* 2016;13:4481-6.

West JA, Davis CP, Sunwoo H, Simon MD, Sadreyev R, Wang P, Tolstorukov MY, Kingston RE. The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites. *Mol Cell* 2014;55:791-802.

Whitehead J, Pandey GK, Kanduri C. Regulation of the mammalian epigenome by long noncoding RNAs. *Biochim Biophys Acta* 2009;1790:936-47.

Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, Rahl PB, Lee TI, Young RA. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* 2013;153:307-19.

Wilusz JE, Freier SM, Spector DL. 3' end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. *Cell* 2008;135:919-32.

Wilusz JE. Long noncoding RNAs: Re-writing dogmas of RNA processing and stability. *Biochim Biophys Acta* 2016;1859:128-38.

Wolff AC, Hammond ME, Schwartz JN, Hagerty KL, Allred DC, Cote RJ, Dowsett M, Fitzgibbons PL, Hanna WM, Langer A, McShane LM, Paik S, Pegram MD, Perez EA, Press MF, Rhodes A, Sturgeon C, Taube SE, Tubbs R, Vance GH, van de Vijver M, Wheeler TM, Hayes DF; American Society of Clinical Oncology/College of American Pathologists. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. *Arch Pathol Lab Med* 2007;131:18-43.

Wolff AC, Hammond ME, Hicks DG, Dowsett M, McShane LM, Allison KH, Allred DC, Bartlett JM, Bilous M, Fitzgibbons P, Hanna W, Jenkins RB, Mangu PB, Paik S, Perez EA, Press MF, Spears PA, Vance GH, Viale G, Hayes DF; American Society of Clinical Oncology; College of American Pathologists. Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. *J Clin Oncol* 2013;31:3997-4013.

Wood AJ, Oakey RJ. Genomic imprinting in mammals: emerging themes and established theories. *PLoS Genet* 2006;2(11):e147.

Wu J, Zhang J, Shen B, Yin K, Xu J, Gao W, Zhang L. Long noncoding RNAnctCF7, induced by IL-6/STAT3 transactivation, promotes hepatocellular carcinoma aggressiveness through epithelial-mesenchymal transition. *J Exp Clin Cancer Res* 2015;34:116. doi: 10.1186/s13046-015-0229-3.

Xiang J-F, Yin Q-F, Chen T, Zhang Y, Zhang X-O, Wu Z, et al. Human colorectal cancer-specific CCAT1-L lncRNA regulates long-range chromatin interactions at the MYC locus. *Cell Res* 2014;24:513-31.

Xiao C, Sharp JA, Kawahara M, Davalos AR, Difilippantonio MJ, Hu Y, Li W, Cao L, Buetow K, Ried T, Chadwick BP, Deng CX, Panning B. The XIST noncoding RNA functions independently of BRCA1 in X inactivation. *Cell* 2007;128:977-89.

Xie H, Ma H, Zhou D. Plasma HULC as a promising novel biomarker for the detection of hepatocellular carcinoma. *Biomed Res Int* 2013;2013:136106.

Xie W, Ren B. Developmental biology. Enhancing pluripotency and lineage specification. *Science* 2013;341:245-7.

Xie H, Rachakonda PS, Heidenreich B, Nagore E, Sucker A, Hemminki K, Schadendorf D, Kumar R. Mapping of deletion breakpoints at the CDKN2A locus in melanoma: detection of MTAP-ANRIL fusion transcripts. *Oncotarget* 2016;7:16490-504.

Xing Z, Lin A, Li C, Liang K, Wang S, Liu Y, Park PK, Qin L, Wei Y, Hawke DH, Hung MC, Lin C, Yang L. lncRNA directs cooperative epigenetic regulation downstream of chemokine signals. *Cell* 2014;159:1110-25.

Xu B, Gerin I, Miao H, Vu-Phan D, Johnson CN, Xu R, Chen XW, Cawthorn WP, MacDougald OA, Koenig RJ. Multiple roles for the non-coding RNA SRA in regulation of adipogenesis and insulin sensitivity. *PLoS One* 2010;5:e14199. doi: 10.1371/journal.pone.0014199.

Xu C, Yang M, Tian J, Wang X, Li Z. MALAT-1: a long non-coding RNA and its important 3' end functional motif in colorectal cancer metastasis. *Int J Oncol* 2011;39:169-75.

Xu S, Sui S, Zhang J, Bai N, Shi Q. Downregulation of long noncoding RNA MALAT1 induces epithelial-to-mesenchymal transition via the PI3K-AKT pathway in breast cancer. *Int J Clin Exp Pathol* 2015;8:4881-4891.

Xue S, Li QW, Che JP, Guo Y, Yang FQ, Zheng JH. Decreased expression of long non-coding RNA NBAT-1 is associated with poor prognosis in patients with clear cell renal cell carcinoma. *Int J Clin Exp Pathol* 2015;8:3765-74.

Yahya SM, Elsayed GH. A summary for molecular regulations of miRNAs in breast cancer. *Clin Biochem* 2015;48:388-96.

Yamazaki T, Hirose T. The building process of the functional paraspeckle with long non-coding RNAs. *Front Biosci (Elite Ed)* 2015;7:1-41.

Yan J, Zhou X, Dang Y, Yin C, Zhang G. Prognostic role of the long non-coding RNA metastasis-associated lung adenocarcinoma transcript 1 in various cancers: A meta-analysis. *Mol Clin Oncol* 2016;4:100-106.

Yang F, Zhang L, Huo X, Yuan J, Xu D, Yuan S, et al. Long noncoding RNA high expression in hepatocellular carcinoma facilitates tumor growth through enhancer of zeste homolog 2 in humans. *Hepatology* 2011a;54:1679-89.

Yang XR, Chang-Claude J, Goode EL, Couch FJ, Nevanlinna H, Milne RL, Gaudet M, Schmidt MK, Broeks A, Cox A et al: Associations of breast cancer risk factors with tumor subtypes: a pooled analysis from the Breast Cancer Association Consortium studies. *J Natl Cancer Inst* 2011b;103:250-263.

Yang Z, Zhou L, Wu L-M, Lai MC, Xie HY, Zhang F, et al. Overexpression of long non-coding RNA HOTAIR predicts tumor recurrence in hepatocellular carcinoma patients following liver transplantation. *Ann Surg Oncol* 2011c;18:1243-50.

Yang L, Lin C, Jin C, Yang JC, Tanasa B, Li W, et al. lncRNA-dependent mechanisms of androgen-receptor-regulated gene activation programs. *Nature* 2013a;500:598-602.

Yang Y, Li H, Hou S, Hu B, Liu J, Wang J. The noncoding RNA expression profile and the effect of lncRNA AK126698 on cisplatin resistance in non-small-cell lung cancer cell. *PLoS One* 2013b;8:e65309.

Yang F, Xue X, Bi J, Zheng L, Zhi K, Gu Y, Fang G. Long noncoding RNA CCAT1, which could be activated by c-Myc, promotes the progression of gastric carcinoma. *J Cancer Res Clin Oncol* 2013c;139:437-45.

Yang F, Huo X, Yuan S, Zhang L, Zhou W, Wang F, et al. Repression of the long noncoding RNA-LET by histone deacetylase 3 contributes to hypoxia-mediated metastasis. *Mol Cell* 2013d;49:1083–96.

Yang J, Lin J, Liu T, Chen T, Pan S, Huang W, Li S. Analysis of lncRNA expression profiles in non-small cell lung cancers (NSCLC) and their clinical subtypes. *Lung Cancer* 2014a;85:110-15.

Yang L, Froberg JE, Lee JT. Long Noncoding RNAs: Fresh Perspectives into the RNA world. *Trends Biochem Sci* 2014b;39:35–43.

Yang YR, Zang SZ, Zhong CL, Li YX, Zhao SS, Feng XJ. Increased expression of the lncRNA PVT1 promotes tumorigenesis in non-small cell lung cancer. *Int J Clin Exp Pathol* 2014c;7:6929–35.

Yang C, Wu D, Gao L, Liu X, Jin Y, Wang D, Wang T, Li X. Competing endogenous RNA networks in human cancer: hypothesis, validation, and perspectives. *Oncotarget* 2016a;7:13479-90.

Yang P, Chen T, Xu Z, Zhu H, Wang J, He Z. Long noncoding RNA GAPLINC promotes invasion in colorectal cancer by targeting SNAI2 through binding with PSF and NONO. *Oncotarget* 2016b. doi: 10.18632/oncotarget.9741.

Yao Y, Li J, Wang L. Large Intervening Non-Coding RNA HOTAIR Is an Indicator of Poor Prognosis and a Therapeutic Target in Human Cancers. *Int J Mol Sci* 2014;15:18985–99.

Yap KL, Li S, Muñoz-Cabello AM, Raguz S, Zeng L, Mujtaba S, Gil J, Walsh MJ, Zhou MM. Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. *Mol Cell* 2010;38:662-74.

Ye S, Yang L, Zhao X, Song W, Wang W, Zheng S. Bioinformatics Method to Predict Two Regulation Mechanism: TF-miRNA-mRNA and lncRNA-miRNA-mRNA in Pancreatic Cancer. *Cell Biochem Biophys* 2014;70:1849-58.

Yen K, Vinayachandran V, Batta K, Koerber RT, Pugh BF. Genome-wide nucleosome specificity and directionality of chromatin remodelers. *Cell* 2012;149:1461-73.

Ying L, Chen Q, Wang Y, Zhou Z, Huang Y, Qiu F. Upregulated MALAT-1 contributes to bladder cancer cell migration by inducing epithelial-to-mesenchymal transition. *Mol Biosyst* 2012;8:2289–94.

Ying L, Huang Y, Chen H, Wang Y, Xia L, Chen Y, et al. Downregulated MEG3 activates autophagy and increases cell proliferation in bladder cancer. *Mol Biosyst* 2013;9:407–11.

Yoon JH, Abdelmohsen K, Srikantan S, Yang X, Martindale JL, De S, et al. LincRNAp21 suppresses target mRNA translation. *Mol Cell* 2012;47:648–55.

Yoon J-H, Abdelmohsen K, Gorospe M. Posttranscriptional gene regulation by long noncoding RNA. *J Mol Biol* 2013;425:3723–30.

Yoshimizu T, Miroglio A, Ripoche MA, Gabory A, Vernucci M, Riccio A, Colnot S, Godard C, Terris B, Jammes H, Dandolo L. The H19 locus acts in vivo as a tumor suppressor. *Proc Natl Acad Sci U S A* 2008;105:12417-22.

Yoshimoto R, Mayeda A, Yoshida M, Nakagawa S. MALAT1 long non-coding RNA in cancer. *Biochim Biophys Acta* 2016;1859:192-9.

Yu W, Gius D, Onyango P, Muldoon-Jacobs K, Karp J, Feinberg AP, et al. Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. *Nature* 2008;451:202–6.

Yu TY, Kao YW, Lin JJ. Telomeric transcripts stimulate telomere recombination to suppress senescence in cells lacking telomerase. *Proc Natl Acad Sci U S A* 2014;111:3377-82.

Yuan SX, Yang F, Yang Y, Tao QF, Zhang J, Huang G, et al. Long noncoding RNA associated with microvascular invasion in hepatocellular carcinoma promotes angiogenesis and serves as a predictor for hepatocellular carcinoma patients' poor recurrence-free survival after hepatectomy. *Hepatology* 2012;56:2231–41.

Yuan SX, Tao QF, Wang J, Yang F, Liu L, Wang LL, Zhang J, Yang Y, Liu H, Wang F, Sun SH, Zhou WP. Antisense long non-coding RNA PCNA-AS1 promotes tumor growth by regulating proliferating cell nuclear antigen in hepatocellular carcinoma. *Cancer Lett* 2014;349:87-94.

Zack TI, Schumacher SE, Carter SL, Cherniack AD, Saksena G, Tabak B, Lawrence MS, Zhsng CZ, Wala J, Mermel CH, Sougnez C, Gabriel SB, Hernandez B, Shen H, Laird PW, Getz G, Meyerson M, Beroukhim R. Pan-cancer patterns of somatic copy number alteration. *Nat Genet* 2013;45:1134-40.

Zeng C, Xu Y, Xu L, Yu X, Cheng J, Yang L, et al. Inhibition of long non-coding RNA NEAT1 impairs myeloid differentiation in acute promyelocytic leukemia cells. *BMC Cancer* 2014;14:693.

Zhang X, Lian Z, Padden C et al. A myelopoiesis-associated regulatory intergenic noncoding RNA transcript within the human HOXA cluster. *Blood* 2009;113:2526–2534.

Zhang X, Gejman R, Mahta A, Zhong Y, Rice KA, Zhou Y, et al. Maternally expressed gene 3, an imprinted noncoding RNA gene, is associated with meningioma pathogenesis and progression. *Cancer Res* 2010;70:2350–8.

Zhang B, Arun G, Mao YS, Lazar Z, Hung G, Bhattacharjee G, Xiao X, Booth CJ, Wu J, Zhang C, Spector DL. The lncRNA Malat1 is dispensable for mouse development but its transcription plays a cis-regulatory role in the adult. *Cell Rep* 2012;2:111-23.

Zhang Z, Zhu Z, Watabe K, Zhang X, Bai C, Xu M, et al. Negative regulation of lncRNA GAS5 by miR-21. *Cell Death Differ* 2013;20:1558–68.

Zhang Y, Yang L, Chen LL. Long non-coding RNAs: a new frontier in the study of human diseases. *Int J Biochem Cell Biol* 2014a;54:338-49.

Zhang Y, Yang L, He Y, Meng XM, Huang C, Wu BM, Zhang L, Lv XW, Li J. Life without A tail: new formats of long noncoding RNAs. *Cancer Lett* 2014b;1;344:20-7.

Zhang E, Yin D, Sun M, Kong R, Liu X, You L, et al. P53-regulated long non-coding RNA TUG1 affects cell proliferation in human non-small cell lung cancer, partly through epigenetically regulating HOXB7 expression. *Cell Death Dis* 2014c;5:e1243.

Zhang EB, Kong R, Yin DD, You LH, Sun M, Han L, Xu TP, Xia R, Yang JS, De W, Chen Jf. Long noncoding RNA ANRIL indicates a poor prognosis of gastric cancer and promotes tumor growth by epigenetically silencing of miR-99a/miR-449a. *Oncotarget* 2014d;5:2276-92.

Zhang X, Weissman SM, Newburger PE. Long intergenic non-coding RNA HOTAIRM1 regulates cell cycle progression during myeloid maturation in NB4 human promyelocytic leukemia cells. *RNA Biol* 2014e;11:777-87.

Zhang Y, Li Z, Zhang Y, Zhong Q, Chen Q, Zhang L. Molecular mechanism of HEIH and HULC in the proliferation and invasion of hepatoma cells. *Int J Clin Exp Med* 2015;8:12956-62.

Zhang S, Zhang G, Liu J. Long noncoding RNA PVT1 promotes cervical cancer progression through epigenetically silencing miR-200b. *APMIS* 2016;124:649-58.

Zhao W, Luo J, Jiao S. Comprehensive characterization of cancer subtype associated long non-coding RNAs and their clinical implications. *Sci Rep* 2014;4:6591. doi: 10.1038/srep06591.

Zhao Y, Sun H, Wang H. Long noncoding RNAs in DNA methylation: new players stepping into the old game. *Cell Biosci* 2016;6:45. doi: 10.1186/s13578-016-0109-3.

Zheng J, Dong P, Mao Y, Chen S, Wu X, Li G, Lu Z, Yu F. lincRNA-p21 inhibits hepatic stellate cell activation and liver fibrogenesis via p21. *FEBS J* 2015a;282:4810-21.

Zheng H, Wang M, Jiang L, Chu H, Hu J, Ning J, Li B, Wang D, Xu J. BRAF-activated Long Non-coding RNA Modulates Papillary Thyroid Carcinoma Cell Proliferation through Regulating Thyroid Stimulating Hormone Receptor. *Cancer Res Treat* 2015b. doi: 10.4143/crt.2015.118.

Zheng S, Chen H, Wang Y, Gao W, Fu Z, Zhou Q, Jiang Y, Lin Q, Tan L, Ye H, Zhao X, Luo Y, Li G, Ye L, Liu Y, Li W, Li Z, Chen R. Long non-coding RNA LOC389641 promotes progression of pancreatic ductal adenocarcinoma and increases cell invasion by regulating E-cadherin in a TNFRSF10A-related manner. *Cancer Lett* 2016;371:354-65.

Zhou XZ, Lu KP. The Pin2/TRF1-interacting protein PinX1 is a potent telomerase inhibitor. *Cell* 2001;107(3):347-59.

Zhou Y, Cheunsuchon P, Nakayama Y. Activation of paternally expressed genes and perinatal death caused by deletion of the Gtl2 gene. *Development* 2010;137:2643–2652.

Zhou Y, Zhang X, Klibanski A. MEG3 noncoding RNA: a tumor suppressor. *J Mol Endocrinol* 2012;48:R45–53.

Zhou P, Sun L, Liu D, Liu C, Sun L. Long Non-Coding RNA lincRNA-ROR Promotes the Progression of Colon Cancer and Holds Prognostic Value by Associating with miR-145. *Pathol Oncol Res* 2016. [Epub ahead of print] PubMed PMID: 27071407.

Zhu J, Fu H, Wu Y, Zheng X. Function of lncRNAs and approaches to lncRNA-protein interactions. *Sci China Life Sci* 2013;56:876-85.

Zhuang J, Lu Q, Shen B, Huang X, Shen L, Zheng X, Huang R, Yan J, Guo H. TGF β 1 secreted by cancer-associated fibroblasts induces epithelial-mesenchymal transition of bladder cancer cells through lncRNA-ZEB2NAT. *Sci Rep* 2015;5:11924. doi: 10.1038/srep11924.

ANNEXES

Annexe 1

Article collaboratif consacré à l'analyse du transcriptome non codant (lncRNAs anti-sens) des carcinomes prostatiques infiltrants.

Characterization of human antisense long non-coding RNA reveals markers for high risk prostate cancer.

Zohra Saci¹, Marina Pinskaya¹, Claire Bertrand¹, Ivan Bièche⁵, Didier Meseure⁵, Virginie Firlej⁴, Alexandre de la Taille⁴, Francis Vacherot⁴, Yves Allory⁴, Arturo Londono², Daniel Gautheret^{3*} and Antonin Morillon^{1*}

¹ ncRNA, epigenetic and genome fluidity, Institut Curie, Centre de Recherche, CNRS UMR 3244, Université Pierre et Marie Curie, 26 rue d'Ulm, 75248 Paris Cedex 05, France

² Telomeres and Genetic Instability, Institut Curie, Centre de Recherche, CNRS UMR 3244, Université Pierre et Marie Curie, 26 rue d'Ulm, 75248 Paris Cedex 05, France

³ Institut de Génétique et Microbiologie, CNRS UMR 8621, Université Paris Sud, bâtiment 400, 91405 Orsay Cedex, France

⁴ Mondor Hospital

⁵ Curie Hospital-translational department, Institut Curie, 26 rue d'Ulm 75248 Paris

* co-corresponding author

Among thousands of long non-coding (lnc)RNAs encoded within the human genome, antisense transcripts have been poorly studied despite their regulatory importance in development and disease¹⁻³. Here, performing RNA-sequencing (RNA-seq) analysis of 18 cell lines and 32 cancer tissues we classified 7812 human antisense (as)lncRNAs including 6984 novel transcripts. Among them, the most robust Class1 aslncRNAs represent cell-type specific nuclear polyA+ transcripts that start mainly at the extremities of their paired-sense genes. A correlation study identified *cis*- and *trans*-associated aslncRNA/mRNA pairs, compatible with either sense/antisense, distal enhancer or *trans*-acting regulatory properties. Strikingly, RNA-seq of prostate tumors revealed an unprecedented signature of 157 aslncRNAs with 3 novel Class1 transcripts that separate high-risk prostate tumors on a cohort of patients. We propose that aslncRNAs are a broad class of transcripts with a high cancer prognosis potential.

High-throughput RNA sequencing produced catalogs of lncRNAs¹⁵⁻¹⁹ with now 58648 lncRNA annotated²⁰. lncRNAs are at least 200 nucleotides long, spliced, cell type-, tissue-specific and poorly conserved during evolution¹. Among different classes of lncRNAs, Natural Antisense Transcripts (NATs) are the less described with 4200 annotated antisense (as)lncRNAs²¹, certainly due to their low expression levels²². In addition, most genome-wide studies were focused on lncRNAs without addressing specifically their directionality explaining a lack of systematic characterization of human aslncRNA. AslncRNAs were shown to regulate gene expression at the post-transcriptional and transcriptional levels. Post-transcriptional regulation is triggered by double-stranded (ds)RNA formation with their paired-sense mRNA affecting its translation, splicing or stability³. Transcriptional regulation was considered as a result of interference between transcribing sense/antisense polymerases^{23,24} but also as a direct, still uncharacterized, effect of nuclear dsRNA formation²⁵. Transcriptional interference also generates *cis*-deposition of epigenetic marks where the nascent aslncRNA serves as recruiting/activating platforms for chromatin modifiers controlling the sense transcription as described for genomic imprinting, X-inactivation, tumor suppressor regulation and cell differentiation³. Interestingly, aslncRNA-mediated *trans*-regulation is poorly described at the notable exception of HOTAIR¹⁸ but might represent an important allele-specific regulatory function of aslncRNA.

To gain insight into antisense transcripts and their potential regulatory role(s), we classified different families of human aslncRNAs. We set up a dedicated bioinformatic pipeline (Fig. 1a, suppl Fig1, methods) to reveal aslncRNA and rank them, according to robust features. The protocol was applied to strand-specific polyA+ and total RNA-seq datasets, available on 14 ENCODE cell lines and our own RNA-seq data obtained on 3 human cell lines (6 conditions) and 4 normal and tumor prostate and breast tissues (table 1). Reads were mapped, assembled and merged, according to their density and proximity (see method). From this, aslncRNAs were extracted, according to the GENCODE V14 ORF annotation. To determine the most robust class of aslncRNA, we filtered segments with a minimum of 0.8 RPKM expression in at least two RNA-seq replicates and merged them with the 0.8 RPKM GENCODE V14 aslncRNAs (n=6984 and 894, respectively, Class3, Fig. 1b)²⁶, all tested for their coding potential. To increase stringency, we defined those with at least 1 junction (1935 segments, Class2a), or with a spliced EST (2940 segments, Class2b). Finally, we considered segments that met all criteria as the most robust class of aslncRNAs (1595, Class1), which were kept for the rest of the study. Of note, only 7% of novel Class3 aslncRNAs (481/6984, Fig. 1b) were

annotated by the recent MiTranscriptome project²⁰ (with at least 50% overlap), reflecting differences in tissue/cell origins of RNA samples but also in transcriptome assembly protocols.

Class1 aslncRNAs are 5 kilobases long in average (Suppl. Fig. 2) with a mean expression around 1 RPKM (Fig. 2a) that is much lower than lncRNAs and mRNAs with 10 and 100 RPKM, respectively. The Specificity Index (SI) calculated independently of their expression levels²⁷, showed that 47.6% of Class1 aslncRNAs are expressed in one cell type (SI>0.9), similar to lncRNAs (50.5%) but dramatically more than mRNAs (25.43%), indicating that aslncRNAs are expressed in a cell-type specific manner (see examples in Suppl. fig. 3). Systematic comparison of HeLa S3 polyA+ and polyA- datasets showed that Class1 aslncRNAs are mainly polyA+ as mRNAs but in contrast to snoRNAs (Fig. 2c). Transcriptome analysis of cytoplasmic and nuclear fractions of the ENCODE HeLa S3 and our HeLa S2 datasets revealed that the Class1 aslncRNAs are mainly nuclear (Fig. 2d, Suppl. Fig. 4a and 4b). Sub-cellular localization of 7 novel aslncRNAs was also validated by quantitative RT-PCR (suppl Fig. 4c).

To determine putative transcriptional start sites (TSS) of the Class1 aslncRNAs, we analyzed the presence of cap structure and epigenetic signals reminiscent of 5'ends of actively transcribed genes. ENCODE CAGE-seq analysis (Fig. 3a) revealed that 59% of Class1 segments have a CAGE signal within 500 nucleotides around their hypothetical TSS comparable with a range of 58% and 80% of detected CAGE signals for lncRNAs and mRNAs, respectively, and in contrast to 10% overlap detected randomly. Within the same region, comparison of CAGE signals distribution for mRNAs, lncRNAs and class1 aslncRNAs (suppl. Fig. 5) showed that Class1 aslncRNA 5'end signals were comparable in location as that of mRNAs and lncRNAs. To further confirm that Class1 aslncRNAs are actively transcribed, we monitored several epigenetic marks (Fig. 3b). As for mRNA and lncRNA, Class1 genes showed sharp peaks of the active transcription marks: Histone H3 trimethylated Lysine 4 (H3K4me3), acetylated Lysine 27 (H3K27ac) and acetylated Lysine 9 (H3K9ac), - just upstream of the hypothetical TSS, albeit less pronounced probably reflecting their low transcription. These signals are nevertheless significant comparing to a random region with no enrichment (Fig. 3b). In contrast, the repressive histone H3 tri-methylated Lysine 27 (H3K27me3) showed a strong depletion around the aslncRNA TSS reinforcing an idea of an active transcription at these loci. These data validated our RNA-seq transcripts prediction and allowed us to determine the relative TSS position of each antisense transcript

with respect to its paired-sense mRNA gene. The distribution of distance between the aslncRNA 5' end and the mRNA TSS showed two populations of transcripts (Fig. 3c) centered around each extremity of the paired mRNA gene: 269 of them start within +/-2kb of the mRNA TSS (promoter-associated, Class1A) and 374 within +/-2kb of the Transcription Termination Site (TTS, Class1B); and the other 952 elsewhere (Class1C). This result indicates that aslncRNA transcription preferentially uses regions proficient for pervasive transcription, *i.e.* near active promoters generating divergent transcripts but also near terminators, both regions devoid of nucleosomes²⁸. We anticipate that regulatory *cis* or *trans* activities of aslncRNAs will strongly depend on their position within mRNA genes.

To further analyze aslncRNA-mediated gene *cis*-regulation, we determined whether the expression of Class1 aslncRNAs correlates with their paired-sense mRNAs. We calculated the Pearson correlation coefficient (r) of expression of each sense/antisense pair within the 14 ENCODE datasets. The majority of pairs did not show any significant correlation (1099 over 1595 aslncRNA) and for $r > 0.5$, 460 were positively (Fig 4a) and only 39 negatively correlated. Interestingly, all anti-correlated pairs fell into Class1B and Class1C non-promoter-associated aslncRNAs (see suppl. Fig6 and Fig.4d), suggesting a bi-directional promoter origin of Class1A aslncRNAs. In this respect, an interesting example is the HOTAIR/HOX11C pair of bi-directional transcripts (suppl. Fig.7) showing a strong correlation of expression. Gene Ontology (GO) analysis revealed enrichment in developmental and embryonic morphogenesis pathways of the positively *cis*-correlated mRNAs (Suppl. table 1), suggesting that this cluster of genes could use aslncRNA or bidirectional promoter for fine-tuning. Altogether, these results point to positive *cis*-regulatory roles of aslncRNAs on their paired mRNA genes (potentially 30% of the cases, 460/1595).

To identify aslncRNAs with potential *trans*-regulatory roles, we calculated Pearson's coefficients between all aslncRNAs and all mRNAs ($n=15540$) of the 14 ENCODE RNA-seq datasets as in Figure 4a and visualized gene clusters with positive or negative correlation on a heatmap (Fig. 4b). Positively correlated pairs were found on the same chromosome more often than expected by chance. Sorting of 308 mRNA/aslncRNA pairs of the same chromosome and $r \geq 0.9$ showed a 641 kb mean distance (Fig. 4c) with their nearest correlated mRNA (suppl. table 2). Interestingly, Bonferroni correction retained 29 aslncRNA/mRNA pairs as significantly correlated (Suppl. Fig. 8), among which 14 positioned on the same chromosome with a mean distance of 120 kb (Fig. 4c), as exemplified for *TAGLN* locus

(suppl. figure 9). Altogether, both correlation studies revealed highly positively correlated aslncRNA/mRNA pairs separated by a distance compatible with a distal *cis*-acting enhancer activity²⁹. In addition, we noted that the Bonferroni filter revealed 15 aslncRNAs associated with 52 mRNAs located on distinct chromosomes (suppl. Table 3), suggesting putative *trans*-acting lncRNAs. GO analysis of *trans*-correlated candidate genes showed a major pathway involved in neuronal sensory and olfactory development (Suppl. Table 4), indicating as for *cis*-regulated genes a strong specialization of a group of genes that could use aslncRNA for fine-tuning.

Since some of the novel aslncRNAs might be involved in cell differentiation, and due to their high cell-type specificity, we asked whether we could define a robust aslncRNA signature of cancer progression. Prostate cancer is the second type of cancer for mortality. Removal of the prostate is generally performed based on a series of diagnostic tests, among them the *PCA3* antisense transcript highly expressed in prostate cancer tissues³⁰, but a powerful prognostic signature remains to be determined. Here, we investigated the presence of aslncRNAs in total RNA samples performing 24 RNA-seq experiments of 8 normal and 16 tumor prostate tissues (Suppl table 5). Differential expression analysis revealed 157 aslncRNAs (Fig. 5a) significantly up-regulated in tumors encompassing 36 and 17 annotated aslncRNA in GENCODE and MiTranscriptome projects, respectively, and 109 novel transcripts with 32 Class1 aslncRNAs (Fig. 5b). Principal component analysis (PCA) showed that this novel signature discriminates prostate tumors from normal tissues and from breast tumor and normal tissues (suppl. Fig. 10). Interestingly, the 3 most up-regulated Class1 aslncRNA showed similar fold-changes as *PCA3* albeit with lower dispersion, revealing a novel robust restricted signature for prostate cancer (suppl. Fig. 11b and Reverse transcriptase assay in suppl. Fig 11c). Strikingly, a prediction analysis on a cohort of 89 patients (suppl table 6) from RNA extraction measured by Nanostring demonstrated that the “restricted signature” discriminates high-risk prostate tumors more efficiently than *PCA3* alone (AUC 0.93 and 0.86, respectively; Fig.5c), exemplified on the *FBXL7* locus (Fig. 5d). In contrast, these novel markers had similar or lower AUC than *PCA3* for low risk and intermediate cases (suppl. Fig. 12a and b, respectively), strongly reinforcing a novel marker of aggressive tumors for prognostic studies.

In summary, our computational screen for antisense RNA defined an unprecedented human landscape of 7812 aslncRNA including 6984 novel candidates. We categorized these transcripts into different classes according to well-accepted criteria applied to lncRNAs, *i.e.* a

minimum of 0.8 RPKM expressions, presence of a spliced junction and/or a spliced EST. We would like to point out that our method is particularly sensitive since it allows identification of novel aslncRNA that were undetectable by the very recent MiTranscriptome and other projects²⁰. In particular, we defined a new prostate cancer signature of 157 aslncRNAs including 109 novel transcripts but only 17 MiTranscriptome lncRNAs. Divergent experimental strategies (assembling processes) and biological origins of samples would explain such discrepancy and highlight the specificity of our methodology for detection of antisense transcripts. Altogether, we propose that both annotations are complementary, revealing all lncRNA features over the human genome that could be used as powerful diagnostic markers.

The second main result is the systematic analysis of a regulatory potential of aslncRNAs based on their genomic position and expression correlations. Among 40% of significantly correlated genes, promoter-associated aslncRNAs (Class1A, Fig 4d) could be considered as bi-directional transcripts with yet unknown function on their associated partner. The others, non-promoter associated aslncRNAs might play a negative role on the sense transcription as proposed in yeast but the majority would rather have a positive activity in *cis* or in *trans* behaving as distal enhancers or *trans*-acting RNAs. It would be of particular interest to address the functional relevance of these candidates in specific normal or pathological cell context.

Online Material and methods

Cell lines, tissues and cell cultures

HEK cells were characterized in²¹. MCF7 was obtained from the American Type Culture Collection (ATCC) and transfected with siRNA against random or the Xrn1 gene sequences for other purpose. Prostate tissues were provided by the Mondor Hospital (suppl. Table 3) and breast tissues by Institut Curie. The histological status of each tumor was determined by respective clinical pathology services.

Cells were maintained in a humidified 5% CO₂ atmosphere at 37°C in DMEM, high glucose media for HeLa S2 and MEMalpha without nucleosides for MCF7 and HEK cells, both supplemented with 10% fetal bovine serum, essential amino acids and sodium pyruvate.

Sub-cellular fractionation

Cytoplasmic and nuclear RNA fractions were isolated from 10^8 HeLa S2 cells using PARIS procedure (Ambion).

RNA extraction and reverse-transcription (RT) PCR quantification

Total RNA was extracted from 10^8 cells using Trizol according to manufacturer's procedure, quantified and quality controlled using a 2100 Bioanalyzer (Agilent). RNA samples with RNA Integrity Number (RIN) above 6 were depleted for ribosomal RNA and converted into cDNA library using a *TruSeq Stranded Total Library Preparation kit* (Illumina). cDNA libraries were normalized using an Illumina Duplex-specific Nuclease (DSN) protocol prior to a paired-end sequencing on HiSeq™ 2500 (Illumina, NGS Curie platform). At least 20x coverage per sample was considered as minimum of unique sequences for further data analysis.

Reverse Transcription was performed on 100 ng of total RNA using gene-specific, strand-specific primers and the SuperScript II (Invitrogen) reverse transcriptase. cDNA for each transcript was quantified relative to POLR2A by the real-time PCR with LightCycler® 480 SYBR Green I Master mix (Roche) on Roche LightCycler480. Error bars represent standard deviations of three independent biological experiments (cell samples) or three independent RT (tissue samples). Primers sequences are available upon request.

NanoString nCounter Gene Expression Assay³²

Total RNA isolated from 89 tumors and 9 normal prostate tissues (suppl. table 6) was used for direct digital detection of 13 different RNAs: 3 housekeeping genes (RPL11, GAPDH and POLR2A), 4 mRNAs deregulated in our or published RNAseq datasets (HPN, OR51E2, PDLIM5 and SLC14A1) and 3 lncRNAs previously reported as deregulated in different cancers (HOTAIR, ANRIL and PCGEM1). Each target gene of interest was detected using reporter and capture probes carrying 35- to 50-base target-specific sequences (available upon request) on Nanotring nCounter V2 (Curie Genomic platform).

Data resource

The raw pair-end strand specific RNA-seq data from the ENCODE project (14 cell lines both polyA+ and polyA-) were downloaded from Gene Expression Omnibus (GEO) or generated

by our laboratory from ribo-depleted total RNA samples of HEK, Hela S2 and MCF7 cell lines (6 conditions), breast (12 conditions) and prostate (24 conditions) normal and tumor tissues (Table 1).

Computational pipeline of RNA-seq data analysis

The human genome version hg19 and the GENCODE V14 annotation were used in this study. First, we performed a quality control of all sequencing data by FastQC Babraham Bioinformatics software. Reads were mapped using TopHat 2.0.4²³, allowing 3 mismatches and requesting uniquely mapped reads. For libraries generated in this study, we assembled mapped reads using the BedTools suite²⁴. For ENCODE datasets, only ENCODE-validated⁴ contigs with npIDR score ≤ 0.01 were used. Overlapping contigs from all libraries were merged and only contigs supported by at least 10 reads in either library were further assembled in segments if mapped in the same strand and separated by less than 100 nucleotides. We compared segments to the GENCODE annotation to extract those oriented antisense to annotated genes, irrespectively to the overlap size, that we named human aslncRNAs.

RPKM¹⁶ expression levels were calculated for GENCODE annotated aslncRNAs, mRNAs, lncRNAs and for our novel aslncRNAs. To classify aslncRNAs, we applied 3 filters: the expression level ≥ 0.8 RPKM in at least two replicates for cell lines or in one condition for tissues, the presence of at least one TopHat-identified exon-exon junction within a contig, and finally, the presence of at least one spliced EST from UCSC mapped contigs. This led to the following three classes of antisense transcripts: Class1 meet all three criteria (expression/junction/EST); Class2 meet expression and junction (Class2a) or expression and EST (Class2b) criteria; and Class3 meet only the expression criterion. Only transcripts with the PhyloCSF score below 100 were considered as non-coding and retained as aslncRNAs^{35 16}.

Characterization of antisense RNAs

The expression specificity index (SI) was calculated according to ¹⁷ and represented as a density plot. Positions of CAGE tags and epigenetic marks were retrieved from ENCODE ChIP-seq data available at UCSC and plotted using a heatmap.2 function in the gplots R package (Version 2.12.1) as tags density in a window of 1, 2 or 4 kb distance around annotated (mRNAs, lncRNAs) or putative TSS (Class1 aslncRNAs) of transcripts with at least 1 RPKM expression level.

Correlation analysis of expression of Class1 aslncRNAs and mRNAs

The Pearson correlation coefficient (r) was calculated for each Class 1 aslncRNA and all mRNAs of the 14 ENCODE datasets and presented as a histogram and a heatmap. AslncRNA/mRNA pairs with $r \geq 0.9$ were further subjected to Bonferroni corrections (Pearson product moment correlation coefficients from the R `cor.test` function) and those with P-values ≤ 0.05 were retained for further analysis.

Gene Ontology enrichment of mRNA genes from significantly correlated pairs was analyzed using DAVID³⁶.

Signature of prostate tumor tissues

Differential expression analysis between prostate normal and tumor samples was performed using the DESeq R package³⁷. Only aslncRNAs with an expression fold change >2 and adjusted P-value below 0.05 were retained as differentially expressed to constitute prostate tumor signature.

Receiver Operating Curve (ROC) analysis

ROC curves assessed the assignment of 89 patients as high-risk versus low or intermediate risk, based on PCA3 only or the sum of 3 novel aslncRNA or the sum of all 4 lncRNAs, using pROC R-package³⁸ and $\log_2(\text{expression})$ scores from Nanostring data.

RNA-seq data visualization

RNA-seq reads of a locus of interest were represented as a strand-specific heatmap using our in-house R script VING (Descrimes et al., submitted).

Figures Legends

Figure 1: Establishment of Human antisense lncRNA catalog.

(a) Computational pipeline for identification and classification of aslncRNAs: RNA-seq data were mapped to the human reference genome hg19, assembled in contigs and merged with the ENCODE validated contigs using BEDtools. Neighbouring by no more than 100 nts contigs on the same strand were joined together in segments. Only segments antisense to the GENCODE annotation and supported by at least 10 reads were retained and tested for a coding potential by PhyloCSF. Three classes of transcripts were further defined according to their expression level $RPKM \geq 0.8$

(Class3), presence of at least one exon-exon junction (Class2a) or spliced EST (Class2b) and the most robust set meeting all three criteria (RPKM, junction and EST; Class1)

(b) Total number of aslncRNAs annotated in GENCODE and identified by our pipeline including those overlapping for at least 50% with MiTranscriptome transcripts (in brackets).

Table 1: Cell lines and tissue samples used to construct antisense catalog.

Figure 2: AslncRNA are lowly expressed nuclear polyA+ transcripts with strong specificity of expression

(a) Global expression level of novel Class1 aslncRNAs (blue) compared to mRNAs (black) and lncRNAs (red) as a mean of RPKM among 14 ENCODE RNA-seq datasets.

(b) Density plot of the expression specificity index (SI) for Class1 aslncRNAs (blue), mRNAs (black) and lncRNAs (red).

(c) 2D plot of Class1 aslncRNAs (blue), lncRNAs (red) and mRNAs (black) representing polyA+ versus polyA- HeLa S3 RNA-seq datasets.

(d) Kernel density plot of expression for mRNAs (black), snoRNAs (yellow) and Class1 aslncRNAs (blue) using ENCODE HeLa S3 datasets from nuclear (nuc) and cytoplasmic (cyt) total RNA.

Figure 3: AslncRNA are actively transcribed and initiate mainly at 5' and 3' extremities of the paired-sense mRNA genes

(a) Class1 aslncRNAs harbor capped transcripts next to their putative TSS: CAGE signals were retrieved from the ENCODE database and counted within 500 bases window around the

TSS defined by the GENCODE V14 annotation (mRNA in black, lncRNA in red) and novel Class1 aslncRNA (blue). Random sequences were used as a control (violet).

(b) Distribution of the histone H3 modification marks around TSS for expressed RNAs in HeLa S3 ENCODE ChIP-seq datasets.

(c) Density plot of relative distances for the putative TSS of Class1 aslncRNAs compared to the paired-sense metagene TSS and transcription termination site (TTS).

Figure 4: Expression of aslncRNAs is positively correlated with *cis* and *trans* mRNA genes

(a) Histogram of positive (red) and negative (green) correlation values distribution between sense/antisense pairs.

(b) Heatmap of Pearson coefficients between Class1 aslncRNAs and paired-sense mRNAs expressed in ENCODE datasets (n=1380 and 16287, respectively) with positive (green, +1) and negative (red, -1) correlation.

(c) Box plot of distances between the middle of aslncRNAs and the TSS of the nearest correlated mRNA on the same chromosome (mean is indicated as a black line) for correlated pairs with $r \geq 0.9$ (pink box, n=308) and Bonferonni adjusted pairs (blue box, n=14).

(d) Classification of Class1 aslncRNAs in positively (pos.) and negatively (neg.) *cis*- and *trans*-correlated pairs in respect to their putative TSS position: within the 2kb region of the paired sense mRNA TSS (Class1A), within the 2kb region of the TTS (Class1B) or elsewhere (Class1C).

Figure 5: AslncRNAs as a novel signature for high-risk prostate tumors

(a) Heatmap of expression (RPKM) of deregulated aslncRNAs in prostate tumors obtained by DESeq with at least 2 fold-change and adjusted P-value ≤ 0.05 (157 candidates over 24 tissues, listed in suppl. Table3).

(b) Summary of aslncRNAs prostate signature annotated by GENECODE and our pipeline, including those identified by MiTranscritome in all datasets (referenced as CAT) and as specific prostate-associated transcripts (PRCAT).

(c) ROC curve detecting high-risk prostate tumors according to clinical classification (Suppl Table 5) within a cohort of 89 patients and expression level of PCA3, sum of novel 3 aslncRNAs. (asFBXL7/asABCC4/asPDIML5) and all 4 transcripts measured by Nanostring. AUC values (Area Under the Curve) indicate the specificity of each set to detect high-risk tumors.

(d) RNA-seq snapshot of asFBXL7 in normal and tumor tissues using in-house VING visualization tool.

Supplementary information

Suppl Figure 1: Computational workflow for detection and classification of aslncRNAs

Suppl Figure 2: AslncRNAs are 10 kb long in average

Density plot of size distribution of aslncRNAs, annotated mRNAs and lncRNAs.

Suppl Figure 3: AslncRNA are highly cell-type specific

Expression (RPKM) of selected RPL11 mRNA, MALAT1 lncRNA and asPZP aslncRNA (asPZP) with low or high specificity index.

Suppl Figure 4: AslncRNA are majorly nuclear transcripts in HeLa S2

(a) Fractionation protocol applied for nuclear (nuc) and cytoplasmic (cyt) fractions isolation prior to RNA extraction and sequencing.

(b) Kernel density plot of expression of mRNAs (black), snoRNAs (yellow) and Class1 aslncRNAs (blue) in HeLa S2 fractionated RNA.

(c) RT-PCR quantification of several aslncRNAs from fractionated HeLa S2 RNA extracts. GAPDH mRNA, MALT1 lncRNA and U1 snoRNA were used as controls.

Suppl Figure 5: Class1 aslncRNAs harbor CAGE signals at proximity to their putative TSS

Mean of signal distribution of CAGE tags around 5'extrimity of aslncRNAs, lncRNAs, mRNAs and random loci in 8 ENCODE datasets.

Suppl Figure 6: Only non-promoter associated aslncRNA show negative correlation with their paired sense mRNA

Cis-correlation between Class1 aslncRNA/mRNA pairs depending on their position within the mRNA genes: Class1A starting within the 2kb of mRNA TSS (a), class1B within the 2kb of the TTS (b), or Class1C initiating elsewhere (c). Note that only pairs with $r > 0.5$ were counted (positive correlation in green and negative in red) with corresponded numbers.

Suppl Figure 7: Visualization of the bidirectional promoter of a HOTAIR-HOXC11 pair
VING representation of HOTAIR-HOXC11 bidirectional expression in 2 ENCODE cell lines that expressed the lncRNA (HSMM and MCF7) and 2 cell lines that do not (GM12878 and H1-hESC).

Suppl Figure 8: The most significant candidates of putative *trans*-acting aslncRNAs
Heatmap of Bonferroni adjusted correlation coefficients of Class1 aslncRNA and mRNAs in 14 ENCODE datasets: P-values of Fig. 4c were corrected using Bonferroni method and only candidates with P-value ≤ 0.05 were represented.

Suppl Figure 9: Representation of a putative distal aslncRNA/mRNA enhancer
VING representation of positively correlated aslncRNA/mRNA module (RP11-109L13.1) and distal TAGLN gene on chromosome 11 separated by 62kb distance. 3 cell lines transcribe both loci (AG04550, BJ, HSMM) and 3 cell lines none (GM12878, HeLa S3, MCF7).

Suppl figure 10: The restricted signature clusters prostate cancer tissues
PCA clustering using 157 deregulated aslncRNAs signature in prostate and breast tumor and normal tissues assessed by RNA-seq (suppl. Table3).

Suppl Figure 11: Expression levels of the restricted prostate cancer signature
(a) Box plot of PCA3 and the restricted prostate cancer signature (asFBXL7, asABCC4 and asPDIML5) in healthy (N) and tumor (T) tissues measured in 24 samples by RNA-seq. c; Quantitative RT-PCR validation of RNA-seq data for the restricted signature in 6 prostate tissues relative to POLR2A.

Suppl Figure 12: ROC analysis of low and intermediate risks prostate tumors
Same as in Figure 5e but with prediction values for low (a) and intermediate (b) risks patients.

Suppl Table 1: Gene Ontology Terms of paired aslncRNAs/mRNA showing positive and negative correlations.

Suppl Table 2: List of mRNA/aslncRNA pairs with $r \geq 0.8$ positioned on the same chromosome.

Suppl Table 3: List of mRNA/aslncRNA pairs significantly correlated after Bonferroni adjustment.

Suppl Table 4: Gene Ontology Terms of mRNA showing positive or negative *trans*-correlations.

Suppl Table 5: Repertoire of prostate and breast normal and tumor tissue samples used for RNA-seq experiments. Paired patients start with the same name.

Acknowledgments

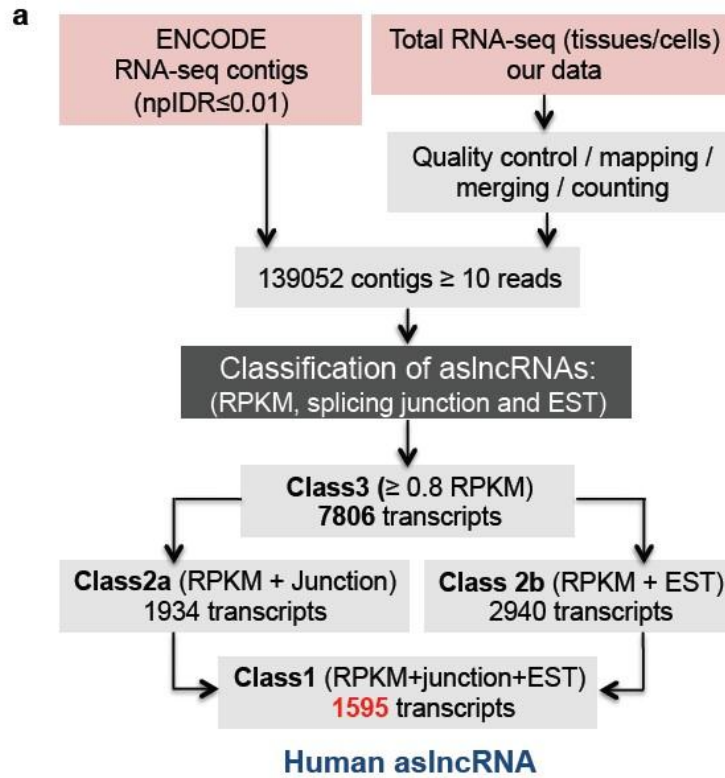
We thank T. Rio Frio and P. Legoix-Né from the Institut Curie (IC) NGS platform, S. Lair and A. Lhermine from the IC bioinformatics platform, D. Gentien and B. Albaud from the IC Genomic platform and the eBio bioinformatics platform (Orsay) for support. We are grateful to all members of our labs for helpful discussions. Fannie Semprez (Mondor) for the prostate tissues RNA preparation. High-throughput sequencing was performed by the NGS platform of IC, supported by the grants ANR-10-EQPX-03 and ANR10-INBS-09-08 from the Agence Nationale de la Recherche (investissements d'avenir) and by the Canceropôle Ile-de-France. Nanostring assays were performed on the genomic platform supported by .A. Morillon's lab is supported by the INSERM and ERC 'EPIncRNA' starting and ERC "DARK" consolidator grants. D. Gautheret lab is supported by INSERM.

References

- 1 Rinn, J. L. & Chang, H. Y. Genome regulation by long noncoding RNAs. *Annu Rev Biochem* **81**, 145-166, doi:10.1146/annurev-biochem-051410-092902 (2012).
- 2 Tisseur, M., Kwapisz, M. & Morillon, A. Pervasive transcription - Lessons from yeast. *Biochimie* **93**, 1889-1896, doi:10.1016/j.biochi.2011.07.001 (2011).
- 3 Pelechano, V. & Steinmetz, L. M. Gene regulation by antisense transcription. *Nat Rev Genet* **14**, 880-893, doi:10.1038/nrg3594 (2013).
- 4 Carrieri, C. *et al.* Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. *Nature* **491**, 454-457, doi:10.1038/nature11508 (2012).
- 5 Volders, P. J. *et al.* LNCipedia: a database for annotated human lncRNA transcript sequences and structures. *Nucleic Acids Res* **41**, D246-251, doi:10.1093/nar/gks915 (2013).
- 6 Volders, P. J. *et al.* An update on LNCipedia: a database for annotated human lncRNA sequences. *Nucleic Acids Res* **43**, D174-180, doi:10.1093/nar/gku1060 (2015).
- 7 Werner, A., Carlile, M. & Swan, D. What do natural antisense transcripts regulate? *RNA biology* **6**, 43-48 (2009).
- 8 Werner, A. Natural antisense transcripts. *RNA biology* **2**, 53-62 (2005).
- 9 Werner, A. Biological functions of natural antisense transcripts. *BMC biology* **11**, 31, doi:10.1186/1741-7007-11-31 (2013).
- 10 Werner, A. & Berdal, A. Natural antisense transcripts: sound or silence? *Physiological genomics* **23**, 125-131, doi:10.1152/physiolgenomics.00124.2005 (2005).
- 11 Werner, A., Preston-Fayers, K., Dehmelt, L. & Nalbant, P. Regulation of the NPT gene by a naturally occurring antisense transcript. *Cell biochemistry and biophysics* **36**, 241-252, doi:10.1385/CBB:36:2-3:241 (2002).
- 12 Preker, P. *et al.* RNA exosome depletion reveals transcription upstream of active human promoters. *Science* **322**, 1851-1854 (2008).
- 13 Lepoivre, C. *et al.* Divergent transcription is associated with promoters of transcriptional regulators. *BMC genomics* **14**, 914, doi:10.1186/1471-2164-14-914 (2013).
- 14 Spicuglia, S. & Vanhille, L. Chromatin signatures of active enhancers. *Nucleus* **3**, 126-131, doi:10.4161/nucl.19232 (2012).
- 15 Djebali, S. *et al.* Landscape of transcription in human cells. *Nature* **489**, 101-108, doi:10.1038/nature11233 (2012).

- 16 Cabili, M. N. *et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev* **25**, 1915-1927, doi:10.1101/gad.17446611 (2011).
- 17 Orom, U. A. *et al.* Long noncoding RNAs with enhancer-like function in human cells. *Cell* **143**, 46-58 (2010).
- 18 Khalil, A. M. *et al.* 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, doi:10.1073/pnas.0904715106 (2009).
- 19 Gibb, E. A., Brown, C. J. & Lam, W. L. The functional role of long non-coding RNA in human carcinomas. *Molecular cancer* **10**, 38, doi:10.1186/1476-4598-10-38 (2011).
- 20 Iyer, M. K. *et al.* The landscape of long noncoding RNAs in the human transcriptome. *Nat Genet* **47**, 199-208, doi:10.1038/ng.3192 (2015).
- 21 Derrien, T. *et al.* The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Res* **22**, 1775-1789, doi:10.1101/gr.132159.111 (2012).
- 22 Van Dijk, E. L. *et al.* XUTs are a class of Xrn1-sensitive antisense regulatory non coding RNA in yeast. *Nature* **475**, 114-117 (2011).
- 23 Latos, P. A. *et al.* Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. *Science* **338**, 1469-1472, doi:10.1126/science.1228110 (2012).
- 24 van Werven, F. J. *et al.* Transcription of two long noncoding RNAs mediates mating-type control of gametogenesis in budding yeast. *Cell* **150**, 1170-1181, doi:10.1016/j.cell.2012.06.049 (2012).
- 25 Portal, M. M., Pavet, V., Erb, C. & Gronemeyer, H. Human cells contain natural double-stranded RNAs with potential regulatory functions. *Nat Struct Mol Biol* **22**, 89-97, doi:10.1038/nsmb.2934 (2015).
- 26 Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature methods* **5**, 621-628, doi:10.1038/nmeth.1226 (2008).
- 27 Yanai, I. *et al.* Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics* **21**, 650-659, doi:10.1093/bioinformatics/bti042 (2005).
- 28 Schones, D. E. *et al.* Dynamic regulation of nucleosome positioning in the human genome. *Cell* **132**, 887-898, doi:10.1016/j.cell.2008.02.022 (2008).

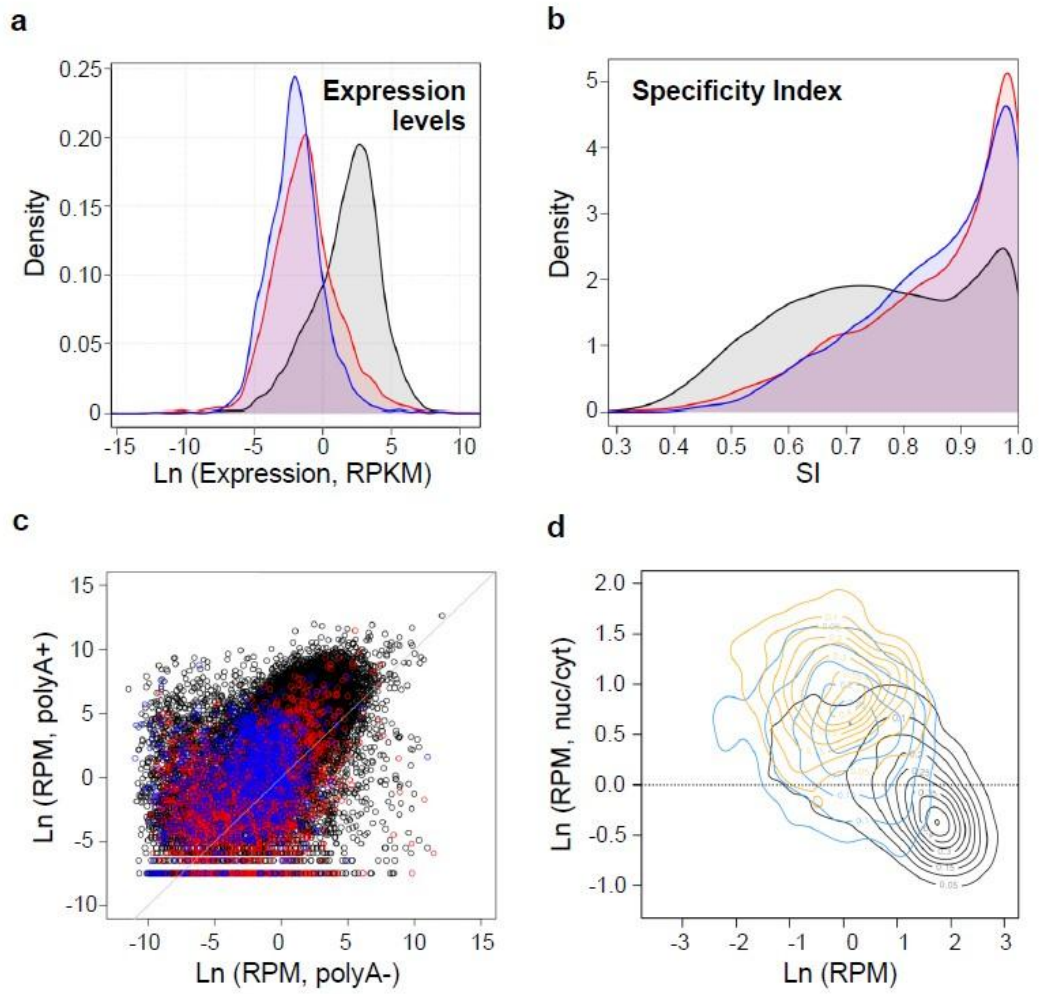
- 29 Sanyal, A., Lajoie, B. R., Jain, G. & Dekker, J. The long-range interaction landscape of gene promoters. *Nature* **489**, 109-113, doi:10.1038/nature11279 (2012).
- 30 de Kok, J. B. *et al.* DD3(PCA3), a very sensitive and specific marker to detect prostate tumors. *Cancer Res* **62**, 2695-2698 (2002).
- 31 Castro-Vega, L. J. *et al.* Telomere crisis in kidney epithelial cells promotes the acquisition of a microRNA signature retrieved in aggressive renal cell carcinomas. *Carcinogenesis*, doi:10.1093/carcin/bgt029 (2013).
- 32 Malkov, V. A. *et al.* Multiplexed measurements of gene signatures in different analytes using the Nanostring nCounter Assay System. *BMC research notes* **2**, 80, doi:10.1186/1756-0500-2-80 (2009).
- 33 Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-1111, doi:10.1093/bioinformatics/btp120 (2009).
- 34 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).
- 35 Lin, M. F., Jungreis, I. & Kellis, M. PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics* **27**, i275-282, doi:10.1093/bioinformatics/btr209 (2011).
- 36 Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* **4**, 44-57, doi:10.1038/nprot.2008.211 (2009).
- 37 Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol* **11**, R106, doi:10.1186/gb-2010-11-10-r106 (2010).
- 38 Robin, X. *et al.* pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC bioinformatics* **12**, 77, doi:10.1186/1471-2105-12-77 (2011).
- 39 Djebali, S. *et al.* Evidence for transcript networks composed of chimeric RNAs in human cells. *PLoS One* **7**, e28213, doi:10.1371/journal.pone.0028213 (2012).



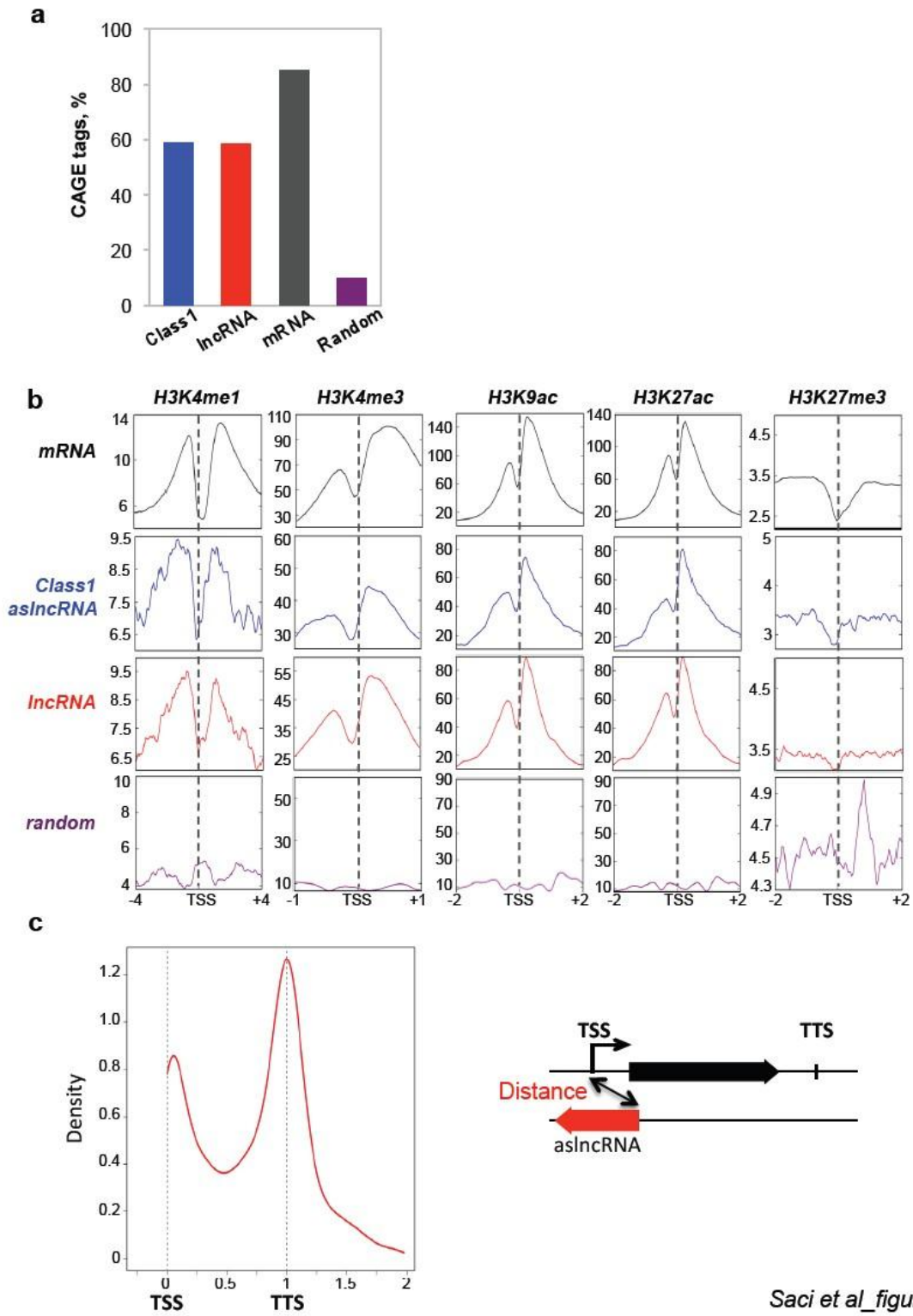
b

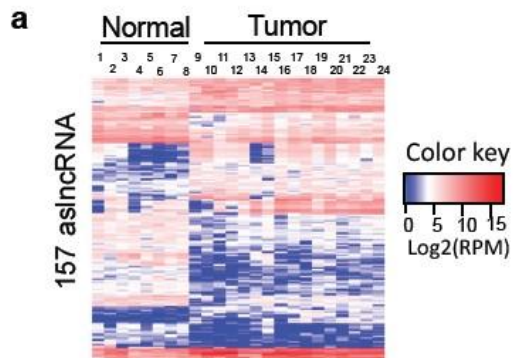
aslncRNA	Class3	Class2a	Class2b	Class1
Novel aslncRNA (MiTranscriptome)	6988 (481)	1216 (353)	2230 (364)	963 (287)
Annotated aslncRNA	828	718	710	362
Total	7806	1934	2940	1595

Saci et al_figure 1



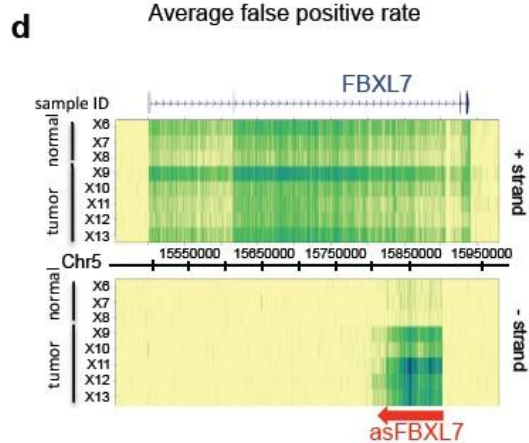
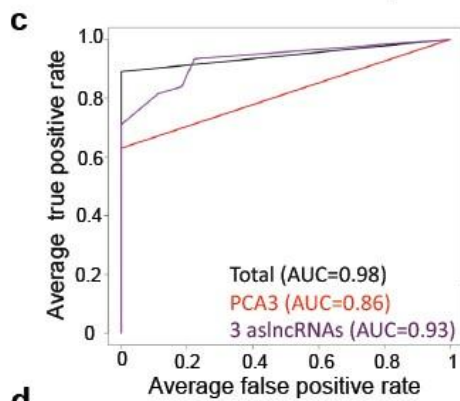
Saci et al_figure 2





b

aslncRNA (n=157)	Class3	Class2a	Class2b	Class1
Novel aslncRNA	12	40	53	32
MiTranscriptome*	17	-	-	14
PRCAT**	2	-	-	2
GENCODE aslncRNAs	36	36	32	32
MiTranscriptome	26	-	-	24
PRCAT**	2	-	-	1
Total	91	76	85	64
PRCAT**	4	-	-	3

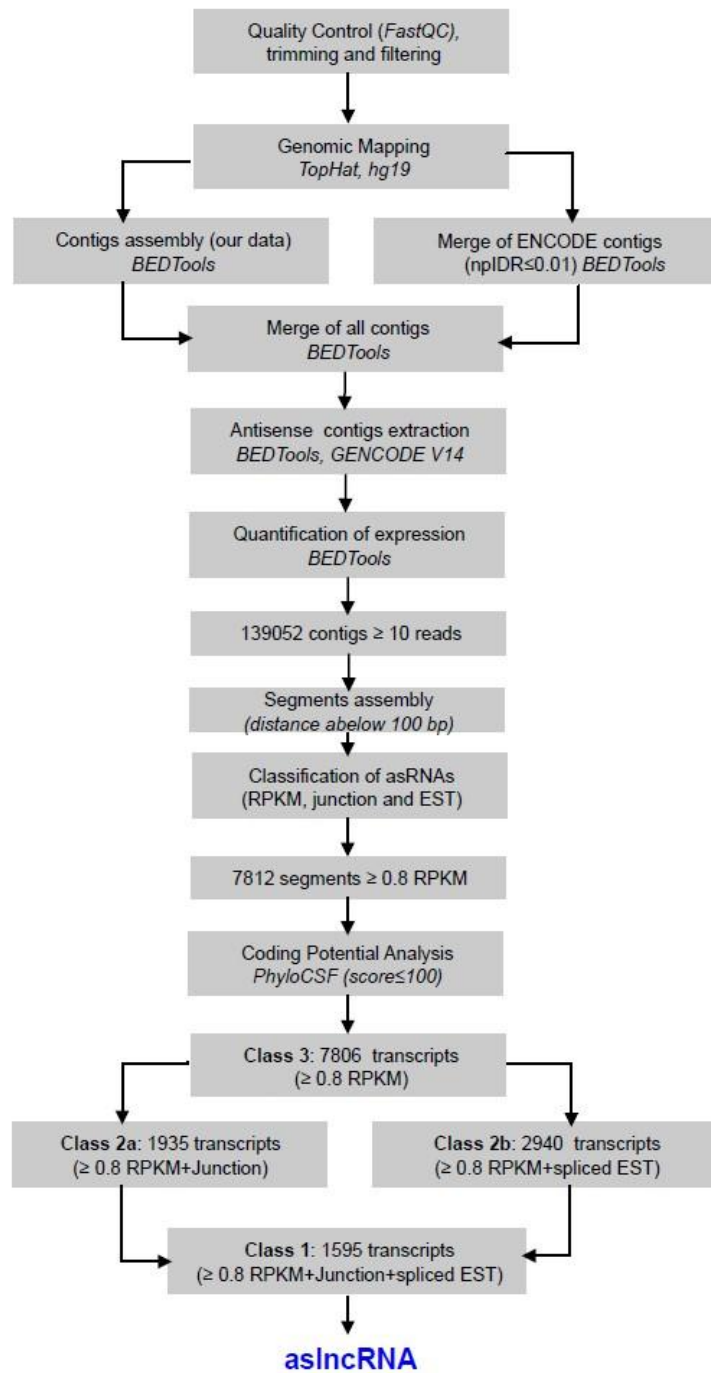


Saci et al_figure 5

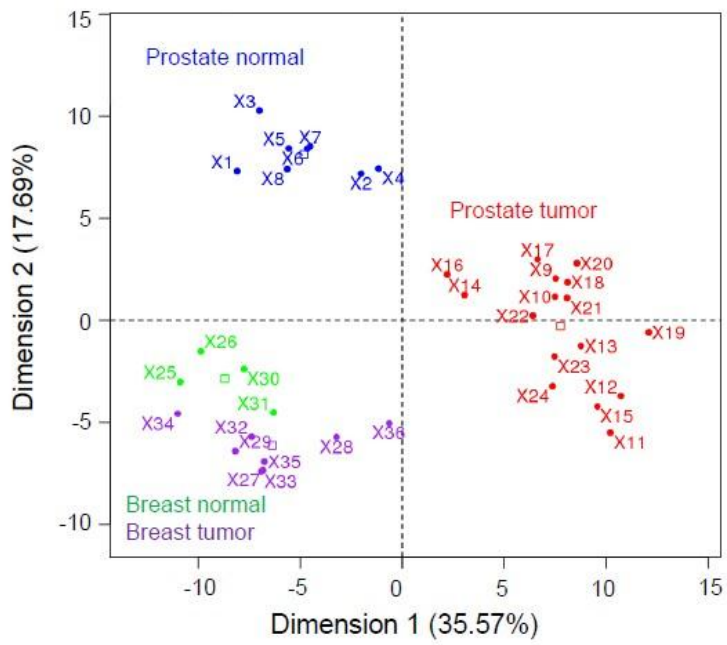
GO term of genes within sense/antisense mRNA/aslncRNA pairs showing positive *cis*-correlation

GO term	Description	Count	P-value	Genes	Fold Enrichment	Bonferroni correction
GO:0043565	sequence-specific DNA binding	33	7,04E-9	ISX, HNF1B, HNF1A, ELF3, HOXA11, WT1, HOXA1, FOXQ1, HOXA3, FLI1, HAND2, POU2F2, HOXA9, MKX, NR2F2, PITX1, FOXD3, ERG, SMG6, TBX2, TBX5, FOXA1, TBX1, ISL1, PROX1, HOXC10, ZFXH4, HOXC11, HOXC13, DLX6, HOXD3, SIX1, LBX2	3,24	3,08E-6
GO:0048736	appendage development	14	7,80E-9	HNF1A, HOXA11, TBX5, GREM1, HOXC10, HOXC11, DLX6, RAB23, HOXA9, NR2F2, PITX1, FLVCR1, TWIST1, BMPR1A	8,59	1,48E-5
GO:0060173	limb development	14	7,80E-9	HNF1A, HOXA11, TBX5, GREM1, HOXC10, HOXC11, DLX6, RAB23, HOXA9, NR2F2, PITX1, FLVCR1, TWIST1, BMPR1A	8,59	1,48E-5
GO:0007389	pattern specification process	21	7,95E-9	HNF1B, NRP1, TBX5, HOXA11, EDN1, TGFB2, CELSR2, TBX1, GREM1, HOXC10, HOXA1, HOXA3, HOXC11, HOXC13, HOXD3, SIX1, RAB23, PKD2, HOXA9, NR2F2, BMPR1A	4,97	1,50E-5
GO:0048598	embryonic morphogenesis	22	1,64E-8	HNF1A, TBX5, HOXA11, EDN1, TGFB2, TBX1, GREM1, PROX1, HOXC10, HOXA1, HOXC11, HOXA3, DLX6, HOXD3, SIX1, RAB23, HOXA9, TFAP2A, TXNRD1, FLVCR1, TWIST1, BMPR1A	4,53	3,11E-5
GO:0035108	limb morphogenesis	13	4,73E-8	HNF1A, HOXA11, TBX5, GREM1, HOXC10, HOXC11, DLX6, RAB23, HOXA9, PITX1, TWIST1, BMPR1A, FLVCR1	8,30	8,95E-5
GO:0035107	appendage morphogenesis	13	4,73E-8	HNF1A, HOXA11, TBX5, GREM1, HOXC10, HOXC11, DLX6, RAB23, HOXA9, PITX1, TWIST1, BMPR1A, FLVCR1	8,30	8,95E-5
GO:0003002	regionalization	16	5,13E-7	HNF1B, HOXA11, EDN1, CELSR2, TBX1, GREM1, HOXC10, HOXA1, HOXA3, HOXC11, HOXC13, HOXD3, RAB23, HOXA9, NR2F2, BMPR1A	5,13	9,71E-4

Saci et al_sup Table 1



Saci et al_sup figure 1



Saci et al_sup figure 10

Annexe 2

Review Article

Long Noncoding RNAs as New Architects in Cancer Epigenetics, Prognostic Biomarkers, and Potential Therapeutic Targets

Didier Meseure,¹ Kinan Drak Alsibai,¹ Andre Nicolas,¹
Ivan Bieche,² and Antonin Morillon³

¹Platform of Investigative Pathology, Institute Curie, 75248 Paris, France

²Department of Pharmacogenomics, Institute Curie and UMR 745 INSERM, University Descartes, 75248 Paris, France

³UMR 3244, Non-Coding RNA, Epigenetics and Genome Fluidity, Institute Curie, 75248 Paris, France

Correspondence should be addressed to Didier Meseure; didier.meseure@curie.fr

Received 25 October 2014; Accepted 6 May 2015

Academic Editor: Konstantinos Arnaoutakis

Copyright © 2015 Didier Meseure et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Recent advances in genome-wide analysis have revealed that 66% of the genome is actively transcribed into noncoding RNAs (ncRNAs) while less than 2% of the sequences encode proteins. Among ncRNAs, high-resolution microarray and massively parallel sequencing technologies have identified long ncRNAs (>200 nucleotides) that lack coding protein function. LncRNAs abundance, nuclear location, and diversity allow them to create in association with protein interactome, a complex regulatory network orchestrating cellular phenotypic plasticity via modulation of all levels of protein-coding gene expression. Whereas lncRNAs biological functions and mechanisms of action are still not fully understood, accumulating data suggest that lncRNAs deregulation is pivotal in cancer initiation and progression and metastatic spread through various mechanisms, including epigenetic effectors, alternative splicing, and microRNA-like molecules. Mounting data suggest that several lncRNAs expression profiles in malignant tumors are associated with prognosis and they can be detected in biological fluids. In this review, we will briefly discuss characteristics and functions of lncRNAs, their role in carcinogenesis, and their potential usefulness as diagnosis and prognosis biomarkers and novel therapeutic targets.

1. Long Noncoding RNAs and Functional Organization of the Genome

1.1. Genome and Noncoding RNA. Large-scale genomic technologies (high-resolution microarray, whole genome, and RNA sequencing) combined with bioinformatics analyses have profoundly changed the genome organization understanding. Unexpectedly, these global transcriptional analyses revealed that 66% of the genome is transcribed and 80% presents biochemical marks of active transcription whereas less than 2% encodes proteins [1]. Discovery of a “second genetic code” consisting of ncRNAs has changed traditional concept of genomic organization characterized by presence of genes encoding islets scattered in a sea of repeats and non-transcribed proteins. Recently, high-throughput techniques have identified 58648 human lncRNAs [2].

1.2. Classification of Noncoding RNAs. ncRNAs are classified into two categories according to their structural or regulatory properties and size. This length, arbitrarily set to 200 nucleotides (nt), corresponds to the threshold of sensitivity of RNA extraction methods and can differentiate lncRNAs from short and medium ncRNAs such as microRNAs (miRNAs), small nucleolar RNAs (snoRNAs), and piwi RNAs (piRNAs) [3] (Table 1). Although functions of most ncRNAs are currently still largely uncharacterized, recent studies have involved them in mechanisms implicated in important biological functions and in various pathologic conditions, including neurodegenerative diseases and cancer [4, 5]. It is likely that development of new diagnostic and prognostic classifications involving ncRNAs refine medical practice and that ncRNAs may be useful potential targets for novel anticancer therapies. In this review, we propose to

Table 1: Type and main function of no-coding RNAs.

Type	Size (nt)	Functions
Short ncRNAs	19–31	
miRNAs		Targeting of mRNAs, regulation of proliferation, differentiation, and apoptosis involved in human development
siRNAs		Posttranscriptional gene silencing; defense against pathogenic nucleic acids
tiRNAs		Regulation of transcription by targeting epigenetic silencing complexes
piRNAs		Transposon repression, DNA methylation, development of germ cell, stem self-renewal, and retrotransposon silencing
tel-sRNAs		Epigenetic regulation
Mid-size ncRNAs	≤200	
snoRNAs		rRNA modifications
PASRs		Regulation of the transcription of protein-coding genes
TSSa-RNAs		Maintenance of transcription
PROMPTs		Activation of transcription
crasiRNAs		Recruitment of heterochromatin and/or centromeric proteins
Long ncRNAs	>200	
lincRNAs		Involvement in biological processes such as dosage compensation and/or imprinting
Intronic lncRNAs		Possible link with posttranscriptional gene silencing
T-UCRs		Regulation of miRNA and mRNA levels and antisense inhibitors for protein-coding genes or other ncRNAs
TERRAs		Negative regulation of telomere length and activity through inhibition of telomerase
Pseudogene RNAs		Regulation of tumor suppressors and oncogenes by acting as microRNA decoys
lncRNAs with dual functions		Modulate gene expression through diverse mechanisms

summarize recent literature data on lncRNAs, their involvement in carcinogenesis, and their value as biomarkers and potential therapeutic targets.

1.3. Definition of lncRNAs. LncRNAs transcripts are defined by a length greater than 200 nts without any potential of translation [6]. This definition is however arbitrary and cannot now be based on a set of physical biochemical, structural or functional criteria. Length is not an absolute criterion because ncRNAs smaller than 200 nts (*BCI*, *snaR*) are included in the lncRNAs subgroup. There is no specific transcription of lncRNAs, the transcriptional machinery being common to lncRNAs and mRNAs. Lack of reading frame cannot also be a discriminating factor between mRNAs and ncRNAs because 50% of lincRNAs contain regions with high translational potential, comparable to those of mRNAs, suggesting that they are actively exported to cytoplasm and translated into short polypeptides, albeit certainly not active. Noncoding characteristic is then not specific to lncRNAs, as bifunctional coding and noncoding lncRNAs and lncRNAs containing other ncRNAs (miRNAs, snoRNAs) have recently been discovered. Several mRNAs lose their ability to encode proteins (*Xist*) while others acquire coding function (*SRA*) [7]. Furthermore, by incorporating multiple exons from coding and noncoding genes, alternative splicing generates ambiguous transcripts beyond all classifications. Nevertheless, this definition appears currently to be the best that could be used waiting for a functional characterization of lncRNAs.

1.4. Investigative Methods and Annotation of lncRNAs. The first lncRNAs were discovered by using functional genetic approaches based on their relation to specific cellular mechanisms: (i) *Ubxn* transcriptional interference-induced repression by trithorax-promoted lncRNA*bdx*, (ii) sex chromosome dosage compensation and *Xist*-induced X chromosome, and (iii) inactivation and genomic imprinting promoted by antisense lncRNAs *Airn* and *Kcnql*. These seminal studies established the concept of functional lncRNAs acting as transcriptional regulators involved in important and diverse biological processes. Over the past two decades, systematic global sequencing of cDNA libraries yielded comprehensive knowledge of the transcribed but noncoding component of the genome. These transcriptomic analyses have identified numerous lncRNAs. RNA immunoprecipitation and siRNA-induced inactivation techniques led to functional validation of many lncRNAs.

Currently, lncRNAs identification and functional characterization are based on an experimental approach combining (i) discovery of new noncoding transcripts using RNA-seq and CHIP-seq, (ii) annotation of transcribed regions extent and quantity of products transcribed through bioinformatic analysis, (iii) lncRNAs quantification in various cellular and tissue types and conditions, (iv) coexpression or coregulation studies, and (v) gain and loss functional tests on cell lines and xenografts (siRNAs, ASO, TALEN, and CRISPR/CAS9). Advantage of RNA-seq technique resides in its sensibility, allowing most weakly expressed lncRNAs to be identified.

However, due to their frequent very low expression and complex exon/intron structures, it can be difficult to identify different transcripts produced from lincRNA gene. It is then necessary to use other techniques through epigenetic analyses (markers of promoter regions or entire gene), targeted transcripts of interest sequencing after capture on a DNA chip (capture-RNAseq) and 5 and 3 extremities identification analyses [8–11].

Systematic large-scale projects have contributed a lot to the systematic characterization of mammalian lincRNAs. For instance, the Fantom (Functional Annotation of the Mammalian genome) is an international project initiated by Japan in 2000 designed to identify and annotate entire transcripts of mouse. Results published identified 15000 lincRNAs in 2002, 23000 in 2005, and 34000 in 2006 with a transcribed genome fraction of 63%. The ENCODE project initiated in 2003 is designed to identify, map, and make public all functional elements of the human genome. This project led to a catalog of lincRNAs with 10000 genes producing 15000 lincRNAs. Very recently, 58648 lincRNAs were identified by using the TCGA database, of which 79% were previously unannotated [2].

1.5. Properties of lincRNAs. lincRNAs genes outnumber those of short ncRNAs and are probably more abundant than mRNA genes. lincRNAs are very heterogeneous in size with some extending over tens of kilobases (kb). They are transcribed in all regions of eukaryotic genome, particularly during development, and are remarkably specific for a given cellular and tissue type.

lincRNAs genes share many biochemical characteristics with proteins coding genes: predominant action of RNA polymerase II, epigenetic profiles, cotranscriptional modifications (5 Cap, pre-lincRNAs alternative splicing, and 3 polyadenylation), and exons and introns size. However, when lincRNAs are compared to mRNAs, they appear less stable and shorter, with fewer exons, a less well defined reading frame, and many repeat sequences (LINE, SINE). A minority group of poorly characterized lincRNAs is represented by nonpolyadenylated lincRNAs synthesized by RNA polymerase III and lincRNAs synthesized via alternative splicing and snoRNAs.

Depending on their relative position to the nearest coding genes, several categories of lincRNAs can be identified as follows.

(i) Intragenic Regions. Intragenic lincRNAs may be subdivided depending on how they overlap protein-coding genes or their orientation [45].

- (a) Long intronic ncRNAs (linRNAs) constitute the major component of ncRNAs transcriptome. Unlike lincRNAs, tens of thousands linRNAs have recently been identified but only few have been analyzed functionally [46].
- (b) Sense lincRNAs are transcribed from the sense strand of protein-coding genes containing exons. They may have some overlap with coding genes or cover entire sequence from an intron. If it has been shown that most have no protein-coding potential, several sense

lincRNAs can function as both ncRNAs and protein-coding genes (*SRA*, *ENOD40*).

- (c) Antisense lincRNAs (NAT) represent 32% of lincRNAs in humans and are poorly defined. They are transcribed from the antisense strand of protein-coding genes coding. They may also have overlapping with exonic or intronic regions or cover entire sequence from an intron. NATs are observed in many species, including animals, plants, yeasts, prokaryotes, and viruses, but have no sequence or conserved structure that could be indicative of a particular function despite several highly conserved lincRNAs [47].

(ii) Intergenic Regions. Intergenic lincRNAs (*lincRNAs*) are located in unannotated genomic regions and represent the best studied class of lincRNAs. lincRNAs are functional lincRNAs characterized by stability, K4H36 active domain of transcription, and tissue-specific expression. They act in trans and are primarily involved in epigenetic regulation of protein-coding genes expression and maintenance of stem cells pluripotency [48].

(iii) Enhancer ncRNAs (eRNAs). Two types of lincRNA (multi-exonic or lacking introns with bidirectional transcription) are derived from enhancer sequences. Enhancer sequences regulate temporal and specific expression of genes via cis or trans mechanisms. The human genome contains approximately one million enhancer sequences most likely controlling the 19000 genes coding for proteins. lincRNAs derived from enhancer sequences play a crucial role in the formation of chromatin loops that allow stabilization and association of enhancer sequences with the promoter regions that initiate transcription of their target genes [49].

(iv) Promoters. PALR (promoter associated long ncRNA).

(v) Telomeric Regions. TERRA (telomeric repeat containing RNAs).

Within cells, lincRNAs are located in the nucleus, according to their main protein-coding genes epigenetic regulatory function. A minority of lincRNAs are involved in nucleocytoplasmic trafficking (*NRON*), mRNA stability, and translational regulation. Several lincRNAs (*MALATI*) are cleaved by RNase P in a structural and regulatory intranuclear component (speckle) and a small cytoplasmic tRNA-like transcript. lincRNAs topography recognition should help in finding new efficient lincRNAs-based targeted therapies [50]. Finally, lincRNAs can be precursors of ncRNAs (snoRNA, miRNA).

1.6. Modular Organization of the Genome and Structural Plasticity of lincRNAs. The genome has a modular architecture composed of complex transcriptional loci characterized by close links between nucleotide sequences organized in sense/antisense and coding/noncoding transcripts. Thus,

over 50% of protein-coding genes are associated with complementary antisense noncoding transcripts cis-regulating chromatin and adjacent genes expression although no systematic study classified them carefully as lncRNAs. Moreover, combined application of alternative splicing and transcriptional initiation and termination utilizes this modular architecture to ensure transcriptional diversification illustrated by the very large number of lncRNAs isoforms [48]. lncRNAs structural plasticity and biochemical properties better explain the great diversity of their mechanisms of action than their mere nucleotide sequence. Their organization in secondary and tertiary structures contributes to the creation of functional domains interacting with proteins (small ligands, multiprotein complexes) or hybridizing with nucleic acids (mRNA, miRNA, and DNA). Their interactions lead in turn to lncRNAs allosteric conformational changes allowing them to bind to other actors involved in gene expression and mRNA translation. Thus, lncRNAs initiate regulatory networks with high complexity at epigenetic, transcriptional, and posttranscriptional levels, in order to transmit and coordinate information flows in signaling pathways required for eukaryotic cells functioning [51].

1.7. Conservation, Evolution, and Origin of lncRNAs. lncRNAs have been observed in many eukaryotes. Conversely to protein-coding sequences, lncRNAs have rapidly evolved, explaining that lncRNAs orthologues to mammals are only found in vertebrates. Even in vertebrates, DNA sequences conservation is low and suggests that most lncRNAs are not functional due to insufficient selection pressure. Nonetheless, others criteria could be taken into account in assessment of their conservation, including genomic localization, transcriptional profile, and tertiary structure.

Current complexity of human physiology cannot be solely explained by expression of 20000 protein-coding genes, comparable to that observed in *Drosophila melanogaster* and *Caenorhabditis elegans*, but rather by parallel development of a noncoding genome. Unlike proteome, amount of ncRNAs has increased during evolution and it seems to be a significant correlation between lncRNAs expression levels and complexity, explaining that primates have most lncRNAs. The number of protein-coding genes cannot explain functioning of very finely regulated organs such as brain. This additional level of complexity observed in vertebrates may be partly related to expression of lncRNAs with a high spatiotemporal specificity and their interactions with DNA, mRNAs, and proteins. lncRNAs could then be considered as molecular sensors of environmental changes conferring evolutionary plasticity that contributes to development of life complexity. During environmental changes, lncRNAs located within intergenic sequences may serve as supports for new functions allowing body to adapt new constraints. Protein-coding genes could generate transcripts from lncRNAs genes. Conversely, proteins may be synthesized from lncRNAs. Recent data have identified 24 human protein-coding genes with noncoding homologous genes in other species.

Recent emergence of numerous lncRNAs suggests that they continue to be actively synthesized (i) from ancestral genes that have lost their coding potential, (ii) by genes or

other lncRNAs duplication, transposition, and mutation and (iii) de novo from intergenic DNA [52, 53].

1.8. lncRNAs Functions and Mechanisms of Action. Growing data have demonstrated that lncRNAs exhibit the greatest diversity among functional ncRNAs and play regulatory and structural roles in embryogenesis, stem cells pluripotency, allelic expression, protein-coding genes regulation, apoptosis, cycle control, growth, differentiation, and senescence. In practice, only a very limited number of lncRNAs (1%) have been well characterized functionally in humans, including *Xist*, *KCNQ10T1*, *AIR*, *HotAir*, *ANRIL*, *HOTTIP*, *MALAT1*, *TERRA*, and *HULC*.

lncRNAs play a role in stem cells and differentiation: maintenance of pluripotency and lineage differentiation in embryonic stem cells (ESCs) and induced pluripotent stem cells (iPSCs). Adult stem cells are finely regulated by lncRNAs networks that are targets of most of master pluripotency transcription factors (Oct4, Sox2, Nanog, cMyc, and KLF4) [54].

lncRNAs are considered as crucial regulators coordinating protein-coding genes expression by numerous mechanisms depending on their cellular localization and leading to modifications of chromatin, transcription, and translation.

These regulatory mechanisms are located at epigenetic, transcriptional, posttranscriptional, and translational levels as follows.

(i) lncRNAs constitute a network of epigenetic modulators, recruiting, guiding, and forming platforms that build ribonucleoprotein complexes at specific genomic sites. 20% of human lncRNAs recruit multimolecular repressor, activator, or chromatin remodeling complexes (PRC1/PRC2 by *ANRIL*, LSD1/PRC2 by *HOTAIR*) whose subcatalytic units (EZH2, EED, BMI1, SUZ12, CBX7, CoREST, and JARID1) interact by altering histone code and methylation profile. Cis-acting lncRNAs adjacent to locus where they are transcribed act by transcriptional interference or by modifying chromatin. Transcriptional interference allows inhibition of preinitialization complexes and interaction with transcription factors. Chromatin modifications result from recruitment of complexes inhibiting (Polycomb by *ANRIL*) or activating (MLL by *HOTTIP*) gene expression. Trans-lncRNAs operate independently of complementarities sequences and act at distance on many genomic loci via specific DNA motifs. They influence genes expression by recruiting modifying chromatin complexes, binding transcriptional elongation factors, and inhibiting RNA polymerase [55].

(ii) lncRNAs associated with promoters (PaRNAs) and lncRNAs of enhancer type (eRNAs) can directly regulate transcription of target genes by transcriptional activation or suppression. PaRNAs function as protein coactivators facilitating gene transcription. PaRNAs of 50–200 nt are involved in repression of polycomb-targeted genes via mechanisms allowing transcription of these lncRNAs from common promoter regions to those of target genes regulated by these noncoding transcripts. Gene cis-repression is the result of PRC2 complex (SUZ12) recruitment by PaRNAs, attachment to promoter regions, and modification of histone methylation (H3K27me3) on target promoters. ERNAs have

transcriptional activation functions of protein-coding genes coding, including genes involved in embryonic development and differentiation (*ncRNAa3 and TAL*, *ncRNAa7 and SNAI1*) [56–58].

(iii) LncRNAs can also act at posttranscriptional level and are widely involved in biogenesis, stability, and transcriptional activity of mRNAs. They regulate alternative splicing, promote trafficking, direct cellular localization, and promote mRNAs degradation. LncRNAs also synthesize miRNAs and build sponge-like structures to prevent binding of miRNAs to their target mRNAs (*CDR1-as/ciRS-7*, *circular RNA sponge for miR-7*) [59].

(iv) LncRNAs can finally bind to inhibiting factors of translation, interact with ribosomes, and allow transport of proteins.

LncRNAs may participate in assembly of specialized intranuclear functional structures, including speckles (*MALAT1*) paraspeckles (*NEAT1*), and polycomb body (*TUG1*).

Several lncRNAs, particularly lincRNAs, interact with factors (Oct4, Sox2, Nanog, c-Myc, Klf4, Smad, and Tcf3) and play key roles in maintaining stem cells pluripotency.

Moreover, lncRNAs have four known molecular functions (Figure 1): (1) Signal: lncRNAs regulate transcriptional activity or pathways (lincRNA-p21); (2) Guide: lncRNAs link specific proteins belonging to chromatin remodeling complexes and recruit them to homology containing target loci where genes silencing is promoted (*HOTTIP*, *XIST*); (3) Decoy: lncRNAs bind and titrate away proteins or RNAs. In the nucleus, they can bind transcription factors or DNA methyltransferase 1 (*PANDA*, *Gas5*, *MALAT1*), whereas in the cytoplasm, they can function like a sponge to attract proteins and miRNA/RISC complexes from their miRNA targets; (4) Scaffold: lncRNAs constitute adaptors that bind molecular complexes and regulate gene expression (*HOTAIR*, *ANRIL*, and *TERC/TERT*) [18, 60].

1.9. lncRNAs and Cancer. Because lncRNAs are involved in various and important physiological processes, their dysfunction should have important consequences for cell homeostasis. Several recent studies have indeed shown that expression of many lncRNAs varies significantly in different conditions compared to healthy tissue. Significance of these deregulations (consequence of a globally altered transcriptional status or causative and driving abnormality) is still a matter of debate. However, transcripts of noncoding genome revealed new dimension of the molecular architecture of cancer and lncRNAs are involved in all stages of oncogenesis. Gene expression profiles analysis of various tumors showed that lncRNAs are deregulated and functional studies have demonstrated that lncRNAs are implicated in general mechanisms of carcinogenesis. Moreover, genetic studies have revealed existence of mutations in their primary sequences. Since most genetic variants identified by genome-wide association studies (GWAS) are located outside coding genes, many of these mutations may therefore affect lncRNAs.

LncRNAs may regulate signaling pathways involved in initiation, tumor progression, and metastatic spread similar

to protein-coding oncogenes and oncosuppressors. *PINC* and *PGEM1* were the first oncogenic lncRNAs found overexpressed in breast and prostate carcinomas. Since then, many other lncRNAs have been identified, with oncogenic properties (*KRAS*, *HULC*, *HOTAIR*, *MALAT1*, *HOTTIP*, *ANRIL*, and *RICTOR*) or oncosuppressive properties (*MEG3*, *GAS5*, *LincRNA-p21*, *PTENP1*, *TERRA*, *CCND1/CyclinD1*, and *TUG1*). Interestingly, several lncRNAs may show both oncogenic and oncosuppressive activities, depending on cellular context. *XIST* noncoding transcript is overexpressed in male tumors and underexpressed in female tumors.

(i) Oncogenic lncRNAs

(a) *SRA* (steroid receptor RNA activator) is a coactivator for steroid receptors and acts as an ncRNA found in the nucleus and cytoplasm. *SRA* regulates gene expression mediated by steroid receptors through complexing with proteins also containing steroid receptor coactivator 1 (SRC-1). The *SRA1* gene can also encode a protein that acts as a coactivator and corepressor. *SRA* levels have been found to be upregulated in breast tumors where it is assumed that increased *SRA* levels change the steroid receptors' actions, contributing to breast carcinogenesis. While expression of *SRA* in normal tissues is low, it is highly upregulated in breast, uterus, and ovary carcinomas [61–63].

(b) *HOTAIR* (HOX antisense intergenic RNA) is a long intergenic noncoding RNA with a length of 2.2 kb in *HOXC* locus and transcribed in antisense manner. *HOTAIR* regulates gene expression by modulating chromatin structures. It was the first lncRNA discovered to be involved in carcinogenesis. Polycomb group proteins mediate repression of transcription of thousands of genes that control differentiation pathways during development, pluripotency, and cancer progression. Target of PRC2 is *HOXD* locus on chromosome 2 where PRC2 in association with *HOTAIR* promotes transcriptional silencing of metastasis suppressor genes. *HOTAIR* acts as a molecular scaffold remodeling chromatin by causing histone modifications on target genes. *HOTAIR* comprises two known chromatin modification complexes with 5 region binding to PRC2 complex responsible for repressive H3K27 methylation and 3 region binding to LSD1, which initiates activating H3K4 demethylation [64, 65].

(c) *ANRIL* (antisense ncRNA in the *INK4* locus) is a natural antisense transcript, which activates the two polycomb repressor complexes PRC1 and PRC2, resulting in chromatin reorganization with silencing of *INK4b-ARF-INK4a*. This results in p15/INK4b, p14/ARF, and p16/INK4a inhibition which are normally implicated in cell cycle negative regulation, senescence, and stress-induced apoptosis. *ANRIL* overexpression in prostate carcinomas has shown silencing of *INK4b-ARF-INK4a* and p15/CDKN2B.

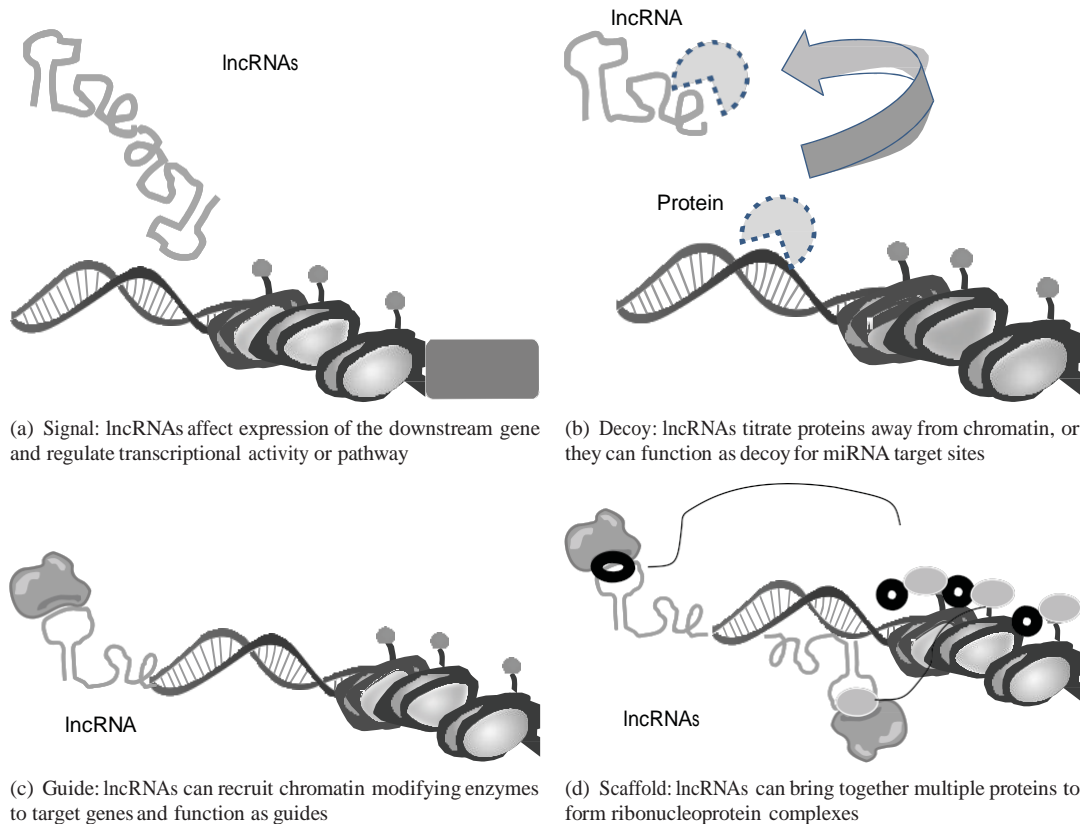


Figure 1: Molecular functions of lncRNAs.

Repression mechanism is mediated by direct binding to CBX 7 and SUZ12, 2 members of PRC1 and PRC2, respectively [66–68].

- (d) *MALAT1* (metastasis-associated lung adenocarcinoma transcript 1) is a lincRNA widely expressed in normal human tissues and overexpressed in a variety of breast, colon, liver, pancreas, prostate, and uterus carcinomas. *MALAT1* locus at 11q13.1 has been identified to harbor chromosomal translocation break points and mutations in breast and bladder carcinomas. It is localized in nuclear speckles and recent studies have implicated *MALAT1* in splicing, EMT, cell mobility, ECM remodeling, and metastatic spread [69].
- (e) *HULC* (highly upregulated in liver cancer) has been suggested to act as a “sponge” that inhibits miR-372 by sequestering it away from potential mRNA targets [70].
- (f) *PCGEM1* (prostate cancer gene expression marker 1) is overexpressed in prostate cancer cells that promotes tumor cells initiation and progression and protects against chemotherapy-induced apoptosis. Moreover, the reciprocal negative control relationship between *PCGEM1* and miR-145 regulates both prostate cancer cells proliferation and tumor growth. These results identify *PCGEM1* and associated regulators as possible targets for prostate cancer therapy [71].

(ii) Tumor Suppressor lncRNAs

- (a) *MEG3* (maternally expressed gene 3) is a transcript of the maternally imprinted gene normally expressed in pituitary cells. *MEG3* loss of expression is observed in pituitary adenomas and meningiomas. *MEG3* acts by regulating p53 pathway. P53 levels are usually extremely low due to its rapid degradation via ubiquitin-proteasome pathway. P53 ubiquitination is mainly mediated by E3 ubiquitin ligase MDM2. *MEG3* downregulates MDM2 expression, increases p53 protein level, stimulates p53-dependent transcription, and enhances p53 binding to target promoters [72].
- (b) *GAS5* (growth arrest-specific 5) is widely expressed in embryonic and adult tissues. *GAS5* acts as a starvation or growth arrest-linked riborepressor for glucocorticoid receptors by inhibiting association of these receptors with their DNA recognition sequence. This suppresses several responsive genes activation including gene encoding cellular inhibitor of apoptosis 2 (cIAP2). *GAS5* low expression levels have been observed in prostate and breast carcinomas [73, 74].
- (c) *CCND1/Cyclin D1* is transcribed from the promoter region of the Cyclin D1 gene. *Cyclin D1* is a cell cycle regulator frequently mutated, amplified, and overexpressed in carcinomas. *CCND1/Cyclin D1* recruits the RNA-binding protein TLS, which undergoes

allosteric modification, resulting in Cyclin D1 gene inhibition [75, 76].

- (d) *LincRNA-p21* expression is directly induced by p53 signaling pathway. This lncRNA is implicated in global repression of genes interfering with p53 function to regulate cellular apoptosis [77].
- (e) *TERRA* (telomeric repeat containing RNA) is expressed at chromosome ends. *TERRA* upregulation upon experimental manipulation or in ICF (immunodeficiency, centromeric instability, and facial anomalies) patients correlates with short telomeres. *TERRA* transcription facilitates the 5-3 nuclease activity of Exo1 at chromosome ends, providing a means to regulate the telomere shortening rate. Thereby, telomere transcription can regulate cellular lifespan through modulation of chromosome end processing activities [78].

(iii) Oncogenic and Tumor Suppressor lncRNAs

- (a) *H19* is expressed from the maternal allele and has a pivotal role in genomic imprinting during cell growth and development. The locus contains *H19* and *IGF2*, which are imprinted. This leads to differential expression of both genes *H19* from maternal and *IGF2* from paternal allele. This lncRNA presents both oncogenic and suppressive properties although the exact mechanism is still elusive [79].

lncRNAs are implicated in cancer epigenetics. Epigenetics refers to events that modulate activity of the genome without changing its sequence. It provides control of genome expression and establishment of tissular and cellular-type specific genes expression profiles. The main epigenetic mechanisms comprise chemical modifications of DNA and histones (cytosine methylation, posttranslational modifications of histones, chromatin remodeling, and nucleosome positioning). These chemical changes are reversible and controlled by enzyme complexes directly connected to metabolic and signaling pathways as well as sensors of extra- and intracellular microenvironments. Deregulation of these epigenetic mechanisms has been demonstrated in cancer cells which have no transcriptional gene or protein expression while being free of DNA damage. Carcinogenesis is very frequently associated with abnormal signaling and epigenetic abnormalities, most of which are considered as significant oncogenic events (aberrant methylation of oncosuppressor genes). lncRNAs interact with numerous proteins and in particular epigenetic regulators, including DNA methyltransferases and enzymatic complexes modifying chromatin and nucleosomes [80].

(i) Methylation of DNA was first observed in cancer epigenetic alterations. Cancer epigenome is characterized by global hypomethylation leading to genomic instability and hypermethylation of CpG islands which cause inactivation of genes oncosuppressors implicated in signaling pathways (DNA repair, apoptosis, and cycle cell regulation) or transcription factors involved in the control of these genes. lncRNAs bind to DNA methyltransferase, guide it to promoter of oncosuppressor genes, and ensure transcriptional silencing.

(ii) Nucleosomes can be destabilized and restructured by remodeling complexes belonging to 4 families (SWI/SNF, ISWI, CHD, and INO80). Nucleosomes positioning and remodeling regulate gene expression by perturbing accessibility of DNA regulatory sequences by transcription factors and transcription machinery. Mutations and silencing of subunits of these complexes in various types of cancer have been observed. lncRNAs could reposition nucleosomes by interacting and guiding these remodeling complexes to specific genomic regions where nucleosome restructuring would ensure repression of oncosuppressive genes [80].

(iii) lncRNAs recruit histone modifying enzymes acting near the transcription site of lncRNAs. Histones modifying enzymes read, add, or remove covalent links with change in accessibility of chromatin and fixation of nonhistone protein effectors that decode modified histone code. Alterations in the expression of histone-modifying enzymes (mutations, overexpression) were observed in various carcinomas. These alterations inhibit oncosuppressive genes transcription (EZH2 histone methyltransferase responsible for repressive H3K27me3 mark and silencing of p15, p16, and p19; histone deacetylase responsible for loss of lysine acetylated H4) or activate oncogenes transcription (demethylases and deacetylases) [81].

Growing numbers of lncRNAs are implicated in numerous mechanisms characterizing hallmarks of cancer. They can act on proliferation via coactivation (*SRA1*) and inhibition of elongation (*RN7SK*). lncRNAs allow escape mechanisms from suppressor pathways via competition (*PFS*), silencing (*ANRIL*), or ribonucleic repression (*GAS5*). Several lncRNAs are involved in replicative immortality by inhibiting telomerase (*TERRA*). lncRNAs promote neoangiogenesis is by inhibiting HIF1A (*AHIF*, *MALAT1*). They induce resistance to cell death by acting on p53 and p21 (*PCGEM1*) and by reducing expression of proapoptotic genes (*PANDA*). Finally lncRNAs are involved in mechanisms of invasion and metastasis via deregulation of alternative splicing (*MALAT1*) or trans-silencing of HOXD locus (*HOTAIR*).

2. lncRNAs as Biomarkers and Therapeutic Targets in Future Medical Practice

lncRNAs are emerging as integral functional components of human genome and are now considered as critical regulators in molecular biology of cancer. Recent data have demonstrated that lncRNAs are associated with cancer initiation, tumor progression, and metastatic spread. Unlike mRNAs used as diagnostic and prognostic biomarkers while being expressed by numerous subtypes of malignant tumors, lncRNAs present cellular and tissular specificity and could serve as biomarkers and therapeutic targets in cancer. Furthermore, lncRNAs secreted or released by apoptotic or necrotic tumor cells can be detected in blood, plasma, and urine. There seems to be a significant correlation between levels of circulating nucleic acids and genomic, epigenetic, or transcriptional alterations associated with malignant tumors (Table 2).

(i) *Esophagus*. Three lncRNA signatures have been identified recently in oesophageal squamous cell carcinoma,

Table 2: Representative lncRNAs involved in carcinogenesis and potential cancer biomarkers.

Cancer type	lncRNAs	References
Esophagus	<i>ENST00000435885.1, XLOC-013014, ENST00000547963.1</i>	[12]
Stomach	<i>GCAT1, H19, SUMO1P3</i>	[13–17]
Colon and rectum	<i>HOTAIR, uc.73</i>	[18, 19]
Liver	<i>HULC, HOTAIR, MALAT1, HOTTIP, HEIH</i>	[20–22]
Lung	<i>MALAT1, TUG1, BANCR, GASS5</i>	[23–28]
Breast	<i>HOTAIR, lincRNA-RoR, UCA1</i>	[29–32]
Ovary	<i>HOTAIR</i>	[33]
Bladder	<i>UCA1, H19, linc-UBC1, MALAT1</i>	[34, 35]
Prostate	<i>PCA3, PCAT1, PCGEM1</i>	[36–39]
Glioma	<i>H19</i>	[40]
Melanoma	<i>BANCR</i>	[41]
Oral cavity and nasopharynx	<i>HOTAIR, linc-C22orf32-1, linc-AL355149.1-1, linc-ZNF674-1</i>	[4, 42–44]

including ENST00000435885.1, XLOC.013014, and ENST00000547963.1. The expression of these lncRNAs classified the patients into two groups with significantly different overall survival [12].

(ii) *Stomach*. *HOTAIR*, *GCAT1* (gastric cancer-associated transcript 1), *H19*, and *SUMO1P3* (small ubiquitin-like modifier SUMO 1 pseudogene 3) are the main lncRNAs reported as overexpressed in gastric carcinomas. They are often associated with lymph node and distant metastasis, suggesting that they might serve as potential diagnostic and prognostic biomarker [13–17]. Moreover, *GCAT1* expression is closely related to c-Myc activation [14]. *GCAT1* functions as an oncogene and may be used as biomarker and potential therapeutic target in gastric carcinoma [15, 16].

(iii) *Lower Digestive Tract*. *HOTAIR* overexpression is observed in colorectal carcinomas with advanced stage and liver metastases [19]. *uc.73* lncRNA is also associated with poor overall survival in patient with colorectal carcinomas [18]. These results place *HOTAIR* and *uc.73* lncRNA as reliable biomarkers for poor prognosis in colorectal cancer.

(iv) *Liver*. In hepatocellular carcinomas (HCC), *HULC* (highly upregulated in liver cancer) was the first lncRNA with highly specific upregulation detected in blood [20]. High plasma *HULC* rates were observed in patients with high grades HCC or with HBV+ status [21]. Recent data have shown that HBx could regulate *HULC* promoter to induce HCC via oncosuppressor *p18* silencing [22]. HBx was found to downregulate lncRNA-*Dreh*, which can inhibit hepatocellular growth and metastasis *in vitro* and *in vivo*.

MALAT-1 and *HOTAIR* have been shown to be overexpressed in large cohorts of HCC patients. Furthermore, *HOTAIR* is a prognostic biomarker for recurrence after liver transplantation. SiRNA-mediated inhibition of

MALAT-1 and *HOTAIR* suppresses cancer cell viability and invasion, sensitizes TNF- α , induces apoptosis, and increases chemotherapeutic sensitivity of HCC to cisplatin and doxorubicin [82, 83]. *HOTTIP* (HOXA transcript at the distal tip) and *HOXA13* were also found to be upregulated in HCC. *HOTTIP* and *HOXA13* levels were associated with HCC tumor progression, metastasis, and survival [84]. *HEIH* (high expression in HCC) is another lncRNA overexpressed in HCC and an independent prognostic factor associated with recurrence [85].

(v) *Lung*. *MALAT1* is a key prognostic biomarker for metastatic spread in lung adenocarcinomas [23]. *TUG1* (taurine upregulated gene 1) is generally downregulated in non-small cell lung carcinomas (NSCLC). In NSCLC patients, *TUG1* low expression was associated with high TNM stage, tumor size, and poorer overall survival. Univariate and multivariate analyses revealed that *TUG1* expression serves as an independent predictor for overall survival [24]. *HOTAIR* was initially reported to be highly expressed in NSCLCs with advanced stage and lymph node metastasis [25]. But more recent meta-analyses study showed that it did not reach statistical significance, and thus it needs further investigations [26]. In NSCLCs, *BANCR* (BRAF activated noncoding RNA) expression is significantly decreased compared to normal tissues. *BANCR* underexpression is considered as an independent prognostic factor and is associated with larger tumor size, advanced pathological stage, metastasis distance, and shorter overall survival. Recently, *BANCR* overexpression was found to play a key role in epithelial-mesenchymal transition [27]. In malignant pleural mesothelioma (MPM), *GASS5* (growth arrest specific transcript 5) underexpression was observed compared to normal mesothelial tissue. Conversely, *GASS5* was upregulated upon growth arrest induced by inhibition of Hedgehog and PI3K/mTOR signaling in MPM models [28].

(vi) *Breast*. Breast cancer progression is correlated with *HOTAIR* activity in numerous recent studies. The study carried out by Chisholm and colleagues demonstrated *HOTAIR* overexpression in primary and metastatic breast carcinoma tissues by using RNA in situ hybridization technique [29]. Another study focusing on *HOTAIR* revealed the dependence of *HOTAIR* expression on oestradiol production, due to its promoter region via several estrogen response elements. This study also demonstrated that *HOTAIR* knock-down induced apoptotic pathways in breast cancer cell lines and suggested estrogen receptors as coregulators for *HOTAIR* expression [30]. Conversely, a recent study focusing on 348 primary breast carcinomas revealed that increased DNA methylation led to *HOTAIR* downregulation and an unfavorable disease state, questioning suitability of *HOTAIR* as negative prognostic biomarker in breast carcinomas [31]. *LincRNA-RoR* suppresses p53 in response to DNA damage through interaction with heterogeneous nuclear ribonucleoprotein I (hnRNP I). Recent data demonstrated that hnRNP I can also form a functional ribonucleoprotein complex with *UCA1* (urothelial carcinoma-associated 1) and increase *UCA1* stability. Of interest, the phosphorylated form of hnRNP I, predominantly located in the cytoplasm, is responsible for the interaction with *UCA1*. Although hnRNP I enhances p27 translation through interaction with the 5'-untranslated region of p27 mRNAs, interaction of *UCA1* with hnRNP I suppresses p27 protein expression via competitive inhibition. In support of this finding, *UCA1* seems to have an oncogenic role in breast carcinogenesis both *in vitro* and *in vivo*. Finally, a negative correlation between p27 and *UCA1* was found. Together, these results suggest an important role of *UCA1* in breast carcinogenesis [32].

(vii) *Glioma*. *H19*/miR-675 signaling was recently identified as a critical pathway in glioma progression. By analyzing gene expression data, *H19* increased levels of expression were found in high grade glioma. SiRNA-induced *H19* depletion inhibited invasion in glioma cells. Furthermore, *H19* expression was positively correlated with miR-675 and *H19* inhibition reduced miR-675 expression. Collectively, these data suggest that *H19* regulates glioma development by deriving miR-675 [40].

(viii) *ENT Tumors*. Recently, *HOTAIR* was suggested to play a part in nasopharyngeal carcinoma (NPC). *HOTAIR* is implicated in NPC progression and patients with high *HOTAIR* levels have poor clinical outcome with tumor recurrence and distant metastasis [42, 43]. In primary NPCs, upregulation of *lnc-C22orf32-1*, *lnc-AL355149.1-1*, and *lnc-ZNF674-1* has been also observed. High levels of *lnc-C22orf32-1* and *lnc-ZNF674-1* are associated with advanced tumor stages. Recurrent NPC displayed a distinctive lncRNA expression pattern with increased expression of *lnc-BCL2L11-3* and decreased expression of *lnc-AL355149.1-1* and *lnc-ZNF674-1* [44]. Interestingly, lncRNAs can be located in whole saliva. *HOTAIR* is differentially expressed in saliva of metastatic oral squamous cell carcinoma patients compared to primary tumors. These findings suggest that detection of lncRNAs in

saliva may be used as a noninvasive and rapid diagnostic tool for diagnosis of oral carcinomas [4].

(ix) *Bladder*. Oncogenic lncRNAs, including *UCA1*, *H19*, *MALAT1*, and *linc-UBC1* (upregulated in bladder cancer 1), are overexpressed in bladder carcinomas and activate PI3K-AKT and Wnt/ β -catenin pathways. A pilot study took advantage of this to evaluate potential application of *UCA1* in urinary sediments from patients with bladder carcinomas. It turned out to be especially valuable for superficial G2/G3 patients at a high risk for muscular invasion, indicating that *UCA1* may be a new promising urinary biomarker for the diagnosis of bladder cancer. *H19* expression levels are remarkably increased in bladder carcinomas compared to normal tissue and could also serve as another biomarker. More recently, a new lncRNA *linc-UBC1* was found to be overexpressed in 60% of invasive bladder carcinomas and correlated with lymph node metastasis and poor survival. *MALAT1* is upregulated in bladder cancer and its expression level is correlated with tumor grade and metastatic stage [34, 35].

(x) *Prostate*. Prostate cancer specific lncRNA *ADD3/PCA3* was identified fifteen years ago [86]. *PCA3* is a prostate-specific lncRNA markedly overexpressed in prostate carcinomas. It can be detected in prostate cancer tissue, urine, and/or urine sediments. In recent studies, *PCA3* specificity was found even higher than prostate biomarker PSA and *PCA3* score can accurately predict tumor volume and pathological features, which may guide treatment [36, 37]. Therefore, *PCA3* can be used as a noninvasive urine-based test for large-scale screening protocols and for predicting prostate carcinomas aggressiveness. This prominent example of rapid translation of lncRNA research into clinical practice offers a prototype for developing different lncRNAs as biomarkers. Other lncRNAs, such as PCGEM1 and PCAT1 (prostate cancer associated ncRNA transcript 1), are also prostate-specific, posing as attractive biomarkers [38, 39].

(xi) *Melanoma*. *BANCR* (lncRNA BRAF-activated noncoding RNA) plays a potentially functional role in melanoma cells proliferation and migration by activating ERK1/2 and JNK MAPK pathways. *BANCR* is upregulated in human malignant melanoma and patients with high levels of expression have a lower survival rate [41].

(xii) *Hemopathies*. Several recent reports have revealed deregulation of lncRNAs in leukemia, including *ANRIL*, *lncRNA-P21*, *MEG3*, *Dleu2*, *HOTAIRM1*, *EGO*, and *lncRNA-a7*. Moreover, *MEG3*, *UCA1*, and *H19* are upregulated in acute myeloid leukemia [87, 88].

(xiii) *Ovary*. *HOTAIR* plays a pivotal role in epithelial ovarian cancer (EOC) metastasis and could represent a novel prognostic marker and potential therapeutic target in patients with EOC. In recent study, *HOTAIR* expression was elevated in EOC tissues, and its level of expression is highly positively

correlated with FIGO stage, histological grade, lymph node metastasis, reduced overall survival, and disease-free survival [33].

Because of their central role in genes expression regulation, lncRNAs could also represent potential therapeutic targets. Better characterization of lncRNAs (structure, functions, polymorphisms, and intracellular topography) could help in faster development of new anticancer strategies modulating expression levels and functions of deregulated lncRNAs. Targeting lncRNAs offers novel exciting opportunity to treat cancer. Currently, nucleic acid-based methods prevail in targeting RNA, by regulating levels of expression and modifying their structures or mature sequences. Among them, RNA interference (RNAi) based techniques are arguably the most popular methods to inhibit lncRNAs in cancer cells. Both siRNAs and shRNAs exhibit great RNA selectivity and knockdown efficiency. Meanwhile, other established methods in inhibiting cancer-associated RNA, including antisense oligonucleotide (ASO), ribozyme, and aptamer, are also effective to modulate lncRNAs, and they show unique features that can have advantages over siRNAs [89].

In principle, targeting of lncRNAs can be achieved using the several approaches, including siRNA-mediated silencing, functional block using small molecules, or oligonucleotide inhibitors to prevent interactions of lncRNAs with proteins and structure disruption via small molecules or oligonucleotide inhibitors to change or mimic their secondary structure to compete for their binding partners [90].

(1) *Small interfering RNAs (siRNA)* are short stretched (19–30 nt) double-stranded RNAs that target RNA molecules via complementary to unpaired lncRNA sequences. The RNA duplex of siRNAs must be unwound into single strands before assembling into the active RNA-induced silencing complex (RISC). SiRNAs are fully complementary to their RNA target that is then cleaved at a single phosphodiester bond located near the centre of the sequence complement to the siRNA sequence [90]. SiRNAs exhibit high knockdown efficiency to many oncogenic lncRNAs in cancer cells and induce anticancer effects both *in vitro* and *in vivo*. For instance, depletion of *HOTAIR* by siRNAs decreases matrix invasiveness of breast cancer cells and inhibits tumor growth of pancreatic cancer xenograft. *HULC* and *MALAT1* siRNA-induced knockdown inhibits HCC cell proliferation and cell cycle progression [91]. Furthermore, siRNA-mediated knockdown of *H19* induces apoptosis and inactivates p53 [92]. Phase I and II clinical therapeutic trials have been evaluated using siRNAs to inhibit critical cancer-associated genes, including siRNA-EphA2-DOPC (targeting *EphA2*), TKM-080301 (targeting *PLK1*), and CALAA-01 (targeting *RRM2*). Therapy combining siRNA drug siG12D LODER and Atu027 with conventional chemotherapy is studied in phase II for its therapeutic effect in advanced pancreatic cancer patients. However, the main obstacle of siRNAs therapeutics remaining is their delivery. SiRNAs using ribonucleic acid as building block are susceptible to be degraded by nuclease and have poor pharmacokinetics.

(2) *Antisense oligonucleotides (ASO)* are short, single-stranded DNAs or RNAs (between 8 and 50 nt) designed

with sequence specific to target lncRNA [90]. ASOs directly hybridize to lncRNA transcripts via base pairing and endogenous RNase H1, which results in cleavage of lncRNA molecules, then recognizes hybrids formed. ASOs modulate lncRNAs function through degradation of lncRNA transcripts. Inhibition of *MALAT1* by ASO attenuates various malignant phenotypes in cancer cells via cycle arrest in cervical cancer cells [93]. Injection of ASO into subcutaneous tumors of nude mice effectively inhibits *MALAT1 in vivo* and blocks metastasis of lung cancer cells [91].

(3) *Ribozymes* are naturally produced RNA molecules that present intracellular catalytic functions. One of their functions is degradation of RNA molecules. Among all types of ribozyme, hammerhead ribozyme (HamRz) has caught major interest as it shows good target inhibitory effect while having the smallest RNA endoribonucleolytic motif as well as function independent to presence of metal ions [94]. HamRz is single-stranded RNA in neutral condition and undergoes folding in cells to expose the binding arms. Binding of HamRz to target sequence depends on complementary match with homologous target site. Both arms of HamRz have to bind with target sites correctly in order to form functional catalytic motif. After binding, HamRz catalyzes cleavage of the flanked RNA region downstream to a NUH site via destabilizing phosphodiester backbone of target RNA [90].

(4) *Aptamers* are short DNA or RNA oligonucleotides or peptides that have a stable 3-dimensional structure *in vivo*. They have broad molecular targets including protein, RNA, and small molecules that rely on fitting 3-dimensional shape of their ligands [95]. They specifically bind to their target lncRNAs that rely on fitting 3-dimensional shape of the lncRNA structures. Aptamers antagonize their lncRNA targets by blocking the interactions between lncRNAs and critical factors [90]. Some reports show promising effects of aptamers to either degrade RNA or inhibit RNA functions, implying the potential of aptamers as therapeutic agents to target lncRNAs. Aptamers are used to modulate viral gene expression by interacting with viral RNAs. A hairpin aptamer is identified to form stable and specific complex with the transactivation response element (TAR) RNA element of HIV-1 mRNA and decrease the TAR-dependent viral protein expression [96]. Another aptamer, selected against TAR element, formed stable loop-loop complexes with the element to regulate the TAR-mediated process [97]. Another study showed that aptamer could target the apical loop domain of pri-microRNA molecules and modulate the biogenesis of mature microRNA [98].

(5) *MicroRNAs (miRNAs)* can interact with lncRNAs. *H19/miR-675* signaling is critical in glioma progression. By analyzing glioma gene expression data sets, *H19* is found to be increased in high grade gliomas. *H19* depletion via siRNA inhibits invasion in glioma cells. Further, *H19* is positively correlated with its derivate miR-675 expression and reduction of *H19* inhibits miR-675 expression. MiR-675 modulates cadherin 13 expression by directly targeting its binding site within 3' UTR. These results demonstrate that *H19* regulates glioma development by deriving miR-675 and provide important clues for understanding key roles of lncRNA-microRNA functional network in glioma.

MicroRNA-based anticancer therapy has great potential as reports show there is apparent lack of adverse event in normal tissues when administrated with microRNA-based agent. Various microRNA delivery strategies, such as cationic lipid and nanoparticle encapsulation, are developed to improve microRNA shuttling into target cancer cells [40, 99, 100]. The potential of microRNA regulation on lncRNAs can be further realized by better understanding of intragenic lncRNA regulating elements.

(6) *Small molecules* are synthesized to specifically bind to RNA binding pockets of lncRNAs. They compete with protein factors or intracellular small ligands for binding lncRNAs. Binding of small molecules may also induce conformational change within lncRNA molecules and disrupt formation of important lncRNA structures [90].

3. Conclusion and Perspectives

Deregulated oncogenic and oncosuppressive lncRNAs are observed at all stages of malignant tumors development of various origins. Mechanisms implicating lncRNAs in carcinogenesis are dominated by deregulation of signaling pathways and altered epigenetic, transcriptional, and post-transcriptional expression of numerous genes. Their use as biomarkers and potential therapeutic targets could appear promising.

Major challenges of the next 10 years remaining are identification, mapping of all lncRNAs belonging to the human genome, and their functional characterization. Although their mechanisms of action are better known, it is still possible that lncRNAs are rather downstream products of particular chromatin structures or deregulated transcriptional processes. Their functionality is currently still debated because of their low level of expression and conservation between species. Unlike proteins, which often have well-defined functional areas, it is currently impossible to predict lncRNAs function from their single sequence. However, many traits exhibited by lncRNAs favor functionality, including 3 polyadenylation, 5 cap, multiple exons, transcriptional activation similar to that of mRNA, K4K36 domain, and alternative splicing. These challenges will only be achieved through combined efforts of functional genomics, epigenomics, and bioinformatics.

Conflict of Interests

No potential conflict of interests was disclosed.

Acknowledgment

This work was supported by Grant INCa-DGOS-4654.

References

- [1] S. Djebali, C. A. Davis, A. Merkel et al., "Landscape of transcription in human cells," *Nature*, vol. 489, no. 7414, pp. 101–108, 2012.
- [2] M. K. Iyer, Y. S. Niknafs, R. Malik et al., "The landscape of long noncoding RNAs in the human transcriptome," *Nature Genetics*, vol. 47, no. 3, pp. 199–208, 2015.
- [3] R. J. Taft, K. C. Pang, T. R. Mercer, M. Dinger, and J. S. Mattick, "Non-coding RNAs: regulators of disease," *Journal of Pathology*, vol. 220, no. 2, pp. 126–139, 2010.
- [4] E. A. Gibb, K. S. S. Enfield, G. L. Stewart et al., "Long non-coding RNAs are expressed in oral mucosa and altered in oral premalignant lesions," *Oral Oncology*, vol. 47, no. 11, pp. 1055–1061, 2011.
- [5] J. M. Perkel, "Visiting 'Noncodarnia,'" *BioTechniques*, vol. 54, no. 6, pp. 301–304, 2013.
- [6] K. C. Pang, M. C. Frith, and J. S. Mattick, "Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function," *Trends in Genetics*, vol. 22, no. 1, pp. 1–5, 2006.
- [7] O. Wapinski and H. Y. Chang, "Long noncoding RNAs and human disease," *Trends in Cell Biology*, vol. 21, no. 6, pp. 354–361, 2011.
- [8] S. Ye, L. Yang, X. Zhao, W. Song, W. Wang, and S. Zheng, "Bioinformatics method to predict two regulation mechanism: TF–miRNA–mRNA and lncRNA–miRNA–mRNA in pancreatic cancer," *Cell Biochemistry and Biophysics*, vol. 70, no. 3, pp. 1849–1858, 2014.
- [9] Q. Jiang, J. Wang, Y. Wang, R. Ma, X. Wu, and Y. Li, "TF2LncRNA: identifying common transcription factors for a list of lncRNA genes from ChIP-seq data," *BioMed Research International*, vol. 2014, Article ID 317642, 5 pages, 2014.
- [10] C. Park, N. Yu, I. Choi, W. Kim, and S. Lee, "lncRNator: a comprehensive resource for functional investigation of long non-coding RNAs," *Bioinformatics*, vol. 30, no. 17, pp. 2480–2485, 2014.
- [11] Z. X. Chen, D. Sturgill, J. Qu et al., "Comparative validation of the D. melanogaster modENCODE transcriptome annotation," *Genome Research*, vol. 24, no. 7, pp. 1209–1223, 2014.
- [12] Z. Chen, J. Li, L. Tian et al., "MiRNA expression profile reveals a prognostic signature for esophageal squamous cell carcinoma," *Cancer Letters*, vol. 350, no. 1–2, pp. 34–42, 2014.
- [13] M. Hajjari, M. Behmanesh, M. Sadeghizadeh, and M. Zeinoddini, "Up-regulation of HOTAIR long non-coding RNA in human gastric adenocarcinoma tissues," *Medical Oncology*, vol. 30, no. 3, article 670, 2013.
- [14] F. Yang, X. Xue, J. Bi et al., "Long noncoding RNA CCAT1, which could be activated by c-Myc, promotes the progression of gastric carcinoma," *Journal of Cancer Research and Clinical Oncology*, vol. 139, no. 3, pp. 437–445, 2013.
- [15] W. Sun, Y. Wu, X. Yu et al., "Decreased expression of long noncoding RNA AC096655.1-002 in gastric cancer and its clinical significance," *Tumor Biology*, vol. 34, no. 5, pp. 2697–2701, 2013.
- [16] B. Xiao and J. Guo, "Long noncoding RNA AC096655.1-002 has been officially named as gastric cancer-associated transcript 1, GACAT1," *Tumor Biology*, vol. 34, no. 5, p. 3271, 2013.
- [17] D. Mei, H. Song, K. Wang et al., "Up-regulation of SUMO1 pseudogene 3 (SUMO1P3) in gastric cancer and its clinical association," *Medical Oncology*, vol. 30, no. 4, article 709, 2013.
- [18] J. Sana, P. Faltejskova, M. Svoboda, and O. Slaby, "Novel classes of non-coding RNAs and cancer," *Journal of Translational Medicine*, vol. 10, no. 1, article 103, 2012.
- [19] R. Kogo, T. Shimamura, K. Mimori et al., "Long noncoding RNA HOTAIR regulates polycomb-dependent chromatin modification and is associated with poor prognosis in colorectal cancers," *Cancer Research*, vol. 71, no. 20, pp. 6320–6326, 2011.
- [20] K. Panzitt, M. M. O. Tschernatsch, C. Guelly et al., "Characterization of HULC, a novel gene with striking up-regulation in

- hepatocellular carcinoma, as noncoding RNA," *Gastroenterology*, vol. 132, no. 1, pp. 330–342, 2007.
- [21] H. Xie, H. Ma, and D. Zhou, "Plasma HULC as a promising novel biomarker for the detection of hepatocellular carcinoma," *BioMed Research International*, vol. 2013, Article ID 136106, 5 pages, 2013.
- [22] J.-F. Huang, Y.-J. Guo, C.-X. Zhao et al., "Hepatitis B virus X protein (HBx)-related long noncoding RNA (lncRNA) down-regulated expression by HBx (Dreh) inhibits hepatocellular carcinoma metastasis by targeting the intermediate filament protein vimentin," *Hepatology*, vol. 57, no. 5, pp. 1882–1892, 2013.
- [23] T. Gutschner, M. Hämmerle, M. Eissmann et al., "The non-coding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells," *Cancer Research*, vol. 73, no. 3, pp. 1180–1189, 2013.
- [24] E. B. Zhang, D. D. Yin, M. Sun et al., "P53-regulated long non-coding RNA TUG1 affects cell proliferation in human non-small cell lung cancer, partly through epigenetically regulating HOXB7 expression," *Cell Death and Disease*, vol. 5, no. 5, Article ID e1243, 2014.
- [25] T. Nakagawa, H. Endo, M. Yokoyama et al., "Large noncoding RNA HOTAIR enhances aggressive biological behavior and is associated with short disease-free survival in human non-small cell lung cancer," *Biochemical and Biophysical Research Communications*, vol. 436, no. 2, pp. 319–324, 2013.
- [26] J. Yang, J. Lin, T. Liu et al., "Analysis of lncRNA expression profiles in non-small cell lung cancers (NSCLC) and their clinical subtypes," *Lung Cancer*, vol. 85, no. 2, pp. 110–115, 2014.
- [27] M. Sun, X.-H. Liu, K.-M. Wang et al., "Downregulation of BRAF activated non-coding RNA is associated with poor prognosis for non-small cell lung cancer and promotes metastasis by affecting epithelial-mesenchymal transition," *Molecular Cancer*, vol. 13, no. 1, article 68, 2014.
- [28] A. Renganathan, J. Kresoja-Rakic, N. Echeverry et al., "GAS5 long non-coding RNA in malignant pleural mesothelioma," *Molecular Cancer*, vol. 13, no. 1, article 119, 2014.
- [29] K. M. Chisholm, Y. Wan, R. Li, K. D. Montgomery, H. Y. Chang, and R. B. West, "Detection of long non-coding RNA in archival tissue: correlation with polycomb protein expression in primary and metastatic breast carcinoma," *PLoS ONE*, vol. 7, no. 10, Article ID e47998, 2012.
- [30] A. Bhan, I. Hussain, K. I. Ansari, S. Kasiri, A. Bashyal, and S. S. Mandal, "Antisense transcript long noncoding RNA (lncRNA) HOTAIR is transcriptionally induced by estradiol," *Journal of Molecular Biology*, vol. 425, no. 19, pp. 3707–3722, 2013.
- [31] L. Lu, G. Zhu, C. Zhang et al., "Association of large noncoding RNA HOTAIR expression and its downstream intergenic CpG island methylation with survival in breast cancer," *Breast Cancer Research and Treatment*, vol. 136, no. 3, pp. 875–883, 2012.
- [32] J. Huang, N. Zhou, K. Watabe et al., "Long non-coding RNA UCA1 promotes breast tumor growth by suppression of p27 (Kip1)," *Cell Death and Disease*, vol. 5, no. 1, p. e1008, 2014.
- [33] J.-J. Qiu, Y.-Y. Lin, L.-C. Ye et al., "Overexpression of long non-coding RNA HOTAIR predicts poor patient prognosis and promotes tumor metastasis in epithelial ovarian cancer," *Gynecologic Oncology*, vol. 134, no. 1, pp. 121–128, 2014.
- [34] Q. Zhang, M. Su, G. Lu, and J. Wang, "The complexity of bladder cancer: long noncoding RNAs are on the stage," *Molecular Cancer*, vol. 12, no. 1, article 101, 2013.
- [35] L. Wang, D. Fu, Y. Qiu et al., "Genome-wide screening and identification of long noncoding RNAs and their interaction with protein coding RNAs in bladder urothelial cell carcinoma," *Cancer Letters*, vol. 349, no. 1, pp. 77–86, 2014.
- [36] G. Ploussard, X. Durand, E. Xylinas et al., "Prostate cancer antigen 3 score accurately predicts tumour volume and might help in selecting prostate cancer patients for active surveillance," *European Urology*, vol. 59, no. 3, pp. 422–429, 2011.
- [37] M. J. Roobol, F. H. Schröder, G. L. J. H. Van Leenders et al., "Performance of prostate cancer antigen 3 (PCA3) and prostate-specific antigen in prescreened men: Reproducibility and detection characteristics for prostate cancer patients with high PCA3 scores (≥ 100)," *European Urology*, vol. 58, no. 6, pp. 893–899, 2010.
- [38] A. L. Walsh, A. V. Tuzova, E. M. Bolton, T. H. Lynch, and A. S. Perry, "Long noncoding RNAs and prostate carcinogenesis: the missing 'linc'?" *Trends in Molecular Medicine*, vol. 20, no. 8, pp. 428–436, 2014.
- [39] G. O. Ifere and G. A. Ananaba, "Prostate Cancer Gene Expression Marker 1 (PCGEM1): a patented prostate-specific non-coding gene and regulator of prostate cancer progression," *Recent Patents on DNA and Gene Sequences*, vol. 3, no. 3, pp. 151–163, 2009.
- [40] Y. Shi, Y. Wang, W. Luan et al., "Long non-coding RNA H19 promotes glioma cell invasion by deriving miR-675," *PLoS ONE*, vol. 9, no. 1, Article ID e86295, 2014.
- [41] R. Li, L. Zhang, L. Jia et al., "Long non-coding RNA BANCR promotes proliferation in malignant melanoma by regulating MAPK pathway activation," *PLoS ONE*, vol. 9, no. 6, Article ID e100893, 2014.
- [42] R. A. Gupta, N. Shah, K. C. Wang et al., "Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis," *Nature*, vol. 464, no. 7291, pp. 1071–1076, 2010.
- [43] Y. Nie, X. Liu, S. Qu, E. Song, H. Zou, and C. Gong, "Long non-coding RNA HOTAIR is an independent prognostic marker for nasopharyngeal carcinoma progression and survival," *Cancer Science*, vol. 104, no. 4, pp. 458–464, 2013.
- [44] W. Gao, J. Y.-W. Chan, and T.-S. Wong, "Differential expression of long noncoding RNA in primary and recurrent nasopharyngeal carcinoma," *BioMed Research International*, vol. 2014, Article ID 404567, 9 pages, 2014.
- [45] A. C. Marques, J. Hughes, B. Graham, M. S. Kowalczyk, D. R. Higgs, and C. P. Ponting, "Chromatin signatures at transcriptional start sites separate two equally populated yet distinct classes of intergenic long noncoding RNAs," *Genome Biology*, vol. 14, no. 11, article R131, 2013.
- [46] R. Louro, A. S. Smirnova, and S. Verjovski-Almeida, "Long intronic noncoding RNA transcription: expression noise or expression choice?" *Genomics*, vol. 93, no. 4, pp. 291–298, 2009.
- [47] T. Kunej, J. Obsteter, Z. Pogacar, S. Horvat, and G. A. Calin, "The decalog of long non-coding RNA involvement in cancer diagnosis and monitoring," *Critical Reviews in Clinical Laboratory Sciences*, vol. 51, no. 6, pp. 344–357, 2014.
- [48] A. M. Khalil, M. Guttman, M. Huarte et al., "Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 28, pp. 11667–11672, 2009.
- [49] W. Xie and B. Ren, "Enhancing pluripotency and lineage specification," *Science*, vol. 341, no. 6143, pp. 245–247, 2013.
- [50] S. Nakagawa, J. Y. Ip, G. Shioi et al., "Malat1 is not an essential component of nuclear speckles in mice," *RNA*, vol. 18, no. 8, pp. 1487–1499, 2012.

- [51] I. Ilik, J. Quinn, P. Georgiev et al., "Tandem stem-loops in roX RNAs act together to mediate X Chromosome dosage compensation in *Drosophila*," *Molecular Cell*, vol. 51, no. 2, pp. 156–173, 2013.
- [52] Y. A. Sytnikova, R. Rahman, G. Chirn, J. P. Clark, and N. C. Lau, "Transposable element dynamics and PIWI regulation impacts lncRNA and gene expression diversity in *Drosophila ovarian* cell cultures," *Genome Research*, vol. 24, no. 12, pp. 1977–1990, 2014.
- [53] J. W. Nam and D. P. Bartel, "Long noncoding RNAs in *C. elegans*," *Genome Research*, vol. 22, no. 12, pp. 2529–2540, 2012.
- [54] Y. Zhang, J. Xia, Q. Li et al., "NRF2/long noncoding RNA ROR signaling regulates mammary stem cell expansion and protects against estrogen genotoxicity," *The Journal of Biological Chemistry*, vol. 289, no. 45, pp. 31310–31318, 2014.
- [55] S. He, H. Zhang, H. Liu, and H. Zhu, "LongTarget: a tool to predict lncRNA DNA-binding motifs and binding sites via Hoogsteen base-pairing analysis," *Bioinformatics*, vol. 31, no. 2, pp. 178–186, 2015.
- [56] J. C. Schwartz, S. T. Younger, N.-B. Nguyen et al., "Antisense transcripts are targets for activating small RNAs," *Nature Structural and Molecular Biology*, vol. 15, no. 8, pp. 842–848, 2008.
- [57] R. J. Taft, C. D. Kaplan, C. Simons, and J. S. Mattick, "Evolution, biogenesis and function of promoter-associated RNAs," *Cell Cycle*, vol. 8, no. 15, pp. 2332–2338, 2009.
- [58] J. Han, D. Kim, and K. V. Morris, "Promoter-associated RNA is required for RNA-directed transcriptional gene silencing in human cells," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 30, pp. 12422–12427, 2007.
- [59] S. Memczak, M. Jens, A. Elefsinioti et al., "Circular RNAs are a large class of animal RNAs with regulatory potency," *Nature*, vol. 495, no. 7441, pp. 333–338, 2013.
- [60] K. C. Wang and H. Y. Chang, "Molecular mechanisms of long noncoding RNAs," *Molecular Cell*, vol. 43, no. 6, pp. 904–914, 2011.
- [61] E. Leygue, H. Dotzlaw, P. H. Watson, and L. C. Murphy, "Expression of the steroid receptor RNA activator in human breast tumors," *Cancer Research*, vol. 59, no. 17, pp. 4190–4193, 1999.
- [62] S. Chooniedass-Kothari, E. Emberley, M. K. Hamedani et al., "The steroid receptor RNA activator is the first functional RNA encoding a protein," *FEBS Letters*, vol. 566, no. 1–3, pp. 43–47, 2004.
- [63] D. Ulveling, C. Francastel, and F. Hubé, "When one is better than two: RNA with dual functions," *Biochimie*, vol. 93, no. 4, pp. 633–644, 2011.
- [64] J. L. Rinn, M. Kertesz, J. K. Wang et al., "Functional demarcation of active and silent chromatin domains in human HOX Loci by noncoding RNAs," *Cell*, vol. 129, no. 7, pp. 1311–1323, 2007.
- [65] M.-C. Tsai, O. Manor, Y. Wan et al., "Long noncoding RNA as modular scaffold of histone modification complexes," *Science*, vol. 329, no. 5992, pp. 689–693, 2010.
- [66] E. Pasmant, I. Laurendeau, D. Héron, M. Vidaud, D. Vidaud, and I. Bièche, "Characterization of a germ-line deletion, including the entire INK4/ARF locus, in a melanoma-neural system tumor family: identification of ANRIL, an antisense noncoding RNA whose expression coclusters with ARF," *Cancer Research*, vol. 67, no. 8, pp. 3963–3969, 2007.
- [67] E. Pasmant, A. Sabbagh, J. Masliah-Planchon et al., "Role of noncoding RNA ANRIL in genesis of plexiform neurofibromas in neurofibromatosis type 1," *Journal of the National Cancer Institute*, vol. 103, no. 22, pp. 1713–1722, 2011.
- [68] E. Pasmant, A. Sabbagh, M. Vidaud, and I. Bièche, "ANRIL, a long, noncoding RNA, is an unexpected major hotspot in GWAS," *FASEB Journal*, vol. 25, no. 2, pp. 444–448, 2011.
- [69] L. Shen, L. Chen, Y. Wang, X. Jiang, H. Xia, and Z. Zhuang, "Long noncoding RNA MALAT1 promotes brain metastasis by inducing epithelial-mesenchymal transition in lung cancer," *Journal of Neuro-Oncology*, vol. 121, no. 1, pp. 101–108, 2015.
- [70] J. Wang, X. Liu, H. Wu et al., "CREB up-regulates long non-coding RNA, HULC expression through interaction with microRNA-372 in liver cancer," *Nucleic Acids Research*, vol. 38, no. 16, pp. 5366–5383, 2010.
- [71] J. H. He, J. Z. Zhang, Z. P. Han, L. Wang, Y. Lv, and Y. G. Li, "Reciprocal regulation of PCGEM1 and miR-145 promote proliferation of LNCaP prostate cancer cells," *Journal of Experimental & Clinical Cancer Research*, vol. 33, no. 1, article 72, 2014.
- [72] K.-H. Lu, W. Li, X.-H. Liu et al., "Long non-coding RNA MEG3 inhibits NSCLC cells proliferation and induces apoptosis by affecting p53 expression," *BMC Cancer*, vol. 13, article 461, 2013.
- [73] M. R. Pickard and G. T. Williams, "Regulation of apoptosis by long non-coding RNA GAS5 in breast cancer cells: implications for chemotherapy," *Breast Cancer Research and Treatment*, vol. 145, no. 2, pp. 359–370, 2014.
- [74] M. R. Pickard, M. Mourtada-Maarabouni, and G. T. Williams, "Long non-coding RNA GAS5 regulates apoptosis in prostate cancer cell lines," *Biochimica et Biophysica Acta*, vol. 1832, no. 10, pp. 1613–1623, 2013.
- [75] R. Kurokawa, "Promoter-associated long noncoding RNAs repress transcription through a RNA binding protein TLS," *Advances in Experimental Medicine and Biology*, vol. 722, pp. 196–208, 2011.
- [76] X. Song, X. Wang, S. Arai, and R. Kurokawa, "Promoter-associated noncoding RNA from the CCND1 promoter," *Methods in Molecular Biology*, vol. 809, pp. 609–622, 2012.
- [77] G. Wu, J. Cai, Y. Han et al., "LincRNA-p21 regulates neointima formation, vascular smooth muscle cell proliferation, apoptosis, and atherosclerosis by enhancing p53 activity," *Circulation*, vol. 130, no. 17, pp. 1452–1465, 2014.
- [78] V. Pfeiffer and J. Lingner, "TERRA promotes telomere shortening through exonuclease I-mediated resection of chromosome ends," *PLoS Genetics*, vol. 8, no. 6, Article ID e1002747, 2012.
- [79] M. Nordin, D. Bergman, M. Halje, W. Engström, and A. Ward, "Epigenetic regulation of the Igf2/H19 gene cluster," *Cell Proliferation*, vol. 47, no. 3, pp. 189–199, 2014.
- [80] S. Marquardt, R. Escalante-Chong, N. Pho et al., "A chromatin-based mechanism for limiting divergent noncoding transcription," *Cell*, vol. 157, no. 7, pp. 1712–1723, 2014.
- [81] R. Backofen and T. Vogel, "Biological and bioinformatical approaches to study crosstalk of long-non-coding RNAs and chromatin-modifying proteins," *Cell and Tissue Research*, vol. 356, no. 3, pp. 507–526, 2014.
- [82] M.-C. Lai, Z. Yang, L. Zhou et al., "Long non-coding RNA MALAT-1 overexpression predicts tumor recurrence of hepatocellular carcinoma after liver transplantation," *Medical Oncology*, vol. 29, no. 3, pp. 1810–1816, 2012.
- [83] Z. Yang, L. Zhou, L.-M. Wu et al., "Overexpression of long non-coding RNA HOTAIR predicts tumor recurrence in hepatocellular carcinoma patients following liver transplantation," *Annals of Surgical Oncology*, vol. 18, no. 5, pp. 1243–1250, 2011.
- [84] L. Quagliata, M. S. Matter, S. Piscuoglio et al., "Long noncoding RNA HOTTIP/HOXA13 expression is associated with disease progression and predicts outcome in hepatocellular carcinoma patients," *Hepatology*, vol. 59, no. 3, pp. 911–923, 2014.

- [85] S.-X. Yuan, F. Yang, Y. Yang et al., "Long noncoding RNA associated with microvascular invasion in hepatocellular carcinoma promotes angiogenesis and serves as a predictor for hepatocellular carcinoma patients' poor recurrence-free survival after hepatectomy," *Hepatology*, vol. 56, no. 6, pp. 2231–2241, 2012.
- [86] M. J. G. Bussemakers, A. Van Bokhoven, G. W. Verhaegh et al., "DD3: a new prostate-specific gene, highly overexpressed in prostate cancer," *Cancer Research*, vol. 59, no. 23, pp. 5975–5979, 1999.
- [87] M. Hajjari, A. Khoshnevisan, and Y. K. Shin, "Long non-coding RNAs in hematologic malignancies: road to translational research," *Frontiers in Genetics*, vol. 4, article 250, 2013.
- [88] E. F. Heuston, K. T. Lemon, and R. J. Arceci, "The beginning of the road for non-coding RNAs in normal hematopoiesis and hematologic malignancies," *Frontiers in Genetics*, vol. 2, article 94, 2011.
- [89] C. H. Li and Y. Chen, "Targeting long non-coding RNAs in cancers: progress and prospects," *International Journal of Biochemistry and Cell Biology*, vol. 45, no. 8, pp. 1895–1910, 2013.
- [90] L. Ma, M.-S. Chua, O. Andrisani, and S. So, "Epigenetics in hepatocellular carcinoma: an update and future therapy perspectives," *World Journal of Gastroenterology*, vol. 20, no. 2, pp. 333–345, 2014.
- [91] T. Gutschner, M. Hämmerle, and S. Diederichs, "MALAT1—a paradigm for long noncoding RNA function in cancer," *Journal of Molecular Medicine*, vol. 91, no. 7, pp. 791–801, 2013.
- [92] F. Yang, J. Bi, X. Xue et al., "Up-regulated long non-coding RNA H19 contributes to proliferation of gastric cancer cells," *FEBS Journal*, vol. 279, no. 17, pp. 3159–3165, 2012.
- [93] V. Tripathi, Z. Shen, A. Chakraborty et al., "Long noncoding RNA MALAT1 controls cell cycle progression by regulating the expression of oncogenic transcription factor B-MYB," *PLoS Genetics*, vol. 9, no. 3, Article ID e1003368, 2013.
- [94] D. E. Ruffner, G. D. Stormo, and O. C. Uhlenbeck, "Sequence requirements of the hammerhead RNA self-cleavage reaction," *Biochemistry*, vol. 29, no. 47, pp. 10695–10702, 1990.
- [95] J. Müller, O. El-Maarri, J. Oldenburg, B. Pöttsch, and G. Mayer, "Monitoring the progression of the in vitro selection of nucleic acid aptamers by denaturing high-performance liquid chromatography," *Analytical and Bioanalytical Chemistry*, vol. 390, no. 4, pp. 1033–1037, 2008.
- [96] G. Kolb, S. Reigadas, D. Castanotto et al., "Endogenous expression of an anti-TAR aptamer reduces HIV-1 replication," *RNA Biology*, vol. 3, no. 4, pp. 150–156, 2006.
- [97] M. Watrin, F. von Pelchrzim, E. Dausse, R. Schroeder, and J.-J. Toulmé, "In vitro selection of RNA aptamers derived from a genomic human library against the TAR RNA element of HIV-1," *Biochemistry*, vol. 48, no. 26, pp. 6278–6284, 2009.
- [98] C. E. Lünse, G. Michlewski, C. S. Hopp et al., "An aptamer targeting the apical-loop domain modulates pri-miRNA processing," *Angewandte Chemie—International Edition*, vol. 49, no. 27, pp. 4674–4677, 2010.
- [99] Y. Wu, Y. Xiao, X. Ding et al., "A mir-200b/200c/429-binding site polymorphism in the 3' untranslated region of the AP-2 gene is associated with cisplatin resistance," *PLoS ONE*, vol. 6, no. 12, Article ID e29043, 2011.
- [100] J. Wu, G. Wu, L. Lv et al., "MicroRNA-34a inhibits migration and invasion of colon cancer cells via targeting to Fra-1," *Carcinogenesis*, vol. 33, no. 3, pp. 519–528, 2012.

Annexe 3

Articles soumis, en révision et acceptés pour publication (2013-2016)

1 Mameri H, Bieche I, **Meseure D**, Marangoni E, Buhagiar-Labarchède G, Nicolas A, Vacher S, Onclercq-Delic R, Rajapakse V, Varma S, Reinhold WC, Pommier Y, Amor-Guélet M. Cytidine deaminase deficiency reveals new therapeutic opportunities against cancer. *Clin Cancer Res* 2016. pii: clincanres.0626.2016. [Epub ahead of print] PubMed PMID: 27601591.

2 Assayag F, Nicolas A, Vacher S, Bieche I, **Meseure D**, Aerts I, Cassoux N, Doz F and Decaudin Didier. Combination of carboplatin and bevacizumab is an efficient therapeutic approach in retinoblastoma Patient-Derived xenografts. *Invest Ophthalmol Vis Sci* 2016;57:4916-4926.

3 **Meseure D**, Vacher S, Drak Alsibai K, Trassard M, Nicolas A, Leclere R, Lerebours F, Guinebretiere JM, Marangoni E, Lidereau R, Bieche I. Biopathological Significance of TLR9 Expression in Cancer Cells and Tumor Microenvironment Across Invasive Breast Carcinomas Subtypes. *Cancer Microenviron*. 2016 Jul 9. [Epub ahead of print] PubMed PMID: 27392414.

4 Wassef M, Rodilla Benito V, Teissandier A, Zeitouni B, Gruel N, Sadacca B, Irondelle M, Charruel M, Michaud A, Caron M, Marangoni ME, Chavier P, Servant N, Reyat F, **Meseure D**, Vincent-Salomon A, Fre S and Margueron R. Relieving Polycomb-mediated repressive chromatin facilitates breast tumorigenesis. *Genes Dev* 2015;29:2547-62.

5 Baldeyron C, Brisson A, Tesson B, Némati F, Koundrioukoff S, Saliba E, De Koning L, Martel E, Ye M, Rigail G, **Meseure D**, Nicolas A, Gentien D, Decaudin D, Debatisse M, Depil S, Cruzalegui F, Pierré A, Roman-Roman S, Tucker GC, Dubois T. TIPIN depletion leads to apoptosis in breast cancer cells. *Mol Oncol* 2015;9:1580-98.

6 **Meseure D**, Drak Alsibai K, Nicolas A. Pivotal role of pervasive neoplastic and stromal cells reprogramming in circulating tumor cells dissemination and metastatic colonization. *Cancer Microenvironment* 2014;7:95-115.

7 Montes de Oca R, Gurard-Levin ZA, Berger F, Rehman H, Martel E, Corpet A, de Koning L, Vassias I, Wilson LO, **Meseure D**, Reyat F, Savignoni A, Asselain B, Sastre-Garau X, Almouzni G. The histone chaperone HJURP is a new independent prognostic marker for luminal A breast carcinoma. *Mol Oncol* 2015;9:657-74.

8 Lefort S, Joffre C, Kieffer Y, Givel AM, Bourachot B, Zago G, Bieche I, Dubois T, **Meseure D**, Vincent-Salomon A, Camonis J, Mehta-Grigoriou F. Inhibition of autophagy as a new means of improving chemotherapy efficiency in high-LC3B triple-negative breast cancers. *Autophagy* 2014;10:2122-42.

9 **Meseure D** and Drak Alsibai K. Cancer Metabolic and Immune Reprogramming: The Intimate Interaction Between Cancer Cells and Microenvironment. *J Cancer Prev Curr Res* 2014;1(2):00006.

10 Chanrion M, Kuperstein I, Barrière C, El Marjou F, Cohen D, Vignjevic D, Stimmer L, Paul-Gilloteaux P, Bièche I, Tavares Sdos R, Boccia GF, Cacheux W, **Meseure D**, Fre S, Martignetti L, Legoix-Né P, Girard E, Fetler L, Barillot E, Louvard D, Zinovyev A, Robine S. Concomitant Notch activation and p53 deletion trigger epithelial-to-mesenchymal transition and metastasis in mouse gut. *Nat Commun* 2014;5:5005. doi: 10.1038/ncomms6005.

- 11 Rocha C, Papon L, Cacheux W, Marques Sousa P, Lascano V, Tort O, Giordano T, Vacher S, Lemmers B, Mariani P, **Meseure D**, Medema JP, Bièche I, Hahne M, Janke C. Tubulin glycosylases are required for primary cilia, control of cell proliferation and tumor development in colon. *EMBO J* 2014;33:2247-60.
- 12 Cizkova M, Vacher S, **Meseure D**, Trassard M, Susini A, Mlcuchova D, Callens C, Rouleau E, Spyrtatos F, Lidereau R, Bièche I. PIK3R1 underexpression is an independent prognostic marker in breast cancer. *BMC Cancer* 2013;13:545.doi:10.1186/1471-2407-13-545.
- 13 **Meseure D**, Vacher S, Drak Alsibai K, Hatem R, Nicolas A, Caly M, Callens C, Lidereau R, Lerebours F, Guinebretiere JM and Bieche I. Altered expression of the three EGFR post-translational regulators MDGI, MIG6 and EIG121 in invasive breast carcinomas. *Article accepté à Biomed Res Int*.
- 14 Silvestre D, Brisson A, Marty-Prouvost B, Vacher S, Dengli F, Lombard B, Martel E, Reyes C, Rigaiil G, **Meseure D**, Nicolas A, Loew D, Gentien D, Bièche I, Tucker GC, Roman-Roman S and Dubois. PRMT1 is an attractive therapeutic target in breast cancers which regulates EGFR expression and activity. *Article soumis*.
- 15 Núñez N, Niborski L, De La Rochere P, Nadan A T, Milder M, Viel S, Loirat D, Pérol L, Richer W, **Meseure D**, Vincent-Salomon A, Sastre-Garau X, Sedlik C, Amigorena S and Piaggio E. Effective T cell activation in invaded lymph nodes from breast cancer patients. *Article soumis*.
- 16 Bougherara H, Némati F, Nicolas A, Massonnet G, **Meseure D**, Ngô C, Le Frère-Belda M A, Leary A, Alexandre J, Barret JM, Roman-Roman S, Prost JF, Donnadiou E and Decaudin D. The humanized anti-human AMHR2 mAb 3C23K directs anti-tumor activity of Tumor-Associated Macrophages against human ovarian cancer. *Article soumis*.

Titre : Evaluation du rôle des longs ARNs non codants dans les carcinomes mammaires infiltrants

Mots clés : sein, lncRNAs, cancer, ANRIL, MALAT1

Résumé : Le cancer du sein représente à l'échelle mondiale le deuxième cancer le plus fréquent et la première des tumeurs malignes de la femme. Actuellement, seuls certains biomarqueurs (RO, RP, récepteur HER2, index Ki67) et la signature transcriptomique PAM50 sont pris en compte dans la classification morphologique et l'orientation thérapeutique. Les analyses transcriptomiques à haut débit ont révélé que plus de 80% du génome humain est transcrit en ARN. Parmi les ARNs non codants, les transcrits dont la longueur est supérieure à 200 nt sont arbitrairement qualifiés de longs ARNs non codants (lncRNAs). Les lncRNAs jouent un rôle crucial dans le maintien de l'homéostasie cellulaire et présentent des profils d'expression anormaux dans diverses pathologies, dont le cancer. L'objectif principal de mon projet de thèse a consisté à analyser l'expression des lncRNAs, leur fonctionnalité et leur rôle dans l'oncogénèse mammaire. La première partie s'est focalisée sur l'étude des gènes ANRIL (ainsi que 10 gènes de la même voie de signalisation) et MALAT1, deux lncRNAs dont les mécanismes d'action et la signification clinique au cours de la cancérogénèse mammaire sont encore controversés. ANRIL et MALAT1 sont respectivement surexprimés dans 20% et 14% des tumeurs de notre série, confirmant leurs rôles prooncogénique dans la cancérogénèse mammaire. La surexpression de MALAT1 se traduit en RNA-FISH par la présence de volumineux speckles intranucléaires. La complexité de leur dérégulation est liée à la présence d'isoformes et de réseaux d'interactions avec les mRNAs et les miRNAs. Concernant les sous-unités appartenant aux complexes Polycomb PRC2 et PRC1 qui interagissent avec ANRIL, EZH2 (PRC2) est normalement ciblé par 3 miRNAs onco-suppresseurs (miR-26A1, miR-125B et miR-214) qui sont sous-exprimés dans notre série de CCI. Les 2 oncomiRs miR-181B1 and miR-181A2 qui ciblent et inactivent CBX7 (PRC1), apparaissent surexprimés dans notre série, en relation avec l'activation de l'oncogène HMGA1. Concernant MALAT1, le complexe de biogénèse des miRNAs Drosha-DGCR8-Microprocesseur régule l'expression d'un variant d'épissage Δ -MALAT1 et ce dernier est impliqué dans l'activation de la voie PI3K/Akt. Des corrélations significatives sont observées entre MALAT1 et des gènes impliqués dans l'épissage alternatif, le cycle cellulaire, l'apoptose, la réparation de l'ADN et la migration cellulaire. Les profils transcriptomiques aberrants de ces 2 lncRNAs semblent caractéristiques des carcinomes mammaires. Ainsi, ANRIL (i) présente une association positive inattendue avec le cluster p16-CDKN2A/p15-CDKN2B/p14-ARF dans notre série, alors que cette association apparaît négative dans les carcinomes de prostate et (ii) inactive épigénétiquement les miRNAs onco-suppresseurs miR-99a/miR-449a dans les carcinomes gastriques et non dans notre série de CCI. D'un point de vue clinique, deux signatures pronostiques indépendantes ont pu être identifiées, l'une intégrant les 2 partenaires protéiques d'ANRIL appartenant aux complexes Polycomb (surexpression d'EZH2 / sous-expression de CBX7), et l'autre représentée par la sous-expression de Δ -MALAT1, observée dans 20% des tumeurs de notre série. La présence de variants d'épissage alternatifs, de réseaux d'interactions multiples et d'une spécificité d'organe devra être prise en compte lors de l'évaluation des thérapies épigénétiques ciblant ANRIL (inhibiteurs des bromodomains et des oncoMIRs) ou MALAT1 (ASOs) dans les cancers du sein. La deuxième partie du projet de thèse a consisté en l'analyse du transcriptome non codant des carcinomes mammaires par une stratégie pangénomique, afin d'identifier de nouveaux types de lncRNAs, tels les nouveaux lncRNAs antisens, circulaires et associés à des séquences ultraconservées ou induisant des résistances médicamenteuses. L'analyse préliminaire réalisée sur une petite cohorte de carcinomes mammaires (n=8) permettra la mise en place de l'étude principale (n=40) pour renforcer la robustesse des signatures identifiées sur la cohorte initiale.

Title : Analysis of lncRNAs in breast carcinomas

Keywords : breast, carcinoma, MALAT1, ANRIL

Abstract : Breast cancer is the second most common cancer and the first malignancy of women. Currently, only few biomarkers (ER, PR, receptor HER2, index Ki67) and transcriptomic signature PAM50 are included in the morphological classification and therapeutic orientation. Transcriptome genome-wide analyses unexpectedly revealed that over 80% of the DNA is transcribed into RNA. Among these noncoding RNAs, transcripts longer than 200 nt are arbitrarily qualified as long noncoding RNAs (lncRNAs). lncRNAs play a crucial role in maintenance of cellular homeostasis and present abnormal expression patterns in various diseases, including cancer. The main objective of my project was to analyze expression of lncRNAs, their functionality and their roles in breast oncogenesis. The first part focused on the study of ANRIL and MALAT1 genes, two lncRNAs whose mechanisms of action and clinical significance in breast carcinogenesis are still controversial. ANRIL and MALAT1 respectively overexpressed in 20% and 14% of tumors in our series, confirming their pro-oncogenic roles in mammary carcinogenesis. MALAT1 overexpression results in RNA-FISH by presence of huge intranuclear speckles. Complexity of their deregulation is associated with presence of various isoforms and interaction networks with miRNAs, mRNAs and other lncRNAs. Concerning PRC2/PRC1 polycomb sub-units interacting with ANRIL, EZH2 (PRC2) is normally targeted by 3 onco-suppressor miRNAs (miR-26A1, miR-125B and miR-214) that are under-expressed in our series of CCIs. The 2 oncomiRs miR-181B1 and miR-181A2 that normally target and inactivate CBX7 (PRC1) appear overexpressed in our series of CCIs, resulting from activation of the oncogene HMGA1. Concerning MALAT1, the miRNAs biogenesis complex Drosha-DGCR8-Microprocessor regulates expression levels of the splicing variant Δ -MALAT1 and the latter is involved in activation of PI3K/Akt pathway. Significant correlations were observed between MALAT1 and genes involved in alternative splicing, cell cycle, apoptosis, DNA repair and migration. Aberrant transcriptomic profiles of these two lncRNAs seem characteristics of mammary carcinomas. Thus, ANRIL (i) presents an unexpected positive association with the p16-CDKN2A/p15-CDKN2B/p14-ARF cluster in our series of CCIs, whereas this association appears negative in prostate carcinomas and (ii) epigenetically inactivates onco-suppressor miRNAs miR99a/miR-449a in gastric carcinomas, but not in our series. From a clinic point of view, two independent prognostic signatures were identified, one incorporating two protein partners of ANRIL belonging to the polycomb complexes (EZH2 overexpression / CBX7 under-expression) and the other represented by under-expression of the variant Δ -MALAT1 observed in 20% of tumors in our series. The presence of alternative splice variants, multiple interactions with mRNAs and miRNAs and organ specificity should be considered when evaluating epigenetic antitumoral drugs designed to target ANRIL (bromodomains and oncoMIRs inhibitors) and MALAT1 (ASOs) in breast cancers. The second part of the project involved analysis of non-coding transcriptome of mammary carcinomas to identify new types of lncRNAs, including new antisens lncRNAs, circular lncRNAs, induced lncRNAs, noncoding ultraconserved transcripts and lncRNAs associated with resistance to systemic treatments. The preliminary analysis performed on a small cohort of breast cancers (n=8) will allow the implementation of the main (n=40) which will enhance robustness of identified signatures.