



UNIVERSITE LIBRE DE BRUXELLES
FACULTE DE MEDECINE
LABORATOIRE D'EPIGENETIQUE DU CANCER

Portrait, caractérisation et intérêt clinique
des ARNs non codants dans le cancer du
sein

Olivier VAN GREMBERGEN

Thèse présentée en vue de l'obtention du grade académique de Docteur en Sciences
Biomédicales et Pharmaceutiques

Sous la direction du Professeur François Fuks

2016

"Tu apprendras à faire tourner la roue de la Vie en ta faveur. Tu accepteras le jeu de la Vie et ce que tu appelais difficultés, tu appelleras cela désormais Opportunités."

William Shakespeare

Résumé

Ces dernières années, les avancées technologiques ont révélé qu'une majeure partie de notre génome est transcrit en ARNs non codants, qui sont impliqués dans de nombreuses pathologies. Dès lors, nous avons voulu, lors de cette thèse, mieux comprendre le rôle des ARNs non codants dans le cancer du sein.

Dans un premier temps, nous avons étudié les micros ARNs non codants (miRs) dans un modèle cellulaire de cancer mammaire. Nous avons identifié deux miRs qui sont exprimés dans les lignées mammaires normales, mais pas dans les lignées cancéreuses. La surexpression de miR-137 entraîne une diminution de la prolifération et de la migration des cellules cancéreuses. De plus, nous avons identifié que miR-137 cible directement la protéine histone déméthylase KDM5B et diminue son expression génique et protéique. Ensuite, nous avons cherché si d'autres histones déméthylases de la famille KDM5 pouvaient être régulées par les miRs. Nous avons découvert que KDM5C, qui est surexprimée dans les lignées mammaires cancéreuses, est une cible directe de miR-138. La surexpression de miR-138 dans les lignées tumorales diminue l'expression de KDM5C et entraîne une chute de la prolifération cellulaire. Globalement, ces résultats révèlent que les miRs peuvent réguler les protéines épigénétiques KDM5 et contrôler la prolifération cellulaire.

La deuxième partie de cette thèse est consacrée à un nouveau sujet de recherche pour la communauté scientifique : l'étude des longs ARNs non codants (lncRNAs). Des études pionnières révèlent que quelques lncRNAs sont impliqués dans les cancers du sein. Des milliers de lncRNAs existent, mais très peu ont été caractérisés. Dès lors, nous avons décidé d'évaluer globalement leur profil d'expression dans une large cohorte de cancers du sein. Nous avons identifié 215 lncRNAs dérégulés dans les tumeurs. Nos résultats révèlent que l'expression des lncRNAs permet de classer les cancers du sein en différents sous-types. Des analyses bioinformatiques ont permis de prédire leurs fonctions et leurs implications dans différentes voies moléculaires clés du cancer du sein telles que les voies PI3K/AKT/mTOR et MAPK. Nous avons également découvert que 210 lncRNAs sont des marqueurs pronostics indépendants du risque de rechute. Enfin, nous avons choisi deux lncRNAs que nous avons étudiés expérimentalement. Nous avons montré que lnc-KIN-2 contrôle la prolifération cellulaire en régulant l'expression des gènes GATA3 et ESR1. Ensuite, nous avons découvert que CYTOR, un lncRNA surexprimé dans les tumeurs mammaires, régule des gènes de la voie EGFR/mTOR et est requis pour la prolifération et la migration cellulaire ainsi que pour le maintien du cytosquelette et de la morphologie normale de la cellule.

En démontrant l'importance biologique des ARNs non codants dans le développement des tumeurs mammaires, ces résultats laissent entrevoir la mise en lumière de nouveaux mécanismes par lesquels ces ARNs particuliers, qui ne codent pas des protéines, contribuent au processus de cancérogénèse et pourraient devenir des nouvelles cibles thérapeutiques.

Remerciements

Ce travail est le fruit de longs efforts et de longues discussions menées par différents acteurs talentueux. Je remercie toutes les personnes qui ont contribué directement ou indirectement au bon déroulement de mon travail de thèse. D'abord François, mon promoteur qui m'a sélectionné pour jouer le rôle principal et qui, lors de nombreuses réunions, m'a donné les idées et les outils nécessaires à l'aboutissement de ce projet. Ensuite, tous les membres du laboratoire qui ont joué des rôles plus ou moins importants, mais qui ont toujours été là pour me conseiller lorsque j'en ressentais le besoin. Je les remercie sincèrement pour tous les bons moments passés dans, et surtout hors du laboratoire. En particulier Eric et nos débats interminables sur d'innombrables sujets, les filles de mon bureau qui ont toujours été présentes pour une bonne tranche de rire et nos formidables techniciennes qui ont réalisé des expériences de grande qualité.

Je n'oublie pas Martin, sans qui cette thèse n'aurait pas été possible. Il a réalisé un travail extraordinaire d'analyses bioinformatiques, et il a surtout pris le temps de me transmettre une (toute petite) partie de son savoir. Je remercie Christos Sotiriou, Jana et Matthieu qui m'ont également apporté toute leur expertise sur le cancer du sein ainsi que nos autres collaborateurs de l'institut Bordet, de l'IB², de l'UGENT et de la KUL. Je pense aussi à toi Ioly, qui a toujours été d'une écoute attentive, ainsi qu'à Maud qui m'a remonté le moral lorsque mon projet patinait.

Merci à Elise pour sa relecture et ses conseils concernant ma rédaction. Merci également à Céline mon épouse, et à ma maman qui ont aussi fait le gros effort de relire ce manuscrit.

Ajoutons évidemment Céline, mes parents, ma famille et mes amis qui ont cru en moi.

Enfin, je remercie le Télévie, le FNRS et la région wallonne qui ont financé mes recherches ainsi que les membres du jury qui ont pris le temps de lire ce manuscrit.

Ici s'achève une belle première partie de ma carrière professionnelle (non sans une pointe de nostalgie) mais de nombreux nouveaux et passionnants défis m'attendent.

Table des matières

Liste des abréviations	13
Liste des figures et tables	15
Préambule.....	17
<u>INTRODUCTION GENERALE</u>	19
1 L'épigénétique	21
1.1 Introduction	21
1.2 Les fonctions physio-pathologiques de l'épigénétique	23
1.3 La chromatine	24
1.4 Les machineries de modifications épigénétiques	26
1.4.1 Les modifications covalentes de l'ADN.....	27
1.4.1.1 Méthylation des cytosines.....	27
1.4.1.1.1 Méthyltransférase de l'ADN	28
1.4.1.1.2 Mécanismes de répression par la méthylation.....	29
1.4.1.1.3 Implications pathologiques	31
1.4.1.2 Autres modifications covalentes de l'ADN	32
1.4.1.2.1 L'hydroxyméthylation des cytosines.....	32
1.4.1.2.2 La méthylation des adénines.....	34
1.4.2 Les modifications d'histones	34
1.4.2.1 Modifications post-traductionnelles des histones.....	35
1.4.2.1.1 Acétylation	36
1.4.2.1.2 Méthylation	37
Les histones déméthylases de la famille KDM5	38
1.4.2.1.3 Autres modifications d'histones	40
1.4.2.2 Mécanismes du code histone et connexion avec la méthylation de l'ADN	41
1.5 L'épigénétique en clinique.....	43
1.5.1 Les inhibiteurs des méthyltransférases de l'ADN.....	43
1.5.2 Les inhibiteurs des désacétylases d'histones.....	44
1.5.3 Thérapies épigénétiques en développement.....	44
2 Les ARNs non codants	47
2.1 Introduction	47
2.1.1 Classification des ARNs non codants.....	49
2.2 Les micro-ARNs (miRs).....	50
2.2.1 Biogenèse des miRs	50
2.2.2 Mode d'action des miRs	51
Inhibition <i>pré</i> -traductionnelle	52
Inhibition <i>co</i> -traductionnelle	53
Inhibition <i>post</i> -traductionnelle	53
2.2.3 Fonctions des miRs dans le développement tumoral.....	53
2.2.3.1 Profil d'expression des miRs dans le cancer.....	53
2.2.3.2 Exemples de miRs oncogènes (onco-miRs)	54
Le cluster miR-17-92.....	54
MiR-21	55
2.2.3.3 Exemples de miRs onco-suppresseurs	55
Le cluster let-7	55
Le cluster miR-15-16.....	55

2.2.3.4	MiR-137 et miR-138	56
	MiR-137	56
	MiR-138	56
2.3	Les longs ARNs non codants	57
2.3.1	Introduction	57
2.3.2	Classification	57
2.3.3	Propriétés des longs ARNs non codants.....	58
2.3.3.1	Caractérisation génomique	58
2.3.3.2	Localisation cellulaire	59
2.3.3.3	Conservation	60
2.3.4	Fonctions et Modes d'action	61
2.3.5	Implications pathologiques	63
2.3.5.1	Implication des lncRNAs dans le cancer.....	64
	<i>HOTAIR</i>	66
	<i>MALAT1</i>	67
	<i>MEG3</i>	67
2.3.5.2	Potentiel thérapeutique.....	68
3	Le cancer du sein	71
3.1	Introduction	71
3.1.1	Epidémiologie	71
3.2	Développement du cancer du sein	72
3.2.1	Anatomie du sein	72
3.2.2	Tumorigénèse mammaire.....	74
3.3	Diversité des cancers du sein.....	74
3.3.1	Classification anatomo-pathologique	74
3.3.2	Classification clinique	75
3.3.2.1	Le statut des récepteurs hormonaux	76
3.3.2.2	Le statut HER2.....	76
3.3.3	Sous-types moléculaires du cancer du sein.....	76
3.3.4	Altérations (épi)génétiques du cancer du sein	79
3.3.4.1	Altérations génétiques du cancer du sein	79
3.3.4.2	Altérations épigénétiques du cancer du sein	80
3.4	Prise en charge thérapeutique.....	80
3.4.1	La chirurgie et la radiothérapie	81
3.4.2	Les traitements systémiques	81
3.4.2.1	La chimiothérapie	81
3.4.2.2	L'hormonothérapie	82
3.4.2.3	Thérapie anti-HER2	82
3.4.3	Conclusions	83
4	But du projet	85
	<u>RESULTATS</u>.....	87
	Chapitre 1 : Les micro-ARNs régulent les histones déméthylases de la famille KDM5 dans les cellules cancéreuses mammaires	89
I.	INTRODUCTION	91
II.	RESULTATS	93
II.1	MiR-137 est sous-exprimé dans les cellules cancéreuses mammaires.	93
II.2	La surexpression de miR-137 inhibe la prolifération et la migration des cellules cancéreuses mammaires.....	94
II.3	KDM5B est une cible directe de miR-137.....	96

II.4 MiR-138 régule KDM5C en modèle cellulaire de cancer du sein	98
II.5 La surexpression de miR-138 inhibe la prolifération des cellules cancéreuses mammaires	100
Chapitre 2 :	103
Description et caractérisation fonctionnelle des longs ARNs non codants dans le cancer du sein	103
I. INTRODUCTION	105
II. RESULTATS – Partie 1.....	109
II.1 Les tumeurs mammaires ont un profil d’expression des lncRNAs différent du tissu mammaire sain.....	109
II.2 L’expression des lncRNAs est associée au récepteur aux œstrogènes	112
II.3 L’expression des lncRNAs est associée aux sous-types moléculaires des cancers du sein	113
II.4 Les lncRNAs sont associés à différents processus biologiques et voies moléculaires du cancer	113
II.4 Lnc-KIN-2 module la prolifération cellulaire et l’expression du récepteur à l’œstrogène.....	115
II.5 Conclusions	117
II.6 Matériels et méthodes.....	140
Préparation des ARNs de patientes.....	140
Quantification par PCR quantitative.	140
Transfection par les LNA.....	140
Analyse bioinformatique et statistiques.....	141
III. RESULTATS – Partie 2.....	118
III.1 Les tumeurs mammaires possèdent un profil d’expression des lncRNAs distinct des tissus sains	118
III.2 L’expression des lncRNAs est associée à la voie ER.....	122
III.3 L’expression des lncRNAs est associée aux sous-types moléculaires de cancer du sein	127
III.4 Prédiction fonctionnelles des lncRNAs associés aux sous-types moléculaires du cancer du sein	128
III.5 Les lncRNAs sont des marqueurs pronostics dans le cancer du sein	131
III.6 CYTOR est requis pour la prolifération, la migration et l’organisation du cytosquelette.	134
<u>DISCUSSION</u>	143
I. Les micro-ARNs régulent les histones déméthylases de la famille KDM5 dans les cellules cancéreuses mammaires	145
II. Description et caractérisation fonctionnelle des longs ARNs non codants dans le cancer du sein	149
Analyse des lncRNAs à l’échelle du génome	149
Comparaison des résultats de la partie 1 (puce à ADN spécifique) et de la partie 2 (données publiques).....	151
Fonctions associées aux lncRNAs	152
Les lncRNAs comme marqueurs pronostics	155
Caractérisation expérimentale des lncRNAs	156
Perspectives.....	160
IV. Conclusion générale	163
<u>BIBLIOGRAPHIE</u>	165

Bibliographie	167
<u>ANNEXE I</u>.....	181
<u>ANNEXE II</u>.....	183

Liste des abréviations

5caC : 5-carboxylcytosine
5fC : 5-formylcytosine
5mC : 5-méthylcytosine
5hmC : 5-hydroxy-méthylcytosine
6mA : N6-méthyladénosine
3'-UTR : région non traduite en 5' (*five prime untranslated region*)
5'-UTR : région non traduite en 5' (*three prime untranslated region*)
5-Aza : 5-aza-2-désoxycytidine
ADN : acide désoxyribonucléique
ARN : acide ribonucléique
ATP : adénosine triphosphate
CpG : dinucléotide cytosine-phosphate-guanine
DNMT : ADN méthyltransférase (*DNA methyltransferase*)
EGFR : récepteur au facteur de croissance épithéliale (*epidermal growth factor receptor*)
EMT : transition épithélio-mésenchymateuse (*epithelial-mesenchymal transition*)
ER : récepteur aux œstrogènes (*estrogen receptor*)
ER+ : tumeurs exprimant le récepteur aux œstrogènes
ER- : tumeurs n'exprimant pas le récepteur aux œstrogènes
FBS : serum fœtal bovin (*fetal bovine serum*)
FDA : agence américaine des produits alimentaires et médicamenteux (*Food and Drug Administration*)
FDR : taux de faux positif (*false discovery rate*)
FT : facteur de transcription
H3K4 : lysine 4 de l'histone 3
H3K4me1 : histone 3 monométhylée sur la lysine 4
H3K4me3 : histone 3 triméthylée sur la lysine 4
H3K27 : lysine 27 de l'histone 3
HAT : histone acétyltransférase
HDAC : histone désacétylase
HDACi : inhibiteur d'HDAC (*HDAC inhibitor*)
HDM : histone déméthylase
HER2 : récepteur au facteur de croissance épidermale humain 2 (*human epidermal growth factor receptor-2*)
HKMT : histone lysine méthyltransférase
HMT : histone méthyltransférase
HR : ratio de risque (*hazard ratio*)
JAK : janus kinase
lncARN : long ARN non codant (*long noncoding RNA*)
LNA : acide nucléique bloqué (*locked nucleic acid*)
miRISC : complexe effecteur de silençage par les miR (*miRNA-induced silencing complex*)
m7G : N-7-méthylguanine
Mb : mégabase
MBP : protéine liant les CpG méthylés (*mCpG binding protein*)
miR : micro-ARN
mRNA : ARN messager
mTOR : gène cible de la rapamycine chez les mammifères (*mamalian target of rapamycin*)
ncRNA : ARN non codant (*noncoding RNA*)
PR : Récepteur à la progestérone
PI3K : phosphatidylinositol-3-kinase
PR : Récepteur à la progestérone
PRC2 : complexe de répression Polycomb 2 (*polycomb repressive complex 2*)
ROC : Receiver operating characteristic

RBP : protéine liant l'ARN (*RNA binding protein*)
RIN: Score d'intégrité de l'ARN (*RNA integrity number*)
RNase : ribonucléase
qPCR : réaction en chaîne par polymérase, quantitative (*quantitative polymerase chain reaction*)
SAM : S-adenosylméthionine
SIT : site d'initiation de la transcription
TSA : trichostatine A

Liste des figures et tables

<i>Fig. 1 : Génome versus épigénomes.</i>	22
<i>Fig. 2 : Composantes de l'épigénétique.</i>	24
<i>Fig. 3 : Organisation de la chromatine.</i>	25
<i>Fig. 4 : Principales modifications épigénétiques.</i>	26
<i>Fig. 5 : Méthylation des cytosines par les DNMTs.</i>	28
<i>Fig. 6 : Méthylation de maintenance et « de novo ».</i>	29
<i>Fig. 7 : Inhibition de la transcription par la méthylation.</i>	30
<i>Fig. 8 : Méthylation aberrante et cancer.</i>	32
<i>Fig. 9 : Fonctions des protéines TETs.</i>	33
<i>Fig. 10 : Modifications post-traductionnelles d'histones.</i>	35
<i>Fig. 11 : Impact des modifications d'histones sur la stabilité des nucléosomes.</i>	36
<i>Fig. 12 : Méthylation d'histones sur les lysines.</i>	37
<i>Fig. 13 : Famille des déméthylases KDM5.</i>	39
<i>Fig. 14 : Recrutement des protéines par les histones modifiées.</i>	41
<i>Fig. 15 : Inter-influence des modifications d'histones.</i>	42
<i>Fig. 16 : Verrouillage de l'inhibition transcriptionnelle.</i>	42
<i>Fig. 17 : Littérature sur les ARNs non codants.</i>	47
<i>Fig. 18 : Génome, ARNs non codants et complexité.</i>	48
<i>Fig. 19 : Biogenèse des miRs.</i>	50
<i>Fig. 20 : Mécanismes d'inhibition par les miRs.</i>	52
<i>Fig. 21 : Classification des longs ARNs non codants.</i>	58
<i>Fig. 22 : Localisation cellulaire des lncRNAs.</i>	60
<i>Fig. 23 : Mode d'action des lncRNAs.</i>	62
<i>Fig. 24 : Implication des lncRNAs dans de nombreuses maladies.</i>	63
<i>Fig. 25 : Implication des lncRNAs dans le cancer.</i>	64
<i>Fig. 26 : Exemples de lncRNAs impliqués dans les voies c-MYC et p53.</i>	65
<i>Fig. 27 : Rôles de HOTAIR dans le cancer.</i>	67
<i>Fig. 28 : Incidence des cancers en Belgique en 2013 chez l'homme et la femme.</i>	72
<i>Fig. 29 : Anatomie du sein.</i>	73
<i>Fig. 30 : Anatomopathologie du cancer du sein.</i>	75
<i>Fig. 31 : Classement moléculaire des cancers du sein.</i>	77
<i>Fig. 32 : Relation entre les sous-types moléculaires et les caractéristiques clinico-pathologiques.</i>	78
<i>Fig. 33 : Relation entre les sous-types moléculaires et les gènes souvent mutés dans le cancer du sein.</i>	79
<i>Fig. 34 : Evolution du taux de mortalité lié au cancer du sein.</i>	83
<i>Fig. 35 : Sous-expression de miR-137 en lignées cancéreuses mammaires.</i>	93
<i>Fig. 36 : Régulation épigénétique de miR-137.</i>	94
<i>Fig. 37 : Surexpression de miR-137 après transfection du précurseur pré-miR-137.</i>	95
<i>Fig. 38 : Régulation de la prolifération et de la migration par miR-137.</i>	96
<i>Fig. 39 : Régulation de KDM5B par miR-137.</i>	97
<i>Fig. 40 : Régulation de KDM5C par miR-138.</i>	99
<i>Fig. 41 : Régulation de la prolifération cellulaire par miR-138.</i>	100
<i>Fig. 42 : Protocole d'analyse des lncRNAs dans le cancer du sein.</i>	110
<i>Fig. 43 : Dérégulation des lncRNAs dans le cancer du sein.</i>	111
<i>Fig. 44 : Association entre l'expression des lncRNAs et le statut ER.</i>	112
<i>Fig. 45 : Association entre l'expression des lncRNAs et les sous-types moléculaires des cancers du sein.</i>	113
<i>Fig. 46 : Prédiction fonctionnelle des lncRNAs dans le cancer du sein.</i>	114
<i>Fig. 47 : lnc-KIN-2 influence la prolifération et l'expression des gènes GATA3 et ESR1.</i>	116

<i>Fig. 48 : Dérégulation des lncRNAs dans une large cohorte de cancers du sein.</i>	119
<i>Fig. 49 : Dérégulation des lncRNAs dans une cohorte indépendante et validation.</i>	121
<i>Fig. 50 : Relation entre l'expression des lncRNAs et le récepteur aux œstrogènes.</i>	123
<i>Fig. 51 : Validation de la signature des lncRNAs associés à ER.</i>	125
<i>Fig. 52 : Expression et analyse fonctionnelle des lncRNAs associés à ER.</i>	126
<i>Fig. 53 : Association entre l'expression des lncRNAs et les sous-types moléculaires des cancers du sein.</i>	128
<i>Fig. 54 : Prédiction fonctionnelles de l'ensemble des lncRNAs et des ARNm associés à chaque sous-type.</i>	129
<i>Fig. 55 : Prédiction fonctionnelles des lncRNAs spécifiques de chaque sous-type de tumeurs mammaires.</i>	131
<i>Fig. 56 : Les lncRNAs sont pronostics du risque de rechute.</i>	132
<i>Fig. 57 : Validation de la signature de 27 lncRNAs pronostics.</i>	133
<i>Fig. 58 : CYTOR est surexprimé dans les cancers mammaires.</i>	134
<i>Fig. 59 : CYTOR influence la prolifération et la migration.</i>	135
<i>Fig. 60 : CYTOR modifie les phases du cycle cellulaire.</i>	136
<i>Fig. 61 : L'état de méthylation de CYTOR est anti-corrélé à son expression.</i>	137
<i>Fig. 62 : Identification des gènes régulés par CYTOR.</i>	138
<i>Fig. 63 : Organisation de l'actine filamenteuse dans les cellules qui sous-expriment CYTOR.</i>	139
<i>Fig. 64 : Modélisation de l'action de CYTOR.</i>	158
<i>Fig. 65 : Expression de CYTOR dans différents cancers.</i>	159

Préambule

La science de la vie est probablement la science la plus complexe de toutes les sciences naturelles. « La vie est compliquée ». Cette courte phrase prend également tout son sens en biologie et particulièrement en génétique. En effet, nous sommes plus que la somme de nos gènes. Toutes les cellules de l'être humain possèdent les mêmes gènes, qui contiennent l'information nécessaire à la vie. Cependant, certains de nos gènes s'expriment alors que d'autres sont éteints, ce qui explique pourquoi certaines cellules sont spécialisées dans le traitement de l'influx nerveux alors que d'autres se contractent pour faire battre le cœur. D'autre part, l'expression aberrante de certains gènes cause diverses maladies, dont le cancer. Ceci démontre qu'une régulation précise de l'expression des gènes est primordiale pour vivre en bonne santé. En biologie, une des questions fondamentales est de comprendre comment les gènes sont exprimés. L'épigénétique est la science qui cherche à répondre à cette question. Ces dernières années, l'épigénétique est devenue un domaine d'étude florissant et fascinant qui permet de comprendre ce que la génétique ne peut expliquer.

Ce travail porte sur l'étude fonctionnelle des ARNs non codants dans le cadre du cancer du sein.

Cette thèse est divisée en trois parties : la première partie est une introduction générale à l'épigénétique, au monde des ARNs non codants au sens large, ainsi qu'au cancer du sein. L'objectif de ce travail conclura notre introduction. La deuxième partie se consacre aux résultats de nos recherches. Cette partie est divisée en deux grands chapitres : nous parlerons d'abord du rôle des petits ARNs non codants dans la régulation de l'enzyme KDM5 qui participe à la tumorigénèse mammaire. Ensuite, nous nous focaliserons sur l'étude des longs ARNs non codants dans le cancer du sein, ce qui constitue la majeure partie de cette dissertation. Enfin, la dernière partie sera consacrée à la discussion des résultats obtenus et aux perspectives en relation avec les données actuelles.

INTRODUCTION GENERALE

1 L'épigénétique

1.1 Introduction

L'épigénétique, c'est d'abord le concept que tout n'est pas simplement codé dans l'ADN du génome. Notre code génétique est simple, composé de 4 bases de l'ADN, et passe d'une manière stable d'une génération à l'autre. Cependant, l'épigénétique infirme le concept de la « fatalité des gènes » et comble la brèche entre l'inné et l'acquis. L'épigénétique permet par exemple d'expliquer pourquoi des jumeaux monozygotes, qui partagent la même séquence d'ADN peuvent avoir des phénotypes différents. En effet, nous sommes plus que la somme de nos gènes, et l'expression des gènes doit être précisément régulée, par des mécanismes épigénétiques, pour maintenir l'homéostasie.

Forgé sémantiquement à partir de $-\text{epi}$ ($-\epsilon\pi\iota$, en grec) qui signifie « au-dessus de, sur » et de $-\text{génétique}$, le terme « épigénétique » a été introduit en 1942 par Conrad Waddington qui le définit, dans le cadre du développement, comme la branche de la biologie qui étudie l'interaction causale entre les gènes et leurs produits, qui est responsable du phénotype (Waddington, 1942). En 1994, Robin Holliday propose une nouvelle définition de l'épigénétique qui inclut l'héritabilité mitotique et l'indépendance de la séquence d'ADN dans les changements de l'expression des gènes (Holliday, 1994). Finalement, l'exploration intime des variations de l'état de la chromatine amène Adrian Bird, en 2007, à redéfinir l'épigénétique comme l'étude des adaptations structurales de régions chromosomiques pour appréhender, marquer ou perpétuer des états modifiés d'activité des gènes. Actuellement, l'épigénétique peut se définir comme l'étude, au niveau des chromosomes, des changements stables, réversibles, et idéalement héréditaires, de l'expression des gènes qui se produisent sans altération de la séquence d'ADN. Cependant, la définition même du terme épigénétique reste un sujet débattu, particulièrement concernant l'héritabilité de ces changements (Deans and Maggert, 2015). L'ensemble des modifications

épigénétiques, qui activent ou répriment l'expression des gènes, forme un code que l'on appelle l'épigénome. Si nous possédons tous un génome qui stocke l'information, nous avons également tous de multiples épigénomes qui organisent l'information (Fig. 1) (Scheen and Junien, 2012).

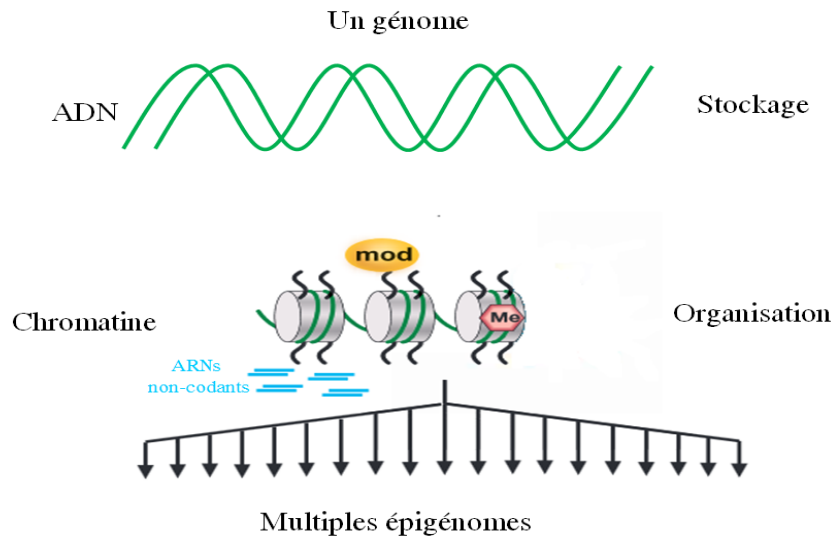


Fig. 1 : Génome versus épigénomes.

Le génome est invariant dans toutes nos cellules, il est le support de l'information. L'organisation de la chromatine, par des modifications épigénétiques telles que les modifications de la queue des histones (jaune), la méthylation de l'ADN (rouge) et les ARNs non codants (bleu), varient et constituent de multiples épigénomes. (Figure adaptée de Allis, Jenuwein, and Reinberg, 2007)

L'étude des modifications épigénétiques constitue un domaine majeur en biologie, car elles permettent à la fois de mieux comprendre le développement physiologique mais également pathologique qui mène à la maladie. Les pages suivantes seront consacrées aux rôles de l'épigénétique ainsi qu'à la définition de la chromatine et aux modifications épigénétiques les mieux décrites : les modifications covalentes de l'ADN et les modifications d'histones. Le deuxième chapitre de l'introduction générale portera sur les ARNs non codants qui peuvent également être considérés comme des régulateurs épigénétiques. Enfin, nous terminerons cette introduction par une brève description du cancer du sein, nécessaire à la compréhension de cette dissertation.

1.2 Les fonctions physio-pathologiques de l'épigénétique

Les mécanismes épigénétiques, qui régulent le profil d'expression des gènes, sont divers et complexes. Brièvement, l'épigénome influence la manière dont les gènes sont exprimés en régulant la structure chromatinienne. La régulation du génome en différentes structures compactes influence l'accessibilité à l'ADN et module donc l'expression des gènes. Trois grandes familles de mécanismes épigénétiques influencent l'état de la chromatine : les modifications covalentes de l'ADN, les modifications d'histones et les ARNs non codants. Ces mécanismes travaillent de concert pour localement, soit compacter la chromatine, empêchant les facteurs de transcription de se lier à l'ADN, soit relâcher la chromatine permettant la transcription. L'expression d'un gène dépend de plusieurs facteurs : l'état de compaction locale (au niveau des promoteurs) et plus distante (au niveau des activateurs « enhancer ») de la chromatine, mais aussi d'autres facteurs environnementaux (hormones, inflammation, nutrition, ...). Les marques épigénétiques, influencées par l'âge, le stade développemental et l'environnement, varient d'une cellule à l'autre, mais également au sein d'une même cellule. Les modifications des marques peuvent être transitoires ou perdurer dans le temps, et peuvent également être transmises aux générations suivantes (hérédité). L'épigénome n'est donc pas statique mais est en remodelage constant pour répondre à divers stimuli, ce qui permet, par exemple, aux cellules de se spécialiser au cours du développement (Scheen and Junien, 2012). Les marques épigénétiques peuvent donc être considérées comme les gardiens du contrôle de l'expression des gènes.

L'épigénétique est impliquée dans de très nombreux phénomènes biologiques incluant la régulation (i) de l'expression des gènes codants et non codants, (ii) de l'interaction entre l'ADN et les protéines, (iii) des éléments transposables, (iv) de l'inactivation du chromosome X et (v) de l'empreinte parentale (Goldberg, Allis, and Bernstein, 2007). L'importance de l'épigénétique dans le maintien de l'homéostasie est démontrée par l'existence de nombreuses observations d'altérations épigénétiques qui mènent à différentes pathologies (Fig. 2). Les aberrations épigénétiques sont très connues dans le cancer, mais sont égale-

ment décrites dans virtuellement toutes les maladies incluant les troubles neurologiques (Alzheimer, Parkinson, maladie d'Huntington,...), les maladies auto-immunes (diabète de type I, arthrite rhumatoïde, ...) et les maladies cardiaques et du sang (α -thalassémie) (Brookes and Shi, 2014; Portela and Esteller, 2010).

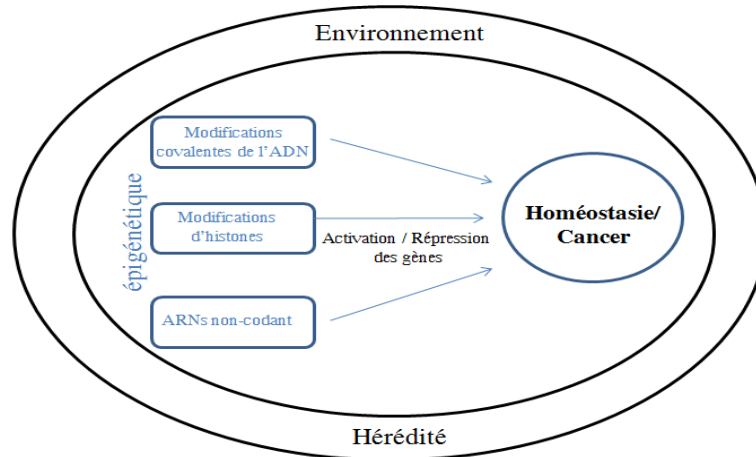


Fig. 2 : Composantes de l'épigénétique.

Les facteurs hérités et environnementaux influencent les modifications épigénétiques qui régulent l'expression des gènes afin de maintenir l'homéostasie. Des dérégulations de l'épigénome induisent de nombreuses maladies dont le cancer.

Les enjeux épigénétiques sont donc larges et complexes. Ils concernent à la fois la médecine, la biologie du développement, la physiologie, les théories de l'évolution et la santé publique.

1.3 La chromatine

Le génome humain diploïde comporte plus de six milliards de paires de bases, qui, dans une conformation totalement déroulée, mesure plus de deux mètres. L'information génétique, qui code pour plus de 20 000 protéines, doit donc être fortement compactée afin de rentrer dans le noyau cellulaire dont le diamètre est approximativement de 10 μ m. De plus, l'information génétique doit être finement organisée pour permettre une régulation précise de l'expression des gènes. Cet exploit remarquable (compaction de plus de 100 000 fois) est accompli par l'emballage de l'ADN autour de protéines histones, en unités répétées de nu-

cléosomes qui, ensemble forment la chromatine (Brookes and Shi, 2014). La structure de la chromatine est dynamique et subit diverses modifications pour influencer l'expression des gènes.

L'unité de base de la chromatine est le nucléosome qui est constitué de 146 paires de base d'ADN, enroulées autour d'un octamère d'histones constitué de deux copies de chaque histone H2A, H2B, H3 et H4, que l'on appelle les histones du « cœur ». L'alignement des nucléosomes forme une fibre en « collier de perles » de 11 nm de diamètre. L'histone H1 quant à lui est une histone de liaison, qui fait le pont inter-nucléosomal pour former une fibre de 30nm (Fig. 3).

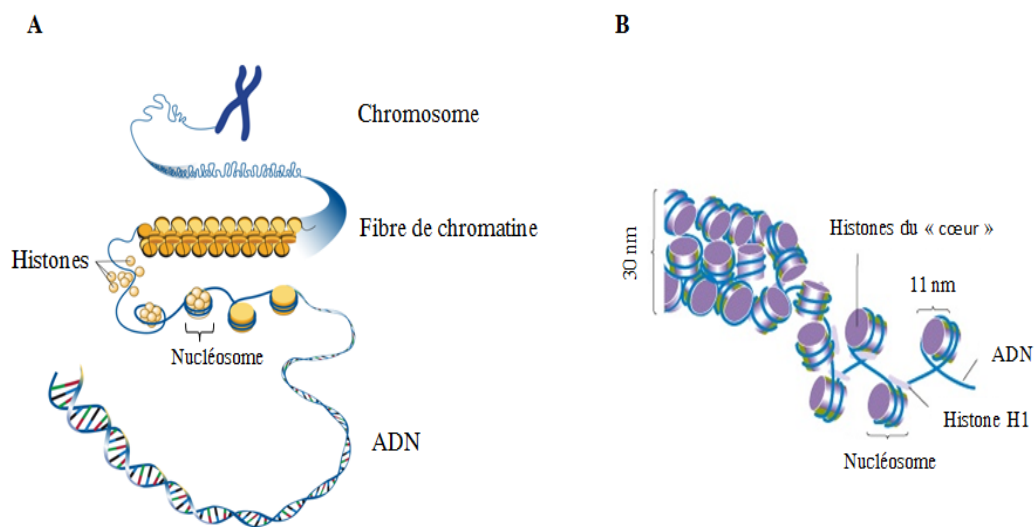


Fig. 3 : Organisation de la chromatine.

(A) La double hélice d'ADN est entourée autour du nucléosome pour former la fibre de chromatine. D'autres processus de compaction génèrent « in fine » le chromosome. (B) Organisation des nucléosomes pour former une fibre de 30 nm. (Figure adaptée de <https://www.genome.gov/dmdl/img.cfm?node=Photos/Graphics&id=85280> et de <http://www.bio.miami.edu/dana/104/nucleosome.jpg>)

Le degré de compaction de l'ADN autour des histones définit deux types de chromatines : l'euchromatine, faiblement condensée et accessible à la machinerie transcriptionnelle et l'hétérochromatine, fortement condensée et donc non transcrite. On distingue encore l'hétérochromatine en hétérochromatine constitutive qui est toujours compacte et l'hétérochromatine facultative qui peut transiter d'un état compact et inactif à un état plus ouvert et transcriptionnellement actif. La régulation de la compaction de la chromatine est donc fon-

damentale pour orchestrer correctement les processus cellulaires. Cette régulation est dirigée par de nombreux mécanismes épigénétiques, incluant les complexes de remodelage dépendant de l'ATP de la famille SWI/SNF, les modifications d'histones et la méthylation de l'ADN (Dawson and Kouzarides, 2012; Tessarz and Kouzarides, 2014). Ces deux derniers mécanismes sont plus amplement décrits ci-dessous.

1.4 Les machineries de modifications épigénétiques

Cette partie se consacre aux deux types de machineries épigénétiques les plus connues : 1) les modifications covalentes de l'ADN (principalement la méthylation) et 2) les modifications des histones (Fig. 4).

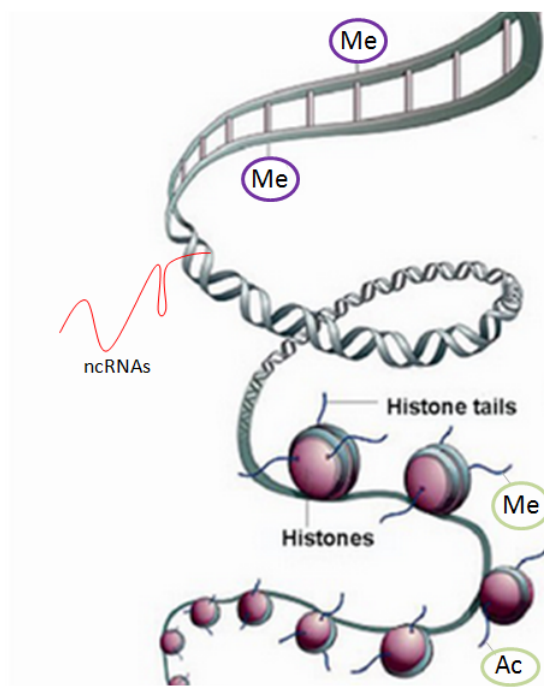


Fig. 4 : Principales modifications épigénétiques.

La méthylation de l'ADN (« Me », entouré de mauve) et les modifications des queues des histones (méthylation et acétylation entourées en vert) représentent les modifications épigénétiques les plus connues. Les ARNs non codants (en rouge) exercent également un contrôle épigénétique de la transcription. (Figure adaptée de Maleszewska and Kaminska, 2013)

Les ARNs non codants, qui font également partie des modifications épigénétiques, feront l'objet d'un chapitre entier plus tard dans cette introduction.

1.4.1 Les modifications covalentes de l'ADN

Nous avons vu qu'une régulation précise de l'expression des gènes est primordiale pour le maintien de l'homéostasie. Un des mécanismes impliqués concerne les modifications covalentes de l'ADN. La modification la mieux décrite est la méthylation des cytosines (5mC), qui est traditionnellement associée à la répression de la transcription. Cependant, d'autres modifications existent telles que l'hydroxyméthylation des cytosines (5hmC) et la méthylation des adénines, qui elles aussi régulent l'expression des gènes (Breiling and Lyko, 2015). Dans la suite de cette dissertation, le terme « modification de l'ADN » recouvre les modifications covalentes de l'ADN sans altération de séquence.

1.4.1.1 Méthylation des cytosines

La méthylation des cytosines de l'ADN se retrouve dans le génome de tous les animaux, ainsi que chez les plantes et les procaryotes. Alors qu'elle joue principalement un rôle de défense immunitaire contre les phages chez les bactéries, chez les vertébrés, la méthylation est impliquée dans la répression de l'expression des gènes, le développement, l'empreinte génomique parentale, l'inactivation du chromosome X, la régulation des éléments transposables et la ségrégation des chromosomes (Jones, 2012; Smith and Meissner, 2013; Breiling and Lyko, 2015).

Dans le règne animal, la méthylation des cytosines se produit généralement au sein de dinucléotides cytosine-phosphate-guanine (CpG), pour former la 5-méthylcytosine (5mC), également appelée 5^{ème} base de l'ADN. Environ 4% des cytosines du génome humain sont méthylées, principalement dans les régions riches en CpG, que l'on nomme îlot CpG. Cependant le niveau de méthylation est très variable d'une cellule à l'autre (Breiling and Lyko, 2015). Comme

nous l'expliquerons en détail plus loin, la présence de 5mC au niveau du promoteur des gènes est associée à la répression des gènes.

1.4.1.1.1 Méthyltransférase de l'ADN

La méthylation des cytosines est une modification chimique covalente catalysée par les enzymes ADN méthyltransférases (DNMTs). Les DNMTs transfèrent un groupe méthyle (-CH₃) de la S-adénosylméthionine (SAM) sur le carbone en position 5 des cytosines (Fig. 5) (H Denis, Ndlovu, and Fuks, 2011).

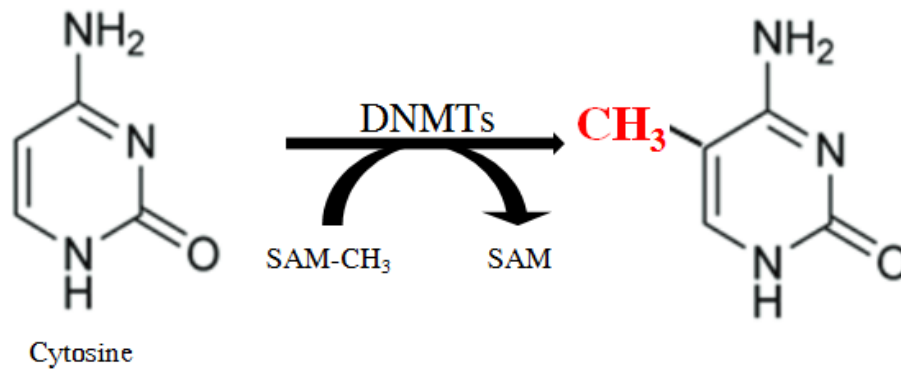


Fig. 5 : Méthylation des cytosines par les DNMTs.

Les ADN méthyltransférases transfèrent un groupe méthyle de la S-adénosylméthionine sur le carbone 5 des cytosines

Il existe trois enzymes catalytiquement actives chez les mammifères : DNMT1, DNMT3A et DNMT3B. DNMT1, première méthyltransférase découverte en 1988, est l'enzyme de maintenance de la méthylation, qui préserve les motifs de méthylation aux travers des divisions mitotiques. En effet, lors de la phase S du cycle cellulaire, l'enzyme DNMT1 se lie préférentiellement sur les substrats hémi-méthylés et méthyle le brin nouvellement synthétisé (Fig. 6).

Les enzymes DNMT3 sont les méthyltransférases « *de novo* », qui sont responsables de l'apposition de nouvelles marques méthyles pendant le développement embryonnaire. En association avec d'autres protéines, les enzymes DNMT3 déposent également, au cours du cycle cellulaire, de nouvelles marques méthyles sur certains loci spécifiques (H Denis, Ndlovu, and Fuks, 2011).

Enfin, notons que les DNMTs sont des enzymes essentielles à la vie, en effet l'invalidation des gènes codants pour les DNMTs entraîne une létalité précoce (H Denis, Ndlovu, and Fuks, 2011).

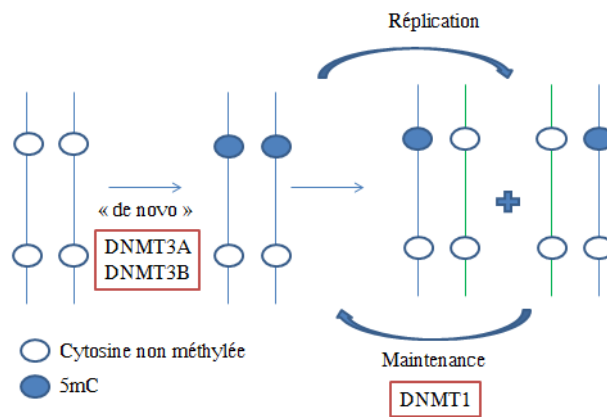


Fig. 6 : Méthylation de maintenance et « de novo ».
 Les enzymes DNMT3 sont responsables de la méthylation « de novo » alors que DNMT1 est responsable de la maintenance du profil de méthylation pendant la réplication.

1.4.1.1.2 Mécanismes de répression par la méthylation

Il est largement admis que les marques 5mC situées dans les îlots CpG, au niveau des promoteurs et des « enhancers », soient associées à la répression des gènes. Au moins deux types de mécanismes permettent d'expliquer la répression par la méthylation. D'abord, la 5mC peut directement empêcher la liaison de facteurs de transcription, tels que CTCF, sur leur locus cible et ainsi inhiber la transcription. Le second mécanisme implique les protéines MBPs (mCpG binding proteins), qui possèdent un domaine de liaison à l'ADN méthylé. Une fois liées, les protéines MBPs recrutent diverses machineries répressives telles que des histones déacétylases (HDACs) et d'autres complexes de remodelage de la chromatine (Fig. 7). Ce réseau d'interactions agit de concert pour stabiliser et verrouiller la répression (Rottach, Leonhardt, and Spada, 2009). Cependant, il n'est pas clairement démontré que la méthylation initie la répression des gènes. En effet, le complexe des DNMTs semble dépendant de la présence de nucléosomes pour déposer la marque méthyle. Or, les sites d'initiation de la transcription (SIT) actifs sont dépourvus de nucléosome. De plus, la présence d'histones

modifiées H3K4me3 (voir sous-chapitre 1.4.2) sur les promoteurs actifs, est incompatible avec la méthylation « de novo ». La méthylation, en association avec d'autres machineries répressives, serait donc plutôt impliquée dans le verrouillage de la transcription (Jones, 2012).

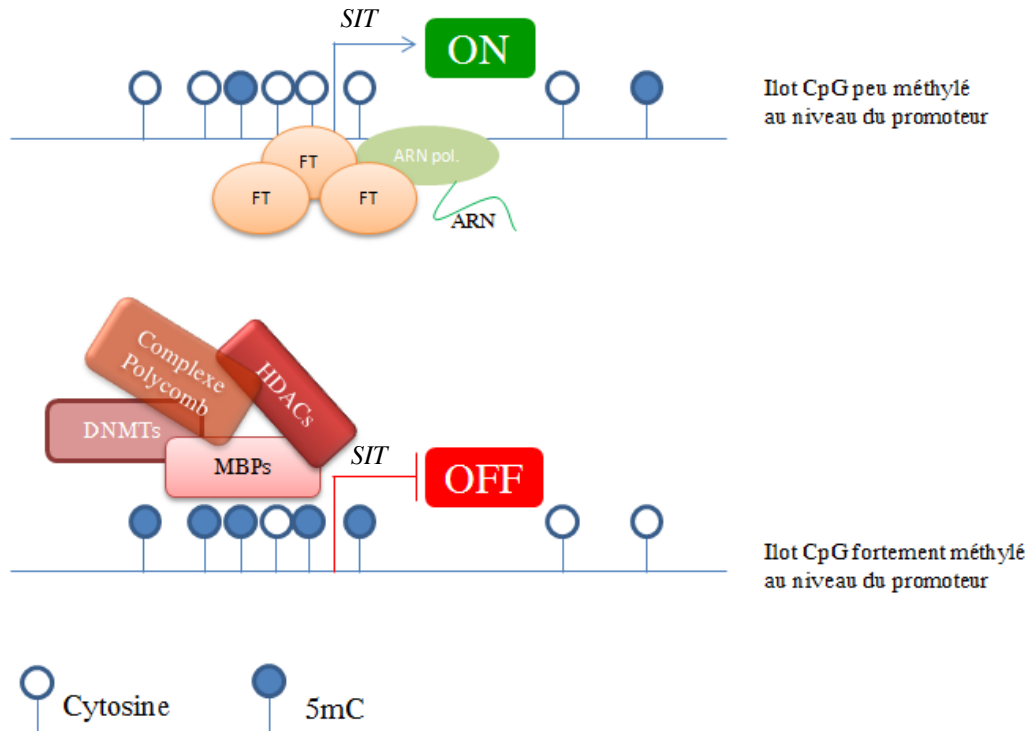


Fig. 7 : Inhibition de la transcription par la méthylation.

Les îlots CpG peu méthylés au niveau du promoteur permettent le recrutement de facteurs de transcription (FT) et de l'ARN polymérase. Ils sont associés à une transcription active. La méthylation des îlots CpG réprime l'expression par le recrutement de complexes inhibiteurs. SIT : Site d'initiation de la transcription, HDAC : désacétylase d'histone.

Ajoutons que les nouvelles techniques permettant d'appréhender le profil de méthylation à l'échelle du génome (telles que le traitement au bisulfite suivi de séquençage ou la technologie de puces « Infinium ») ont montré un enrichissement de la méthylation dans le corps des gènes exprimés (Jones, 2012; Breiling and Lyko, 2015). Ce serait donc au niveau des promoteurs et des sites d'initiation de la transcription que la méthylation participe à la répression transcriptionnelle. Au niveau du corps du gène, la méthylation influencerait l'épissage des transcrits (Jones, 2012).

1.4.1.1.3 Implications pathologiques

Il est désormais clairement établi qu'une méthylation aberrante participe au développement de nombreuses pathologies, telles que les maladies neurodégénératives (Alzheimer, Parkinson, ataxie de Friedreich, ...), les maladies immunitaires (syndrome ICF (*immunodeficiency, centromeric instability and facial anomalies*), lupus érythémateux, ...) et le cancer (Portela and Esteller, 2010). Dans la suite de ce sous-chapitre, nous nous focaliserons uniquement sur le lien entre méthylation et cancer.

Les cellules cancéreuses sont caractérisées par une perte globale de la méthylation de l'ADN et par une hyperméthylation localisée de gènes suppresseurs de tumeurs. L'hypométhylation globale se produit principalement dans les régions répétées et augmente l'instabilité chromosomique. De plus, les éléments transposables sont réactivés et peuvent s'intégrer dans le génome de manière aléatoire. Une hypométhylation localisée est également décrite sur certains promoteurs de gènes pro-cancéreux, tels que le gène R-RAS dans le cancer gastrique et le gène SNCG dans le cancer du sein (Portela and Esteller, 2010). Ajoutons que le génome cancéreux est caractérisé par une hyperméthylation localisée et donc l'extinction transcriptionnelle de gènes suppresseurs de tumeurs, incluant p16 (aussi connu sous le nom CDKN2A), MLH1 et BRCA1 (Fig. 8) (Jones, 2012). D'une manière intéressante, plusieurs études ont démontré le rôle diagnostique et pronostic de la méthylation dans l'évolution de différents cancers (Zhuang et al., 2012; Dedeurwaerder et al., 2011). La détection de l'hyperméthylation du gène GSTP1 dans les urines permet, par exemple, d'identifier les patients avec un cancer de la prostate (Woodson et al., 2008).

Enfin, les DNMTs peuvent également être la cible de mutations, qui les rendent non fonctionnelles et produisent un profil de méthylation aberrant. Notons que la méthylation des cytosines augmente le taux de mutation de transition $C \rightarrow T$, ce qui serait responsable de nombreuses maladies génétiques (Jones, 2012).

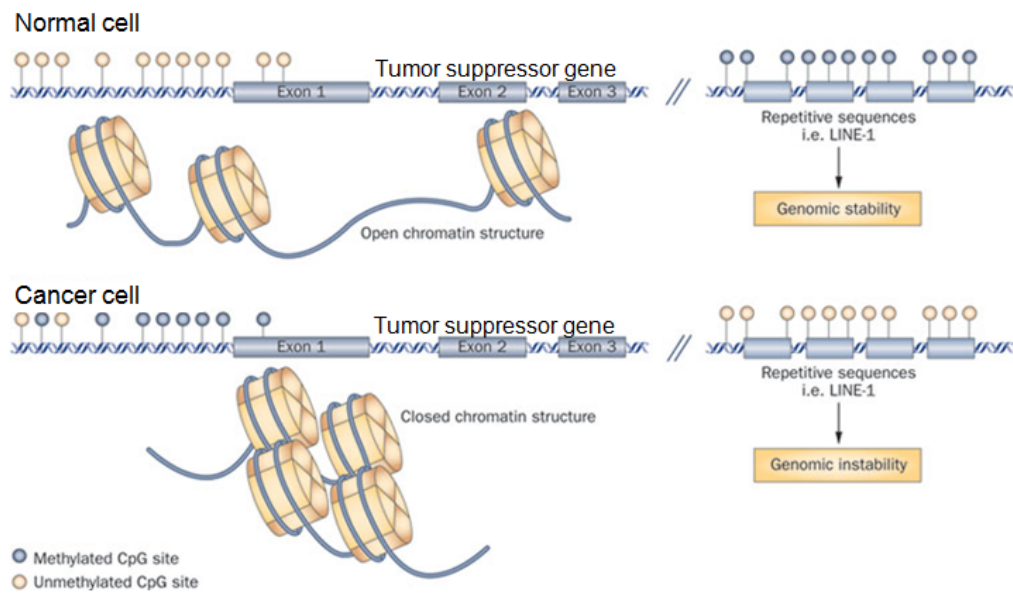


Fig. 8 : Méthylation aberrante et cancer.

Dans les cellules cancéreuses (bas), une hyperméthylation localisée (à gauche), qui inhibe les gènes suppresseurs de tumeurs, ainsi qu'une hypométhylation globale (à droite), qui induit l'instabilité du génome, sont caractéristiques. (Adapté de Lao and Grady, 2011)

1.4.1.2 Autres modifications covalentes de l'ADN

1.4.1.2.1 L'hydroxyméthylation des cytosines

En 2009, deux groupes décrivent une nouvelle modification de l'ADN : l'hydroxyméthylation des cytosines (5hmC) (Tahiliani et al., 2009; Kriaucionis and Heintz, 2009), que certains qualifient de 6^{ème} base de l'ADN. Le niveau d'hydroxyméthylation de l'ADN est estimé à 0.1% des cytosines, cependant de fortes variations existent d'un tissu à l'autre (Breiling and Lyko, 2015). C'est la famille des protéines TETs (*Ten-eleven translocation methylcytosine dioxygenase*), au nombre de trois chez l'homme, qui catalyse l'oxydation des 5mC pour produire les 5hmC. Les protéines TETs peuvent encore oxyder les 5hmC pour produire, de manière itérative, la 5-formylcytosine (5fC) et la 5-carboxylcytosine (5caC) (Delatte, Deplus, and Fuks, 2014). Les protéines TETs jouent un rôle dans la déméthylation de l'ADN, soit par la dilution passive des 5hmC lors de la réplication de l'ADN, soit activement via l'action d'enzymes de

réparation de l'ADN: les glycosylases TDG (thymine DNA glycosylase). Celles-ci peuvent enlever les 5fC et les 5caC et les remplacer par une cytosine non modifiée via l'action de la machinerie BER (*base excision repair*). Alternativement 5hmC peut être désaminée en 5-hydroxyméthyluridine (5hmU) par les enzymes AID/APOBEC. 5hmU est prise en charge par les enzymes SMUG1 et la machinerie BER pour déméthyle l'ADN (Fig. 9A) (Delatte, Deplus, and Fuks, 2014).

Cependant, les 5hmC semblent également correspondre à une marque épigénétique stable qui influence l'expression des gènes. De nombreuses protéines (MeCP2, MBD1, ...) peuvent se lier aux marques 5hmc et les « interpréter » pour réguler l'expression génique. De plus, la présence de protéines TETs sur le génome, indépendamment de leur fonction oxydative, influence la conformation de la chromatine via l'interaction avec des complexes activateurs (OGT et SET1/COMPASS) ou répresseurs (EZH2, SIN3A et NuRD) (Fig. 9B) (Delatte, Deplus, and Fuks, 2014).

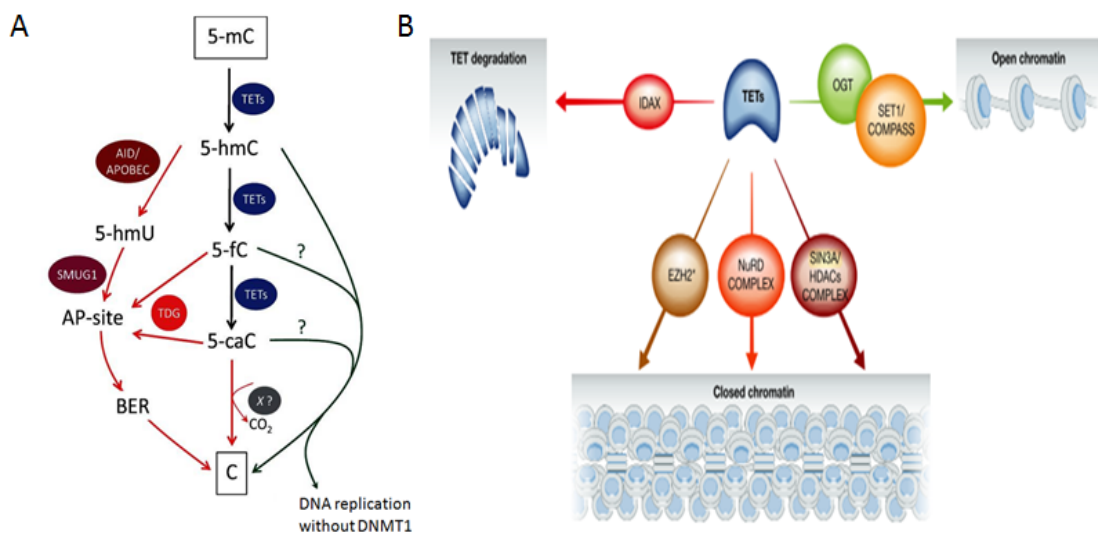


Fig. 9 : Fonctions des protéines TETs.

(A) Les protéines TETs peuvent déméthyle les cytosines d'une manière passive (en vert) ou active (en rouge). 5mC peut être oxydée par TET en 5hmC, 5fC et 5caC. DNMT1 a peu d'affinité pour ces modifications, ce qui produit une dilution progressive de la marque. Le complexe AID/APOBEC désamine les 5hmC pour produire des 5hmU, qui peuvent être reconnues par la glycosylase SMUG1. SMUG1 supprime les bases 5hmU et forme un site apurinique/aprimidique (AP) pris en charge par la machinerie BER. Alternativement, Les enzymes glycosylases TDG reconnaissent 5fC et 5caC pour former un site AP pris en charge par BER. Enfin, 5caC pourrait être directement décarboxylée par une enzyme en-

core méconnue appelée X. (B) Différents complexes protéiques interagissent avec les protéines TETs pour moduler la conformation de la chromatine. (Adapté de Delatte, Deplus, and Fuks, 2014; Delatte and Fuks, 2013)

Ajoutons que les protéines TETs sont impliquées dans de nombreux cancers tels que les leucémies, les mélanomes et le cancer du sein. Elles doivent donc être minutieusement régulées par leurs partenaires protéiques, pour cibler et modifier des régions spécifiques de l'ADN.

1.4.1.2.2 La méthylation des adénines.

Pour clore cette partie sur les modifications covalentes de l'ADN, nous évoquons brièvement la méthylation des adénines. Alors que la méthylation des cytosines a été fortement étudiée au cours des dernières décennies, peu de choses sont connues sur la méthylation des adénines chez les mammifères. En 2015, un groupe a décrit la distribution des N⁶-méthyladénosines (6mA) sur l'ADN murin (Koziol et al., 2015). Il en résulte que la marque 6mA est peu abondante (0.00009% des adénines), semble spécifique de chaque tissu et est enrichie dans des régions dépourvues de gènes. Cependant, le rôle fonctionnel de la méthylation des adénines reste à élucider. Notons que chez les bactéries, la méthylation des adénines par les méthylases Dam est bien décrite, et est associée à la protection de l'ADN contre des endonucléases lors d'infection par des bactériophages (système de défense immunitaire), à la réplication, la transcription et aux systèmes de réparation de l'ADN (Koziol et al., 2015).

1.4.2 Les modifications d'histones

Pour être complet dans cette première partie de l'introduction consacrée à l'épigénétique au sens large, ajoutons que parallèlement aux modifications de l'ADN, l'épigénome est constitué d'un ensemble de modifications post-traductionnelles des queues amino-terminales des histones.

Ce sous-chapitre sera volontairement bref car les modifications d'histones ne constituent pas l'objet principal de ce travail de thèse. Cependant, nous consacrerons une partie plus importante aux déméthylases d'histones de la

famille KDM5, que nous avons étudiée plus avant lors de ce travail de thèse (voir Résultat – Chapitre 1).

1.4.2.1 Modifications post-traductionnelles des histones

Plus de 100 modifications post-traductionnelles d’histones ont été décrites incluant l’acétylation, la méthylation, la phosphorylation et l’ubiquitination qui ensemble forment un « code histone » (Fig. 10) (Bernstein, Meissner, and Lander, 2007). De nombreuses enzymes peuvent ajouter, lire ou supprimer ces modifications, ce qui démontre le caractère labile de ces modifications. Nous décrirons d’abord l’acétylation des histones, corrélée avec une transcription active, puis la méthylation qui peut être activatrice ou inhibitrice de la transcription et enfin d’autres modifications moins étudiées chez les mammifères.

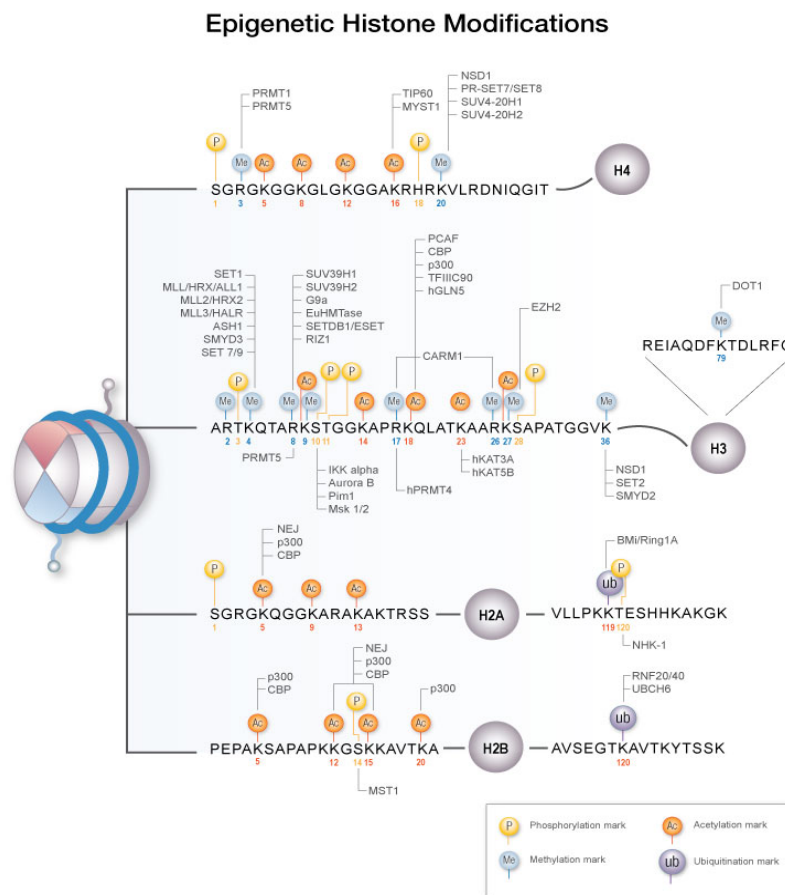


Fig. 10 : Modifications post-traductionnelles d’histones.

Représentation simplifiée de la queue N-terminale des histones et des modifications qui y sont associées. Les protéines responsables des modifications sont également représentées. (d'après *Epigenetic Histone Modifications*, <http://www.epitomics.com/products/pathways/epigenetic-modification.html>)

1.4.2.1.1 Acétylation

L'acétylation des histones est la modification la mieux décrite à ce jour. L'acétylation est dynamique et régulée d'une part par les acétyltransférases d'histones (HATs) et d'autre part par les désacétylases d'histones (HDACs). Les HATs catalysent le transfert d'un groupe acétyle sur les lysines des histones, ce qui neutralise leur charge, et diminue l'interaction entre l'ADN et les histones (Fig. 11). Il existe deux grandes classes de HATs : les HATs de classe A qui contiennent trois familles d'enzymes : GNAT, MYST et CBP/p300, qui acétylent les histones en différentes positions sur la chromatine, alors que les HATs de classe B sont cytoplasmiques et acétylent les histones libres nouvellement synthétisés. Par opposition, quatre grandes classes de HDACs forment différents complexes qui désacétylent les histones et inhibent la transcription (Bannister and Kouzarides, 2011).

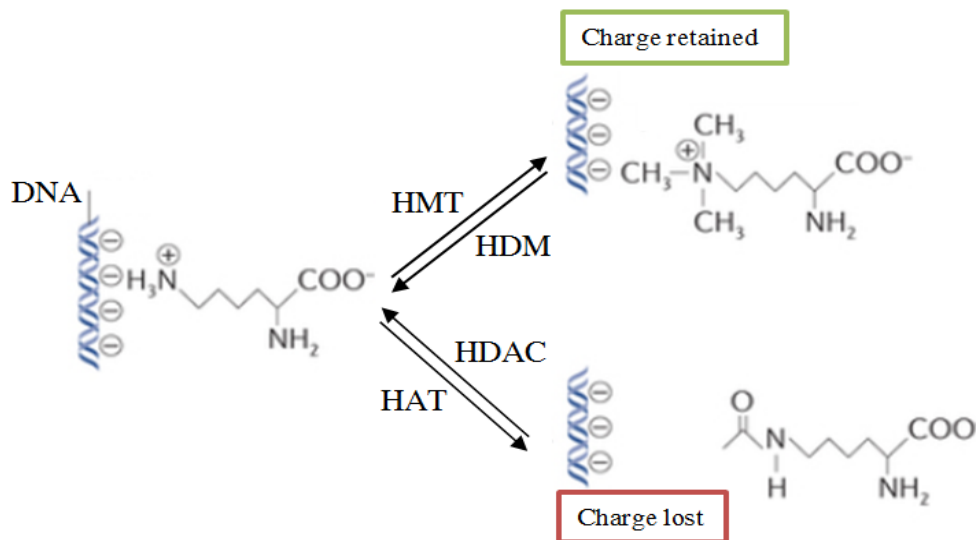
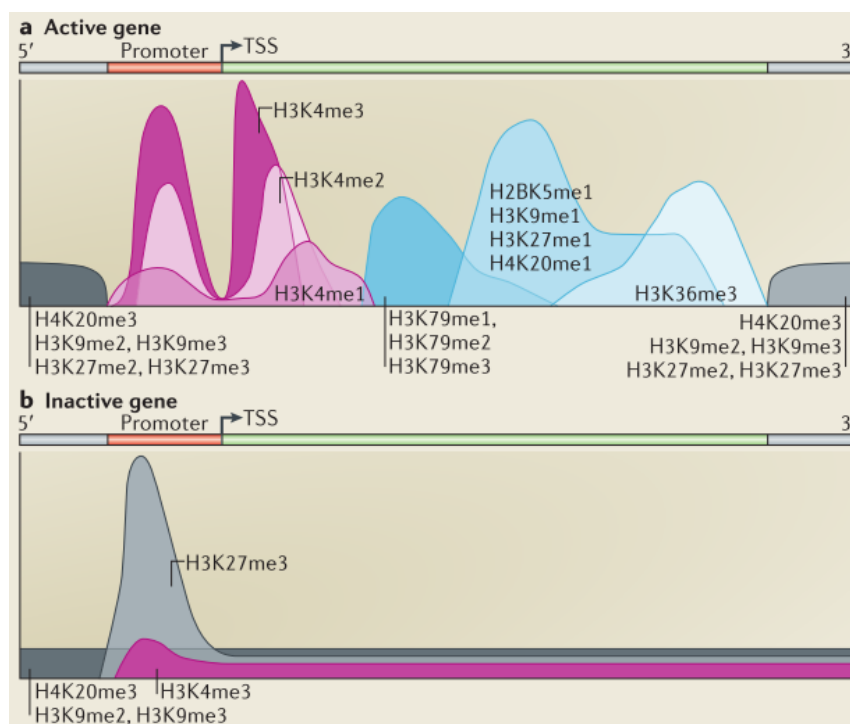


Fig. 11 : Impact des modifications d'histones sur la stabilité des nucléosomes. Les résidus lysines des histones possèdent une charge positive qui interagit avec la charge négative de l'ADN. La méthylation des histones par les HMTs ne modifie pas la charge, alors que l'acétylation par les HATs modifie la charge et diminue la stabilité des nucléosomes. (Figure adaptée de Tessarz and Kouzarides, 2014)

1.4.2.1.2 Méthylation

La méthylation des histones se produit principalement sur les lysines et les arginines et ne neutralise pas les charges des histones (Fig. 11). Différents états de méthylation sont décrits, en effet les résidus peuvent être mono-, bi- ou triméthylés. Les méthyltransférases d'histones (HMTs) sont classées selon l'acide aminé modifié. La méthylation des histones sur les arginines, catalysée par les PRMTs (proteine arginine methyltransférase) est associée à une activation de la transcription alors que la méthylation des lysines par les HKMTs (lysine methyltransferase) peut avoir des fonctions activatrices ou inhibitrices sur la transcription. Par exemple, la méthylation de la lysine sur l'histone H3 (H3K9) est associée à l'inhibition de la transcription alors que la triméthylation de la lysine 4 (H3K4) ou de la lysine 36 (H3K36) de l'histone H3 sont des marques activa-



trices (Fig. 12) (Kooistra and Helin, 2012).

Fig. 12 : Méthylation d'histones sur les lysines.

Les gènes actifs (a) ou inactifs (b) possèdent des marques de méthylation d'histones différentes. (D'après Kooistra and Helin, 2012)

Les HKMTs possèdent un domaine SET (*Su(var)3-9-Enhancer of zeste-Trithorax*) qui est responsable de leur activité catalytique. Parmi les HKMTs, on retrouve le complexe de répression Polycomb 2 (PRC2) souvent dérégulé dans les cancers. PRC2, qui contient la protéine EZH2 (KMT6), est capable de méthyle spécifiquement H3K27 pour former une marque caractéristique de l'hétérochromatine facultative (Bannister and Kouzarides, 2011).

Les déméthylases d'histones (HDMs) forment deux grandes classes : les enzymes amines oxydases de la famille LSD (lysin-specific demethylase), incluant LSD1 et LSD2 et les di-oxygénases de la famille JMJC qui contiennent un domaine jumonji. Les HDMs sont très spécifiques des résidus méthylés et plus de trente HDMs ont été décrites (Kooistra and Helin, 2012). Dans la section suivante, nous décrirons plus en détail la famille des déméthylases KDM5 étudiée au cours de cette thèse.

Les histones déméthylases de la famille KDM5

Les déméthylases de la famille KDM5 peuvent activer ou inhiber la transcription. La famille KDM5, qui déméthyle les histones H3K4 bi- et triméthylés, comprend 4 paralogues : KDM5A (également connue comme JARID1A ou RBP2), KDM5B (JARID1B, PLU1), KDM5C (JARID1C, SMCX) et KDM5D (JARID1D, SMCY). Les différentes KDM5 sont relativement similaires et contiennent toutes un domaine JmjN et un domaine JmjC, un domaine ARID, un domaine C5HC2 en doigt de zinc et deux ou trois domaines PHD (Fig. 13). Elles sont dépendantes de la présence de Fe^{2+} et de 2-oxo-glutarate pour leur fonction déméthylase. Cependant, les enzymes KDM5 ont des fonctions distinctes, d'une part elles ont des profils d'expression très variables et d'autre part elles agissent au sein de complexes protéiques différents (Rotili and Mai, 2011; X. Liu and Secombe, 2015).

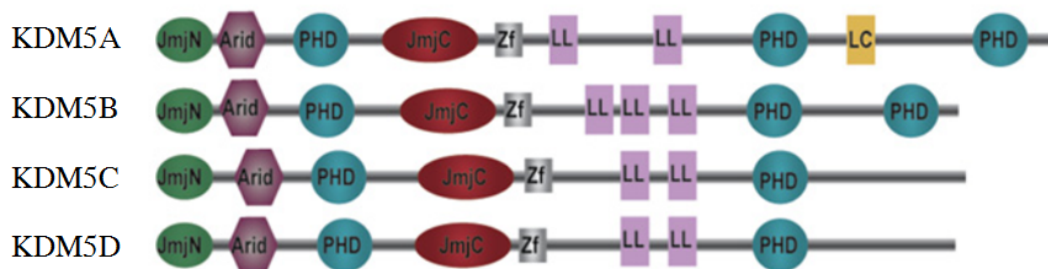


Fig. 13 : Famille des déméthylases KDM5.

Réprésentation schématique des membres de la famille KDM5. ARID, AT-rich interactive domain; JmjN, jumonji N domain; JmjC, jumonji C domain; PHD, plant homeotic domain; Zf, C5HC2 zinc finger domain; LL, motif LXXLL. (D'après Christensen et al., 2007)

KDM5A est exprimée dans de nombreux tissus, et régulerait le cycle circadien, les fonctions mitochondriales et la différenciation en inhibant la voie Notch. KDM5A promeut la croissance cellulaire, diminue l'apoptose et régule l'expression des gènes p21 et p27. Les interactants de KDM5A incluent Sin3, c-Myc, pRB et le complexe PRC2 (Blair et al., 2011; Dimitrova, Turberfield, and Klose, 2015).

KDM5B est presque uniquement exprimée dans le cerveau, les testicules et les ovaires où elle s'associe aux récepteurs aux androgènes et aux œstrogènes pour réprimer ou activer ses gènes cibles. Cependant, son expression augmente significativement dans de nombreux cancers dont le cancer du sein, où elle jouerait un rôle oncogène. A contrario, d'autres études lui adressent une fonction de suppresseur de tumeurs (Klein et al., 2014). KDM5B semble donc avoir des rôles opposés en fonction du type cellulaire. Néanmoins, les mécanismes impliqués dans ses fonctions opposées restent à élucider (Horton et al., 2016; Rotili and Mai, 2011).

KDM5C et KDM5D ont été moins étudiées et sont moins bien documentées. Notons tout de même que KDM5C a été décrite pour influencer la prolifération et avoir un rôle inhibiteur au niveau des promoteurs, alors qu'elle stimule l'activité des activateurs (enhancers) (Outchkourov et al., 2013). KDM5D, quant à elle, s'associe avec le complexe Polycomb-like Ring6a/MBLR pour réguler la transcription en déméthylant les histones H3K4 au niveau des SIT (M. G. Lee et al., 2007).

Ajoutons que les enzymes KDM5 possèdent également un rôle régulateur indépendant de leur activité déméthylase. Elles peuvent, par exemple, augmenter l'acétylation des histones en inhibant des HDACs (DiTacchio et al., 2011).

Les KDM5 sont impliquées dans de nombreuses maladies, incluant des désordres neuroaux et certains cancers (X. Liu and Secombe, 2015). KDM5C a par exemple été associée à l'épilepsie et aux retards mentaux liés au chromo-

some X (Rotili and Mai, 2011). KDM5A et KDM5B sont clairement impliquées dans la tumorigenèse. KDM5A est surexprimée dans le cancer gastrique et son inhibition résulte en l'augmentation des protéines p16, p21 et p27 qui mène à la sénescence (Rotili and Mai, 2011). Un autre exemple que nous avons déjà évoqué implique KDM5B qui est surexprimée dans certains cancers du sein où elle inhibe les gènes suppresseurs de tumeurs BRCA1 et HOXA5. De plus, son inhibition supprime la croissance de tumeurs mammaires en modèle murin (Yamane et al., 2007; H. Denis et al., 2015).

1.4.2.1.3 Autres modifications d'histones

Concluons cette partie en abordant brièvement d'autres modifications d'histones décrites chez les mammifères.

La phosphorylation des histones est une marque hautement dynamique que l'on retrouve sur de nombreux résidus incluant les sérines, les thréonines, les tyrosines, les arginines, les histidines et les lysines. Les histones kinases et les histones phosphatases sont respectivement responsables du dépôt et de la suppression d'un groupe phosphate sur les histones. La phosphorylation modifie la charge des histones et donc influence la structure de la chromatine. La phosphorylation est entre autres impliquée lors de dommages à l'ADN où les kinases ATR/ATM phosphorylent le variant d'histone H2A.X pour faciliter le recrutement d'enzymes de réparation de l'ADN (Bannister and Kouzarides, 2011; Biterge, 2016).

La désimination implique la conversion d'une arginine méthylée en citruline par l'enzyme PADI4. On ne peut cependant pas parler de PADI4 comme d'une « vraie » déméthylase d'histone, car il n'y a pas de régénération d'une arginine non modifiée (Bannister and Kouzarides, 2011).

La mono-ubiquitylation des histones est une grosse modification (ajout d'ubiquitine de 76 acides aminés) qui se produit sur les histones H2A (associée à la répression de la transcription) et sur les histones H2B (associée à l'initiation de la transcription) (Bannister and Kouzarides, 2011).

Pour être complet, citons les autres modifications d'histones décrites jusqu'à présent : la β -N-acétylglucosamination, l'ADP ribosylation, la sumoylation, l'isomérisation des prolines et les modifications des lysines (propionylation, butyrylation, crotonylation, malonylation, succinylation) (H. Huang et al., 2014).

1.4.2.2 Mécanismes du code histone et connexion avec la méthylation de l'ADN

Le code histone fonctionne de deux manières : 1) les modifications d'histones modifient le contact entre histone et ADN ou entre nucléosomes successifs. L'acétylation permet par exemple de dérouler la chromatine, car elle neutralise les charges des lysines. 2) Les modifications permettent le recrutement de protéines non-histones qui modifient la chromatine. Ces protéines de remodelage sont souvent dépendantes de l'énergie d'hydrolyse de l'ATP pour déplacer les nucléosomes. Leur recrutement se fait via des domaines protéiques spécifiques. L'acétylation est reconnue par les bromodomains, la méthylation par des domaines « chromo-like »/MBT et par des domaines PHD et la phosphorylation par un domaine spécifique des protéines 14-3-3 (Fig. 14A). Citons par exemple, la marque H3K4 triméthylée, reconnue par le domaine PHD des protéines BPTF. BPTF, membre du complexe de remodelage NURF, recrute la protéine SNF2L pour activer l'expression de gènes cibles (Fig. 14B) (Kouzarides, 2007).

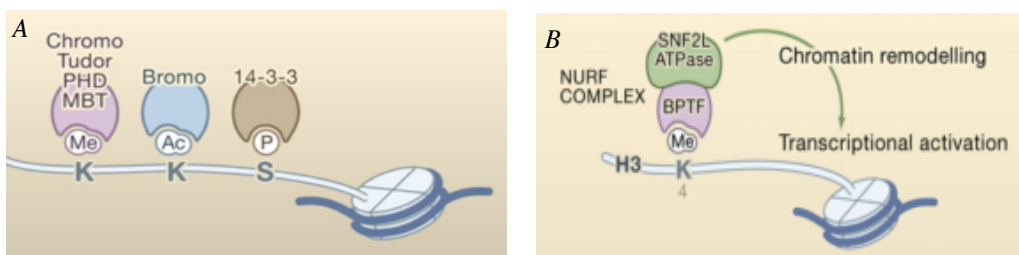


Fig. 14 : Recrutement des protéines par les histones modifiées.

(A) Des domaines protéiques spécifiques sont impliqués dans le recrutement des protéines qui lisent les histones modifiées. (B) La marque H3K4me3 recrute les membres du complexe de remodelage de la chromatine NURF pour activer la transcription. (D'après Kouzarides, 2007)

Les différentes modifications se modulent l'une l'autre pour contrôler l'accessibilité de la chromatine. La marque H3K27me peut, par exemple, recruter la protéine PC2 du complexe Polycomb 2 (PRC2), qui possède une activité de ligase de l'ubiquitine sur les histones H2A. D'autres exemples impliquent la marque H3K9me qui recrute la protéine HP1 qui possède une activité désacétylase et méthyl-transférase, ou encore l'ubiquitination de l'histone H2B qui est requise pour la triméthylation d'H3K4 (Fig. 15) (Kouzarides, 2007).

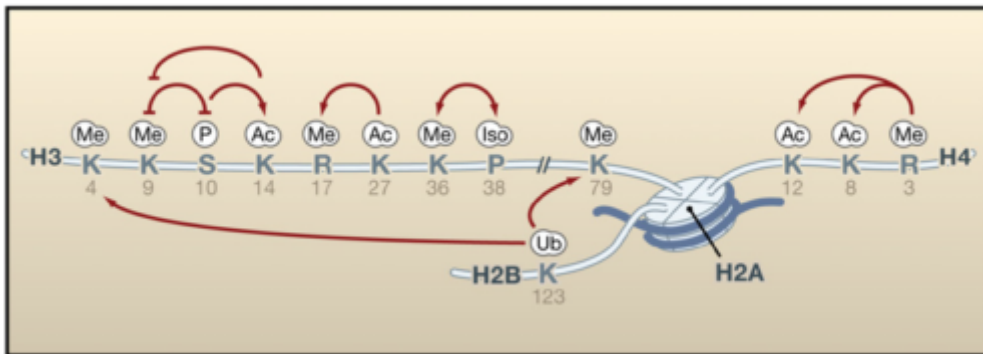


Fig. 15 : Inter-influence des modifications d'histones.

Les différentes modifications d'histones s'influencent l'une l'autre pour générer le code histone. Les influences positives sont montrées par une flèche, les influences négatives par un trait. (D'après Kouzarides, 2007)

D'autre part, la méthylation de l'ADN et les modifications d'histones s'influencent mutuellement. Par exemple, les DNMTs peuvent s'associer avec des HDACs qui désacétylent les histones et des HMTs qui méthylent la lysine 9 de l'histone H3 pour augmenter le phénomène de répression de la transcription (Fuks et al., 2003). Par ailleurs, les DNMTs agissent « main dans la main » avec les protéines EZH2 du complexe PRC2 pour verrouiller l'inhibition transcriptionnelle (Fig. 16) (Viré et al., 2006).

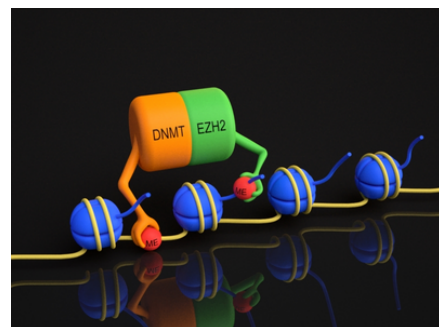
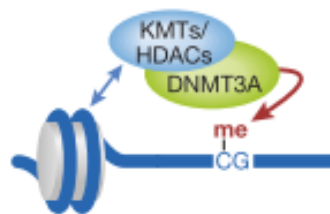


Fig. 16 : Verrouillage de l'inhibition transcriptionnelle.

Gauche : Association des DNMTs et des HDACs/HMTs (KMT : lysine methyl transférase). Droite : Les protéines DNMTs et EZH2 agissent « main dans la main » pour verrouiller la transcription. (D'après Denis, Ndlovu, and Fuks, 2011)

1.5 L'épigénétique en clinique

Pour clore cette première partie de l'introduction, nous nous intéresserons aux applications cliniques associées à l'épigénétique. De part leur caractère réversible, les changements épigénétiques sont des cibles thérapeutiques rationnelles qui font l'objet d'intenses recherches depuis plusieurs années. Cependant il n'y a que six molécules qui ont été approuvées par la FDA (food and drug administration) : les inhibiteurs des méthyltransférases de l'ADN (azacitidine et décitabine), les inhibiteurs des désacétylases d'histones (vorinostat, belinostat et la romidepsin) et les inhibiteurs de JAK1/2 (Janus Kinase) (ruxolitinib). Les protéines JAK1/2, souvent mutées ou amplifiées dans les cancers hématologiques, sont, entre autres, responsables de la phosphorylation de l'histone H3 sur la tyrosine en position 41, menant à l'expression d'oncogènes (Dawson and Kouzarides, 2012; Egger and Arimondo, 2015).

1.5.1 Les inhibiteurs des méthyltransférases de l'ADN

Le profil de méthylation de l'ADN est dérégulé dans de nombreuses pathologies, particulièrement dans les cancers, dû à une activité aberrante des DNMTs. Les inhibiteurs de méthyltransférases de l'ADN sont des analogues des nucléosides cytosines qui entraînent l'hypométhylation des promoteurs et donc la réactivation de gènes suppresseurs de tumeurs. Ils s'incorporent dans l'acide nucléique pendant la phase S du cycle cellulaire et forment un lien covalent irréversible avec les DNMTs. Dès lors, les DNMTs sont bloquées, ce qui déclenche la signalisation de dommage de l'ADN et mène à leur dégradation (Stresemann and Lyko, 2008).

Les inhibiteurs de DNMTs ont été approuvés dans le traitement des syndromes myélo-dysplasiques dont les leucémies aiguës myéloïdes. Cependant leur efficacité est limitée (50% des patients ont une réponse thérapeutique par-

tielle et seuls 15% des patients ont une réponse clinique complète). Par conséquent, des approches pour prédire la réponse au traitement ainsi que la synthèse de nouveaux inhibiteurs plus performants et l'utilisation d'oligonucléotides antisens aux DNMTs sont en cours d'investigation. Enfin, leur utilisation dans d'autres types de tumeurs fait actuellement l'objet de différents essais cliniques (Egger and Arimondo, 2015; Medina-Franco, 2016).

1.5.2 Les inhibiteurs des désacétylases d'histones

Les inhibiteurs d'HDACs (HDACi) augmentent l'acétylation des résidus lysines sur les histones, ce qui diminue le degré de compaction de la chromatine et permet le recrutement de différents facteurs de transcription et donc la réactivation de gènes suppresseurs de tumeur. Citons par exemple le gène CDKN1A encodant la protéine p21, qui peut inhiber l'activité des complexes cyclines/CDK et enrayer la prolifération cellulaire (Gray, 2015).

Les HDACi sont actuellement utilisés dans le traitement des lymphomes T cutanés. Cependant, ils sont peu sélectifs (par exemple : le vorinostat est un inhibiteur à large spectre qui cible les HDACs de classe I à IV et la romidepsin cible toutes les HDACs de classe I) et amènent de nombreux effets secondaires. Des inhibiteurs de seconde génération, plus sélectifs, sont en cours d'essais cliniques dans le traitement de différents cancers, dont le cancer du sein (Arrowsmith et al., 2012).

1.5.3 Thérapies épigénétiques en développement

Les thérapies épigénétiques pionnières ayant démontré leur potentiel thérapeutique, un nombre grandissant de nouvelles drogues épigénétiques sont à l'étude. En effet, outre les essais cliniques portant sur l'étude de nouveaux inhibiteurs des DNMTs et des HDACs, de nombreuses thérapies épigénétiques voient le jour, incluant les inhibiteurs des HATs, des HMTs, des HDMs et d'autres protéines épigénétiques telles que EZH2 (Gray, 2015). A titre d'exemple, deux inhibiteurs sont décrits plus en détails.

La protéine LSD1 déméthylant les histones H3 en position 4 est surexprimée dans le cancer du poumon. Des études « in vitro » ont montré que l'inhibition de LSD1 par le système d'interférence à l'ARN réduit la prolifération et la migration de cellules pulmonaires cancéreuses. Un inhibiteur de LSD1, appelé pargyline, mimique ces effets in vitro, et démontre le potentiel de LSD1 comme cible épigénétique (Huffman and Martinez, 2013).

La protéine EZH2, qui fait partie du complexe Polycomb 2 (responsable de la marque répressive H3K27me3), est surexprimée dans les cancers du sein, de la prostate et du cou. Un inhibiteur d'EZH2, dénommé 3-deazaneplanicon A (DZNep), induit la différenciation et l'apoptose en modèle cellulaire. Ces résultats précliniques intéressants doivent encore être confirmés en clinique (Huffman and Martinez, 2013).

Enfin, le futur thérapeutique repose plus que probablement sur la combinaison de différents agents tels que des agents de chimiothérapie, d'immunothérapie et de thérapies épigénétiques. Il a en effet été démontré que, dans un modèle murin de cancer du sein et de cancer colorectal, une immunothérapie (anticorps anti-PD-1 et anti-CTLA-4) isolée ne donne pas de résultat encourageant, alors qu'en combinaison avec deux drogues épigénétiques (HDACi et inhibiteur de DNMTs), les résultats thérapeutiques sont spectaculaires (éradication de la tumeur primaire chez 10 souris sur 11) (Kim et al., 2014; Weintraub, 2016). Des résultats similaires ont également été montrés dans le traitement des leucémies prolymphocytiques à cellules T (Hasanali et al., 2015). Des nombreux essais précliniques et cliniques portant sur la combinaison d'immunothérapies et de thérapies épigénétiques sont en cours pour divers cancers (leucémies, mélanomes, cancers colorectaux, pulmonaires et rénaux) (Maio et al., 2015).

2 Les ARNs non codants

2.1 Introduction

Parallèlement aux modifications d'histones et à la méthylation de l'ADN, un autre mécanisme épigénétique implique le vaste monde des ARNs non codants (ncRNAs). L'implication des ncRNAs dans la régulation des gènes avait déjà été évoquée en 1961 par Jacob et Monod (Jacob and Monod, 1961). Mais leur étude a été délaissée pendant de nombreuses années. En 1993, la découverte des micro-ARNs (miR) chez *C. Elegans* (R. C. Lee, Feinbaum, and Ambros, 1993) et la description de *let-7*, le premier micro-ARN (miR) chez l'homme en l'an 2000, relancent l'intérêt pour les ncRNAs (Pasquinelli et al., 2000). Ces dernières années, l'émergence des nouvelles technologies de séquençage a démontré, qu'alors que deux pourcents de notre génome codent pour les 23 000 protéines connues à ce jour, plus de 80% du génome est transcrit en ncRNAs (Djebali et al., 2012; Hangauer, Vaughn, and McManus, 2013). Dès lors, l'attrait pour les ncRNAs n'a cessé de croître, afin d'identifier le rôle de ces nombreux transcrits non codants (Fig. 17).

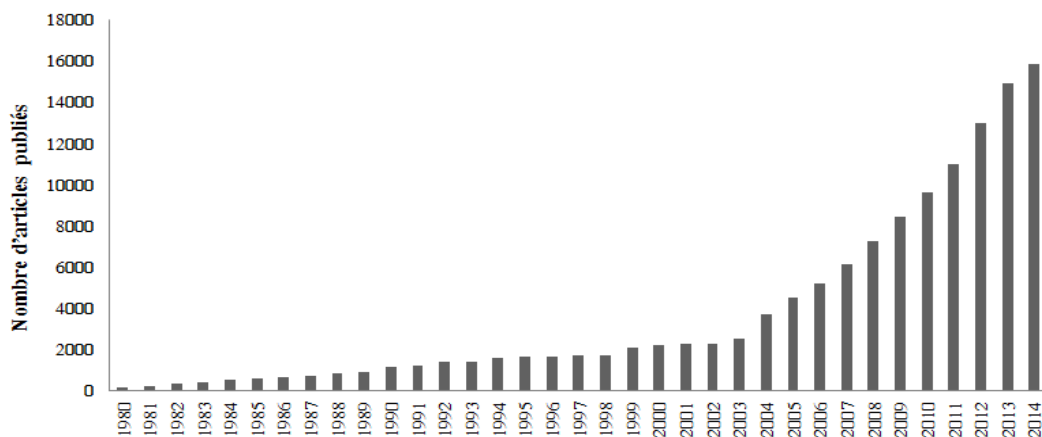
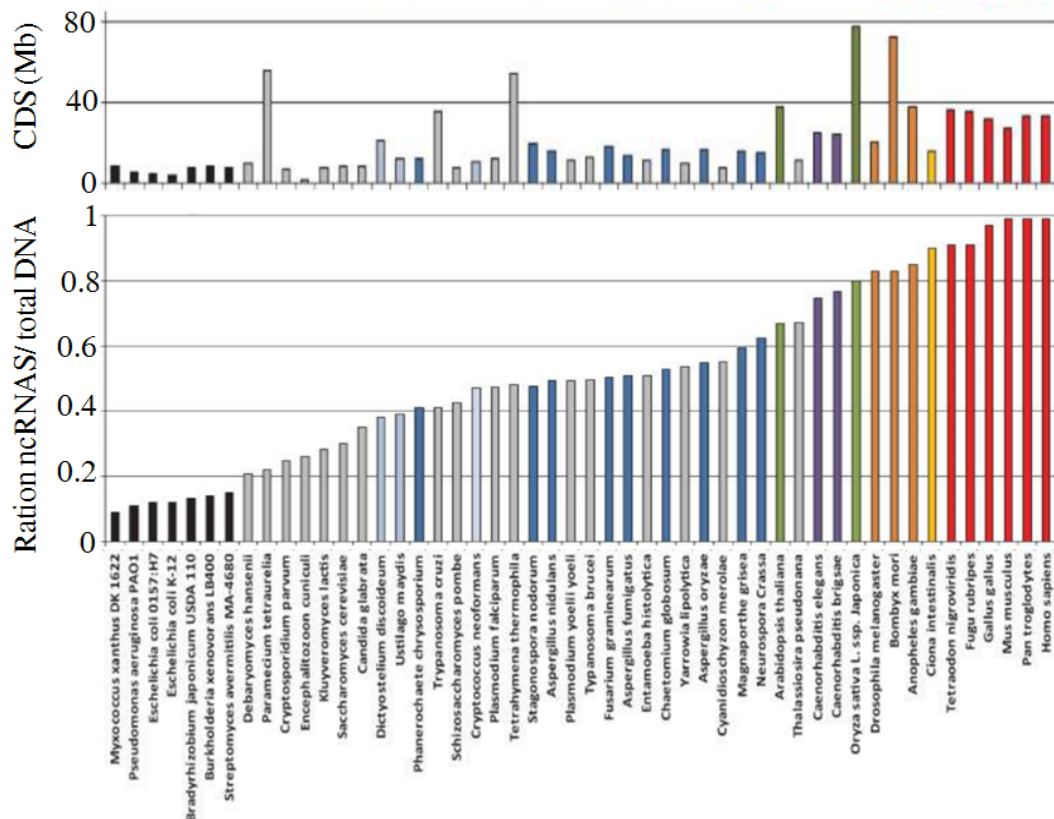


Fig. 17 : Littérature sur les ARNs non codants.

Evolution du nombre d'articles référencés sur Pubmed par les mots « non coding RNA » entre 1980 et 2014.

D'une manière intéressante, on ne retrouve pas de corrélation entre la taille du génome ou le nombre de protéines d'un organisme et sa complexité. Les ncRNAs pourraient expliquer ce paradoxe. En effet, une expansion des séquences ne codant pas pour des protéines est associée avec le degré de complexité d'un organisme (Fig. 18). Ceci suggère l'implication fascinante des



ARNs non codants dans l'évolution (Taft, Pheasant, and Mattick, 2007).

Fig. 18 : Génome, ARNs non codants et complexité.

(haut) Quantité de séquences codantes (CDS) exprimées en mégabase (Mb) dans différents organismes. (bas) Ratio (ordonné) entre le nombre de bases des ncRNAs et le nombre de bases total du génome dans différents organismes. Les procaryotes sont en noir, les unicellulaires stricts en gris, les unicellulaires facultatifs en bleu clair, les organismes multicellulaires basiques en bleu foncé, les plantes en vert, les nématodes en mauve, les arthropodes en orange, les cordés en jaune et les vertébrés en rouge. (D'après Taft, Pheasant, and Mattick, 2007)

2.1.1 Classification des ARNs non codants

Les ARNs non codants sont classés selon leur taille (petit ou long), leur fonction (ARN ribosomale, ARN de transfert, ...) et leur localisation génomique et cellulaire (les petits ARNs nucléaires et nucléolaires par exemple). Pour simplifier, nous avons décidé de classer les ncRNAs d'abord selon leur taille et ensuite selon leur fonction.

Les petits ncRNAs font moins de 200 paires de bases et comprennent : les ARNs ribosomiaux (rRNAs) et les ARNs de transfert (tRNAs) impliqués dans la synthèse des protéines; les petits ARN nucléaires (snRNAs) impliqués dans l'épissage et la poly-adénylation; les petits ARNs nucléolaires (snoRNAs) impliqués dans les modifications de l'ARN; les ARNs interagissant avec les protéines Piwi (piRNAs) qui sont impliqués dans l'inhibition des transposons dans les cellules germinales; ainsi que les micro-ARNs (miRs), acteurs majeurs du système d'interférence par l'ARN (RNAi), qui se lient à des ARNs messagers pour empêcher leur traduction (voir chapitre 2.2). D'autres sous-classifications plus détaillées existent, mais sortent du cadre de ce travail (Dozmorov et al., 2013).

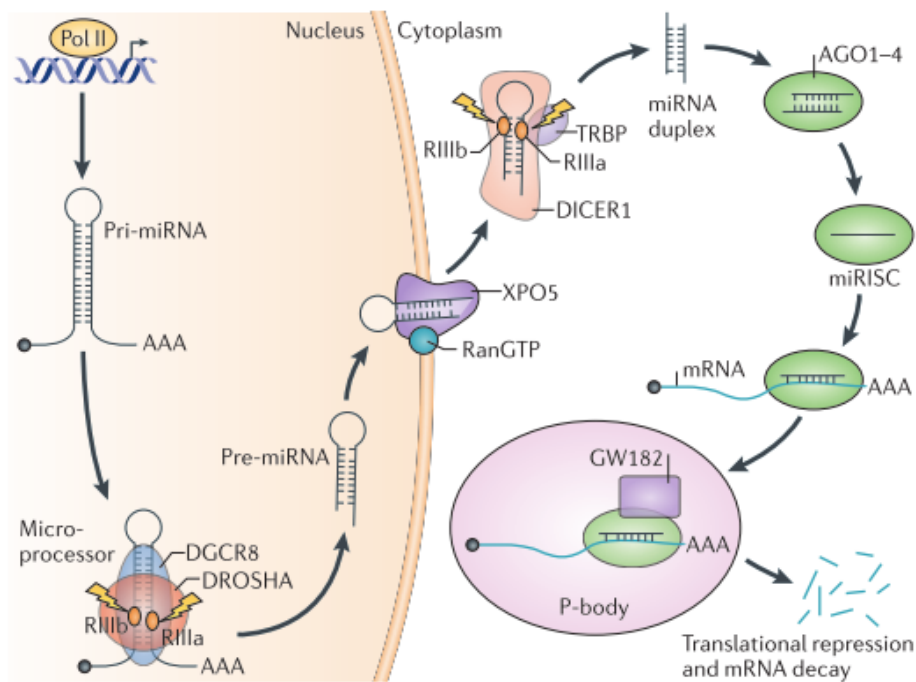
Les analyses transcriptionnelles à l'échelle du génome ont mené à la découverte de dizaines voire d'une centaine de milliers de longs ARNs non codants (lncRNAs) dont la taille est supérieure à 200 paires de bases (Derrien et al., 2012; Volders et al., 2013). Les lncRNAs peuvent être situés dans des régions dépourvues de gènes (inter-géniques) ou à proximité des gènes codants (anti-sens, introniques, exoniques ou chevauchants). Ils sont impliqués dans pratiquement tous les systèmes biologiques en régulant différents mécanismes tels que la transcription, la formation de complexes protéiques et la transduction du signal inter-cellulaire (voir chapitre 2.3).

Dans la suite, nous nous focaliserons sur les deux types de ncRNAs que nous avons étudiés au cours de cette thèse : à savoir les miRs et les lncRNAs.

2.2 Les micro-ARNs (miRs)

2.2.1 Biogenèse des miRs

De loin les plus étudiés des petits ARNs non codants, les miRs se distinguent par la conformation en épingle à cheveux de leur précurseur primaire : les pri-miRs. Les pri-miRs, généralement transcrits par l'ARN polymérase II, sont épissés et possèdent une coiffe N-7-méthylguanine (m7G) en 5' et une queue poly-A en 3'. Leur structure typique est composée d'une région simple brin et d'une région double brin qui forme une boucle. Dans le noyau, les pri-miRs sont clivés par le complexe « microprocesseur » qui comprend DROSHA (une ribonucléase de type III (RNase III)) et DGCR8 qui permet la fixation de DROSHA au pri-miRNA (Fig. 19). Une fois fixée, DROSHA, qui contient deux domaines RNase III, clive les pri-miRs pour former un autre précurseur plus court, appelé pre-



miR.

Fig. 19 : Biogenèse des miRs.

Dans le noyau, les pri-miRs sont clivés par le complexe « microprocesseur » pour former les pre-miRs d'une longueur de 60 à 70 nucléotides, qui sont exportés dans le cytoplasme où ils sont pris en charge par DICER1 pour former un duplex de miRs matures. Au sein du

complexe miRISC, un brin du miR mature dirige le complexe vers sa cible pour dégrader l'ARNm et inhiber la traduction. (D'après Lin and Gregory, 2015)

Les pre-miRs sont exportés dans le cytoplasme par les exportines 5 (XPO5), où ils sont pris en charge par une autre RNase III nommée DICER1. DICER1, en association avec la protéine TRBP, clive les pre-miRs pour former un duplex de miRs matures d'une longueur de 22 paires de bases. Les protéines TRBP-DICER1, en association avec les miRs, se lient alors aux protéines Argonaute (AGO1, AGO2, AGO3 ou AGO4) pour former le complexe effecteur de silençage miRISC (miRNA-induced silencing complex, également connu sous miRNP). Dans le complexe miRISC, un brin du duplex miR est éliminé, et l'autre brin, appelé brin guide, (Ha and Kim, 2014; Lin and Gregory, 2015) va permettre l'association du complexe miRISC et des protéines GW182 sur l'ARNm cible. Ce phénomène se produit au sein de structures appelées p-bodies (processing bodies) (Fig. 19).

Notons l'existence d'autres mécanismes de génération de miRs, indépendants de DROSHA et de DGCR8. Ils sont soit basés sur l'épissage de l'ARN et la production de mirtons à partir d'introns, soit dérivés d'autres sortes de ncRNAs telles que les lncRNAs (Ha and Kim, 2014).

A ce jour, plus de 1800 miRs ont été découverts chez l'homme (Kozomara and Griffiths-Jones, 2014).

2.2.2 Mode d'action des miRs

Au cours d'un processus appelé interférence de l'ARN, les miRs répriment l'expression protéique en se liant à une séquence complémentaire des ARNs messagers (mRNAs), la plupart du temps sur la région non traduite en 3' (3'-UTR). Différents mécanismes ont été proposés et résumés dans (Orang, Safaralizadeh, and Kazemzadeh-Bavili, 2014) :

Inhibition pré-traductionnelle

L'inhibition de l'initiation de la traduction par les miRs peut résulter de la désa-dénylation du côté 3' ou par la suppression de la coiffe en 5' du mRNA, qui sera alors exposé à la dégradation par des exonucléases telles que Xrn1p ou des endonucléases telles que PMR1 (Fig. 20a). Parallèlement, les protéines Argonautes, au sein du complexe miRISC, inhibent le recrutement du facteur d'initiation de la traduction eIF4E, et modifient la conformation des mRNAs par compétition avec les protéines PABP (poly-A binding protein) et CBP (cap binding protein) (Fig. 20b). Enfin, le complexe miRISC et les protéines GW82 peuvent séquestrer le mRNA, le rendre temporairement inaccessible à la machinerie de traduction (Fig. 20c).

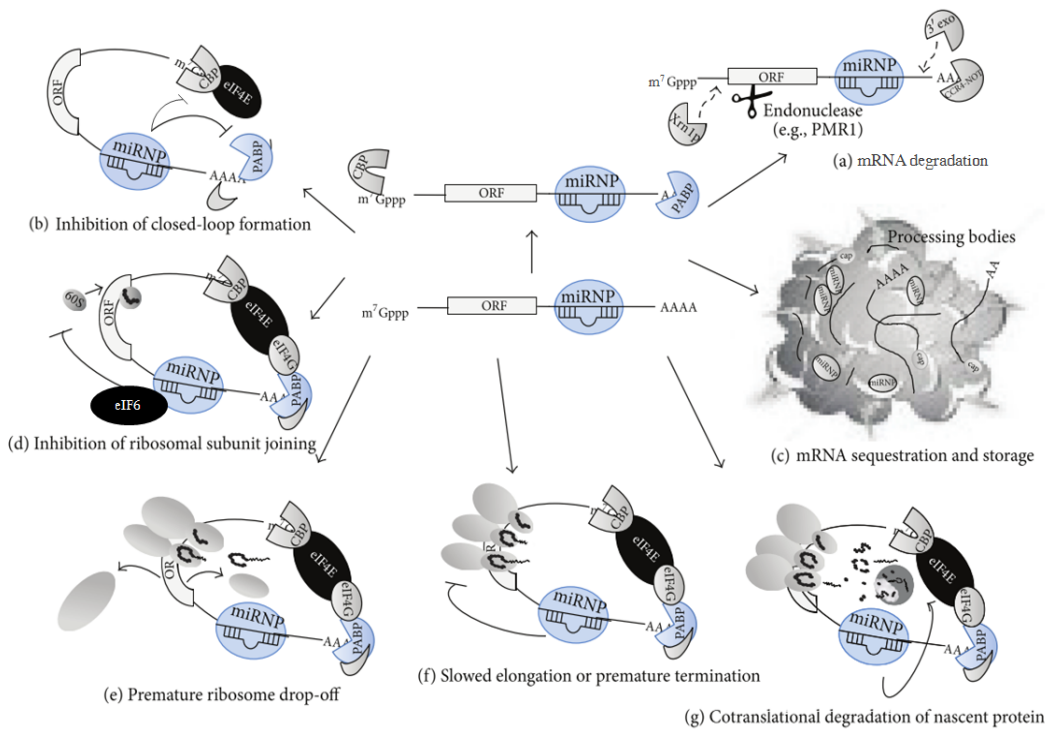


Fig. 20 : Mécanismes d'inhibition par les miRs.

Le complexe miRISC (nommé ici miRNP) cible son mRNA d'intérêt par complémentarité incomplète de séquences. (a) Déstabilisation et dégradation du mRNA par les nucléases. (b) Le complexe miRISC modifie la conformation du mRNA. (c) Les mRNAs peuvent être séquestrés et stockés au sein des p-bodies en vue d'une traduction ultérieure. (d) Le complexe miRISC et eIF6 empêchent la grande sous-unité ribosomiale 60S de s'associer avec la petite sous-unité 40S. (e) Le complexe miRISC dissocie les ribosomes. (f) Compétition du complexe miRISC avec les facteurs d'élongation qui inhibent l'élongation. (g) Recrutement de protéases qui dégradent les polypeptides nouvellement synthétisés. (Adapté d'après Orang, Safaralizadeh, and Kazemzadeh-Bavili, 2014)

Inhibition *co*-traductionnelle

Les protéines Argonautes recrutent les protéines eIF6, ce qui empêche la grande sous-unité ribosomiale de se lier à l'ARNm (Fig. 20d). De plus, le complexe miRISC promeut la dissociation ribosomiale (Fig. 20e) et entre en compétition avec les facteurs d'élongation pour ralentir ou terminer prématurément la traduction (Fig. 20f).

Inhibition *post*-traductionnelle

Le complexe miRISC permet le recrutement de protéases qui dégradent la protéine naissante (Fig. 20g).

Considérons qu'il y a plus de 1800 miRs, que chaque miR peut cibler des centaines de mRNAs, et que chaque mRNA peut être ciblé par différents miRs, nous comprenons dès lors le rôle majeur et complexe joué par les miRs dans le contrôle de la production de protéines (Lin and Gregory, 2015).

2.2.3 Fonctions des miRs dans le développement tumoral

De nombreuses études ont démontré l'implication des miRs dans les cancers. Les miRs peuvent être onco-suppresseurs ou au contraire activer le développement tumoral (onco-miRs). Nous aborderons d'abord l'expression des miRs dans les cancers, puis nous nous focaliserons sur quelques exemples bien connus d'onco-miRs et de miRs suppresseurs de tumeurs. Enfin, nous décrirons les miR-137 et miR-138 que nous avons étudiés dans ce travail.

2.2.3.1 Profil d'expression des miRs dans le cancer

Alors que certains miRs sont spécifiquement surexprimés dans les tumeurs, les miRs sont globalement sous-exprimés dans les cancers, probablement car leur biogenèse est altérée (Lin and Gregory, 2015). En effet, la machinerie de maturation des miRs (DROSHA, DICER) est régulièrement mutée et/ou réprimée dans les cancers. Cette inhibition globale des miRs mène au développe-

ment rapide de tumeurs invasives (Kumar et al., 2007) et est associée à un mauvais pronostic vital. De plus, de nombreux miRs sont situés sur des régions fragiles du génome, souvent délétées, amplifiées, mutées ou transloquées dans les cancers.

L'expression des miRs est également régulée par des facteurs de transcription spécifiques. Citons par exemple, la protéine p53, qui active la transcription de la famille des miRs-34. Ces miRs sont impliqués dans l'inhibition de facteurs pro-mitotiques. Dans les cancers, on observe une perte fonctionnelle de p53 accompagnée d'une suppression des miRs-34 résultant en une prolifération cellulaire anarchique (Lin and Gregory, 2015). D'autre part, les onco-protéines MYC peuvent activer l'expression de miRs oncogéniques tels que les miRs du cluster miR-17-92, qui activent le cycle cellulaire et l'angiogenèse.

L'expression des miRs est aussi régulée par l'état de méthylation de leur ADN. Ainsi, le promoteur de nombreux miRs onco-suppressifs tels que miR-148a, miR-34b et miR-9, est hyperméthylé dans les tumeurs, menant à leur inhibition transcriptionnelle (Lin and Gregory, 2015).

Chaque type de cancer possède une signature spécifique de l'expression des miRs, appelée miRNome. Le miRNome est associé avec l'origine et l'état d'avancement de la maladie, ainsi qu'avec son pronostic. De plus, le profil d'expression des miRs permet l'identification de sous-types de tumeurs, comme par exemple dans le cancer du sein ou les leucémies lymphoblastiques aiguës (Di Leva, Garofalo, and Croce, 2014; Reddy, 2015).

2.2.3.2 Exemples de miRs oncogènes (onco-miRs)

Le cluster miR-17-92

Le cluster miR-17-92, qui contient 7 miRs homologues est fréquemment amplifié dans les tumeurs hématopoïétiques. Ce cluster de miRs, activé par l'onco-protéine MYC, réprime le facteur de transcription E2F et les protéines p21 et PTEN. En conséquence, le cycle cellulaire s'accélère et l'apoptose est inhibée (Reddy, 2015).

MiR-21

MiR-21 est surexprimé dans la majorité des tumeurs solides, incluant le cancer du sein. Ce miR régule de nombreux gènes tels que p53, p21, TGFB et PTEN. Le rôle de miR-21 dans le contrôle du cycle cellulaire, de la migration et de l'invasion est clairement établi dans de nombreux modèles tumoraux. Il s'agit donc d'un acteur majeur dans l'apparition de métastases (Reddy, 2015).

2.2.3.3 Exemples de miRs onco-suppresseurs

Le cluster let-7

Le cluster let-7, composé de 12 miRs, est situé dans des régions fragiles du génome et souvent sous-exprimé dans les tumeurs. Le cluster let-7 est impliqué dans la régulation d'oncogènes incluant RAS ou MYC. Il cible également des acteurs du cycle cellulaire tels que CDC25A, CDK6 et la cycline D2 pour inhiber la progression de la phase G1 vers la phase S (Malumbres, 2013). Dans les cellules mammaires, let-7b cible les gènes PAK1, DIAPH2, RDX et ITGB8. Ces gènes sont impliqués dans la régulation du cytosquelette d'actine, et leur inhibition diminue le potentiel migratoire et invasif (X. Hu et al., 2013).

Le cluster miR-15-16

Ce cluster de miR est associé à une fonction de suppresseur de tumeur dans divers cancers tels que les leucémies, le cancer pancréatique et de la prostate (Malumbres, 2013). La voie p53 est impliquée dans la régulation du cluster miR-15-16. *In vitro*, la surexpression de ces miRs induit l'arrêt du cycle cellulaire. MiR-15 et miR-16 promeuvent également l'apoptose en inhibant le gène anti-apoptotique BCL2. D'autres cibles du cluster miR-15-16 incluent BCL1, WNT3A et les cyclines D1, D2 et E1 (Malumbres, 2013). Dans les cellules cancéreuses mammaires, la surexpression du complexe miR-15-16 conduit à l'arrêt des cellules en phase G1 et à une augmentation de l'apoptose, ainsi qu'à l'inhibition des gènes RPS6KB1, mTOR, EZH2 et du gène TWIST1 impliqué dans la transition épithélio-mésenchymateuse (EMT) (Janaki Ramaiah et al., 2014).

2.2.3.4 MiR-137 et miR-138

Focalisons-nous maintenant sur deux miRs que nous avons étudiés durant ce travail de thèse : miR-137 et miR-138.

MiR-137

MiR-137 possède une fonction de suppresseur de tumeur et est sous-exprimé dans de nombreux cancers tels que le carcinome squameux bucal (Kozaki et al., 2008), le cancer colorectal (Balaguer et al., 2010), gastrique (Steponaitiene et al., 2016) et pancréatique (Neault, Mallette, and Richard, 2016). MiR-137 possède de nombreux îlots CpG dans sa région promotrice et est régulé par la méthylation (Yin et al., 2014). D'un point de vue fonctionnel, miR-137 peut activer les voies onco-suppressives pRB et p53, qui régulent la prolifération, l'invasion et l'apoptose (Neault, Mallette, and Richard, 2016).

Enfin, miR-137 joue également un rôle majeur dans le système nerveux, en contrôlant la maturation neuronale, le développement dendritique et la transduction du signal synaptique. De plus, MiR-137 est impliqué dans le développement de la schizophrénie (Yin et al., 2014).

MiR-138

MiR-138 est également un onco-suppresseur impliqué dans les cancers de la gorge et du cou (Jin et al., 2013), du poumon (T. Yang et al., 2015) et le gliome (J. Wei et al., 2015). Il intervient aussi dans d'autres processus tels que le développement (Morton et al., 2008). MiR-138 peut diminuer l'expression des protéines BCR-ABL (*breakpoint cluster region-abelson*) et CCND3 (*cycline-D3*), et cibler la voie BCR-ABL/GATA1 pour inhiber la prolifération cellulaire (Xu et al., 2014). De plus, miR-138 est impliqué dans le contrôle de l'EMT en ciblant les gènes GIT1 (*G-protein-coupled receptor kinase interacting protein 1*) et SEMA4C (*semaphorin-4C*) (Jiefang Li et al., 2015).

Notre travail a permis de définir de nouvelles cibles et de nouvelles fonctions des miR-137 et miR-138 dans le cancer du sein, impliquant la machinerie épigé-

nétique des histones déméthylases KDM5 et le contrôle de la prolifération (Hélène Denis et al., 2015) (voir section résultat, chapitre 1).

2.3 Les longs ARNs non codants

2.3.1 Introduction

Ces dernières années, une découverte inattendue a révolutionné le domaine de la génomique : l'existence de très nombreux longs ARNs non codants (lncRNAs) fonctionnels. Les lncRNAs sont définis comme des transcrits de plus de 200 paires de bases (seuil arbitraire) qui ne sont pas traduits en protéines (Schmitt and Chang, 2016). L'expression des lncRNAs est précisément régulée et est spécifique des différents types cellulaires. Il est à présent accepté que les lncRNAs jouent un rôle important dans de nombreux processus biologiques critiques tels que le remodelage de la chromatine et la modification des histones, le contrôle de la transcription et de la stabilité des ARNm, le contrôle du cycle cellulaire et le transport des protéines (Jing Li, Xuan, and Liu, 2013). La surexpression, l'inhibition ou la présence des mutations dans les lncRNAs sont impliquées dans de nombreuses maladies. Cependant, en comparaison aux gènes codants, les lncRNAs sont encore fort peu étudiés.

2.3.2 Classification

La manière la plus courante de classifier les lncRNAs se base sur leur contexte génomique et leur relation avec les gènes codants à proximité. Les lncRNAs encodés dans des régions dépourvues de gènes codants sont dits « intergéniques ». Ils peuvent être convergents ou divergents en fonction de leur orientation par rapport au gène codant le plus proche. Les lncRNAs à proximité des gènes codants peuvent être divisés en trois catégories : (i) les exoniques, s'ils partagent une région exonique codante, (ii) les introniques s'ils résident dans les introns de gènes codants et (iii) les chevauchants qui chevauchent littéralement les

gènes codants. Ils peuvent encore être catégorisés par leur sens de transcription. On parle de lncRNAs « sens », s'ils sont transcrits sur le même brin que le gène codant voisin, ou d'« antisens », s'ils sont transcrits sur le brin opposé (Fig. 21) (Derrien et al., 2012; Mercer and Mattick, 2013). Ajoutons que des lncRNAs circulaires ont également été décrits (Quinn and Chang, 2015).

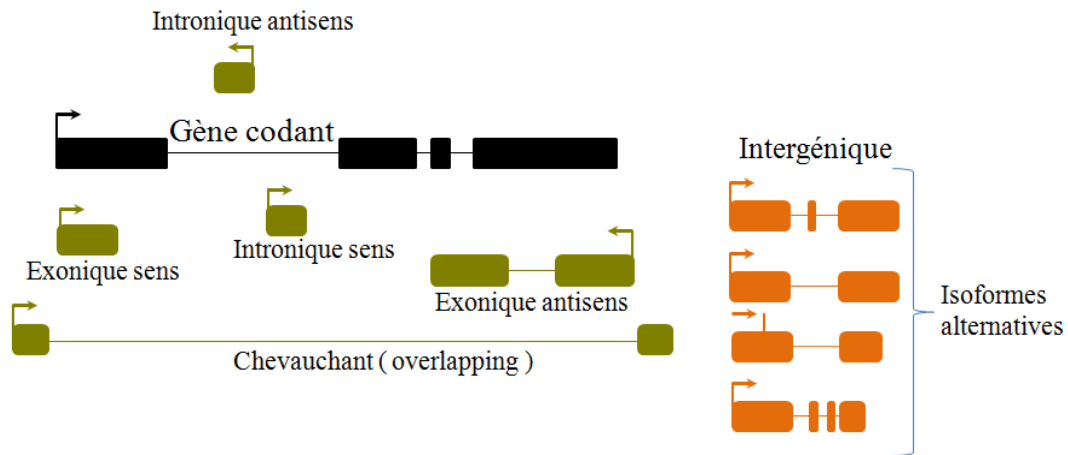


Fig. 21 : Classification des longs ARNs non codants.

Les lncRNAs se classifient en fonction de leur position relative aux gènes codants, ils peuvent être géniques (en vert) ou intergéniques (en orange) s'ils ne partagent pas de séquence avec un gène codant. (Inspiré de Derrien et al., 2012)

2.3.3 Propriétés des longs ARNs non codants

2.3.3.1 Caractérisation génomique

A l'instar des ARNm, les lncRNAs sont le plus souvent multi-exoniques, transcrits par l'ARN polymérase II et peuvent subir un épissage alternatif. Ils possèdent également une queue poly-A et une coiffe méthylguanosine en 5'. Cependant, ils sont généralement dépourvus d'une fenêtre de lecture ouverte traduite, leur séquence et leurs exons sont généralement plus courts et leur expression est plus faible et plus spécifique à chaque tissu que les ARNm (Mercer and Mattick, 2013; Quinn and Chang, 2015).

La transcription des lncRNAs est modulée par des facteurs de transcription et par l'état de la chromatine, elle-même influencée par les modifications d'histones et la méthylation de l'ADN (Wu et al., 2014). L'expression des lncRNAs peut être associée (i) aux marques d'histones H3 triméthylées sur la

lysine 4 (H3K4me3) et sur la lysine 36 (H3K36me3) et acétylées sur les lysines 9 et 27 qui définissent une région promotrice ou (ii) aux marques amplificatrices « *enhancer* » H3 monométhylées sur la lysine 4 (H3K4me1). Le rapport des marques H3K4me3/H3K4me1 permet de définir deux classes de lncRNAs : les lncRNAs associés aux promoteurs et les lncRNAs associés aux marques amplificatrices « *enhancer* » (Marques et al., 2013).

Au niveau post-transcriptionnel, les lncRNAs sont globalement stables, avec des demi-vies comparables aux demi-vies des ARNm (Quinn and Chang, 2015). Ils sont régulés par des protéines particulières se liant aux ARNs et par les miRs (Wu et al., 2014).

2.3.3.2 Localisation cellulaire

En comparaison avec l'ARNm, les lncRNAs sont enrichis dans la fraction nucléaire, où ils régulent la compaction de la chromatine et l'expression des gènes. Cependant, de très nombreux lncRNAs sont majoritairement présents dans le cytoplasme (Quinn and Chang, 2015). In fine, la localisation des lncRNAs semble dépendre de la présence de certains motifs de nucléotides. Par exemple, le motif AGCCC induit une localisation nucléaire (B. Zhang et al., 2014).

Une large étude sur la localisation des lncRNAs par hybridation « *in situ* » en fluorescence a analysé la localisation de 34 lncRNAs. Il en ressort que les lncRNAs peuvent être catégorisés en 5 sous-types : (i) un ou deux larges foyers dans le noyau, (ii) larges foyers dans le noyau et plusieurs petits foyers nucléaires plus éparses, (iii) principalement nucléaires, sans foyer particulier, (iv) cytoplasmiques et nucléaires et (v) majoritairement cytoplasmiques (Fig. 22) (Cabili et al., 2015).

Certains lncRNAs ont une localisation nucléaire précise : citons par exemple MALAT1 qui est présent dans les speckles nucléaires où il se lie à des facteurs d'épissage et NEAT1 qui est présent dans les paraspeckles où sont séquestrés les ARNs édités, c'est-à-dire possédant des conversions d'adénosines vers les inosines (Mercer and Mattick, 2013).

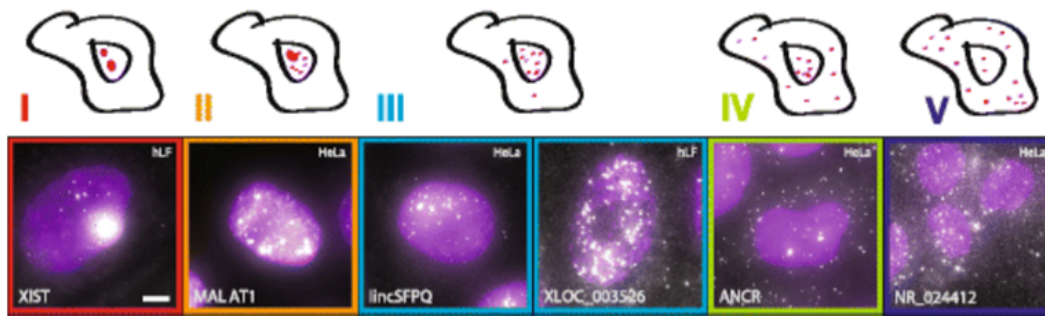


Fig. 22 : Localisation cellulaire des lncRNAs.

Les lncRNAs peuvent être situés majoritairement dans le noyau (I à III), dans le noyau et le cytoplasme (IV) ou majoritairement dans le noyau (V). (Adapté de Cabili et al., 2015)

Dans le cytoplasme, les lncRNAs sont souvent associés aux ribosomes, ce qui pose la question suivante : ces lncRNAs n'encodent-ils pas des polypeptides ? Des données de spectrométrie de masse n'ont globalement pas permis d'identifier des peptides correspondants à ces lncRNAs. Les ribosomes sont donc capables de discriminer l'ARN codant et non codant, et bien qu'associés aux ribosomes, ces lncRNAs n'encodent généralement pas de polypeptides (Quinn and Chang, 2015).

2.3.3.3 Conservation

Les lncRNAs semblent évoluer rapidement, et la présence d'orthologues est limitée. Par exemple, un tiers des lncRNAs humains ne seraient apparus que chez les primates (Derrien et al., 2012). Les séquences exoniques des lncRNAs sont globalement moins conservées, d'un point de vue évolutif entre différentes espèces, que les séquences codantes mais plus conservées que les introns ou des régions intergéniques aléatoires (Diederichs, 2014). Notons que les promoteurs des lncRNAs sont presque aussi bien conservés que les promoteurs de gènes codants.

Si on ne se limite pas à la séquence primaire, mais que l'on considère la structure secondaire des lncRNAs ou leur fonction, de nombreux lncRNAs sont cependant conservés. Par exemple, la conservation de la structure secondaire de MEG3 est plus importante que la structure primaire pour que ce lncRNA puisse exercer son rôle de suppresseur de tumeur (X. Zhang et al., 2010).

2.3.4 Fonctions et Modes d'action

La conformation spatiale et la présence de domaines d'interaction dictent les fonctions des lncRNAs. Ils peuvent se lier à d'autres ARNs par appariement des bases, à de l'ADN et à des protéines RBP (*RNA binding protein*) liant l'ARN (Fig. 23A). Dès lors, pour appréhender leur fonction, différentes techniques ont été mises au point pour identifier leurs interactants. Le ChIRP (*Chromatin isolation by RNA purification*) et ses variantes (CHART (*capture hybridisation analysis of RNA targets*), RAP (*RNA antisense purification*)) utilisent des sondes anti-sens pour capturer un lncRNA d'intérêt et ensuite identifier ses partenaires protéiques par spectrométrie et sa localisation génomique par séquençage à haut débit. A l'inverse, le RIP (*RNA Immunoprecipitation*) et ses variantes (PAR-CLIP (*Photoactivatable Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation*), iCLIP (*Individual-nucleotide resolution UV crosslinking and immunoprecipitation*)) permettent d'identifier par séquençage les ARNs liés à une protéine d'intérêt immuno-précipitée (Chu, Spitale, and Chang, 2015; Y. Yang et al., 2015).

Dans le noyau, les lncRNAs régulent la transcription de différentes manières : ils peuvent agir comme activateur en modulant la conformation de facteurs de transcription (FT) pour les activer, comme répresseur en inhibant l'interaction des FT avec l'ADN, comme guide pour recruter des FT ou encore comme échafaudage entre différentes protéines pour, par exemple, moduler les modifications d'histones et la méthylation de l'ADN au niveau de promoteur (Fig. 23B) (Guodong Yang, Lu, and Yuan, 2014; Mercer and Mattick, 2013). Les lncRNAs peuvent agir en « trans », c'est-à-dire à distance de leur site de transcription (par exemple HOTAIR, localisé sur le chromosome 12 qui influence l'expression des gènes du cluster HOXD sur le chromosome 2) ou en « cis » pour modifier localement l'environnement épigénétique (par exemple AIR qui recrute G9a pour méthyler les résidus H3K9 adjacents sur une région de 300 kb) (Mercer and Mattick, 2013).

Au niveau post-transcriptionnel, ils peuvent moduler l'édition, l'épissage, la traduction et la dégradation des ARNm. Ils peuvent également

jouer le rôle de précurseurs des miRs ou au contraire empêcher l'action d'autres miRs (Fig. 23B).

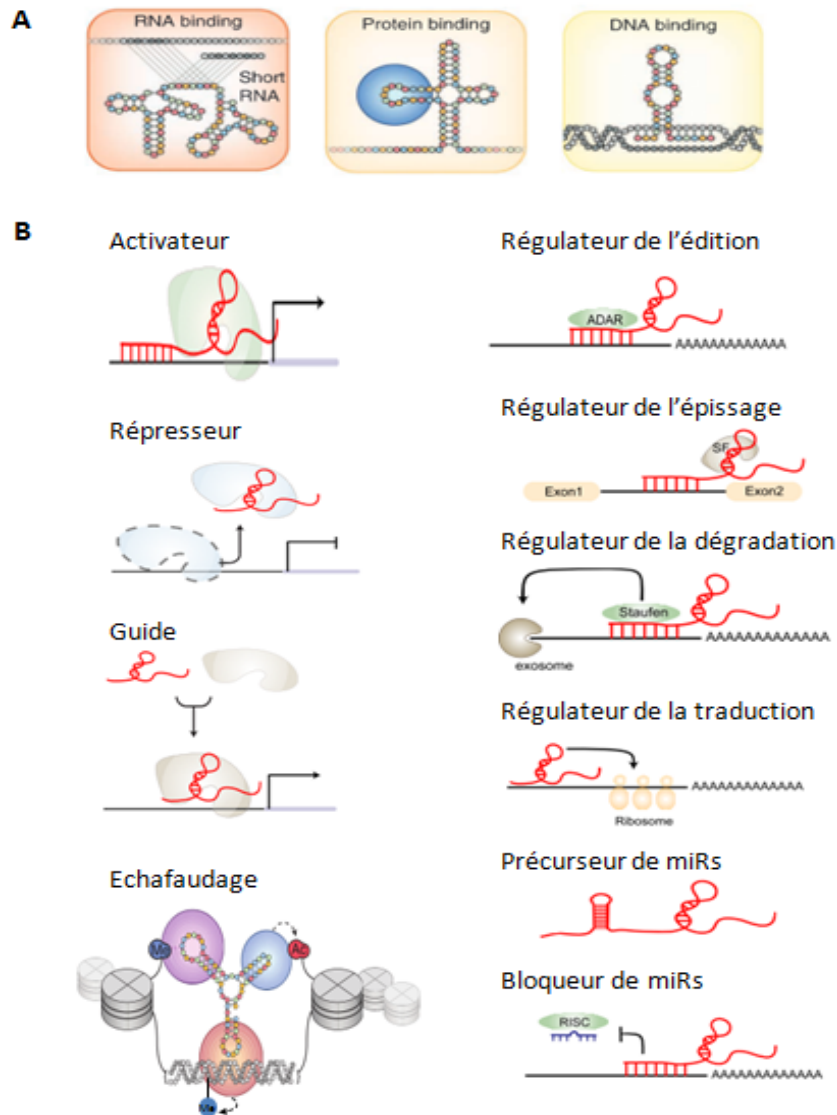


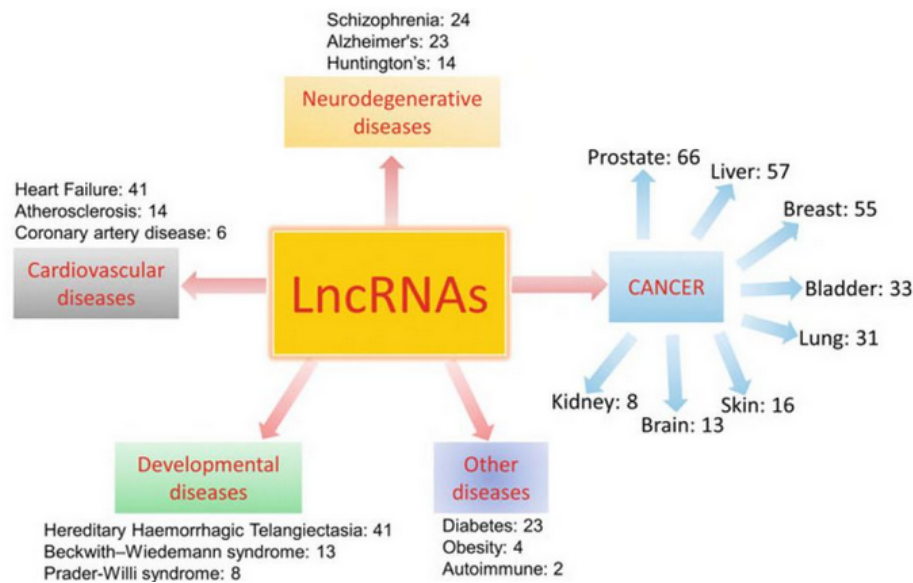
Fig. 23 : Mode d'action des lncRNAs.

(A) Les lncRNAs peuvent se lier à d'autres ARNs, à des protéines et à de l'ADN. (B) Dans le noyau, les lncRNAs régulent la transcription (à gauche) en activant, en éloignant ou en guidant les FT. Ils servent également d'échafaudage entre différentes protéines pour coordonner leurs actions. Au niveau post-transcriptionnel (à droite), les lncRNAs régulent l'édition en recrutant les protéines ADAR responsables de la conversion des adénosines en inosines. Ils peuvent se lier à des sites d'épissage et prévenir l'épissage alternatif, ils modulent la stabilité et la dégradation des ARNs par les exosomes, ils peuvent aussi réguler positivement ou négativement la traduction. Enfin les lncRNAs peuvent également être des précurseurs de miRs ou inhiber leur action en imitant une séquence d'ARNm cible. (Adapté de Guodong Yang, Lu, and Yuan, 2014; Mercer and Mattick, 2013)

Notons que la présence de modifications chimiques, souvent réversibles, sur la séquence des lncRNAs peut également affecter et réguler leur activité (Mercer and Mattick, 2013).

2.3.5 Implications pathologiques

Alors que les fonctions de la plupart des lncRNAs restent méconnues, de très nombreux lncRNAs sont impliqués dans diverses pathologies. En compilant plus de 500 publications, une étude indique que 370 lncRNAs différents sont expérimentalement impliqués dans 247 maladies (Chen et al., 2013, mise à jour en juillet 2015). Certains lncRNAs semblent plus spécifiques d'un type de pathologie, alors que d'autres sont retrouvés dans différentes maladies tels que MALAT1 qui est associé aux cancers colorectaux, du sein et du poumon ainsi



qu'aux maladies neuronales (Fig. 24) (Tang et al., 2013).

Fig. 24 : Implication des lncRNAs dans de nombreuses maladies.

Les principales maladies associées avec les lncRNAs sont illustrées. Le nombre à côté de chaque maladie correspond au nombre de lncRNAs expérimentalement impliqués dans cette maladie, selon les données de juin 2014. (D'après Morris, 2016)

Afin d'identifier de nouveaux lncRNAs associés à une maladie et de découvrir leurs interactants et leurs fonctions potentielles, de nombreux algorithmes de prédiction ont été publiés (ncPred (Alaimo, Giugno, and Pulvirenti,

2014), starBase (J.-H. Li et al., 2014), KATZLDA (X. Chen et al., 2015), LncDisease (Wang et al., 2016), etc).

Ajoutons les études d'association pangénomique (GWAS, *Genome-Wide Association Study*) qui ont révélé que de nombreux lncRNAs encodés par des régions d'ADN avec un nombre de copies altérées (CNA, *Copy Number Alteration*) ou comportant des polymorphismes SNPs (*Single Nucleotide Polymorphisms*) sont associés à des maladies (Gong et al., 2015; Schmitz, Grote, and Herrmann, 2016).

2.3.5.1 Implication des lncRNAs dans le cancer

Considérant la grande diversité des lncRNAs et de leurs fonctions, il n'est pas étonnant de retrouver de nombreux lncRNAs associés aux cancers. Ils ont été décrits dans les leucémies, le cancer du sein, de la prostate, du foie, des reins, du colon, de la vessie, de la thyroïde, des poumons et des ovaires (X. Chen et al., 2016). D'une manière générale, les lncRNAs sont associés à l'ensemble des caractéristiques du cancer (Fig. 25) (Parasramka et al., 2016).

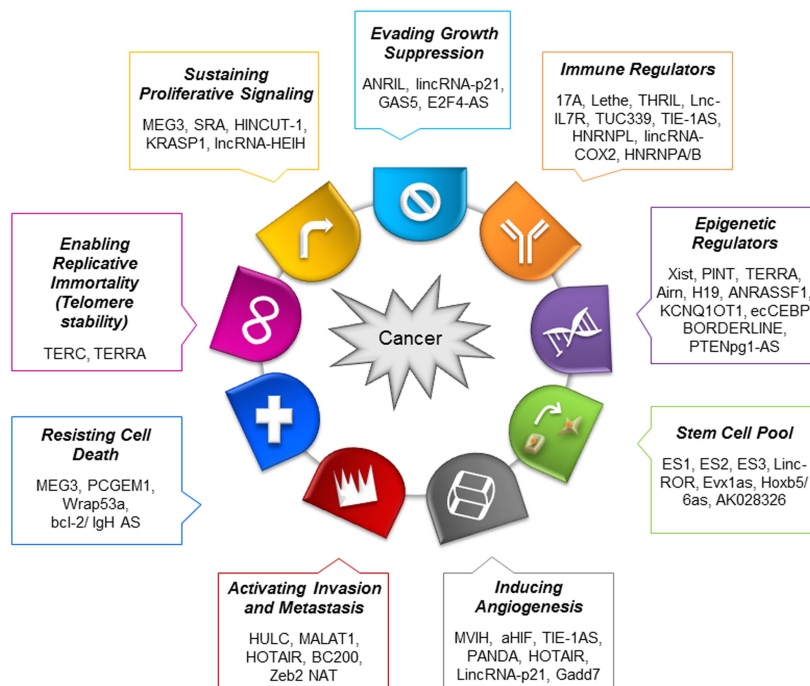


Fig. 25 : Implication des lncRNAs dans le cancer.

Les lncRNAs sont associés avec les caractéristiques intrinsèques du cancer et contribuent au développement et à la progression tumorale. (D'après Parasramka et al., 2016)

A l'instar des gènes codants et des miRs, les lncRNAs peuvent être suppresseurs de tumeur ou oncogènes. Ils sont impliqués dans de nombreuses voies moléculaires du cancer, telles que les voies c-MYC ou p53 (Fig. 26).

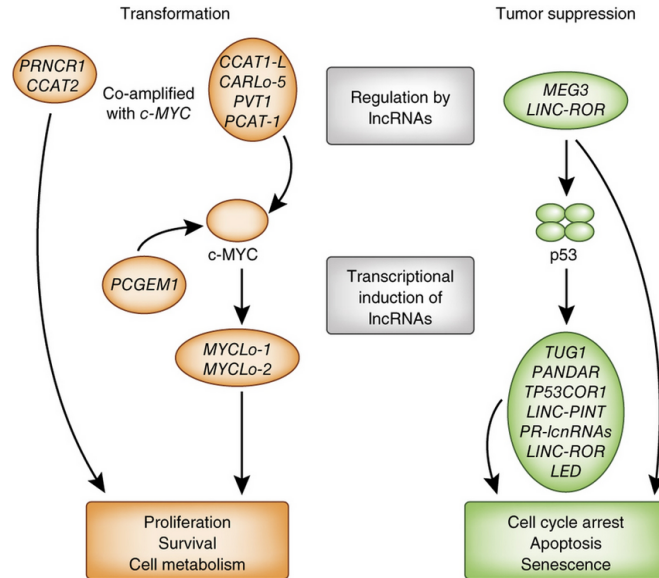


Fig. 26 : Exemples de lncRNAs impliqués dans les voies c-MYC et p53.

De nombreux lncRNAs régulent l'oncogène c-MYC ou l'onco-suppresseur p53 en affectant leur expression ou leur activité protéique. Certains lncRNAs sont localisés sur le même locus que c-MYC et sont co-amplifiés avec lui. En aval, des lncRNAs sont les cibles transcriptionnelles des voies c-MYC et p-53 et contribuent à la réponse oncogénique ou onco-suppressive. (D'après Huarte, 2015)

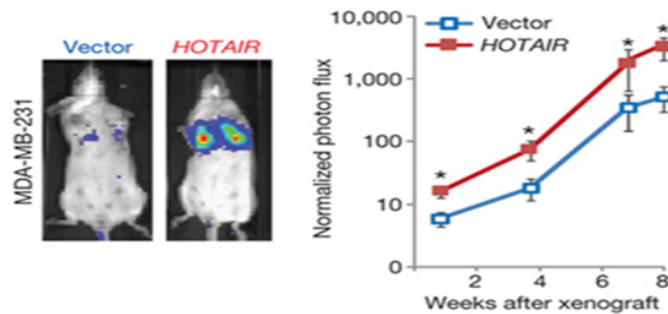
L'expression des lncRNAs peut également servir pour diagnostiquer certains cancers (par exemple le test d'expression de PCA3, commercialisé pour déceler le cancer de la prostate (J. T. Wei et al., 2014)) ou pour estimer l'apparition de métastases et la survie telle que sChLAP1 (*second chromosome locus associated with prostate 1*) dans le cancer de la prostate (Prensner et al., 2013). Ci-dessous nous détaillons quelques exemples clés de lncRNAs impliqués dans les cancers.

HOTAIR

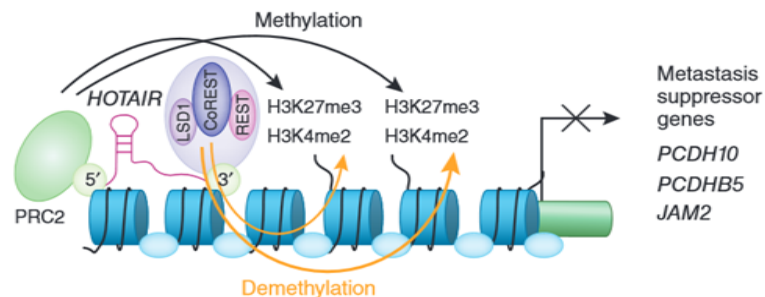
HOTAIR (*HOX transcript antisense RNA*) est surexprimé dans de nombreux cancers. Il est de mauvais pronostic pour l'apparition de métastases dans les cancers mammaires, gastro-intestinaux, hépatiques et colorectaux (Gutschner and Diederichs, 2012). Dans un modèle de xéno greffe de cellules mammaires MDA-MB-231, la surexpression de HOTAIR induit la formation de métastases dans le poumon (Fig. 27A) (Gupta et al., 2010).

D'un point de vue fonctionnel, HOTAIR est transcrit à partir du locus HOXC sur le chromosome 12 et agit en « trans » pour réprimer le locus HOXD sur le chromosome 2 en agissant sur la machinerie épigénétique. D'une part, la région 5' de HOTAIR recrute le complexe de répression Polycomb 2 (PRC2) qui peut dès lors triméthyliser la lysine 27 de l'histone H3 (H3K27) et inhiber l'expression génique. D'autre part, la région 3' de HOTAIR s'associe avec le complexe coREST/LSD1 qui déméthyle les lysines 4 de l'histone H3 (H3K4) (Fig. 27B). HOTAIR sert donc de guide et d'échafaudage pour réprimer la transcription de gènes suppresseurs de métastases tels que les gènes d'adhésion

A



B



JAM2 et PCDH10.

Fig. 27 : Rôles de HOTAIR dans le cancer.

(A) La surexpression de HOTAIR induit l'apparition de métastases pulmonaires en modèle de xénogreffes (injection dans la veine caudale) (B) HOTAIR sert d'échafaudage entre le complexe PRC2 et le complexe CoREST/LSD1 pour réguler la méthylation des histones H3K27 et la déméthylation des histones H3K4 afin d'inhiber l'expression de gènes suppresseurs de métastases. (D'après Gupta et al., 2010; Croce, 2010)

MALAT1

MALAT1 (*metastasis-associated lung adenocarcinoma transcript 1*) est un lncRNA fortement conservé qui peut être très abondant dans certains sous-types cellulaires. Il est exprimé dans de nombreux tissus sains et surexprimé dans diverses tumeurs telles que le cancer du poumon, du sein, de la prostate, du colon, du foie et de l'utérus (Gibb, Brown, and Lam, 2011). La surexpression de MALAT1 est associée à la prolifération, la migration et la transition épithélio-mésenchymateuse en modèle cellulaire, ainsi qu'à un mauvais pronostic de survie dans de nombreux cancers (Tian and Xu, 2015). Toutefois une autre étude décrit une fonction onco-suppressive de MALAT1 dans les tumeurs gliales et associe l'expression de MALAT1 à un bon pronostic (Y Han et al., 2016), ce qui laisse supposer que les fonctions de MALAT1 sont hautement spécifiques des différents tissus.

MALAT1 est localisé dans les speckles nucléaires où il est associé avec les protéines SR du splicéosome (protéines riches en sérines et arginines) et module l'épissage (Fatima et al., 2015; Huarte, 2015; Tripathi et al., 2010). MALAT1 régule également la transcription en modulant la localisation de la protéine Polycomb 2, et en s'associant avec le FT E2F (Yoshimoto, Mayeda, and Yoshida, 2016). Cependant, l'invalidation génique de MALAT1, en modèle murin, n'entraîne pas de changement phénotypique important.

MEG3

Au contraire des deux exemples ci-dessus, MEG3 (*Maternally expressed gene 3*) joue un rôle suppresseur de tumeur. Il est exprimé dans de nombreux tissus sains, mais pas ou peu dans les cellules tumorales en raison d'une hyperméthylation de son promoteur. Son expression est pronostic dans le cancer du foie et du poumon, dans l'adénome pituitaire et le carcinome de la langue (Juan

Li et al., 2016). En modèle cellulaire, la surexpression de MEG3 empêche la prolifération et induit l'apoptose. MEG3 est soumis à l'empreinte parentale et n'est exprimé que par l'allèle maternel (Gibb, Brown, and Lam, 2011).

D'un point de vue fonctionnel, différents mécanismes d'action sont décrits pour ce lncRNA qui est très majoritairement nucléaire: (i) MEG3 interagit et stimule la voie p53. En fait, MEG3 diminue l'expression de MDM2, ce qui empêche la poly-ubiquitination de p53. De plus, MEG3 augmente l'affinité de p53 pour le promoteur de GDF15 afin d'activer ce gène qui inhibe la prolifération (Y. Zhou, Zhang, and Klibanski, 2012). (ii) L'expression des gènes VEGFA et de son récepteur VEGFR1 est diminuée dans le cerveau de souris invalidées pour MEG3. Ces cerveaux sont également caractérisés par la formation très élevée de micro-vaisseaux. Ceci suggère que MEG3 module l'angiogenèse. Notons que les voies Notch et Wnt sont également dérégulées dans ce modèle. (Y. Zhou, Zhang, and Klibanski, 2012). (iii) MEG3 peut réguler la phosphorylation de la protéine RB directement ou indirectement par l'activation du régulateur p16 et moduler le cycle cellulaire. (iv) MEG3 peut se lier et servir de compétiteur endogène de mir-181 et ainsi inhiber la prolifération, la migration et l'invasion et induire l'apoptose par augmentation de protéine Bcl-2 (Peng et al., 2015). (v) MEG3 peut former une triple hélice RNA-DNA et, en interagissant avec EZH2, inhibe les gènes de la voie TGF- β (Mondal et al., 2015). Ajoutons que l'inactivation génique de MEG3, en modèle murin, induit une mortalité périnatale (Y. Zhou, Zhang, and Klibanski, 2012).

Ces quelques exemples illustrent donc la complexité des mécanismes impliquant les lncRNAs, qui peuvent avoir des fonctions pro- ou anti-tumorales.

2.3.5.2 Potentiel thérapeutique

Une étude, basée sur 5000 tumeurs provenant du projet TCGA (*The Cancer Genome Atlas*), a montré que 60% des lncRNAs dérégulés dans les cancers sont spécifiques d'un type de tumeur uniquement, ce qui en fait des cibles thérapeutiques spécifiques très intéressantes (Schmitz, Grote, and Herrmann, 2016). Comme illustré précédemment, les lncRNAs régulent certaines activités particu-

lières des protéines. Dès lors, les médicaments ciblant ces lncRNAs pourraient être plus spécifiques et moins toxiques que les drogues ciblant les protéines (Huarte, 2015).

Une stratégie thérapeutique ciblant les lncRNAs implique l'utilisation d'oligonucléotides antisens tels que les LNA (*Locked Nucleic Acid*). Les LNA sont des oligonucléotides modifiés qui sont protégés contre les nucléases présentes dans le noyau. Les LNA vont cibler spécifiquement (par complémentarité des bases) un lncRNA d'intérêt, qui sera dès lors dégradé par la RNase H. Des études pionnières chez la souris ont montré que l'utilisation de LNA permet d'inhiber *in vivo* l'expression de MALAT1 et ainsi de diminuer la prolifération cellulaire et la formation de métastases (Michalik et al., 2014; Gutschner et al., 2013). SAMMSON est un autre exemple qui illustre le potentiel thérapeutique des lncRNAs. SAMMSON est un lncRNA spécifiquement impliqué dans le développement de mélanomes. Des injections intraveineuses de LNA ciblant SAMMSON diminuent significativement la formation de tumeurs et induit l'apoptose dans un modèle murin de mélanome (Leucci et al., 2016).

D'autres stratégies incluant l'utilisation des petites molécules inhibitrices (qui ciblent spécifiquement le domaine d'interaction RNA-protéine) et des vecteurs viraux sont en cours de développement (Fatemi, Velmeshev, and Faghihi, 2014).

3 Le cancer du sein

3.1 Introduction

Le cancer du sein trouve son origine dans une transformation maligne des lobules (qui produisent le lait) ou des canaux galactophores (qui transportent le lait). Le cancer du sein est une maladie complexe. En effet, il existe de nombreux sous-types de cancer du sein, de morphologies et de pronostics distincts, qui nécessitent des traitements particuliers. L'étiologie du cancer du sein n'est pas claire, mais de nombreux facteurs de risque ont été identifiés tels que l'obésité, les antécédents familiaux, l'exposition aux œstrogènes et à la progestérone et la consommation d'alcool et de tabac ("ESMO Breast Cancer: A Guide for Patients" 2013).

3.1.1 Epidémiologie

Le cancer du sein est une des premières causes de mortalité prématurée chez les femmes dans le monde et constitue donc un problème majeur de santé publique. Le cancer du sein touchera une femme sur huit au cours de sa vie et son incidence est en augmentation. En 2016, on estime que 246 000 femmes seront diagnostiquées avec une tumeur mammaire aux États-Unis, et plus de 40 000 en mourront ("American Cancer Society. Cancer Facts & Figures 2016" 2016). Cependant, grâce au traitement et à la détection précoce, le taux de mortalité a diminué de 36% entre 1987 et 2012.

En Belgique, les derniers chiffres disponibles évoquent 10 695 nouveaux cas de cancer mammaire chez la femme en 2013, ce qui en fait de loin le cancer le plus fréquent chez la femme (Fig. 28). Plus de 2000 décès par an dus au cancer du sein sont également répertoriés sur notre territoire. Enfin, 75% des cas de cancer sein apparaissent après 50 ans (La Fondation Registre du Cancer, 2013).

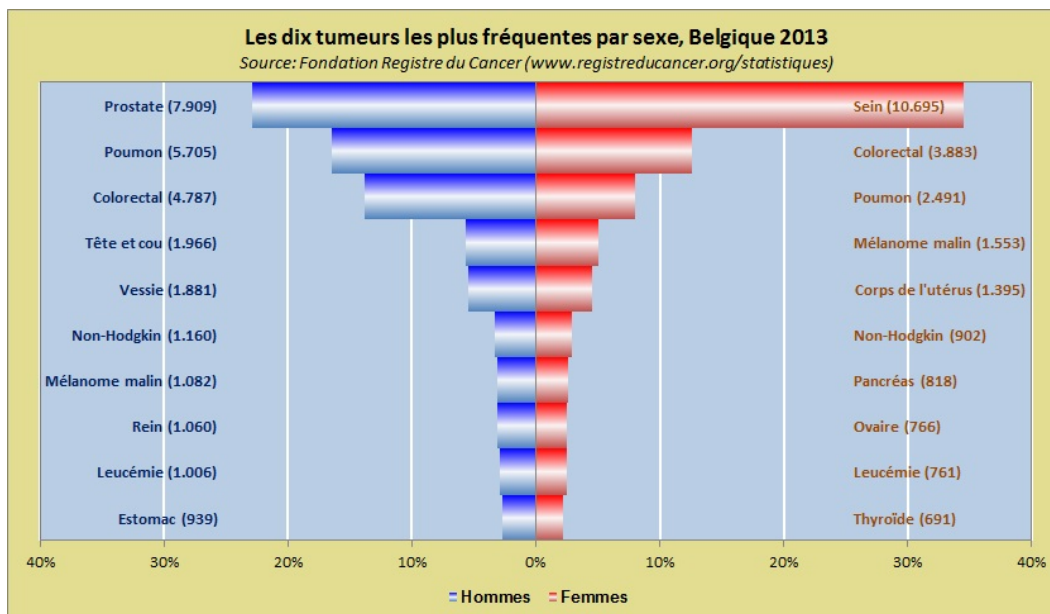


Fig. 28 : Incidence des cancers en Belgique en 2013 chez l'homme et la femme. Nombre de nouveaux cas de cancer diagnostiqués en Belgique en 2013 par sexe. (source : Belgian Cancer Registry, www.kankerregister.org)

3.2 Développement du cancer du sein

Le cancer du sein est une maladie hétérogène, qui trouve son origine dans des entités distinctes et possède des caractéristiques biologiques et cliniques très variables. En fait, nous devrions parler des cancers du sein, car il s'agit d'une collection de maladies distinctes qui affectent la même structure anatomique. Par simplicité, nous parlerons du cancer du sein au singulier pour représenter l'ensemble des cancers du sein. Pour mieux comprendre le cancer du sein et sa complexité, nous aborderons d'abord brièvement l'anatomie du sein et ensuite nous détaillerons les différents types de cancer du sein.

3.2.1 Anatomie du sein

Le sein est une masse de tissus glandulaires, conjonctifs et gras qui n'est fonctionnel qu'en période de lactation. La glande mammaire, qui constitue la partie fonctionnelle du sein, est une glande exocrine composée de 10 à 20 lobes. Chaque lobe comprend de nombreux lobules, eux-mêmes constitués de tubulo-

alvéoles qui sécrètent le lait (Fig. 29). Le drainage du lait est réalisé par les canaux galactophores collecteurs qui s'ouvrent au niveau du mamelon. Les canaux galactophores sont composés de deux types cellulaires : les cellules épithéliales luminales qui forment la couche interne et les cellules myoépithéliales basales qui sont responsables de la contraction et qui forment la couche externe (Hoffmann-Cucuz, 2010). Différentes études montrent l'existence de cellules souches mammaires multipotentes qui peuvent former une glande mammaire à part entière (Ercan, van Diest, and Vooijs, 2011). D'autres études montrent que l'homéostasie de la glande mammaire est maintenue par deux types distincts de cellules souches plus spécialisées : les cellules souches unipotentes luminales et les cellules souches unipotentes myoépithéliales (Van Keymeulen et al., 2011).

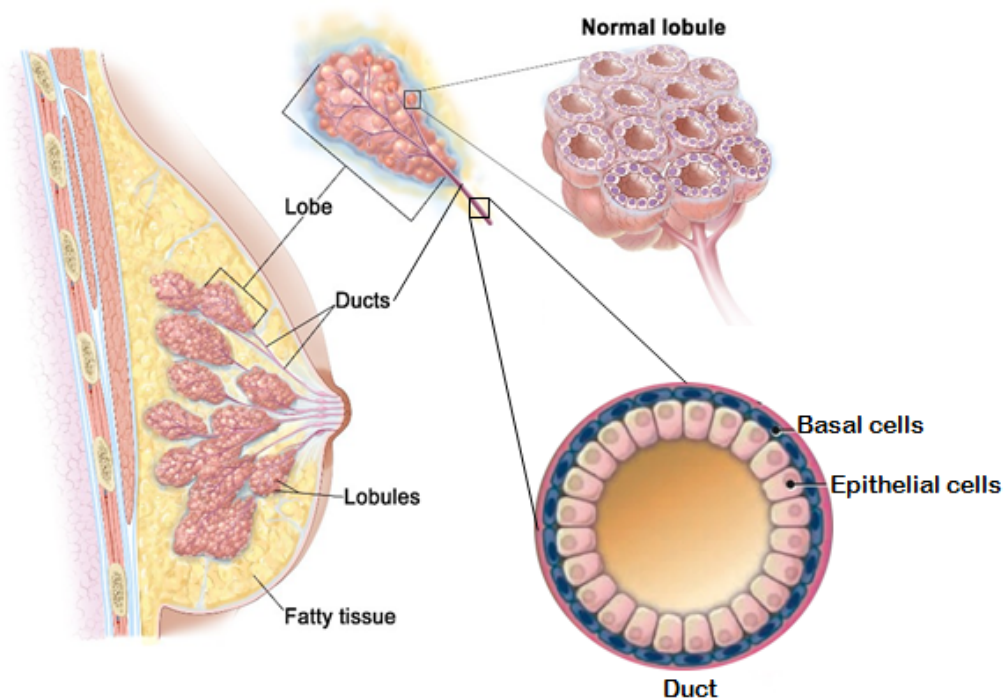


Fig. 29 : Anatomie du sein.

La structure fonctionnelle du sein est la glande mammaire qui est formée de lobes, eux-mêmes composés de lobules produisant le lait et de canaux lactifères qui amènent le lait vers le mamelon. (Adapté de <http://breastcancerhelp.net/lobular-carcinoma-lobular-carcinoma-in-situ-and-invasive-lobular-carcinoma/>)

3.2.2 Tumorigénèse mammaire

Il existe de nombreuses similitudes entre les cellules cancéreuses et les cellules souches. Les cellules souches, qui ont une durée de vie très étendue, peuvent emmagasiner de nombreuses mutations génétiques pour mener à la tumorigénèse. Ces cellules vont alors donner naissance de manière anormale à de nombreuses cellules filles pour former une hyperplasie, qui peut évoluer en carcinomes in situ, dans les lobules ou les canaux galactophores. Le développement de l'hyperplasie peut alors se propager, rompre la membrane basale, envahir le tissu conjonctif et atteindre les ganglions lymphatiques. Les cellules cancéreuses peuvent également envahir d'autres organes via le système circulatoire.

3.3 Diversité des cancers du sein

3.3.1 Classification anatomo-pathologique

Une première distinction des cancers du sein se base sur leur étendue : on parle d'une part de cancers « *in situ* » c'est-à-dire confiné dans leur emplacement d'origine, sans rupture de la membrane basale et d'autre part de cancers infiltrants qui se propagent.

Les carcinomes mammaires se classifient également selon leur origine en deux types de carcinomes principaux : les carcinomes canaux (les plus fréquents) et les carcinomes lobulaires (Fig. 30).

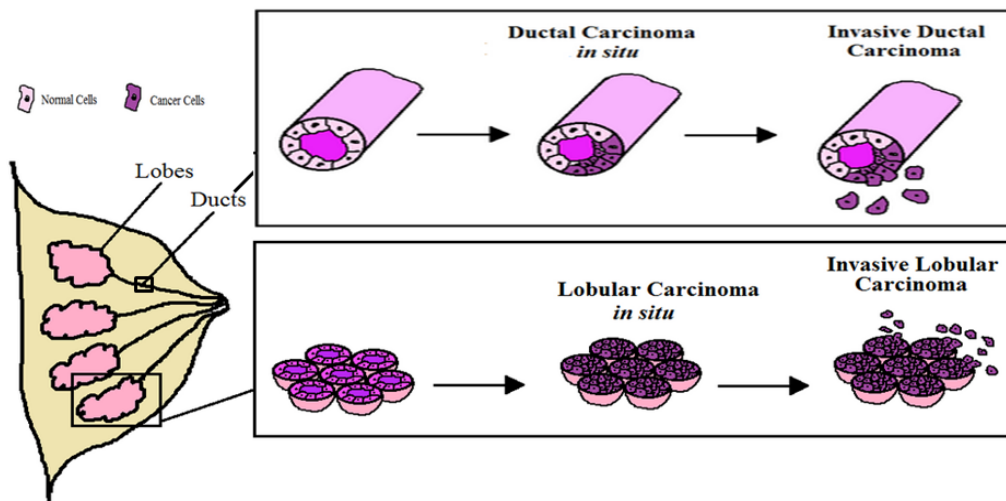


Fig. 30 : Anato-pathologie du cancer du sein.

Le cancer du sein peut se développer à partir des canaux galactophores (haut) ou des lobes. La tumeur peut être localisée (« in situ ») ou invasive.

(Adapté de <http://www.cubocube.com/dashboard.php?a=1643&c=1>)

Une autre classification histologique, basée sur l’aspect et l’hétérogénéité des cellules cancéreuses permet de définir le grade. Les tumeurs de grade 1 sont bien différenciées et présentent peu de mitoses. Les tumeurs de grade 3 sont non différenciées, avec une architecture anormale et un taux de mitoses élevé. Le grade 2 est intermédiaire entre les grades 1 et 3 (“ESMO Breast Cancer: A Guide for Patients” 2013; Inserm Expertise collective, 2008).

3.3.2 Classification clinique

En clinique, les cancers du sein sont caractérisés par la stadification utilisant le système TNM (taille de la tumeur, nombre de ganglions lymphatiques contenant des cellules cancéreuses et la présence de métastases). En outre, quatre marquages sont classiquement réalisés sur une biopsie mammaire afin de déterminer le statut des récepteurs hormonaux à (i) l’œstrogène et à (ii) la progestérone, (iii) le statut HER2 et (iv) le statut de prolifération par le marqueur de prolifération Ki-67 (“ESMO Breast Cancer: A Guide for Patients” 2013).

3.3.2.1 Le statut des récepteurs hormonaux

Les cellules mammaires peuvent exprimer des récepteurs hormonaux à l'œstrogène (ER) et à la progestérone (PR). Ces hormones, produites par les ovaires, jouent un rôle clé dans le développement des seins à la puberté et pendant la grossesse. Lorsque l'œstrogène ou la progestérone se lie sur le ER ou sur le PR respectivement, le récepteur change de conformation et s'associe avec des cofacteurs pour se lier à l'ADN et moduler l'expression de nombreux gènes impliqués dans la prolifération, la différenciation, l'invasion et l'angiogenèse (Tokunaga et al., 2014).

Environ deux tiers des tumeurs mammaires expriment les récepteurs hormonaux et sont dépendantes de ces hormones pour leur croissance. Dès lors, ce type de tumeur est susceptible de répondre aux hormonothérapies. De plus, ces tumeurs sont généralement bien différenciées, moins agressives et associées à un bon pronostic (Dai et al., 2016).

3.3.2.2 Le statut HER2

Le gène HER2 (ou ERBB2) encode un récepteur transmembranaire tyrosine kinase de la famille des récepteurs EGFR. Le récepteur HER2 fonctionne en stimulant, entre autres, la prolifération par l'activation de la voie des phosphatidylinositol-3-kinases (PI3K). L'amplification du gène HER2 est observée dans environ 20% des cancers mammaires. Ce type particulier de cancer mammaire peut être traité par une thérapie spécifique : un anticorps (Trastuzumab) se lie sur un domaine extracellulaire d'HER2 et empêche l'activation du récepteur (Gajria and Chandarlapaty, 2011).

3.3.3 Sous-types moléculaires du cancer du sein

En l'an 2000, les avancées technologiques dont l'avènement de la puce à ADN, ont permis d'analyser le profil d'expression de l'ensemble des gènes dans de nombreuses tumeurs. Il en ressort une classification moléculaire du cancer du sein en cinq sous-types : (i) les tumeurs de type luminal A, généralement

hormono-positives, peu prolifératives et de bon pronostic représentent 40 à 60% des tumeurs mammaires, (ii) les tumeurs de type luminal B (5 à 20% des tumeurs), généralement hormono-positives et fortement prolifératives, (iii) les tumeurs HER2+, qui possèdent le plus souvent une amplification du gène HER2 sont peu fréquentes (7 à 12 % des tumeurs), (iv) les tumeurs de type « basal-like » représentent 15 à 20% des tumeurs; ces tumeurs agressives n'ont généralement pas d'amplification du gène HER2 et n'expriment pas les récepteurs hormonaux, par contre elles expriment des gènes caractéristiques des cellules mammaires épithéliales basales, (v) les tumeurs de type « normal-like », relativement rares, ont un profil d'expression similaire au tissu mammaire sain, mais représentent tout de même un sous-type agressif et de mauvais pronostic (Fig. 31) (Perou et al., 2000; Sorlie et al., 2003). Une signature de l'expression de 50 gènes, appelée signature (ou classification) PAM50 a été développée pour permettre de classer les tumeurs en leurs différents sous-types moléculaires (Parker et al., 2009). La prise en charge thérapeutique et le pronostic varient d'un sous-type à l'autre (Sotiriou et al., 2003).

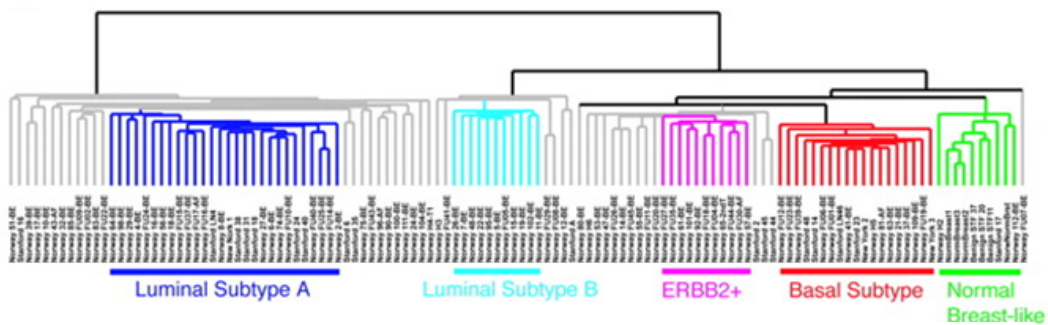


Fig. 31 : Classement moléculaire des cancers du sein.

Le profil d'expression génique permet de classer les tumeurs mammaires en cinq sous-types distincts. (D'après Sorlie et al., 2003)

Une bonne corrélation existe entre les sous-types moléculaires et les marquages histologiques classiques (Fig. 32).

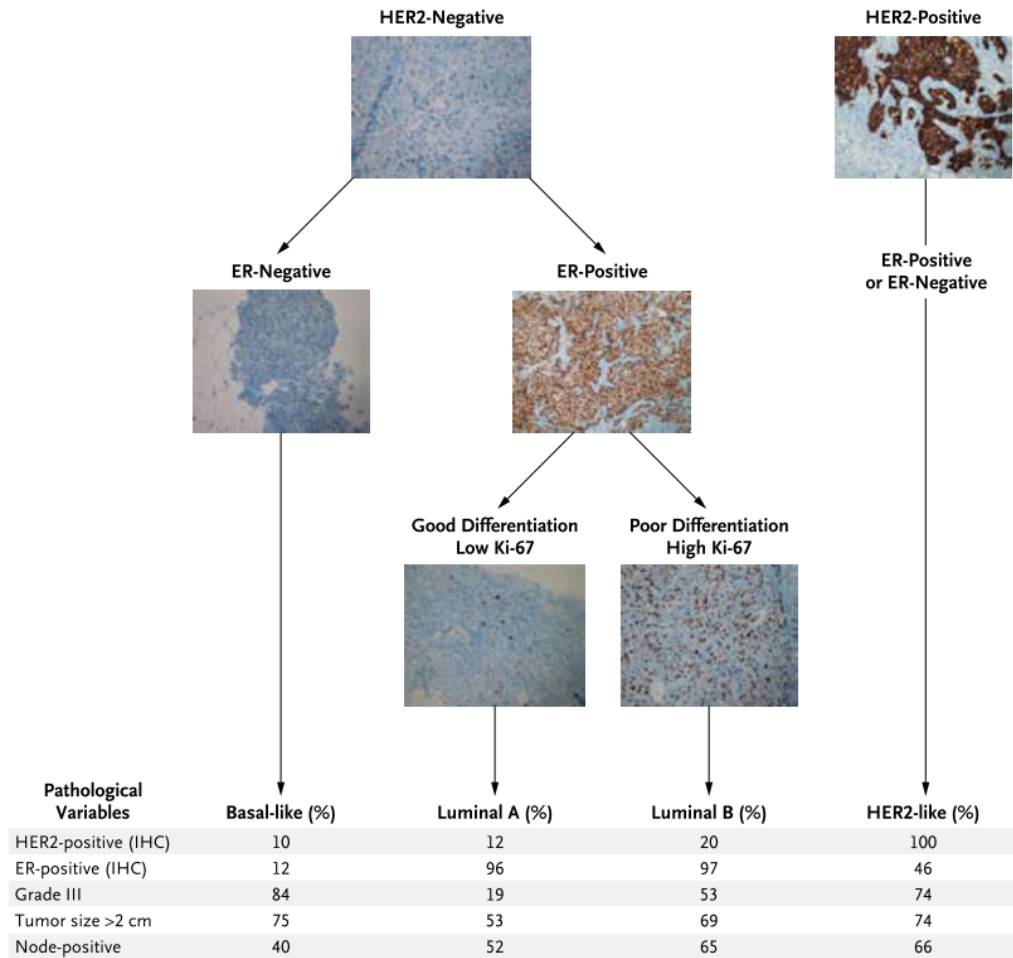


Fig. 32 : Relation entre les sous-types moléculaires et les caractéristiques clinico-pathologiques. Correspondance, en pourcentage, entre les sous-types basés sur l'expression des gènes et les critères cliniques utilisés en routine. (D'après Sotiriou and Pusztai, 2009)

Ces dernières années, de nouvelles classifications basées sur l'expression des gènes émergent, décrivant jusqu'à 12 sous-groupes (The Cancer Genome Atlas, 2012). Ces données ont modifié la manière d'appréhender le cancer du sein et démontrent encore une fois le caractère hétérogène du cancer du sein.

3.3.4 Altérations (épi)génétiques du cancer du sein

3.3.4.1 Altérations génétiques du cancer du sein

De nombreuses mutations somatiques sont décrites dans le cancer du sein, souvent en relation avec les sous-types moléculaires. Par exemple, la mutation du gène PIK3CA est présente dans 45% des tumeurs lumineales A, dans 29% des lumineales B et dans 39% des tumeurs HER2+. Le gène TP53 est, quant à lui, muté dans 80% des tumeurs de type « basal-like » pour seulement 12% des tumeurs lumineales A (Fig. 32) (The Cancer Genome Atlas, 2012).

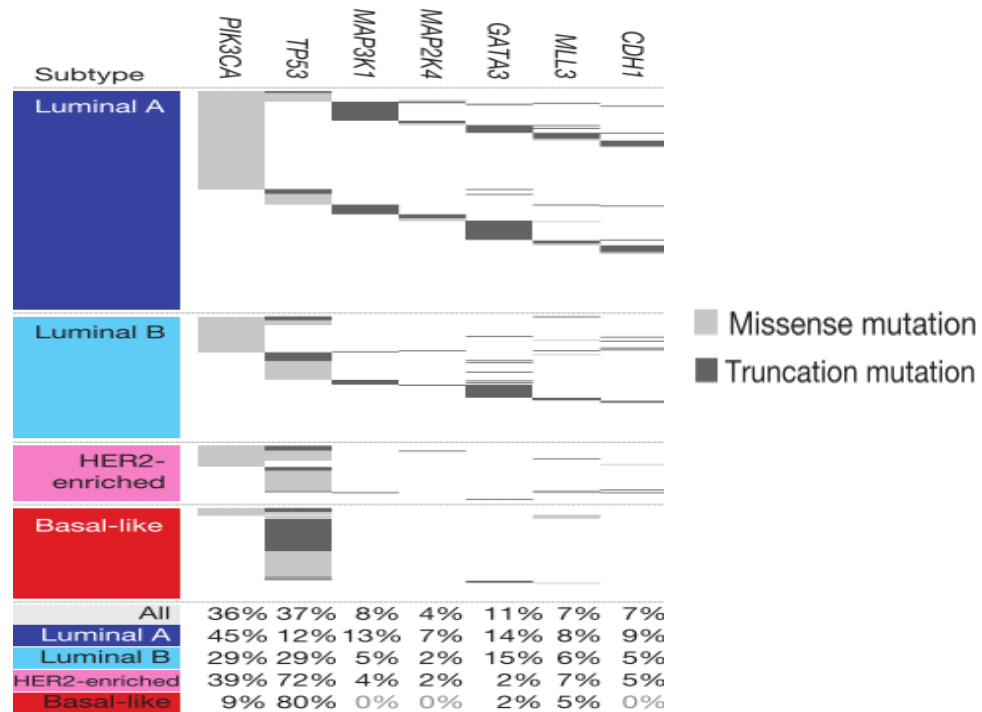


Fig. 33 : Relation entre les sous-types moléculaires et les gènes souvent mutés dans le cancer du sein.

Le pourcentage représente la fréquence des mutations dans les différents sous-types moléculaires de cancer mammaire. (D'après The Cancer Genome Atlas, 2012)

Ajoutons que la présence de la mutation germinale des gènes BRCA1 et BRCA2 (relativement rare dans la population) augmente drastiquement le risque de développer un cancer du sein et peut expliquer, en partie, les causes héréditaires du cancer du sein.

3.3.4.2 Altérations épigénétiques du cancer du sein

Le cancer du sein est souvent associé à des modifications aberrantes d'histones qui mènent à une instabilité génomique et à une répression des gènes suppresseurs de tumeurs. La protéine LSD1, qui déméthyle les histones H3K4, est par exemple fortement exprimée dans les tumeurs ER négatives et est un marqueur prédictif des cancers mammaires agressifs (Y. Huang et al., 2011). Un autre exemple implique la sous-unité EZH2 des complexes Polycomb 2. EZH2 est surexprimée dans le cancer du sein et associée aux sous-types HER2+ et basal-like (Holm et al., 2012). Cependant les mécanismes sous-jacents à la surexpression d'EZH2 restent à élucider.

Notre laboratoire a contribué à l'étude du profil de méthylation dans les cancers du sein. Six groupes de cancers du sein ont été identifiés sur base de leur profil de méthylation. Trois de ces groupes montrent une association forte avec les sous-types moléculaires basal-like, HER2+ et luminal A. Les trois autres groupes sont composés d'une mixture de sous-types moléculaires, suggérant que le profil de méthylation peut améliorer la taxonomie du cancer du sein. De plus, le profil de méthylation peut refléter l'origine cellulaire (progéniteurs) des tumeurs. Enfin, cette étude montre également que le profil de méthylation peut déterminer la composition cellulaire du micro-environnement tumoral, et plus particulièrement l'infiltration immunitaire de cellules T (Dedeurwaerder et al., 2011). Plus récemment, il a été démontré que le profil de méthylation permet de catégoriser les tumeurs de types basal-like en trois sous-types de pronostic variable (Stirzaker et al., 2015). Ceci démontre que le profil de méthylation peut servir de biomarqueur, et que la méthylation est une cible prometteuse de nouvelle thérapie.

3.4 Prise en charge thérapeutique

La prise en charge du cancer du sein est complexe et implique la participation d'une équipe pluridisciplinaire de médecins. Elle va dépendre de l'étendue, de la position, du stade, des statuts hormonaux et HER2 de la tumeur et du nombre de

ganglions atteints. L'âge de la patiente et son statut ménopausique sont également pris en compte. (Senkus et al., 2015; "ESMO Breast Cancer: A Guide for Patients" 2013).

3.4.1 La chirurgie et la radiothérapie

La chirurgie est le traitement de choix pour les tumeurs non invasives, mais peut également être envisagée dans les tumeurs plus avancées. La chirurgie varie en fonction du profil du cancer du sein et consiste en l'ablation localisée de la tumeur (lumpectomie) ou en l'ablation de la tumeur et d'une partie importante de tissus proches (quadrantectomie) ou en l'ablation de la totalité du sein (mastectomie). Notons que la reconstruction mammaire est systématiquement proposée et fait partie de la prise en charge du cancer du sein. La chirurgie peut être accompagnée d'une radiothérapie et/ou d'une chimiothérapie.

La radiothérapie est recommandée pour presque tous les cancers du sein invasifs. Elle a généralement lieu après la chirurgie et diminue drastiquement le risque de rechute. En routine, une dose d'irradiation de 45-50 Grays fractionnée en 25 séances est recommandée (Senkus et al., 2015).

3.4.2 Les traitements systémiques

Les traitements systémiques consistent en l'hormonothérapie, la chimiothérapie et un traitement ciblant la protéine HER2.

3.4.2.1 La chimiothérapie

Différentes chimiothérapies sont prescrites dans le cancer du sein, telles que des agents alkylants de l'ADN qui rendent la réplication impossible (par exemple le cyclophosphamide), des anthracyclines (doxorubicine) qui s'intercalent dans l'ADN et inhibent la topoisomérase II et des taxanes (docetaxel) qui inhibent la dépolymérisation des microtubules pour empêcher la division cellulaire. Différentes combinaisons de chimiothérapies sont prescrites en fonction des caractéristiques de la tumeur et de l'historique de la patiente (Table 1). Cependant

l'utilisation d'anthracyclines est recommandée pour la majorité des patientes (Aebi et al., 2011).

<i>Combinaisons</i>	<i>No. de cycles</i>	<i>Durée (semaines)</i>
AC	4	3
CMF (oral ou i.v. jours 1+8)	6	4
FEC	6	3
CEF	6	4
A (ou E)→CMF	4→4 (-8)	3→4
AP→CMF	4→4	3→4
DC	4	3
AC→P(H)	4→4	3→3
AC→D(H)	4→4	3→3
DCarboH	6	3
ddAC→ddP	4→4	2→2
DAC	6	3
FEC→D	3→3	3→3

Table 1 : Différents régimes de chimiothérapie dans le cancer du sein

A, doxorubicin; C, cyclophosphamide; D, docetaxel; E, epirubicin; F, fluorouracil; M, methotrexate; P, paclitaxel; Carbo, carboplatin; H, trastuzumab ; dd, dose dense; i.v., intra-veineuse; →, suivi de. D'après (Aebi et al., 2011)

3.4.2.2 L'hormonothérapie

Les tumeurs mammaires qui expriment les récepteurs hormonaux ER et PR sont traitées par hormonothérapie. Le traitement le plus classique consiste en une prise de 20 mg de tamoxifène chaque jour pendant 5-10 ans, éventuellement suivi par un inhibiteur d'aromatase (letrozole) qui empêche la synthèse d'oestrogène. Le tamoxifène est métabolisé par le foie en ses métabolites actifs : le 4-hydroxy-tamoxifène et le N-desméthyltamoxifène qui sont des antagonistes aux ERs et qui entrent en compétition avec l'oestrogène pour empêcher l'activation des récepteurs.

3.4.2.3 Thérapie anti-HER2

Les patientes qui surexpriment le récepteur HER2 bénéficient d'un traitement à base de trastuzumab. Il s'agit d'un anticorps monoclonal recombinant se liant à la partie extracellulaire du récepteur HER2 et qui bloque le récepteur, indépen-

damment de la présence du ligand. Le temps de traitement recommandé est d'un an (Aebi et al., 2011; Senkus et al., 2015).

3.4.3 Conclusions

Comme nous l'avons vu, de nombreuses possibilités thérapeutiques existent dans le cancer du sein. Différents outils en ligne existent pour aider le praticien à choisir la meilleure combinaison de traitements. L'apparition et l'amélioration des traitements ces dernière décennies ont permis de faire chuter drastiquement la mortalité liée au cancer du sein (Fig. 34). La survie à 5 ans est passée de 75% en 1975 à plus de 90% en 2010 aux Etats-Unis (NIH, 2016).

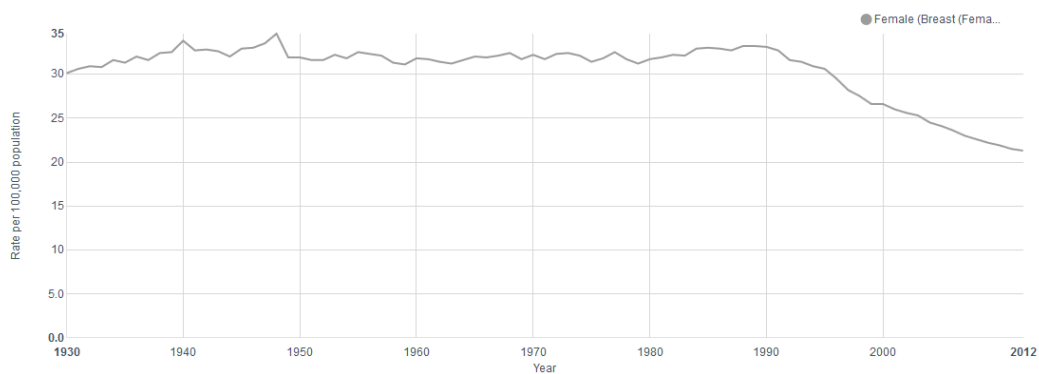


Fig. 34 : Evolution du taux de mortalité lié au cancer du sein.
Taux de mortalité lié au cancer du sein, par 100 000 habitants, chez la femme aux États-Unis.
(D'après <https://cancerstatisticscenter.cancer.org/#/>)

4 But du projet

Comme énoncé dans l'introduction, l'émergence des nouvelles technologies, en particulier le séquençage, a permis de révéler l'importance croissante des ARNs non codants dans tous les processus biologiques, ainsi que leur implication dans de nombreuses maladies telles que le cancer. Cette révolution scientifique permet de mieux comprendre le fonctionnement des êtres vivants à l'échelle moléculaire. Elle devrait aussi, à terme, permettre d'améliorer la prise en charge des patientes, par la découverte de nouveaux marqueurs diagnostics et pronostics, mais surtout par l'apparition de nouvelles stratégies thérapeutiques. Cependant, l'étude des ARNs non codants, et plus particulièrement des longs ARNs non codants n'est qu'à ses balbutiements. En effet, moins d'un pourcent des longs ARNs non codants connus a été caractérisé expérimentalement.

D'autre part, nous avons vu que le cancer du sein reste un problème majeur de santé publique. Malgré les efforts considérables qui ont été faits pour mieux comprendre cette maladie et développer de nouvelles thérapies, de nombreuses patientes ne répondent pas aux traitements, dû à l'hétérogénéité de cette maladie.

Nous avons donc voulu, lors de cette thèse, mieux comprendre le rôle des ARNs non codants dans le cancer du sein.

Dans un premier temps, nous avons étudié deux micro ARNs dérégulés dans le cancer du sein, recherché leur cible respective et caractérisé leur impact sur la prolifération cellulaire.

La seconde et majeure partie de ce travail s'est focalisée sur l'étude des longs ARNs non codants dans le cancer du sein. Au début de cette thèse, seuls quelques longs ARNs non codants étaient décrits dans le cancer du sein. Dès lors, nous avons décidé d'évaluer globalement leur profil d'expression dans les différents sous-types de cancers du sein. Ensuite, des analyses bioinformatiques nous ont permis de prédire leurs fonctions, leurs implications dans différentes voies moléculaires et leur valeur pronostic. Enfin, nous avons choisi quelques longs ARNs non codants que nous avons étudiés expérimentalement.

RESULTATS

Chapitre 1 : Les micro-ARNs régulent les histones déméthylases de la famille KDM5 dans les cellules cancéreuses mammaires

Les résultats de ce travail sont publiés dans le manuscrit (voir Annexe I) :

MicroRNAs regulate KDM5 histone demethylases in breast cancer cells,

Molecular Biosystems, 2015

Hélène Denis, **Olivier Van Grembergen**, Benjamin Delatte, Sarah Dedeurwaerder, Pascale Putmans, Emilie Calonne, Françoise Rothé, Christos Sotiriou, François Fuks and Rachel Deplus

I. INTRODUCTION

L'altération de l'épigénome est une caractéristique majeure du cancer, et les modifications aberrantes des histones permettent d'influencer le profil d'expression des gènes et jouent un rôle clé dans la tumorigénèse (Dawson and Kouzarides, 2012). Les modifications d'histones sont régulées de manière dynamique par des enzymes. Par exemple, les histones lysine méthyltransférases ajoutent un groupe méthyle sur les lysines, qui peut être enlevé par les histones déméthylases (KDMs). Dans cette étude, nous nous sommes focalisés sur les histones déméthylases de la famille KDM5 et leurs fonctions potentielles dans le cancer. Les protéines KDM5 sont capables d'enlever les marques d'activation de la transcription di- et tri-méthyles de la lysine 4 de l'histone H3. La famille KDM5 comprend 4 membres KDM5A, KDM5B, KDM5C et KDM5D. Les deux premiers ont été clairement associés au développement de tumeurs (X. Liu and Secombe, 2015), tandis que l'implication tumorale de KDM5C et KDM5C est moins claire.

Les micro-ARNs (miRs) constituent un autre acteur important dans le développement du cancer. Les miRs sont une classe de petits ARNs non codants d'environ 22 paires de bases qui contrôlent l'expression des gènes en ciblant l'ARN messenger (ARNm) par appariement des nucléotides. Ils mènent à la dégradation de l'ARNm ou à une répression de la traduction. De nombreuses études ont montré que les miRs sont dérégulés dans le cancer du sein et contribuent à la tumorigénèse et au développement de métastases (Takahashi, Miyazaki, and Ochiya, 2015). Par exemple, l'expression de miR-137 est dérégulée dans de nombreux cancers (Balaguer et al., 2010; Steponaitiene et al., 2016; Neault, Mallette, and Richard, 2016). La surexpression de miR-137 inhibe la prolifération cellulaire, il a donc une fonction onco-suppressive. Cependant, le rôle de miR-137 dans le cancer du sein n'est pas connu.

Dans cette étude, nous avons donc investigué le rôle de miR-137 dans un modèle de cellules cancéreuses mammaires. Nous nous sommes intéressés à sa régulation épigénétique ainsi qu'à son influence sur la croissance et la migration cellulaire. Une analyse de la littérature et des prédictions bioinformatiques nous ont permis d'identifier ses cibles potentielles en modèle mammaire. Nous avons confirmé expérimentalement que miR-137 cible directement KDM5B. De plus, nous avons montré que KDM5C est surexprimé dans les cellules mammaires et est régulé par miR-138.

Contribution personnelle à ce travail :

- *Analyse de l'expression des miRs par TaqMan RT-qPCR et de l'expression des KDM par RT-qPCR au SYBR Green*
- *Traitement à la 5-Aza et à la TSA des cellules mammaires*
- *Surexpression des miRs par transfection de progéniteurs pre-miR.*
- *Evaluation de la fonction des miRs sur la prolifération et la migration par la technologie xCELLigence*
- *Recherche et analyses de données publiques concernant l'expression des miRs en cellules mammaires*
- *Implication dans le processus de révision (texte et figures)*

Ce chapitre résume les résultats majeurs obtenus. Plus de détails ainsi que la section « matériels et méthodes » sont fournis dans le manuscrit joint en annexe (Annexe I).

II. RESULTATS

II.1 MiR-137 est sous-exprimé dans les cellules cancéreuses mammaires.

MiR-137 est sous-exprimé dans de nombreux cancers tels que le carcinome squameux buccal (Kozaki et al., 2008) , le cancer colorectal (Balaguer et al., 2010) , le cancer gastrique (Steponaitiene et al., 2016) et le cancer du pancréas (Neault, Mallette, and Richard, 2016). Pour évaluer si l'expression de miR-137 est également diminuée dans le cancer du sein, nous avons utilisé la technologie TaqMan qPCR (quantitative polymerase chain reaction). Nous avons montré que l'expression de miR-137 est diminuée dans toutes les lignées mammaires cancéreuses testées en comparaison à la lignée mammaire normale MCF10A (Fig. 35).

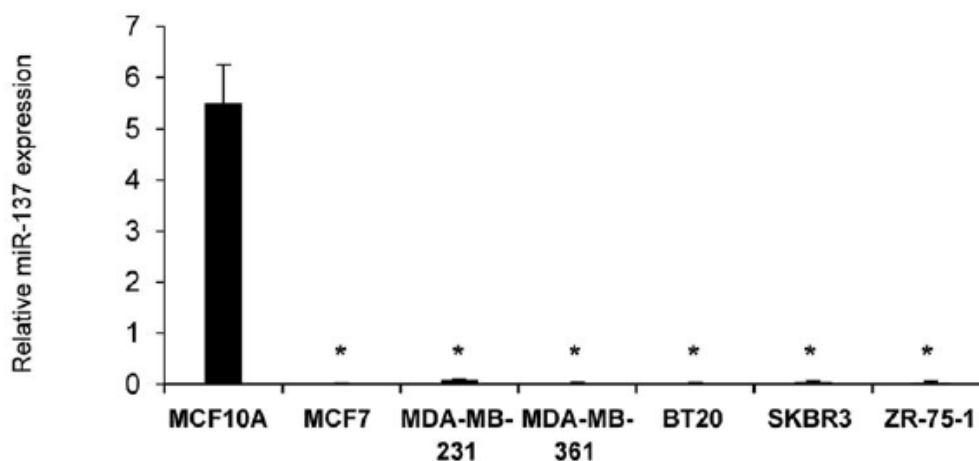


Fig. 35 : Sous-expression de miR-137 en lignées cancéreuses mammaires.

*Expression de miR-137 dans la lignée mammaire non cancéreuse MCF10A et dans les lignées cancéreuses MCF7, MDA-MB-231, MDA-MB-361, BT20, SKBR3, ZR-75-1. Les résultats, obtenus par qPCR, sont normalisés sur base de l'expression du gène RNU6B (n=3, * = $p < 0.05$).*

Dès lors, nous avons investigué la régulation épigénétique de miR-137. Nos résultats révèlent que le promoteur de miR-137 est peu méthylé dans les cellules mammaires normales en comparaison avec les cellules cancéreuses (Fig. 36A). Nous avons également montré qu'un traitement avec l'agent déméthylant

5-aza-2-désoxycytidine (5-Aza) restaure l'expression de miR-137. D'une manière intéressante, un traitement avec un inhibiteur des histones désacétylases (trichostatine A, TSA) réactive également l'expression de miR-137 (Fig. 36B). Ceci suggère donc que miR-137 est régulé à la fois par la méthylation de son promoteur et par l'acétylation des histones.

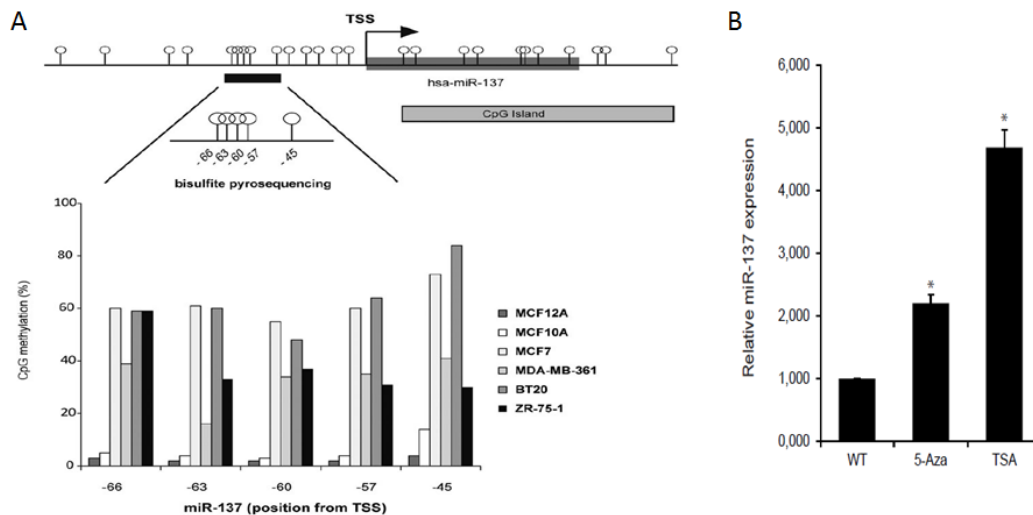


Fig. 36 : Régulation épigénétique de miR-137.

(A) Méthylation du promoteur de miR-137. Le haut de la figure décrit les CpG présents dans le promoteur dont le profil de méthylation a été déterminé. Le bas de la figure illustre le niveau de méthylation des différents CpG dans différentes lignées mammaires normales (MCF10A et MCF12A) et cancéreuses (MCF7, MDA-MB-361, BT20, ZR-75-1). (B) Expression de miR-137 après traitement à la 5-Aza (2,5 µM, 72h) et à la TSA (200mM, 48h) dans les cellules MCF7 (n=2, * = p < 0.05).

II.2 La surexpression de miR-137 inhibe la prolifération et la migration des cellules cancéreuses mammaires

Nous avons ensuite évalué si, à l'instar d'autres cancers, miR-137 influence la prolifération cellulaire. Pour ce faire, nous avons utilisé le système xCELLigence, basé sur l'impédance, qui permet de suivre en temps réel la prolifération. Nos résultats montrent que, les cellules mammaires transfectées avec un précurseur de miR-137, qui donc surexpriment miR-137 (Fig. 37A), voient leur taux de prolifération réduit de plus de 70% (Fig. 38A-B).

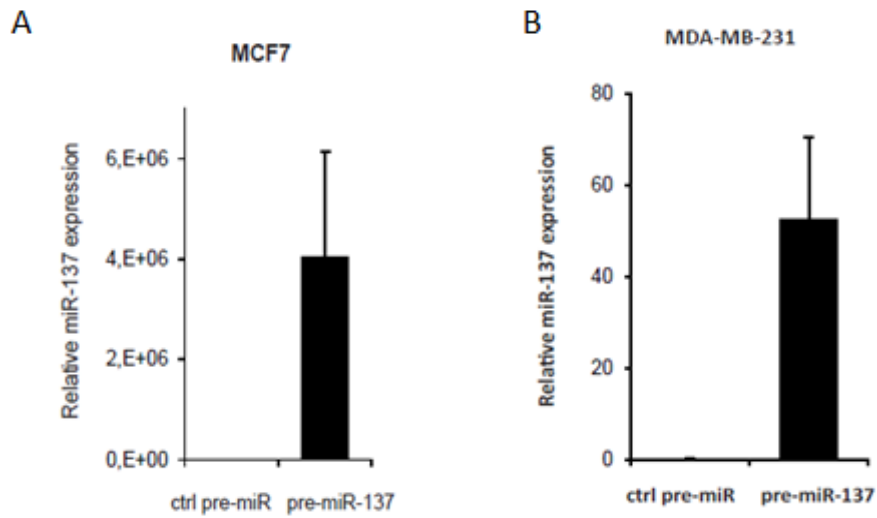


Fig. 37 : Surexpression de miR-137 après transfection du précurseur pré-miR-137.
 (A) Expression de miR-137 en cellules MCF7 après transfection d'un précurseur de miR-137 (pre-miR-137) ou d'un précurseur control (ctrl pre-miR). Les résultats, obtenus par qPCR, sont normalisés sur base de l'expression du gène RNU6B (n=3). (B) Idem pour les cellules MDA-MB-231.

Nous avons ensuite évalué l'influence de miR-137 sur la migration cellulaire en faisant à nouveau appel à la technologie xCELLigence qui fournit en temps réel la cinétique de migration. Nous avons utilisé le sérum fœtal bovin (FBS) comme chémo-attractant. Nous observons que les cellules MDA-MB-231 qui surexpriment miR-137 (Fig. 37B) ont un taux de migration réduit de 40% (Fig. 38C-D). Nos résultats suggèrent donc que miR-137 réduit à la fois la prolifération et la migration dans un modèle cellulaire de cancer mammaire.

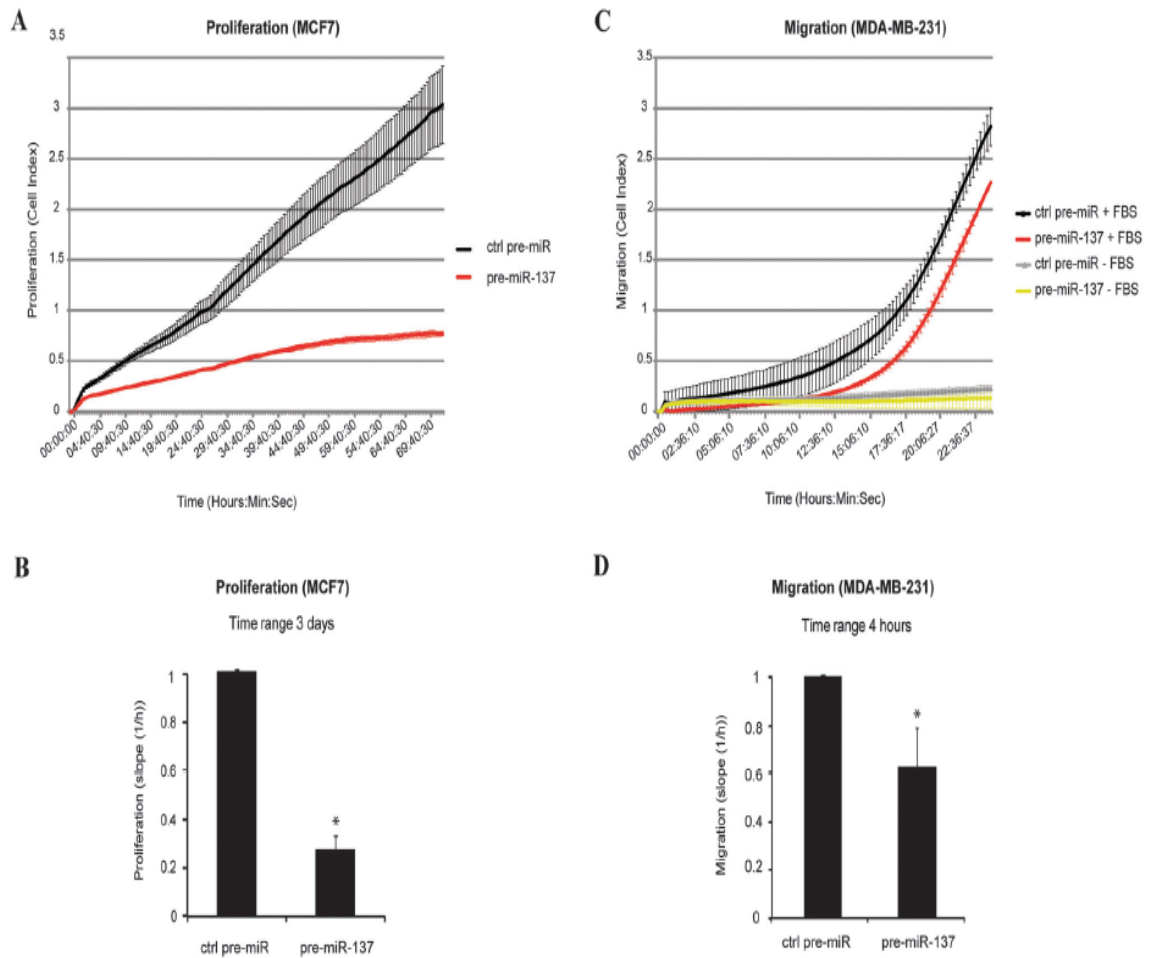


Fig. 38 : Régulation de la prolifération et de la migration par miR-137.
 (A) Profil de prolifération d'une expérience représentative de MCF7 surexprimant miR-137 (rouge) ou un précurseur de miR control (noir). Les mesures sont prises toutes les 30 minutes pendant 72h. (B) Taux de prolifération des cellules mammaires MCF-7 surexprimant miR-137 ($n=3$, $* = p < 0.05$). (C) Profil de migration d'une expérience représentative de MDA-MB-231 surexprimant miR-137 (rouge) ou un précurseur de miR control (noir) en présence ou en absence (gris et jaune) de FBS. Les mesures sont prises toutes les 30 minutes pendant 24h. (D) Taux de migration des cellules mammaires MDA-MB-231 surexprimant miR-137 ($n=3$, $* = p < 0.05$)

II.3 KDM5B est une cible directe de miR-137

Pour identifier les cibles de miR-137, nous nous sommes basés sur les données de la littérature. Une étude en cellules souches ES a montré que miR-137 réprime KDM5B (Tarantino et al., 2010). Des prédictions bioinformatiques réalisées dans notre laboratoire, à l'aide des programmes Miranda, MiRBase et TargetScan (Bartel, 2009) ont confirmé que KDM5B est une cible potentielle de

miR-137. KDM5B est une protéine épigénétique intéressante qui est souvent surexprimée dans le cancer du sein (Barrett et al., 2002) et est associée à la prolifération cellulaire (Yamane et al., 2007). Dès lors, nous avons émis l'hypothèse que miR-137 affecte la prolifération en ciblant KDM5B dans un modèle cellulaire de cancer mammaire. Pour tester cette hypothèse, nous avons évalué, par qPCR, l'expression de KDM5B dans des cellules mammaires MCF7 transfectées pour surexprimer miR-137. Dans ces cellules, le transcrite KDM5B est réduit de 30% (Fig. 39A). Des résultats similaires ont été obtenus en cellules MDA-MB-231 et MDA-MB-361. Enfin, nous avons également observé, par western-blot, une diminution de l'expression protéique de KDM5B dans les cellules MCF7 qui surexpriment miR-137 (Fig. 39B).

Pour confirmer que miR-137 régule KDM5B de manière directe, nous avons réalisé un essai luciférase. Nous avons cloné la séquence cible de miR-137, c'est-à-dire 25 paires de bases de la région 3' non traduite du gène KDM5B, dans un vecteur rapporteur de luciférase. La surexpression de miR-137 conduit à une diminution de l'activité du vecteur rapporteur (Fig. 39C). Ceci suggère que, dans un modèle cellulaire de cancer mammaire, miR-137 régule KDM5B directement, en ciblant la région 3' non traduite de KDM5B.

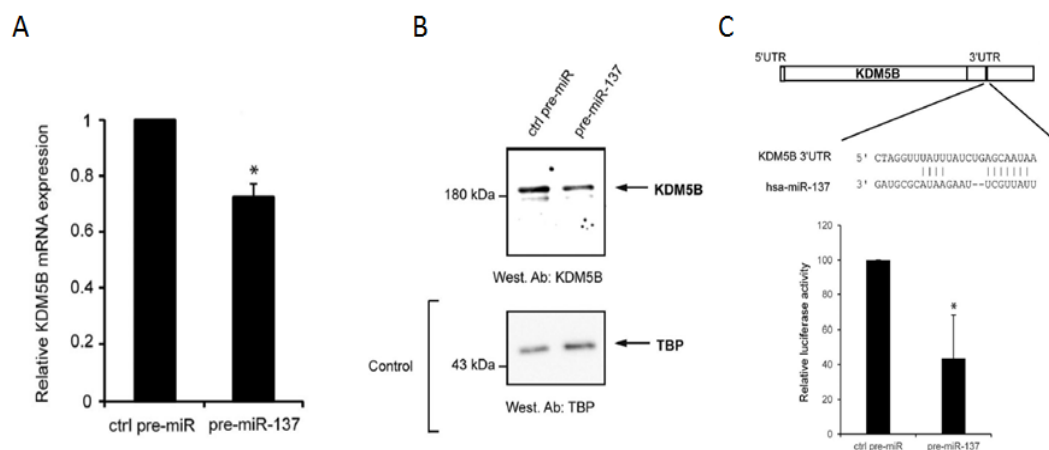


Fig. 39 : Régulation de KDM5B par miR-137.

(A) Analyse par qPCR de l'expression du gène KDM5B après transfection du précurseur de miR-137 en cellules MCF7 (n=3, * = p < 0.05). (B) Analyse du niveau protéique de KDM5B par western-blot après transfection du précurseur de miR-137 en cellules MCF7. La protéine TBP est utilisée comme contrôle de chargement. (C) Séquence de 25 paires de bases de la région 3' non traduite (3'UTR) de KDM5B clonée dans un vecteur rapporteur

*de la luciférase (haut). L'histogramme (bas) illustre l'activité relative de la luciférase en cellules MCF7 transfectées par le précurseur de miR-137 (n=3, * = p < 0.05).*

II.4 MiR-138 régule KDM5C en modèle cellulaire de cancer du sein

Vu le rôle de KDM5B dans le cancer du sein, nous nous sommes ensuite demandé si d'autres membres de la famille KDM5 étaient également impliqués dans le développement de tumeurs. Nous avons observé que KDM5C est surexprimé dans différentes lignées mammaires par rapport aux lignées normales (Fig. 40A). Nous nous sommes demandé si, à l'instar de KDM5B, KDM5C est également régulé par des miRs. Des outils de prédictions bioinformatiques (Bartel, 2009) ont révélé que miR-138 peut potentiellement cibler KDM5C. De plus, l'expression de miR-138 est globalement réduite dans toutes les cellules cancéreuses mammaires testées, excepté les MDA-231 (Fig. 40B). Pour tester l'implication de miR-138, nous l'avons surexprimé en cellules mammaires MCF-7 et nous observons une diminution de KDM5C (Fig. 40C) tant au niveau de l'ARNm qu'au niveau protéique (Fig. 40D).

Pour confirmer que miR-138 régule KDM5C de manière directe, nous avons à nouveau utilisé un vecteur rapporteur de luciférase dans lequel nous avons inséré la séquence de KDM5C prédite pour être ciblée par miR-138. Nous observons une diminution de l'activité du vecteur rapporteur après surexpression de miR-138 (Fig. 40E). Ceci suggère que, dans un modèle cellulaire de cancer mammaire, miR-138 régule directement KDM5C.

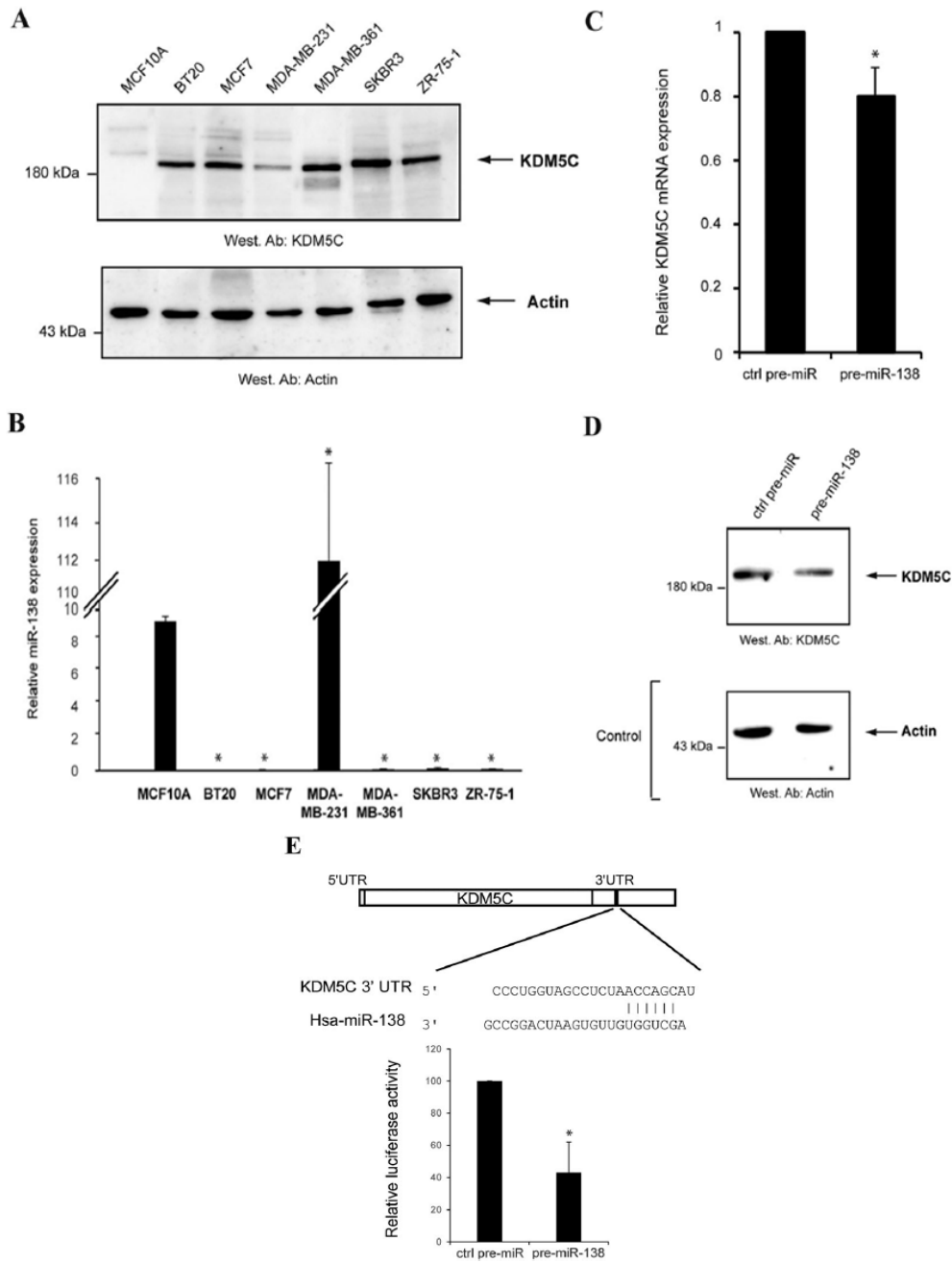


Fig. 40 : Régulation de KDM5C par miR-138.

(A) Expression protéique de KDM5C dans différentes cellules mammaires. (B) Expression de miR-138 dans différentes cellules mammaires. Les résultats sont normalisés sur base de l'expression du gène RNU6B ($n=3$, $* = p < 0.05$). (C) Analyse par qPCR de l'expression du gène KDM5C après transfection du précurseur de miR-138 en cellules MCF7 ($n=3$, $* = p < 0.05$). (D) Analyse du niveau protéique de KDM5C par western-blot après transfection du précurseur de miR-138 en cellules MCF7. L'actine est utilisée comme contrôle de chargement. (E) Séquence de 25 paires de bases de la région 3' non traduite (3'UTR) de KDM5C clonée dans un vecteur rapporteur de la luciférase (haut). L'histogramme (bas) illustre l'activité relative de la luciférase en cellules MCF7 transfecées par le précurseur de miR-138 ($n=3$, $* = p < 0.05$).

II.5 La surexpression de miR-138 inhibe la prolifération des cellules cancéreuses mammaires

Enfin, nous avons examiné si la surexpression de miR-138 affecte la prolifération cellulaire. Nos résultats montrent que, les cellules mammaires transfectées avec un précurseur de miR-138, voient leur taux de prolifération réduit de plus de 50% (Fig. 41)

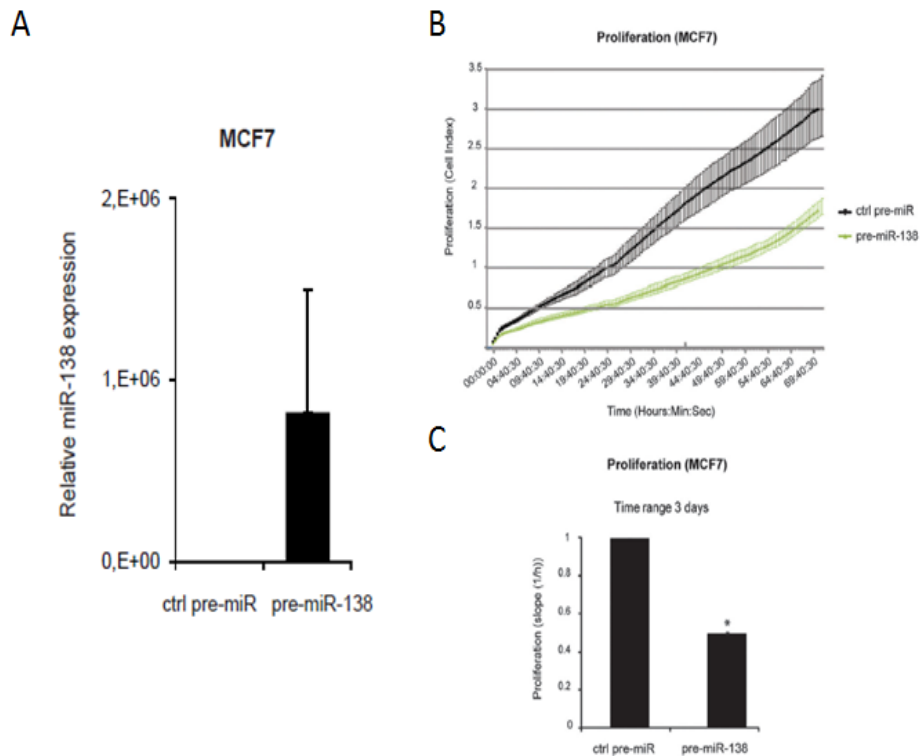


Fig. 41 : Régulation de la prolifération cellulaire par miR-138.

(A) Expression de miR-138 en cellules MCF7 après transfection d'un précurseur de miR-138 (pre-miR-138) ou d'un précurseur contrôle (ctrl pre-miR). Les résultats, évalués par RT-qPCR, sont normalisés sur base de l'expression du gène RNU6B (n=3). (B) Profil de prolifération d'une expérience représentative de MCF7 surexprimant miR-138 (vert) ou précurseur de miR contrôle (noir). Les mesures sont prises toutes les 30 minutes pendant 72h. (C) Taux de prolifération des cellules mammaires MCF-7 surexprimant miR-138 (n=3, * = $p < 0.05$).

Dans son ensemble, notre travail a démontré l'implication de deux miRs dans un modèle cellulaire de cancer du sein. Nos résultats ont dévoilé la fonction de miR-137 et miR-138 dans la régulation des protéines épigénétiques KDM5B et KDM5C, respectivement. Nos données suggèrent que ces miRs se

fixent directement sur leur cible pour diminuer leur expression au niveau de l'ARNm ainsi qu'au niveau protéique. Enfin, nous avons montré que la sur-expression de miR-137 et de miR-138 est associée à une réduction de la prolifération cellulaire.

Pour conclure cette première partie, nous avons démontré que les miRs peuvent cibler les histones déméthylases KDMs. Nos résultats, en concordance avec les données de la littérature (J.-M. Lee et al., 2015; C. Zhang and Hua, 2016), suggèrent que miR-137 et miR-138 ont une fonction onco-suppressive dans le cancer du sein et que leurs sous-expressions contribuent à la tumorigénèse mammaire.

Chapitre 2 :

Description et caractérisation
fonctionnelle des longs
ARNs non codants
dans le cancer du sein

Les résultats de ce travail sont publiés dans le manuscrit (voir Annexe II) :

Portraying breast cancers with long noncoding RNAs,

Science Advance, 2016

Olivier Van Grembergen, Martin Bizet, Eric J. de Bony, Emilie Calonne, Pascale Putmans, Sylvain Brohée, Catharina Olsen, Mingzhou Guo, Gianluca Bontempi, Christos Sotiriou, Matthieu Defrance, François Fuks

I. INTRODUCTION

Le cancer du sein, cancer le plus fréquemment diagnostiqué chez la femme (Siegel and Naishadham, 2013), est responsable de plus de 500 000 décès chaque année et constitue donc un problème majeur de santé publique (Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray, 2012). Il s'agit d'une maladie complexe et différents sous-types de cancers du sein ont été décrits. En clinique, le cancer du sein est subdivisé en trois groupes en fonction de l'expression de récepteurs hormonaux : (i) les cancers qui expriment le récepteur à l'œstrogène mais pas HER2, (ii) les cancers qui surexpriment le récepteur HER2 et (iii) les cancers dits triples négatifs, qui n'expriment ni les récepteurs à l'œstrogène et à la progestérone, ni le récepteur HER2 (Senkus et al., 2015). Plus récemment, l'analyse de l'expression des gènes à large échelle a permis un raffinement de la classification : 4 sous-types moléculaires de tumeurs ont été identifiés : les cancers de types luminal A (peu prolifératifs) et luminal B (hautement prolifératifs) qui expriment généralement les récepteurs à l'œstrogène, les cancers HER2 associés à une surexpression du récepteur HER2 et les cancers de type basal-like qui comprennent la majorité des cancers triples négatifs (Perou et al., 2000; The Cancer Genome Atlas, 2012). Ces différents sous-types de cancers du sein sont associés à un pronostic et une approche thérapeutique différents. Malgré les progrès réalisés ces dernières années, nous sommes loin d'appréhender pleinement la complexité de cette maladie. En effet, certaines tumeurs appartenant au même sous-type moléculaire répondent différemment aux thérapies (Pusztai, Rouzier, and Symmans, 2015). Il est donc primordial de perfectionner nos connaissances sur cette maladie complexe. L'étude des longs ARNs non codants (lncRNAs) constitue une nouvelle approche qui pourrait améliorer la compréhension des mécanismes moléculaires sous-jacents à l'hétérogénéité des cancers du sein et mener à la découverte de nouvelles thérapies.

Les lncRNAs sont devenus, ces dernières années, de nouveaux acteurs majeurs dans de nombreuses maladies, en particulier dans le cancer. En régulant l'expression génique et l'activité protéique, les lncRNAs influencent, entre autres, le cycle cellulaire, la réponse immunitaire et le maintien de la pluripotence. Ils ont été impliqués dans la tumorigénèse et le développement de métastases. A l'image des gènes encodant les protéines, les lncRNAs peuvent être suppresseurs de tumeur ou oncogènes (Huarte, 2015). Cependant, leur expression, précisément régulée, est plus spécifique de cellules ou de tissus particuliers que les ARNm ce qui en fait des cibles thérapeutiques prometteuses et leur confère des valeurs diagnostics et pronostics dans de nombreuses pathologies, dont le cancer du sein (Huarte, 2015; Shore and Rosen, 2014).

Au début de ce travail de thèse en 2011, seul un lncRNA, dénommé HOTAIR, a été décrit en détail dans le cancer du sein et aucune étude ne décrivait, dans leur ensemble, l'expression des lncRNAs dans les tumeurs mammaires. Notre objectif initial était donc d'obtenir une vision, à l'échelle du génome, de la transcription des lncRNAs dans le cancer du sein. Depuis lors, de récentes études ont décrit l'expression globale des lncRNAs dans le cancer du sein et ont établi une association entre leur expression et les différents sous-types de cancer du sein (Su et al., 2014; Zhao, Luo, and Jiao, 2014). Cependant, ces études se sont limitées à décrire les lncRNAs corrélés à certains sous-types, mais n'ont pas investigué précisément les fonctions et les voies moléculaires associées à ces lncRNAs. Notre second objectif consistait donc en la détermination du rôle fonctionnel de ces lncRNAs et de leur implication dans différentes voies de signalisation.

Deux approches ont été utilisées pour l'étude des lncRNAs dans les cancers mammaires. Premièrement, nous avons analysé le profil transcriptionnel de 45 tumeurs mammaires ainsi que de neuf tissus mammaires sains grâce à l'utilisation de puces à ADN uniques, spécialement développées, par le groupe du Dr. Vandesompele (VIB, Université de Gand), pour l'analyse des lncRNAs. La seconde approche repose sur l'exploitation et l'analyse de nombreuses données publiques d'expression par puces à ADN, qui permettent également

d'appréhender l'expression des lncRNAs. Des analyses bioinformatiques nous ont ensuite permis de définir les fonctions potentielles des lncRNAs dans le cancer du sein. Nous avons également investigué leur potentiel pronostic et associé nombre d'entre eux avec le risque de rechute du cancer. Enfin, nous avons caractérisé expérimentalement certains lncRNAs d'intérêt.

Dans son ensemble, ce travail permet de mieux appréhender la complexité du cancer du sein et illustre le rôle clé joué par les lncRNAs dans cette maladie. Il fournit la description la plus complète des lncRNAs dans le cancer du sein à ce jour, et met en évidence l'implication des lncRNAs dans de nombreuses voies moléculaires dérégulées dans le cancer ce qui pourrait, à terme, déboucher sur de nouvelles approches cliniques en termes de pronostics et de traitements.

Contribution personnelle à ce travail :

- *Coordination du projet et développement des stratégies expérimentales*
- *Préparation des échantillons d'ARN, traitements, analyses et interprétation des données de puce à ADN générées à l'université de Gand.*
- *Traitements, analyses bioinformatiques (clustering, prédictions fonctionnelles, analyses de survie, validations, statistiques) et interprétation des données publiques d'expression.*
- *Etude « in vitro » des lncRNAs d'intérêt (prolifération et migration, microscopie en fluorescence, puce à ADN, cytométrie en flux).*
- *Rédaction du manuscrit et révision.*

Ce chapitre est divisé en deux parties. La première partie (non publiée) reprend les résultats que nous avons générés par puce à ADN. Les résultats sont décrits en détail et accompagnés d'une section « matériels et méthode ». La seconde partie, basée sur l'exploitation de données publiques d'expression, résume les

résultats majeurs obtenus ; plus de détails ainsi que la section « matériels et méthodes » sont fournis dans le manuscrit joint en annexe (Annexe II).

II. RESULTATS – Partie 1

Cette première partie des résultats sur l'étude des lncRNAs dans le cancer du sein se base sur des données d'expression que nous avons générées en étroite collaboration avec l'équipe du professeur Jo Vandesompele du centre de génétique médicale de l'université de Gand.

II.1 Les tumeurs mammaires ont un profil d'expression des lncRNAs différent du tissu mammaire sain

En collaboration avec l'équipe du Dr Sotiriou, nous avons obtenu de l'ARN provenant de 45 tumeurs mammaires et de 9 tissus mammaires sains. Nous avons utilisé le Bioanalyseur Agilent 2100 afin de vérifier l'intégrité des ARNs. Cette méthode se base sur le profil électrophorétique des ARNs et fournit un score de qualité dénommé RIN (RNA Integrity Number) allant de zéro (ARN fortement dégradé) à dix (ARN de grande qualité). L'ensemble de nos échantillons d'ARN possèdent un RIN > 7 et a été inclus pour l'analyse de leur transcriptome.

Pour analyser le profil d'expression des lncRNAs, nous avons utilisé une puce à ADN particulière, spécialement mise au point pour étudier les lncRNAs par le laboratoire du Dr Jo Vandesompele à l'université de Gand (Volders et al., 2013). Cette puce, basée sur la technologie « Agilent Sureprint 60k » comporte 33 000 sondes qui ciblent l'ensemble des gènes codants et 23 000 sondes qui ciblent spécifiquement les lncRNAs et qui permettent d'appréhender l'expression de 10 976 lncRNAs (Fig. 42).

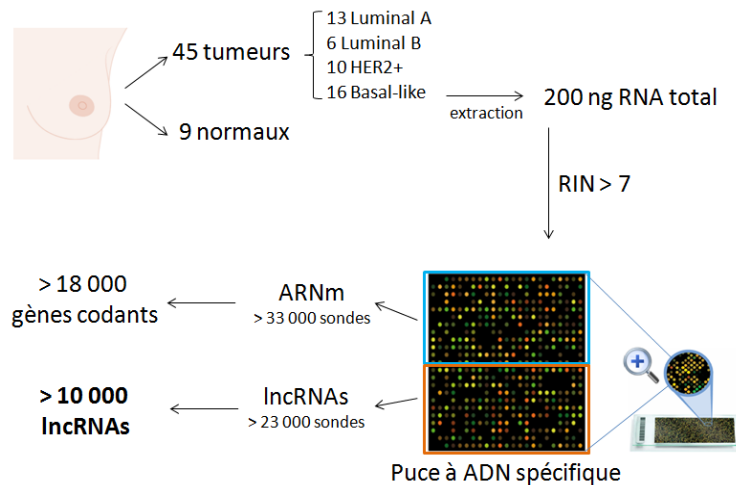


Fig. 42 : Protocole d'analyse des lncRNAs dans le cancer du sein.

200 ng d'ARN ont été extraits de 45 tissus mammaires cancéreux et de 9 tissus sains. Les échantillons, qui possèdent tous un score d'intégrité (RIN) supérieur à 7, ont été analysés sur une puce à ADN spécifique, qui permet l'analyse de l'ensemble des gènes codants et de plus de 10 000 lncRNAs.

Pour analyser les résultats, nous avons d'abord utilisé une approche de clustering hiérarchique non supervisée. Nos résultats démontrent que, à l'instar des ARNm, les tissus normaux ont un profil d'expression de lncRNAs distinct des tissus cancéreux (Fig. 43A), ce qui suggère que de nombreux lncRNAs sont différentiellement exprimés entre les échantillons normaux et tumoraux. Dès lors nous avons recherché les gènes significativement dérégulés dans les tumeurs mammaires. Nous avons identifié 1768 ARNm et 222 lncRNAs statistiquement dérégulés entre les tissus sains et cancéreux (différence d'expression d'un facteur > 2 , taux de faux positif (FDR, false discovery rate) < 0.05 , test t) (Fig. 43B). La majorité de ces lncRNAs sont sous-exprimés dans les tumeurs (72%), dans une proportion relativement semblable aux ARNm (63%).

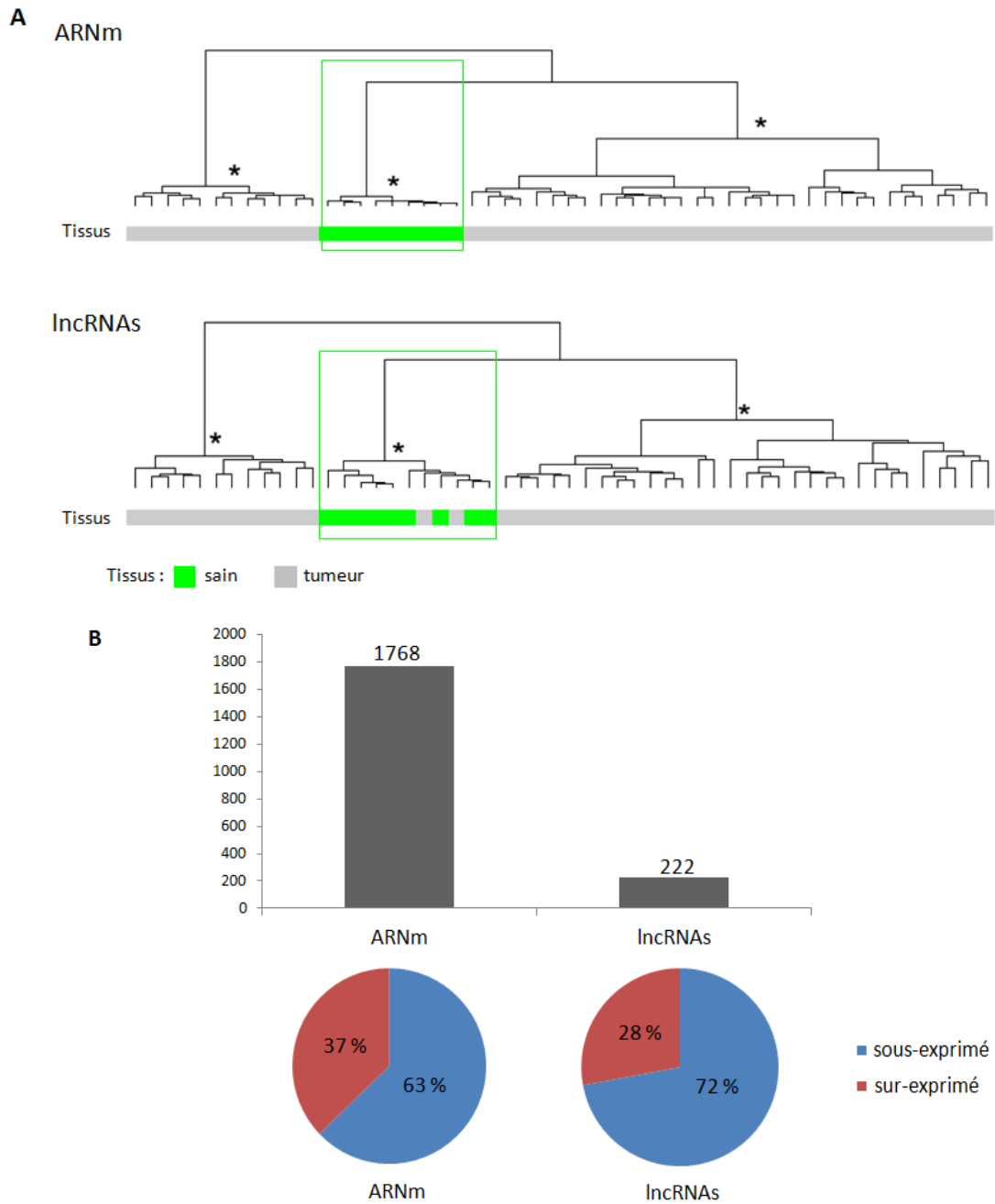


Fig. 43 : Dérégulation des lncRNAs dans le cancer du sein.

(A) Clustering hiérarchique non supervisé de l'expression des ARNm et des lncRNAs (500 plus variants, * cluster significatif (AU score > 70) selon l'algorithme « pv-clust »). (B) Nombre d'ARNm et de lncRNAs significativement dérégulés (variation d'un facteur > 2, FDR < 0.05, test t) dans le cancer du sein et proportion d'ARNm et de lncRNAs sur- ou sous- exprimés.

II.2 L'expression des lncRNAs est associée au récepteur aux œstrogènes

Nous nous sommes ensuite focalisés sur l'expression des lncRNAs au sein des tumeurs mammaires, en excluant les tissus normaux. Une analyse par clustering hiérarchique non supervisé a montré que les tumeurs forment deux groupes distincts qui sont fortement corrélés à l'expression du récepteur à l'œstrogène (Fig. 44A). Nous avons identifié 50 lncRNAs dont l'expression est significativement différente entre les tumeurs ER-positives et ER-négatives (différence d'expression d'un facteur > 2 , FDR < 0.05 , test t) (Fig. 44B). La majorité (64 %) d'entre eux est sous-exprimée dans les tumeurs ER+ (Fig. 44C). Ces résultats suggèrent donc que des lncRNAs différents sont impliqués dans les tumeurs ER+ et ER-.

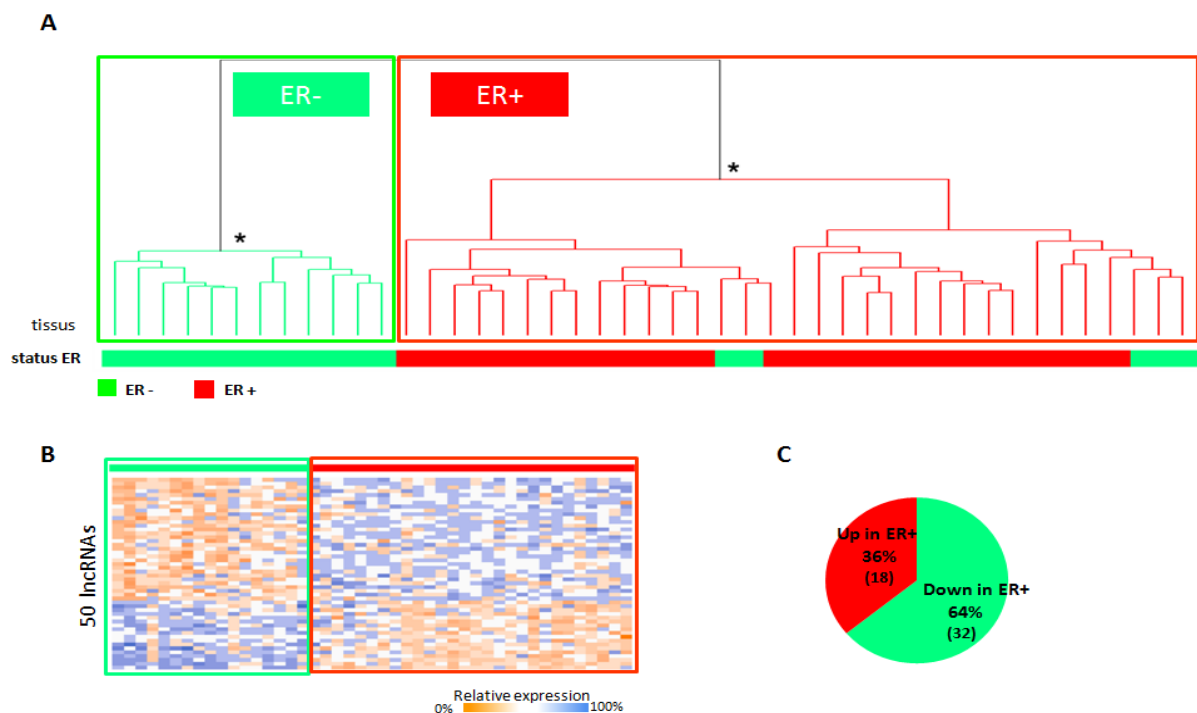


Fig. 44 : Association entre l'expression des lncRNAs et le statut ER.

(A) Clustering hiérarchique non supervisé de l'expression des lncRNAs au sein des tumeurs mammaires (500 plus variant, * cluster significatif (AU score > 70) selon l'algorithme « pv-clust »). (B) Carte d'expression des 50 lncRNAs dérégulés entre les ER+ et les ER-. (C) Proportion de lncRNAs sur- ou sous- exprimés.

II.3 L'expression des lncRNAs est associée aux sous-types moléculaires des cancers du sein

Nous nous sommes également intéressés à l'association entre les sous-types moléculaires du cancer du sein et l'expression des lncRNAs. Nos résultats ont montré, pour la première fois, que le profil transcriptomique des lncRNAs permet de classer les tumeurs mammaires en différents sous-groupes qui sont fortement corrélés aux sous-types moléculaires des cancers mammaires (Fig. 45) suggérant donc une inter-communication intime entre les mRNAs et les lncRNAs.

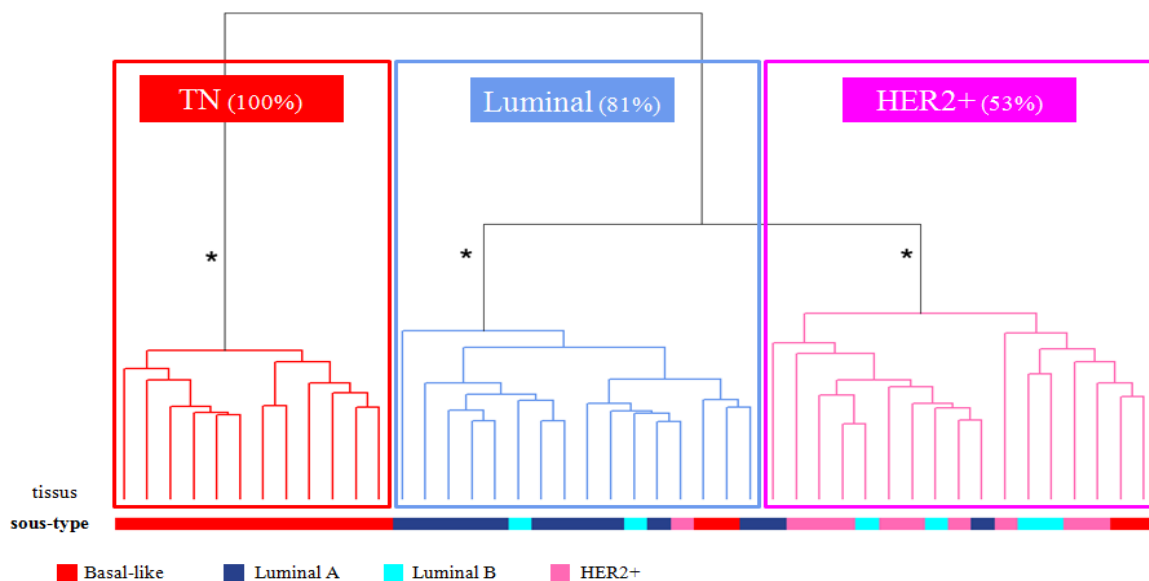


Fig. 45 : Association entre l'expression des lncRNAs et les sous-types moléculaires des cancers du sein.

*Clustering hiérarchique non supervisé de l'expression des lncRNAs au sein des tumeurs mammaires (500 plus variant, * cluster significatif (AU score > 70) selon l'algorithme « pv-clust »).*

II.4 Les lncRNAs sont associés à différents processus biologiques et voies moléculaires du cancer

Un point majeur dans l'étude des lncRNAs consiste en l'identification de leur fonction. Pour ce faire, nous avons utilisé une méthode de prédiction bioinformatique, appelée méthode de « culpabilité par association » (Rinn and Chang, 2012), qui consiste à attribuer un rôle aux lncRNAs en fonction des gènes codants co-régulés (Fig. 46A). Nos analyses montrent que de nombreux lncRNAs sont associés à divers processus biologiques tels que l'adhésion cellulaire,

l'angiogénèse, la réparation de l'ADN et la mitose (Fig. 46B). Nous avons également mis en évidence certains lncRNAs associés à des voies moléculaires clés telles que les voies mTOR, E2F, Myc ainsi des lncRNAs associés aux cibles de la protéine épigénétique EZH2, ce qui suggère que ces derniers soient impliqués dans le phénomène de répression génique par le complexe PolyComb 2 (Fig. 46C).

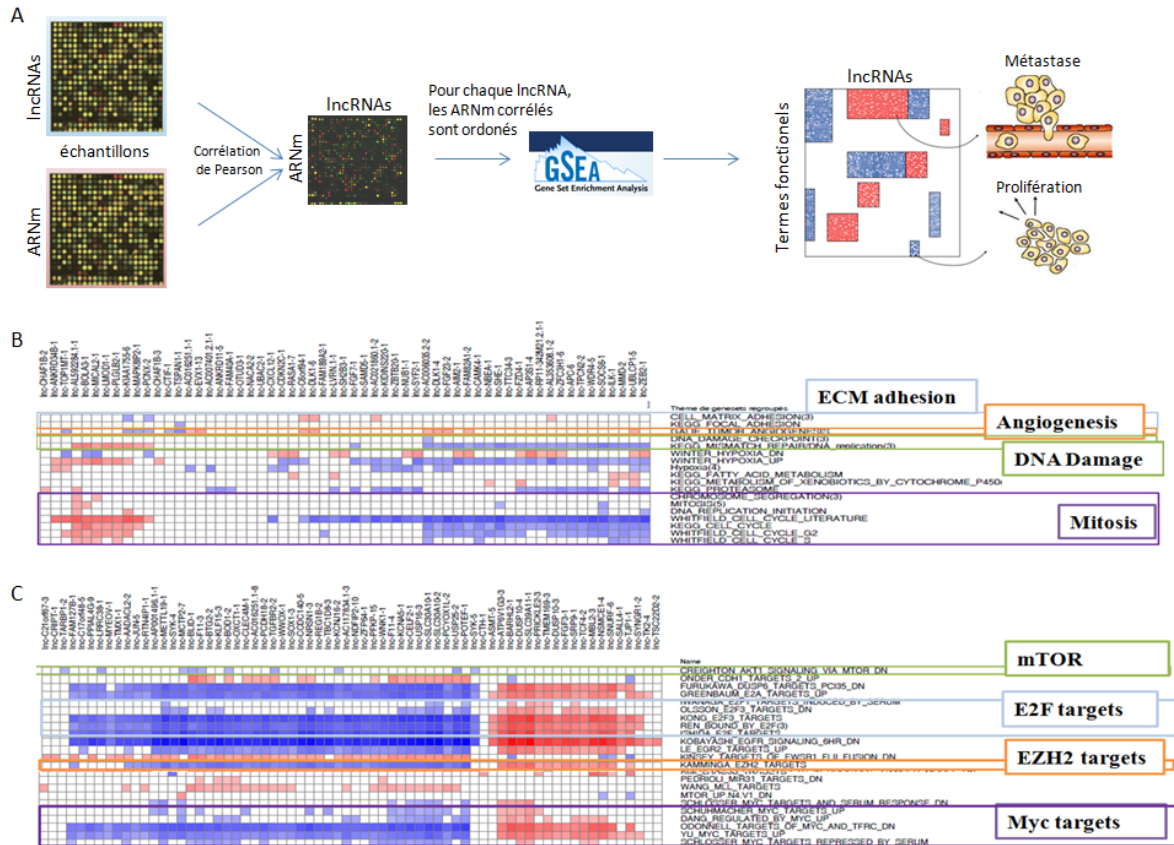


Fig. 46 : Prédiction fonctionnelle des lncRNAs dans le cancer du sein.

(A) Explication de la méthode de prédiction (culpabilité par association) utilisée. Le coefficient de corrélation de Pearson est calculé pour chaque paire de données d'expression lncRNAs-mRNA à travers l'ensemble des échantillons. Pour chaque lncRNA, les ARNm sont ensuite ordonnés selon leur coefficient de corrélation. Cette liste ordonnée d'ARNm est fournie au logiciel GSEA, qui permet d'extraire les fonctions biologiques et moléculaires statistiquement associées à cette liste. Ceci permet de prédire pour chaque lncRNA, ses fonctions potentielles basées sur les ARNm co-régulés. (B) Exemple de lncRNAs associés à des fonctions biologiques. (C) Exemple de lncRNAs associés à des voies moléculaires.

II.4 Lnc-KIN-2 module la prolifération cellulaire et l'expression du récepteur à l'œstrogène

Pour compléter cette analyse descriptive, nous avons choisi un lncRNA, appelé lnc-KIN-2, que nous avons étudié expérimentalement. Nous avons choisi ce lncRNA pour différentes raisons : (i) lnc-KIN-2 est surexprimé dans les tumeurs de type luminal A et sous-exprimé dans les tumeurs de type basal-like (Fig 47A, en haut à gauche). De plus, l'expression de lnc-KIN-2 dans les différentes lignées mammaires représentatives de chaque sous-type de cancer correspond globalement à l'expression en tissus. Plus particulièrement, on observe qu'il est fortement exprimé en cellules MCF7, un modèle cellulaire de cancer mammaire de type luminal A, (ii) nos prédictions fonctionnelles associent lnc-KIN-2 avec les cancers du sein de type luminal et avec la signalisation du récepteur à l'œstrogène ainsi qu'avec la présence de métastases, (iii) lnc-KIN-2 est situé à environ 2 kb, en amont et sur le brin opposé, du gène codant GATA3 et leur expression est fortement corrélée (Fig 47A, droite). GATA3 est un gène important dans le cancer du sein, il est impliqué dans la maturation des glandes mammaires, influence la prolifération et régule l'expression du gène ESR1 (Shan et al., 2014).

Pour étudier lnc-KIN-2, nous avons utilisé la technologie des LNA qui permet de sous-exprimer un lncRNA d'intérêt. Nous observons que les cellules MCF-7 qui sous-expriment lnc-KIN-2 ont un taux de prolifération drastiquement réduit (Fig. 47B-C). D'une manière intéressante, l'expression des gènes GATA3 et ESR1 est également réduite dans ces cellules (Fig. 47D). Ces résultats suggèrent que lnc-KIN-2 influence la prolifération cellulaire, en modulant la voie du récepteur à l'œstrogène.

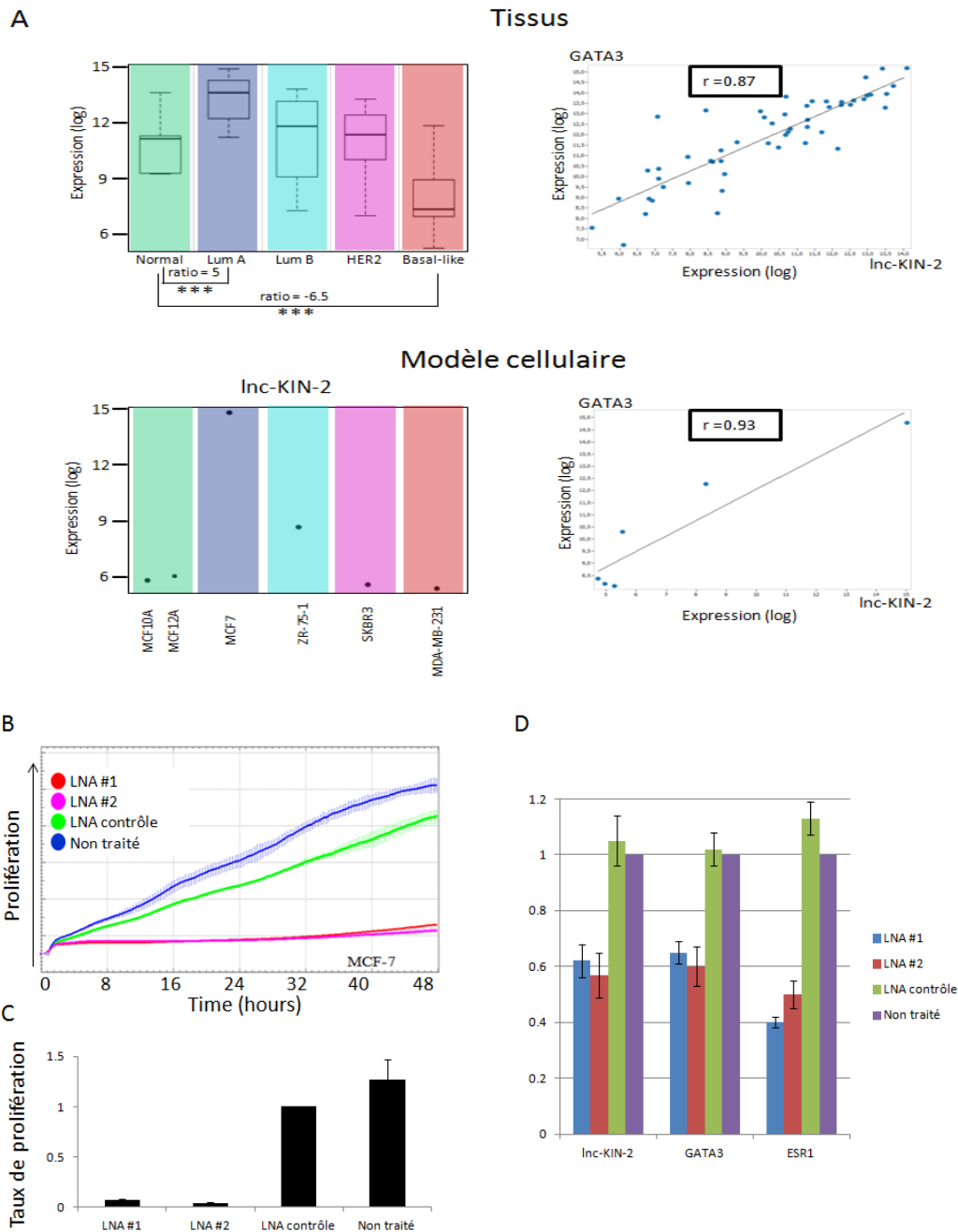


Fig. 47 : lnc-KIN-2 influence la prolifération et l'expression des gènes GATA3 et ESR1.
 (A) En haut, expression de lncRNAs dans les tissus mammaires sains et dans les différents sous-types de cancers mammaires (gauche). Corrélation entre l'expression de lnc-KIN-2 et de son gène voisin GATA3 (droite). En bas, idem dans différentes lignées mammaires représentatives des différents sous-types de cancers du sein. (B) Un profil représentatif du taux de prolifération des cellules MCF7 traitées par deux LNA ciblant lnc-KIN-2. (C) Taux de prolifération sur 48h des cellules MCF-7 qui sous-expriment lnc-KIN-2, $n = 3$. (D) Expression de lnc-KIN-2, GATA3 et ESR1 dans les cellules sous-exprimant lnc-KIN-2 ($n = 3$).

II.5 Conclusions

Nous avons débuté l'étude des lncRNAs dans le cancer du sein en analysant le transcriptome de 54 tissus mammaires (45 tumeurs et 9 normaux) par le biais d'une puce à ADN particulière qui permet d'étudier plus de 10 000 lncRNAs. Nous avons identifié 222 lncRNAs dérégulés dans les tumeurs, qui sont donc potentiellement impliqués dans la tumorigénèse. Nous avons également montré que l'expression des lncRNAs discrimine les tumeurs ER+ et ER-, mais aussi les sous-types moléculaires de cancers du sein. Ceci démontre que l'expression des lncRNAs permet de classer les tumeurs mammaires en sous-groupes cliniquement pertinents qui possèdent des pronostics de guérison variables. L'expression des lncRNAs pourrait permettre d'identifier de nouveaux groupes à l'intérieur d'un sous-type moléculaire. Cependant, notre cohorte de patients étant relativement restreinte, nous ne pouvons pas répondre à cette question à ce stade. Nous ne possédons pas non plus de données de survie exploitables. Dès lors, nous avons décidé de compiler des données publiques d'expression pour pouvoir analyser les lncRNAs dans une large cohorte, avec des données de survie sur une longue période (voir III Résultat – partie 2 ci-dessous).

Enfin, nous avons caractérisé expérimentalement lnc-KIN-2, surexprimé dans les tumeurs de type luminal A. Nous avons montré que ce lncRNA est impliqué dans la prolifération, mais également qu'il contrôle l'expression de deux gènes clés du cancer du sein : ESR1 et GATA3. Dès lors, lnc-KIN-2 est une cible thérapeutique potentielle dans les cancers ER+.

III. RESULTATS – Partie 2

La seconde approche que nous avons utilisée pour analyser les lncRNAs dans le cancer du sein se base sur l'exploitation de données publiques d'expression par puce à ADN que nous avons compilées et analysées. Au cours de ce travail de thèse, nous avons découvert que ces puces, qui ont été très largement utilisées pour étudier les ARNs messagers codants dans les cancers, permettent également l'étude des lncRNAs. Cette partie confirme et étend les résultats de la première partie sur une cohorte de patientes beaucoup plus large. Le fruit de ce travail a été publié dans la revue *Science Advances* dont le manuscrit est joint en annexe. Ce chapitre résume les résultats majeurs obtenus; plus de détails ainsi que la section « matériels et méthodes » sont fournis dans le manuscrit.

III.1 Les tumeurs mammaires possèdent un profil d'expression des lncRNAs distinct des tissus sains

Afin d'extraire de nouvelles informations à partir de données publiques existantes, nous avons développé un pipeline bioinformatique de réannotation des puces à ADN Affymetrix Human Genome U133 Plus 2.0, qui permet d'appréhender simultanément l'expression des ARNs messagers et de 3053 lncRNAs.

Nous avons analysé le profil transcriptionnel de 823 cancers du sein (219 luminal A, 273 luminal B, 124 HER2+, 195 basal like) et de 172 tissus mammaires sains. Un clustering hiérarchique consensus non supervisé révèle trois groupes distincts, dont un (le cluster II, Fig. 48A) regroupe 95 % des tissus sains ce qui suggère que le profil d'expression des lncRNAs est différent entre tumeurs et tissus mammaires sains et valide nos résultats précédents. Dès lors, nous avons voulu identifier les lncRNAs exprimés de manière aberrante dans les tumeurs. Dû à l'hétérogénéité des cancers du sein, nous n'avons pas utilisé le test *t* classique, car nous souhaitions pouvoir identifier les lncRNAs qui ne sont dérégulés que dans une fraction des tumeurs. Dès lors, nous avons adapté une méthode (Yao et al., 2015) qui permet l'identification des lncRNAs significativement dérégulés dans au moins 10% des tumeurs. Ainsi, nous avons identifié 124

lncRNAs surexprimés et 91 lncRNAs sous-exprimés dans le cancer du sein (Fig. 48B). Certains d'entre eux ont déjà été décrits dans le cancer du sein, tels que HOTAIR ou MALAT1, mais la majorité d'entre eux représente de nouveaux lncRNAs associés aux tumeurs mammaires.

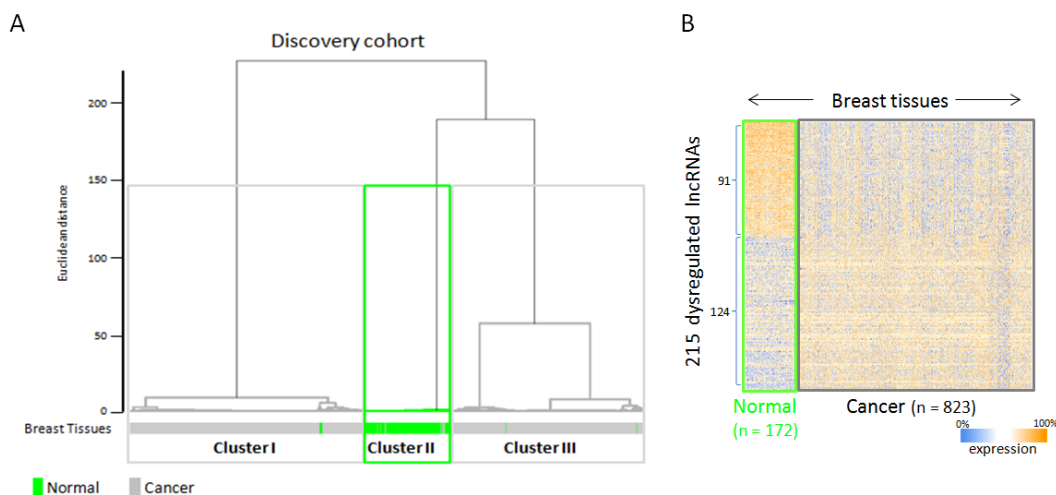


Fig. 48 : Dérégulation des lncRNAs dans une large cohorte de cancers du sein.

(A) Clustering hiérarchique consensus non supervisé de l'expression des 500 lncRNAs les plus variants selon leur déviation standard. (B) Carte d'expression des 215 lncRNAs significativement dérégulés dans le cancer du sein (variation d'un facteur > 1.5, FDR < 0.05, test de Mann-Whitney).

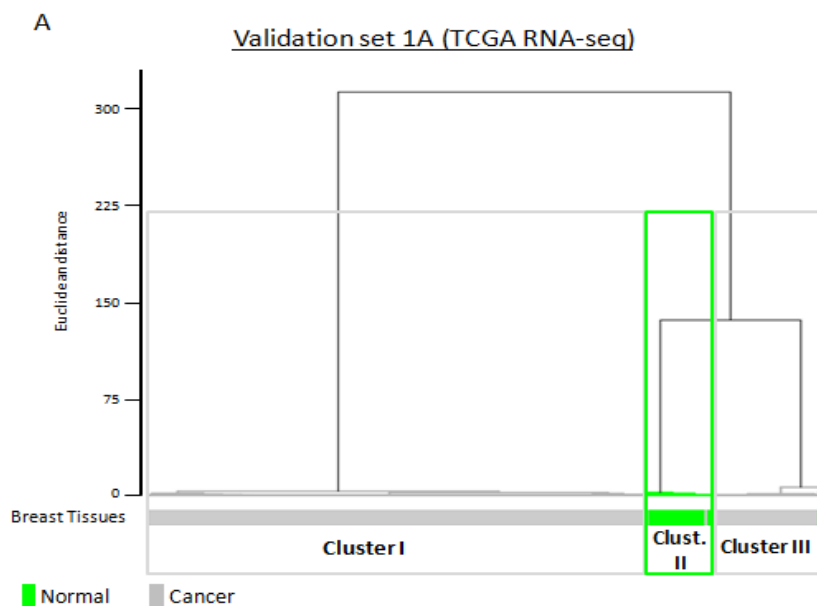
Nous avons ensuite examiné si ces lncRNAs pouvaient, à l'instar des ARNm, servir de biomarqueur diagnostique. Nous avons généré un métagène en utilisant la première composante d'une analyse en composantes principales des 215 lncRNAs dérégulés et analysé son pouvoir de discrimination entre les tissus sains et cancéreux. Le métagène permet d'identifier les tissus sains et les tissus cancéreux avec une sensibilité de 0.95 et une spécificité de 0.96 ($p < 0.0001$).

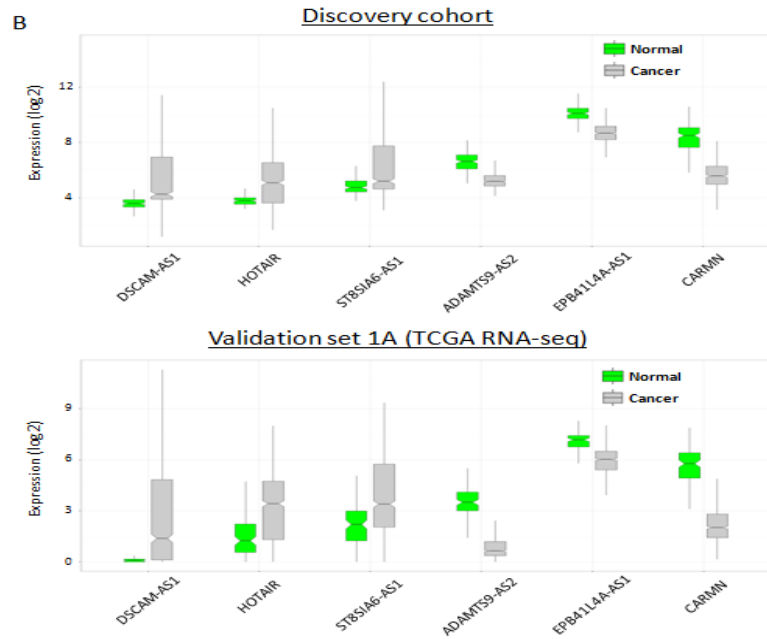
Pour valider nos résultats sur une cohorte indépendante, nous avons utilisé les données de séquençage à haut débit de l'ARN (RNA-seq) du cancer du sein provenant du consortium TCGA (The Cancer Genome Atlas, 2012) que nous avons appelé cohorte de validation 1A. Nous n'avons pas utilisé les données brutes qui sont très complexes à analyser mais les données traitées par Rahman *et al.* (Rahman et al., 2015) qui permettent d'étudier l'expression de 1161 lncRNAs dans 1052 tumeurs et 113 tissus mammaires sains. A nouveau, une analyse hiérarchique consensus non supervisé de l'expression des lncRNAs

permet l'identification de trois groupes de tissus, dont un contient la majorité des tissus sains (Fig 49A).

Parmi les 215 lncRNAs initialement identifiés comme dérégulés dans notre cohorte de découverte, 87 d'entre eux sont présents dans les données de RNA-seq dont 82 sont également dérégulés dans la cohorte de validation 1A (Fig. 49B). Un métagène représentant l'expression de ces 87 lncRNAs permet une classification des tissus sains et cancéreux avec une sensibilité de 0.93 et une spécificité de 0.98 ($p < 0.0001$) (Fig. 49C).

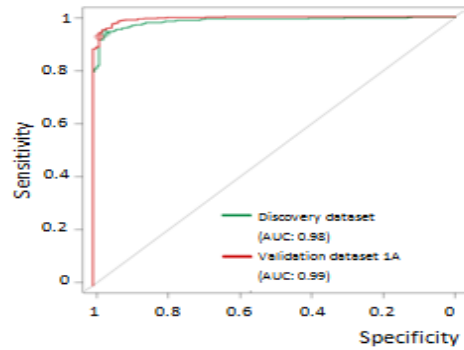
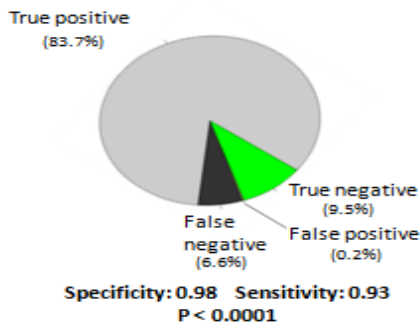
Le consortium TCGA fournit également des données d'expression de puce à ADN « custom Agilent® 244k » sur les tissus mammaires, que nous avons appelé la cohorte de validation 1B (qui comprend certains échantillons en commun avec la cohorte de validation 1A). Nous avons réannoté cette puce à ADN dans le but d'analyser le profil d'expression des lncRNAs. Parmi les 215 lncRNAs dérégulés dans notre cohorte de découverte, 167 sont détectables dans la cohorte de validation 1B. Un métagène représentant l'expression de ces 167 lncRNAs permet une classification des tissus sains et cancéreux avec une sensibilité de 0.98 et une spécificité de 0.97 ($p < 0.0001$) (Fig. 49D). Dans l'ensemble, nous avons donc identifié un ensemble de lncRNAs dérégulés dans le cancer du sein, qui permet de discriminer les tissus sains et cancéreux, de manière reproductible.





C Validation set 1A (TCGA RNA-seq)

(1052 tumors vs 113 normal)



D Validation set 1B (TCGA microarray)

(524 tumors vs 63 normal)

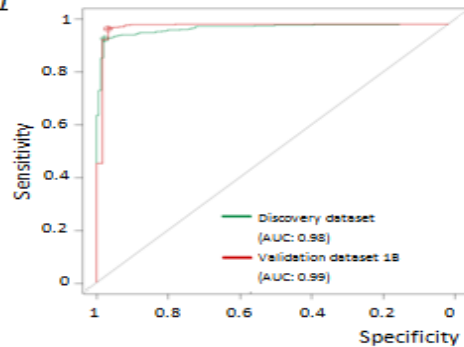
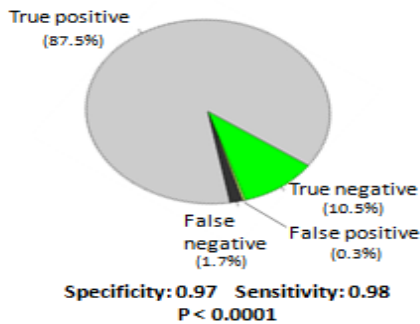


Fig. 49 : Dérégulation des lncRNAs dans une cohorte indépendante et validation.

(A) Clustering hiérarchique consensus non supervisé de l'expression des 500 lncRNAs les plus variants selon leur déviation standard. (B) Profil d'expression des lncRNAs les plus dérégulés dans la cohorte de découverte et dans la cohorte de validation 1A. (C) Validation de la signature de lncRNAs dérégulés dans la cohorte de validation 1A, basée sur les 87 lncRNAs (parmi les 215 dérégulés) présents dans les données de RNA-seq. Le diagramme en camembert, à gauche, illustre la qualité de la prédiction. La p-valeur est définie comme

la proportion de signature aléatoire de même taille (dans ce cas-ci 87 lncRNAs) qui possède un taux d'erreur balancé inférieur ou égal à celui de la signature réelle. La courbe de performance « Receiver operating characteristic » (ROC), à droite, illustre la spécificité et la sensibilité de la signature dans la cohorte de découverte (vert) et dans la cohorte de validation (rouge). Le seuil (croix) a été déterminé sur la cohorte de découverte et est utilisé pour déterminer la sensibilité et la spécificité durant l'étape de validation. (D) Idem que (C) dans la cohorte de validation 1B, basé sur les 167 lncRNAs (parmi les 215 dérégulés) présents sur cette puce à ADN. AUC : Aire sous la courbe.

III.2 L'expression des lncRNAs est associée à la voie ER

Le clustering hiérarchique (Fig. 48A), basé sur l'expression des lncRNAs, montre que les tumeurs mammaires sont séparées en deux groupes, nommés cluster I et cluster III. Nous avons recherché leurs propriétés cliniques et avons observé une association significative entre ces groupes et l'expression de ER. Le cluster I contient 91% de tumeurs exprimant ER et le cluster III contient 76% de tumeurs n'exprimant pas ER (Fig. 50A), suggérant une différence d'expression importante entre les tumeurs hormono-positives et hormono-négatives. Dès lors, nous avons utilisé une approche supervisée pour identifier les lncRNAs associés à ER. Nos résultats révèlent que 38 lncRNAs sont différentiellement exprimés entre les tumeurs ER positives et ER négatives (différence d'expression d'un facteur > 1.5 , $FDR < 0.05$, test t). La majorité (76 %) d'entre eux est surexprimée dans les tumeurs ER positives. Alors que certains d'entre eux ont déjà été associés à la signalisation ER tels que NEAT1 (Chakravarty et al., 2014), MALAT1 (Hansji et al., 2014) et Xist (Turner et al., 2007), la majorité de ces 38 lncRNAs sont nouveaux dans le contexte du cancer du sein. Citons par exemple LINC01297 qui est le lncRNA le plus significativement surexprimé dans les tumeurs ER positives et RP11-303E16.2, le plus significativement sous-exprimé dans ces tumeurs (Fig. 50B). Notons que 32 lncRNAs parmi les 38 associés à ER sont également dérégulés dans les tumeurs mammaires en comparaison avec les tissus sains.

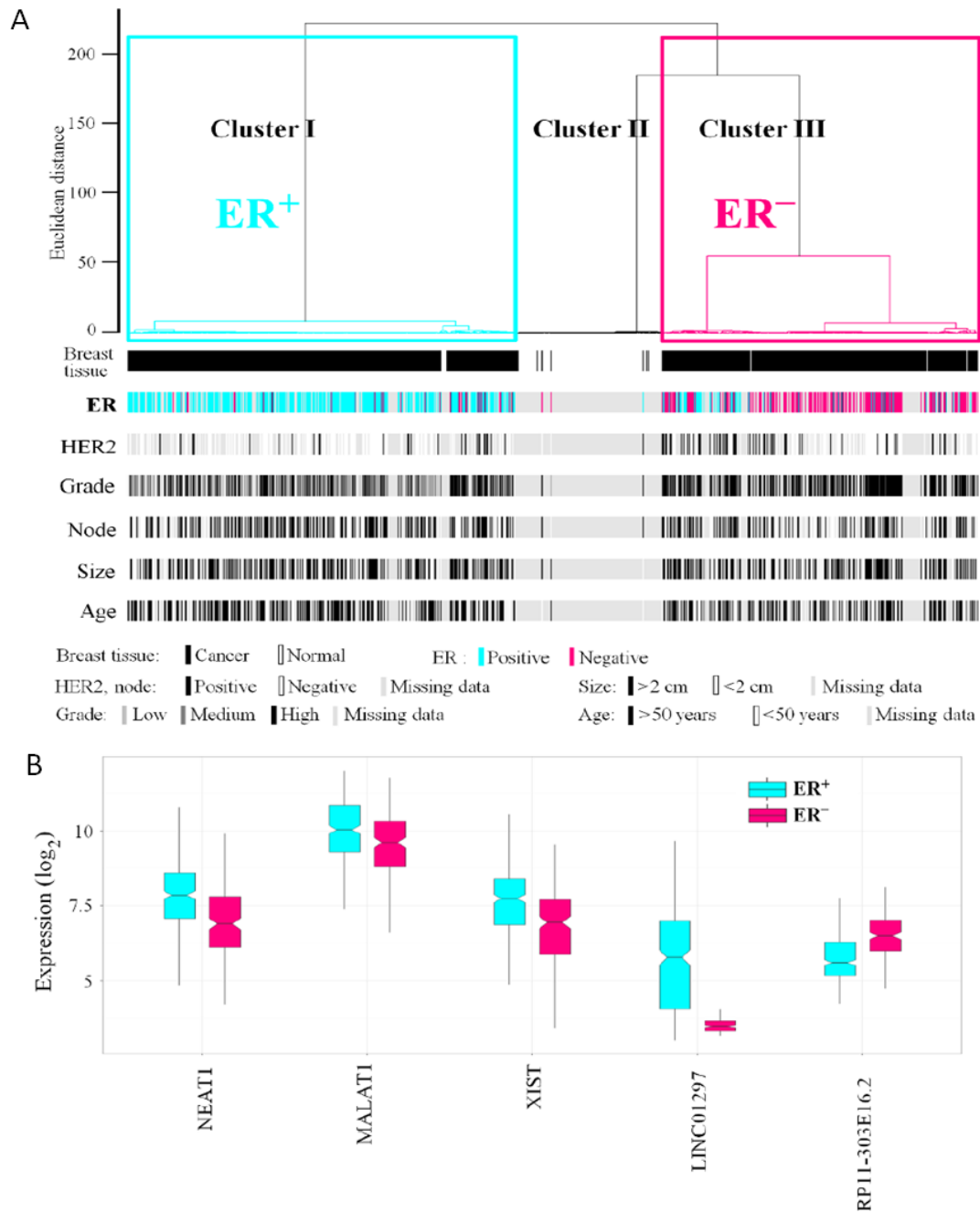
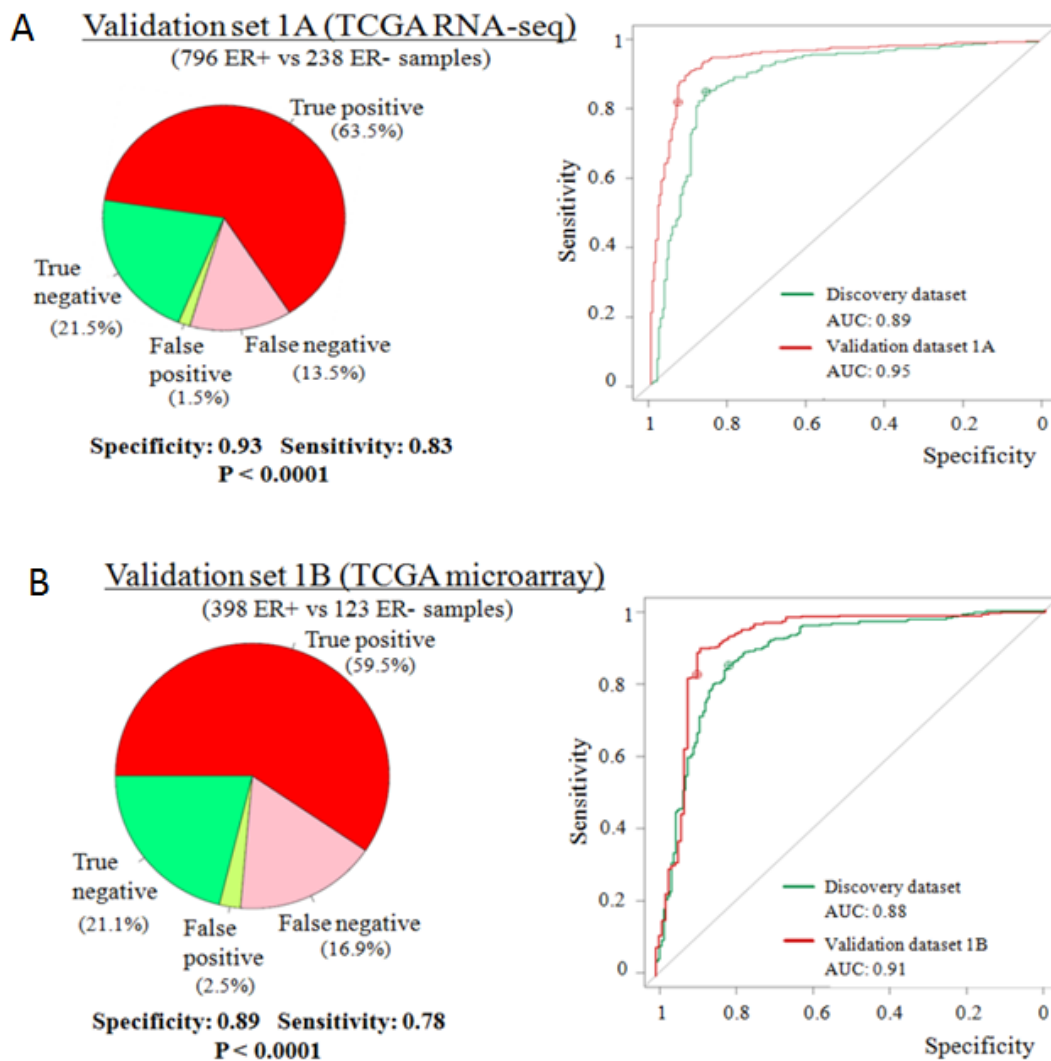


Fig. 50 : Relation entre l'expression des lncRNAs et le récepteur aux œstrogènes.

(A) Clustering hiérarchique consensus non supervisé de l'expression des 500 lncRNAs les plus variants selon leur déviation standard. Les clusters I et III, qui contiennent la majorité des tumeurs, sont associés à l'expression de ER. (B) Profil d'expression de 5 lncRNAs différemment exprimés entre les tumeurs ER positives et ER négatives.

Nous avons ensuite utilisé différentes cohortes de validation (cohorte de validation 1A (TCGA-RNAseq), 1B (TCGA-puce à ADN) et cohorte de validation 2 (GSE20685, puce à ADN similaire à notre cohorte de découverte)) pour examiner si cette signature de lncRNAs associés pouvait prédire le statut ER dans des cohortes indépendantes. Dans ces trois cohortes, le pouvoir prédictif de la signature ER est très significatif ($P < 0.0001$, Fig.51). La prédiction du statut ER basé sur l'expression de ces lncRNAs associés à ER est donc reproductible.



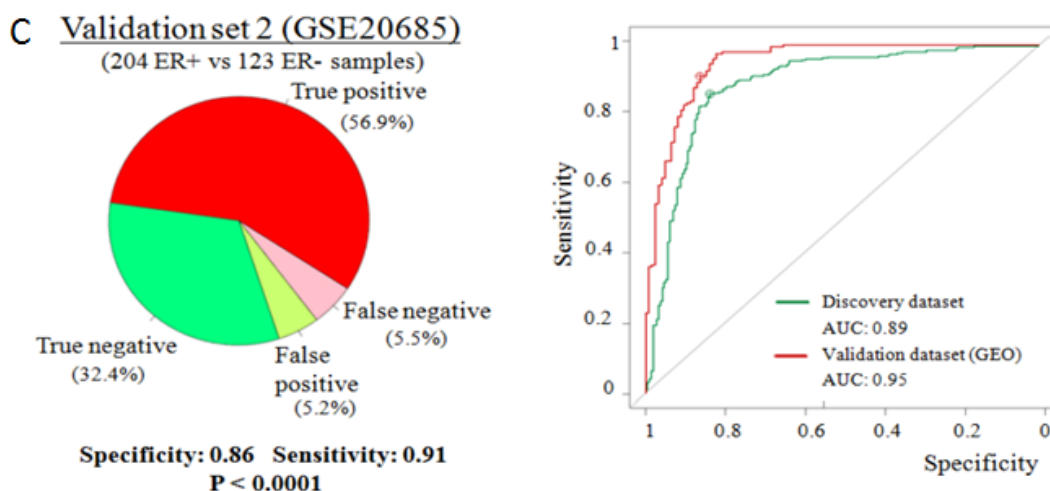


Fig. 51 : Validation de la signature des lncRNAs associés à ER.

(A) Validation de la signature de lncRNAs associés à ER dans la cohorte de validation 1A, basée sur les 12 lncRNAs (parmi les 38 dérégulés) présents dans les données de RNA-seq. Le diagramme en camembert, à gauche, illustre la qualité de la prédiction. La p-valeur est définie comme la proportion de signature aléatoire de même taille (dans ce cas-ci 12 lncRNAs) qui possède un taux d'erreur balancé inférieur ou égal à celui de la signature réelle. La courbe de performance (ROC), à droite, illustre la spécificité et la sensibilité de la signature dans la cohorte de découverte (vert) et dans la cohorte de validation (rouge). Le seuil (croix) a été déterminé sur la cohorte de découverte et est utilisé pour déterminer la sensibilité et la spécificité durant l'étape de validation. (B) Idem que (A) dans la cohorte de validation 1B, basée sur les 27 lncRNAs présents sur cette puce à ADN. (C) Idem que (A) dans la cohorte de validation 2, basée sur les 38 lncRNAs présents sur cette puce à ADN. AUC : Aire sous la courbe.

Pour mieux comprendre le rôle biologique de ces lncRNAs, nous avons utilisé l'approche de culpabilité par association (Rinn and Chang, 2012) qui permet d'associer chaque lncRNA avec ses fonctions potentielles. Nous avons subdivisé la liste des 38 lncRNAs associés à ER en trois groupes, en fonction de leur profil d'expression (par clustering hiérarchique). Nos analyses révèlent que chacun de ces groupes est enrichi pour des termes fonctionnels clés dans le cancer du sein (Fig. 52). Le premier groupe, qui contient des lncRNAs surexprimés dans les tumeurs ER positives, est le plus fortement associé à la voie cible de la rapamycine chez les mammifères (mTOR, mammalian target of rapamycin). Le second groupe est associé à l'activation de la voie E2F1. Enfin le troisième groupe est associé à une inhibition de l'immunité. Globalement, ces résultats démontrent que les lncRNAs permettent de discriminer les tumeurs ER positives

des tumeurs ER négatives et suggèrent que les lncRNAs sont impliqués dans divers processus au-delà de la biologie de ER.

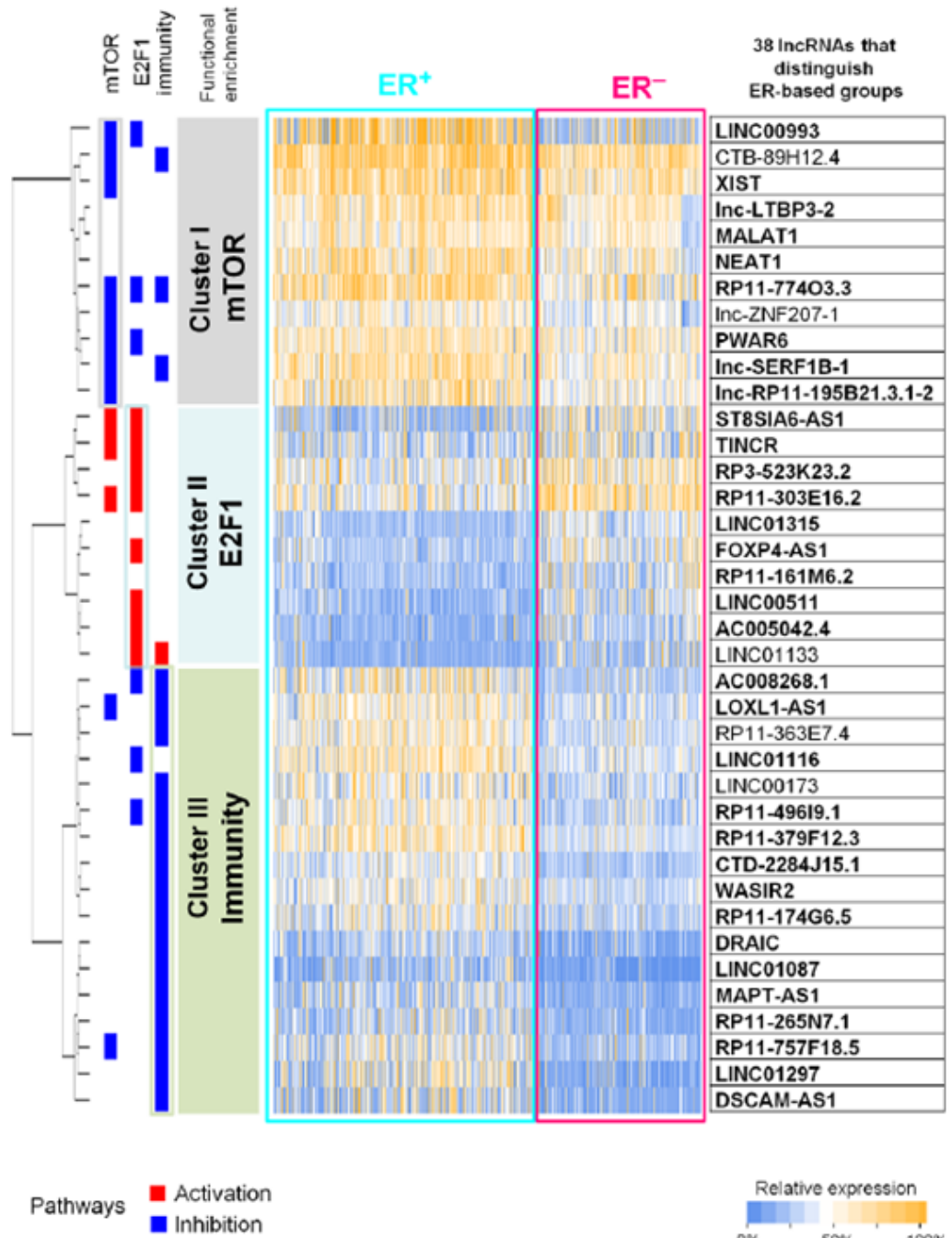
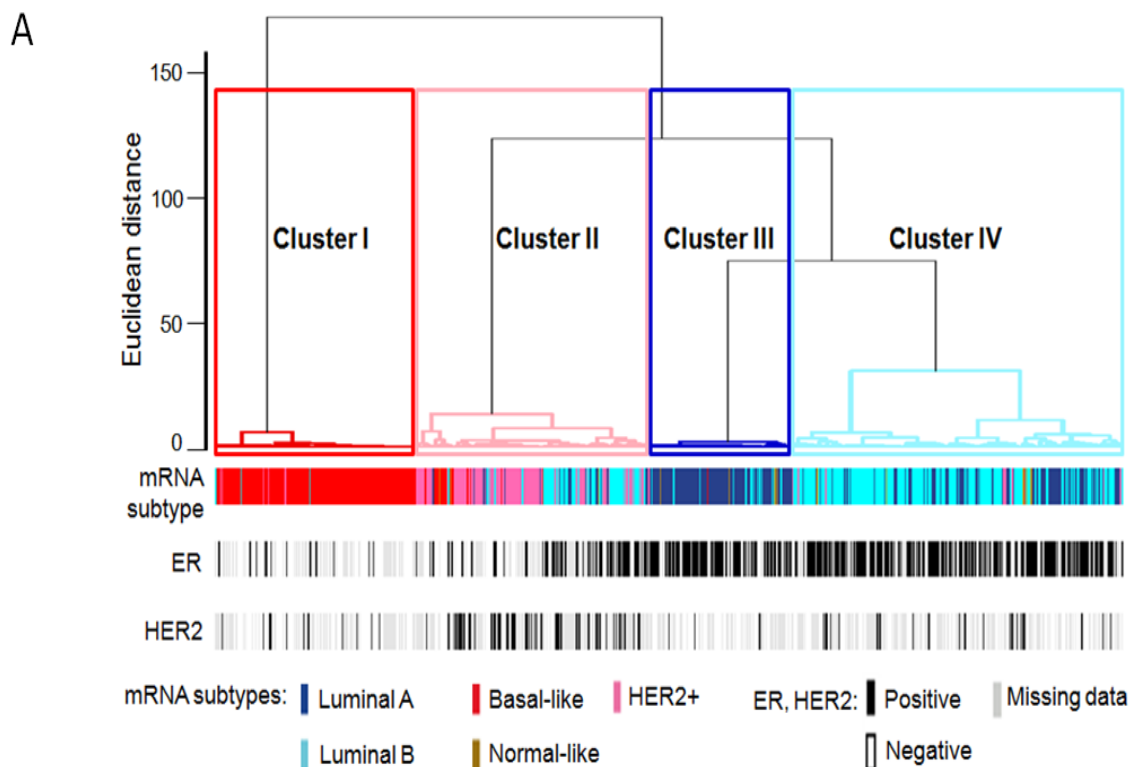


Fig. 52 : Expression et analyse fonctionnelle des lncRNAs associés à ER.

Carte d'expression des 38 lncRNAs associés à ER. Ces lncRNAs ont été séparés en trois clusters par clustering hiérarchique sur base de leur expression. Pour chaque cluster, le terme fonctionnel associé le plus significativement est représenté. Les lncRNAs en gras représentent les lncRNAs dérégulés entre tissus sains et cancéreux.

III.3 L'expression des lncRNAs est associée aux sous-types moléculaires de cancer du sein

Nous avons ensuite examiné si, à l'instar de nos résultats précédents (voir Résultats II.3), nous pouvions mettre en évidence une relation entre l'expression des lncRNAs et les sous-types intrinsèques dans cette plus large étude portant sur 823 tumeurs mammaires. Une analyse par clustering hiérarchique consensus de l'expression des lncRNAs dans les tumeurs mammaires identifie 4 clusters robustes (Fig. 53A). Chacun de ces clusters est associé de manière très significative avec les 4 sous-types moléculaires principaux du cancer du sein : le cluster I contient 96% de tumeurs basal like, le cluster II contient 52% de tumeurs HER2+, le cluster III contient 84% de luminal A et le cluster IV contient 64% de luminal B. Dès lors, nous avons utilisé la classification PAM50 comme référence pour définir une signature de lncRNAs spécifique de chaque sous-type (Fig 53B).



B

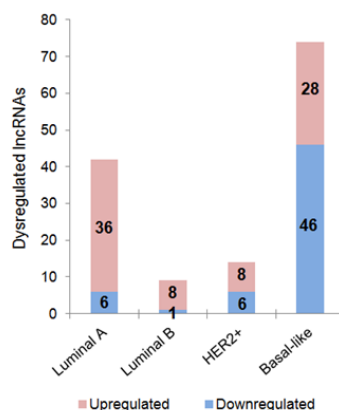


Fig. 53 : Association entre l'expression des lncRNAs et les sous-types moléculaires des cancers du sein.

(A) Dendrogramme des 823 tumeurs obtenu par clustering hiérarchique consensus, basé sur l'expression des 500 lncRNAs les plus variants. La classification moléculaire ainsi que l'expression de ER et de HER2 sont également montrées. (B) Histogramme illustrant le nombre de lncRNAs spécifiques de chaque sous-type qui sont sur- (rouge) ou sous-exprimés (bleu).

Nous avons à nouveau exploré si ces signatures de lncRNAs permettent de classifier les tumeurs en sous-types moléculaires. Tant dans notre cohorte de découverte, que dans nos cohortes de validation, les signatures de lncRNAs spécifiques permettent d'identifier, avec un faible taux d'erreur, à quel sous-type appartient une tumeur (voir figure supplémentaire 3 du manuscrit joint en annexe II).

III.4 Prédiction fonctionnelles des lncRNAs associés aux sous-types moléculaires du cancer du sein

Pour déterminer si chaque signature de lncRNAs spécifique des sous-types est globalement associée à certaines fonctions, nous avons à nouveau utilisé l'approche de prédiction de culpabilité par association. Nous avons généré un méta-score d'enrichissement et une valeur statistique associée, pour chaque fonction biologique (ou voie moléculaire) associée à chacune des signatures. Nous avons comparé ces résultats avec les fonctions enrichies basées sur l'expression des gènes codants, identifiées par le logiciel GSEA (Subramanian et al., 2005). Comme attendu, nous observons une concordance générale entre les fonctions prédites sur base des lncRNAs et des mRNAs (Fig. 54). Par exemple,

nous observons que les ARNs codants et non codants du sous-type basal like sont associés avec une activation du cycle cellulaire et des voies moléculaires RAF (MAPK) et MYC. Cependant, certains lncRNAs spécifiques de certains sous-types sont plus statistiquement associés à une fonction que les ARNm spécifiques de ce sous-type. C'est le cas de la signature des lncRNAs spécifiques des tumeurs luminal A, qui est plus associée à l'angiogénèse et à l'activation des voies FGF et PI3K que les ARNm. Un autre exemple implique les lncRNAs spécifiques des tumeurs basal like qui sont plus associés à l'activation des voies EGFR et LEF1 (lié à la transition épithélio-mésenchymateuse) que les ARNm.

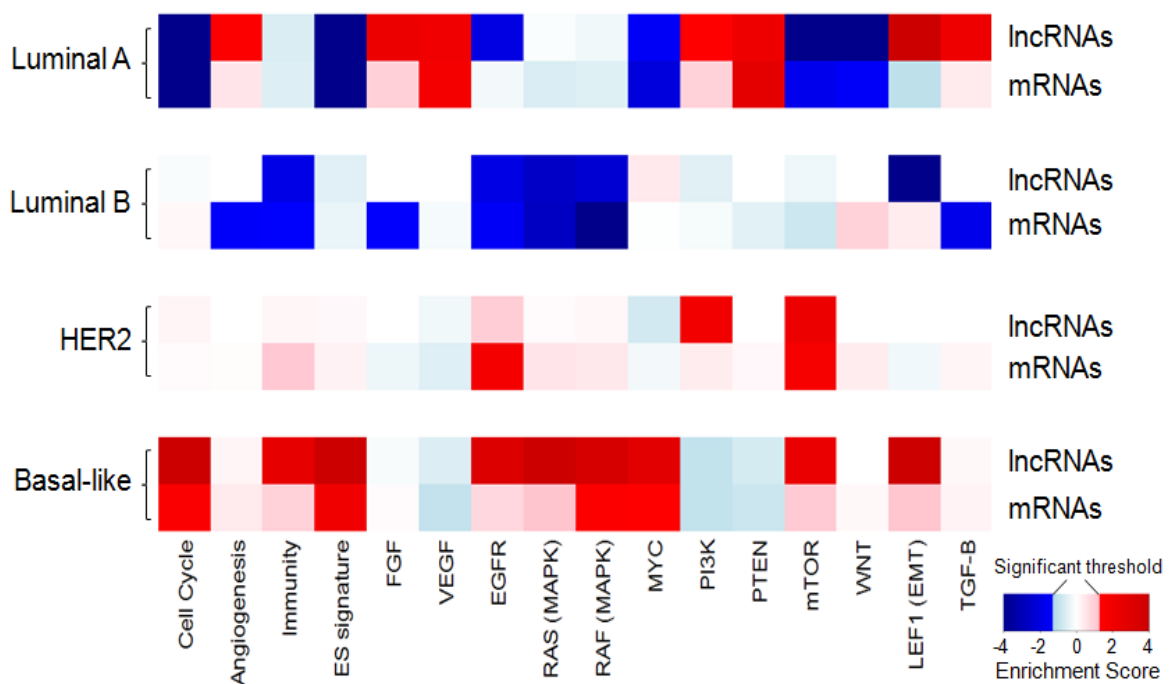
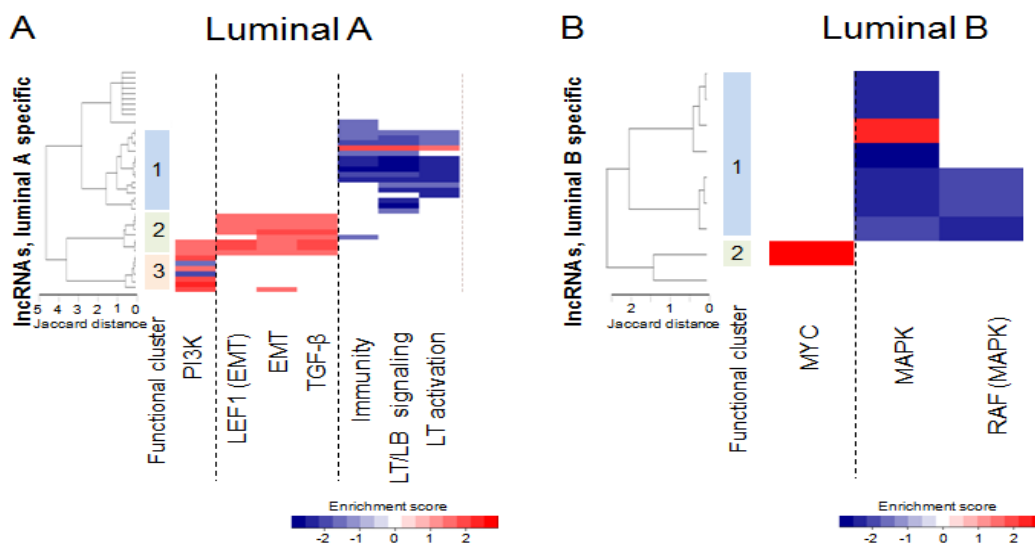


Fig. 54 : Prédiction fonctionnelle de l'ensemble des lncRNAs et des ARNm associés à chaque sous-type.

Carte illustrant les fonctions biologiques et les voies moléculaires qui sont directement (rouge) ou inversement (bleu) associées aux lncRNAs et aux ARNm spécifiques de chaque sous-type.

Nous avons ensuite analysé plus en détail les fonctions prédites pour chaque lncRNA présent dans les signatures spécifiques des sous-types. Nous les avons regroupés, par clustering hiérarchique, sur base de leur expression. Pour chaque cluster de lncRNAs, nous avons recherché les voies biologiques les plus spécifiquement associées (Fig. 55). Les lncRNAs de la signature luminal A sont regroupés en trois clusters : le premier comprend 18 lncRNAs associés à

l'immunité, dont 17 sont prédits pour inhiber cette voie et 1 pour l'activer (Fig. 55A) ; le deuxième cluster contient 8 lncRNAs prédits pour activer la transition épithélio-mésenchymateuse ; et le troisième cluster contient 7 lncRNAs associés à la voie PI3K, dont 5 sont prédits pour activer cette voie et 2 pour l'inhiber. Les lncRNAs spécifiques des tumeurs de type luminal B forment un cluster majeur associé à la voie des MAPK (Fig. 55B). Les lncRNAs spécifiques des tumeurs HER2+ forment deux clusters : le premier contient 7 lncRNAs associés à la voie mTOR dont 4 sont prédits pour activer cette voie et 3 sont prédits pour l'inhiber ; le second cluster est composé de 5 lncRNAs associés à la voie PI3K et/ou à la voie p53 (Fig. 55C). Enfin, les lncRNAs spécifiques des tumeurs basal-like forment trois clusters : le premier, composé de 26 lncRNAs, est associé à la voie E2F ; le second cluster, composé de 20 lncRNAs, est associé à la production d'interleukine 6 et partiellement à la voie EGFR ; enfin le dernier cluster, composé de 9 lncRNAs, est prédit pour moduler plus spécifiquement la voie EGFR. (Fig. 55B)



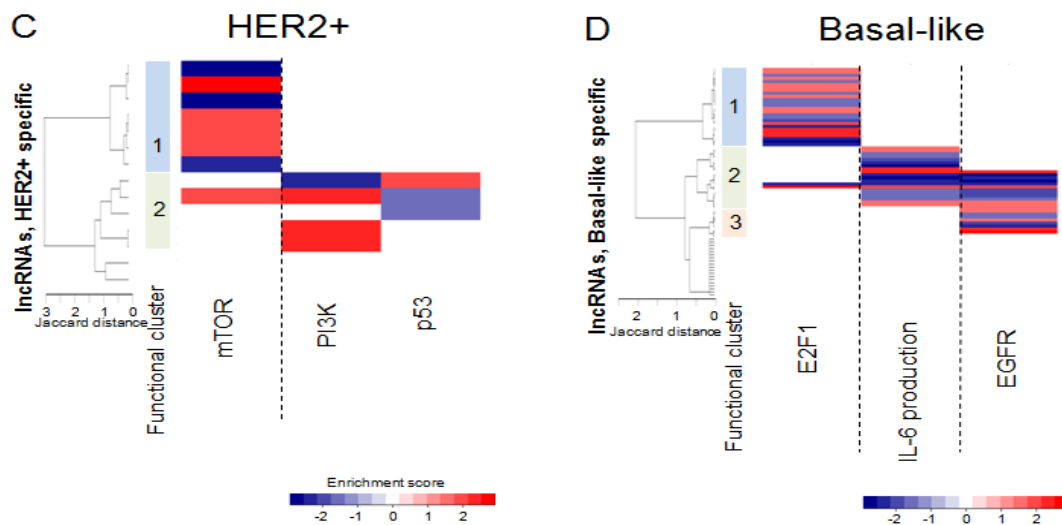


Fig. 55 : Prédictions fonctionnelles des lncRNAs spécifiques de chaque sous-type de tumeurs mammaires.

Association des lncRNAs spécifiques de chaque sous-type et de leur fonctions enrichies au sein des tumeurs luminal A (A), luminal B (B), HER2 (C) et basal-like (D).

Dans son ensemble, ces résultats mettent au jour l'implication potentielle de ces lncRNAs dans de nombreuses voies moléculaires spécifiquement (dé)régulées dans chaque sous-type de cancer du sein. Ce travail fournit donc une liste de lncRNAs d'intérêt à étudier plus avant pour mieux comprendre ces voies qui dysfonctionnent lors de la cancérogenèse mammaire.

III.5 Les lncRNAs sont des marqueurs pronostics dans le cancer du sein

A l'instar des ARNm, certains lncRNAs ont été associés à l'évolution clinique de différentes maladies (Du et al., 2013). Dès lors, nous avons recherché, par une approche de régression de Cox uni-variée si l'expression des lncRNAs corrèle avec le risque de rechute de cancer du sein. Notre analyse, dans notre cohorte de découverte avec un long suivi clinique, révèle que pas moins de 300 lncRNAs sont des marqueurs pronostics du risque de rechute, dont 41 sont dérégulés dans le cancer du sein. Nous avons ensuite analysé, par une approche de régression de Cox multi-variée, si l'expression des lncRNAs (variable continue) reste pronostic en tenant compte des facteurs confondants suivants (variables discontinues) : le grade et la taille de la tumeur, la propagation aux

ganglions lymphatiques, et les statuts ER et HER2. L'expression de 210 lncRNAs est statistiquement corrélée au risque de rechute dans cette analyse (Fig. 56A), suggérant donc que ces lncRNAs sont des facteurs pronostics indépendants. La majorité d'entre eux sont nouveaux dans le contexte du cancer mammaire dont RP11-863K10.2 qui a le ratio de risque HR (hasard ratio) le plus élevé (HR = 8,5, $p = 0.007$) (Fig. 56B) et LINC00152, que nous avons appelé CYTOR (Fig. 56C), qui est surexprimé dans tous les sous-types de cancer mammaire.

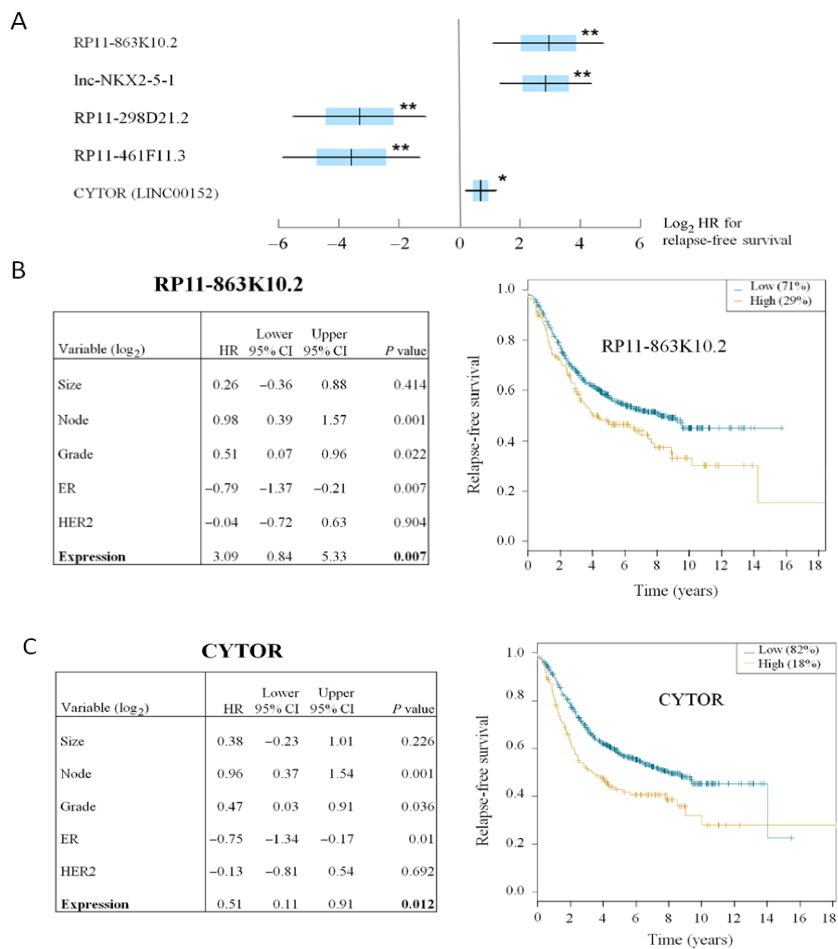


Fig. 56 : Les lncRNAs sont pronostics du risque de rechute.

(A) Graphique en forêt (forest plot) illustrant le HR accompagné de la déviation standard (bleu) et de l'intervalle de confiance (barres) de 5 lncRNAs exemplatifs des résultats de l'analyse de Cox multi-variée (* $p < 0.05$, ** $p < 0.01$). Dans nos analyses de régression de Cox multi-variée, l'expression est considérée comme une variable continue, les facteurs confondants (grade, taille, ganglions lymphatique, ER, HER2) sont considérés comme des variables discontinues. (B) Résultat de l'analyse multi-variée pour RP11-863K10.2 (gauche) et la courbe de Kaplan-Meier (droite) qui montre que les patientes exprimant

beaucoup ce lncRNA (orange) ont un risque de rechute plus important que les patientes qui expriment peu ce lncRNA (turquoise) ($p = 0.01$). (C) Idem que (B) pour CYTOR ($p = 0.000776$ pour la courbe de Kaplan-Meier).

Parmi les 210 lncRNAs associés à la survie par l'analyse multi-variée, 27 sont dérégulés dans le cancer du sein, suggérant leur importance biologique. Nous avons construit un méta-gène sur base de l'expression de 27 lncRNAs et vérifié si ce méta-gène est pronostic dans une cohorte indépendante. Nous observons des HR similaires pour la cohorte de découverte et de validation (p de validation = 0.0029) (Fig. 57) suggérant que ces 27 lncRNAs peuvent servir, de manière reproductible, de nouveaux marqueurs indépendants prédisant le risque de rechute.

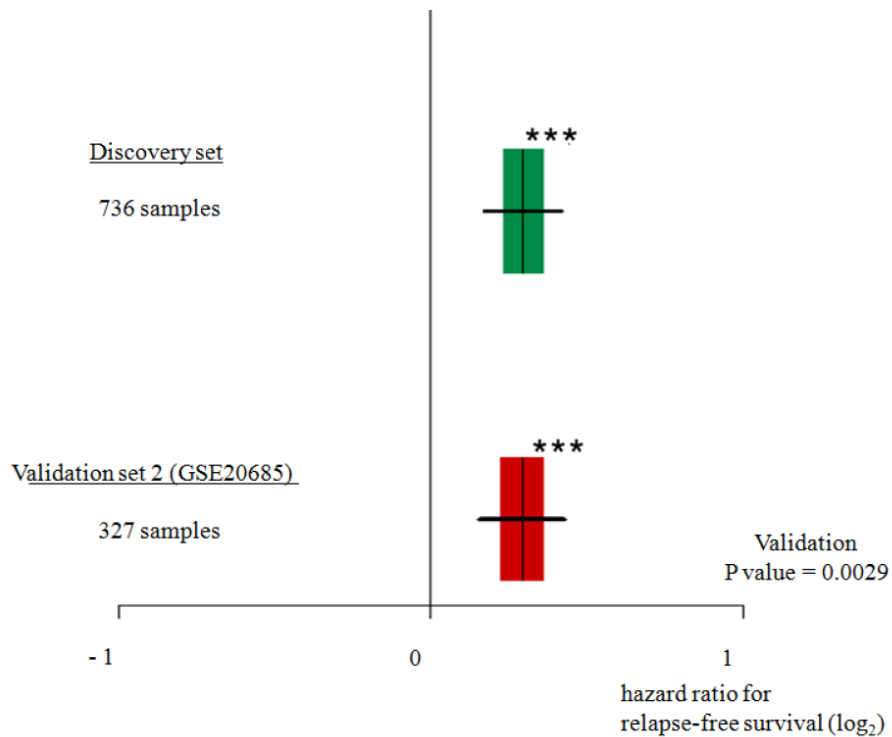


Fig. 57 : Validation de la signature de 27 lncRNAs pronostics.

Graphique en forêt (forest plot) illustrant le HR accompagné de la déviation standard (vert et rouge) et de l'intervalle de confiance (barres) associé au méta-gène (représentant l'expression des 27 lncRNAs) dans la cohorte de découverte et de validation (***) ($p < 0.001$).

III.6 CYTOR est requis pour la prolifération, la migration et l'organisation du cytosquelette.

Nous avons caractérisé expérimentalement LINC00152, que nous avons renommé CYTOR pour « cytoskeleton regulator » selon les règles du comité pour la nomenclature du génome HUGO (Wright, 2014). CYTOR est un lncRNA intergénique, situé à plus de 100 kilo-bases du gène codant le plus proche (PLGLB2) sur le chromosome 2p11.2. CYTOR fut choisi pour les raisons suivantes (i) il est surexprimé dans tous les sous-types de cancers mammaires dans notre cohorte de découverte et de validation 1A (Fig. 58), (ii) il est pronostic du risque de rechute et (iii) il est associé, par l'approche de prédiction de culpabilité par association, avec la prolifération, la migration, la transition épithélio-mésenchymateuse et les voies EGFR, mTOR et MAPK.

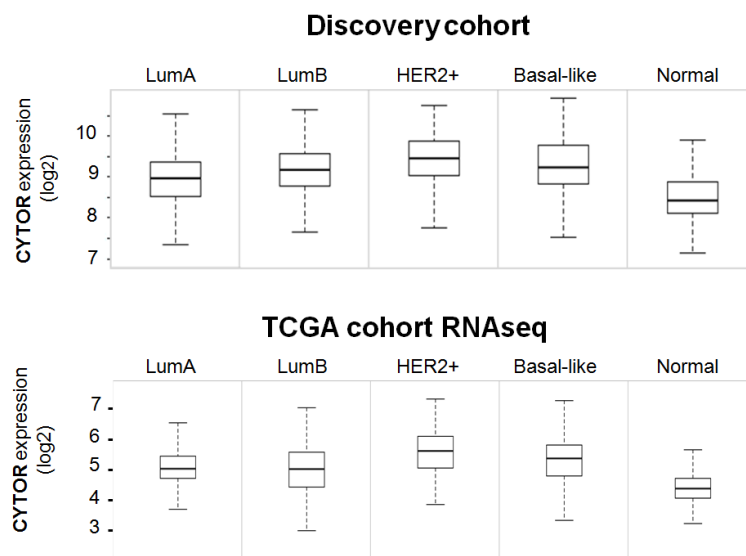


Fig. 58 : CYTOR est surexprimé dans les cancers mammaires.
Expression de CYTOR dans les différents sous-types de cancers du sein, dans la cohorte de découverte (haut) et la cohorte de validation (bas).

Pour analyser les fonctions de CYTOR expérimentalement, nous avons utilisé la technologie des LNA qui permettent de diminuer l'expression d'un lncRNA d'intérêt. Les cellules MDA-MB-231 transfectées voient l'expression de CYTOR diminuer d'environ 80% (Fig 59A). La sous-expression de CYTOR dans les cellules MDA-MB-231 induit une forte diminution de la prolifération (Fig. 59B-C) et de la migration (Fig. 59 D-E). Des résultats similaires ont été obtenus en lignée ZR-75-1.

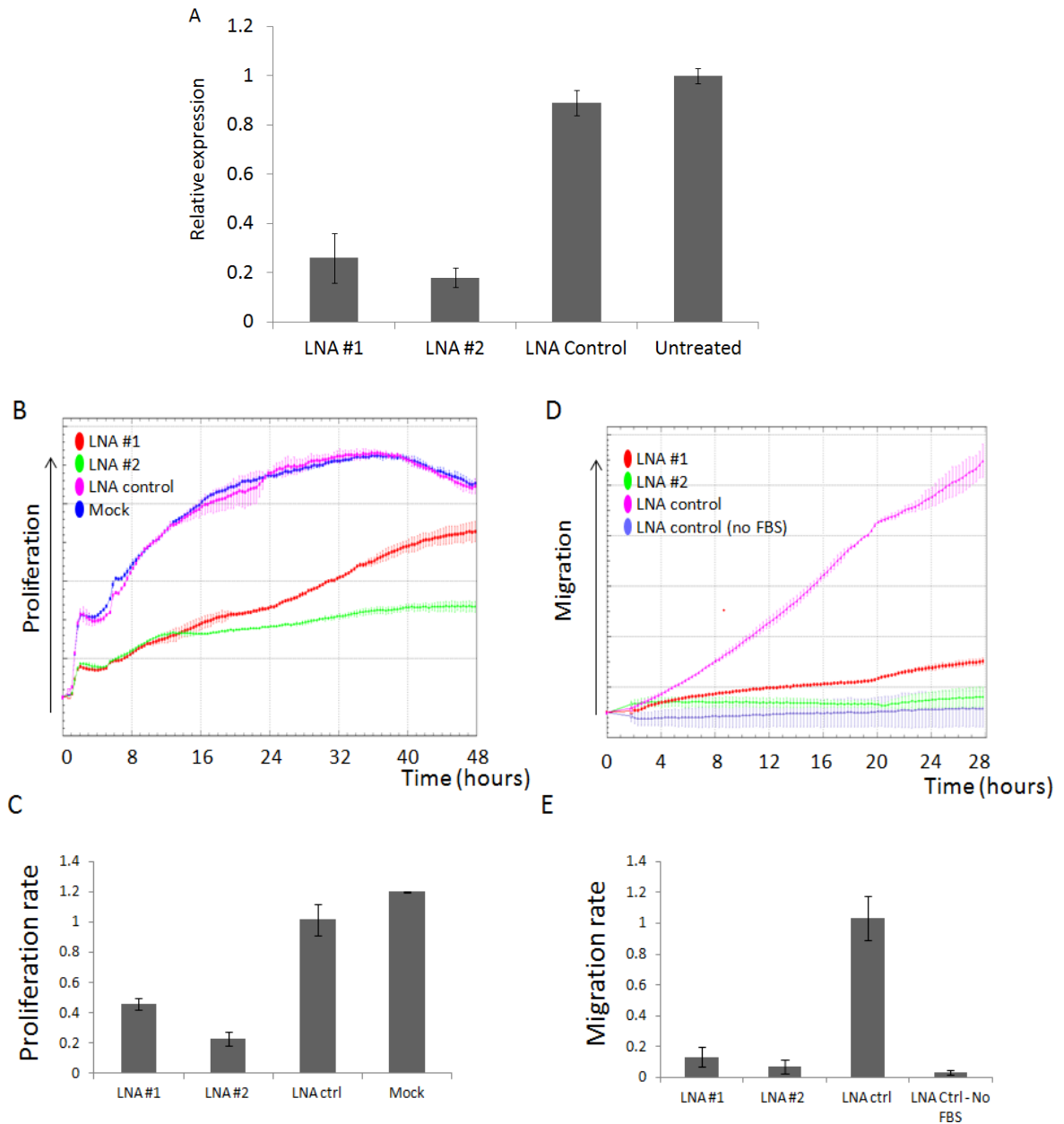


Fig. 59 : CYTOR influence la prolifération et la migration.

(A) Expression de *CYTOR*, analysé par RT-qPCR, dans les cellules MDA-MB-231 transfectées par des LNA ciblant *CYTOR* (LNA #1 : 5' AACTGTGCTGTGAAGA-3' ; LNA #2 : 5'-TCATAGAGCTTCCTGT-3') (n = 3). (B) Un profil représentatif du taux de prolifération des cellules MDA-MB-231 transfectées par des LNA ciblant *CYTOR*. (C) Taux de prolifération sur 48h des cellules MDA-MB-231 qui sous-expriment *CYTOR*, n = 3). (D) Un profil représentatif du taux de migration des cellules MDA-MB-231 transfectées par des LNA ciblant *CYTOR*. (E) Taux de migration sur 24h des cellules MDA-MB-231 qui sous-expriment *CYTOR*, n = 3). En l'absence de FBS chémo-attractant, les cellules ne migrent pas (contrôle négatif)

Nous avons confirmé l'influence de *CYTOR* sur le cycle cellulaire par des expériences de cytométrie de flux qui montrent que les cellules qui sous-expriment *CYTOR* s'accumulent en phase G2/M (marquage au 5-Bromo-2'-deoxyuridine/7-aminoactinomycine D) (Fig. 60A), mais ne meurent pas (marquage à l'annexine-V et au 7-aminoactinomycine D, suivant les recommandations du fournisseur BD Biosciences) (Fig. 60B).

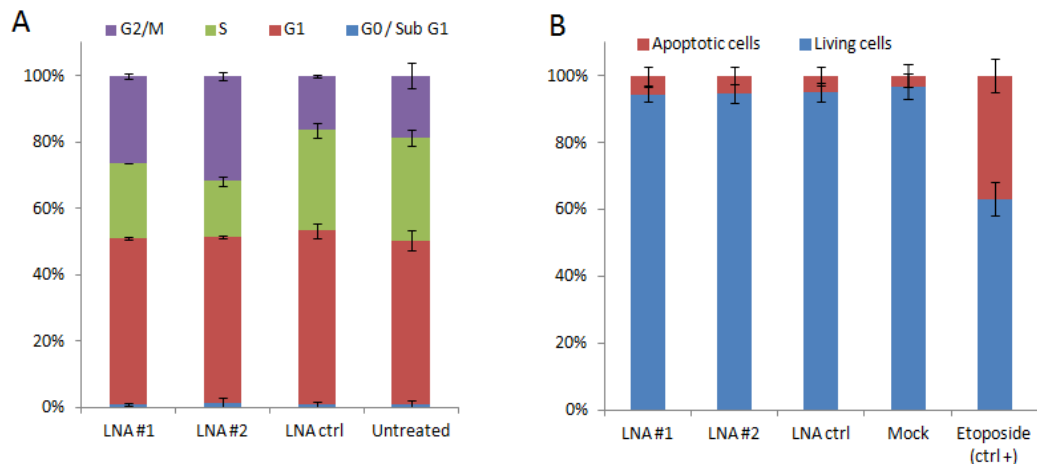


Fig. 60 : *CYTOR* modifie les phases du cycle cellulaire.

(A) Quantification (%), par cytométrie de flux, de la distribution cellulaire dans chaque phase du cycle. Les cellules MDA-MB-231 ont été transfectées par des LNA ciblant *CYTOR*. Les résultats de cytométrie ont été analysés 48h après la transfection (n=2). (B) Quantification (%) des cellules vivantes et apoptotiques, 48h après transfection des LNA ciblant *CYTOR*. L'étoposide (30 μ M, 24h), un inhibiteur de la topoisomérase II qui induit la mort cellulaire est utilisé comme contrôle positif (n = 2).

Au niveau de la régulation de *CYTOR*, nous avons examiné l'état de méthylation de son promoteur. Se basant sur les données de TCGA, nous observons une anti-corrélation entre le niveau de méthylation du promoteur de *CYTOR* et son expression, et plus particulièrement que les tissus qui montrent une forte méthylation du promoteur de *CYTOR* l'expriment peu (Fig. 61A). Ajou-

tons que dans nos modèles cellulaires, nous observons également une forte anti-corrélation entre le niveau de méthylation et l'expression (Fig. 61B). Ceci suggère que CYTOR est, au moins partiellement, régulé par l'état de méthylation de l'ADN de son promoteur.

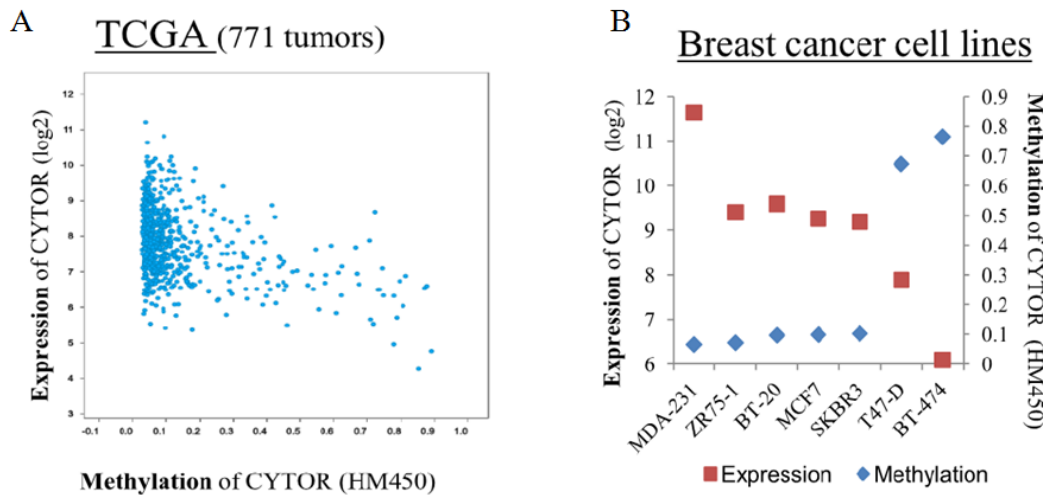


Fig. 61 : L'état de méthylation de CYTOR est anti-corrélé à son expression.

(A) Relation entre l'état de méthylation du promoteur de CYTOR et son expression dans une cohorte de 771 patientes provenant du TCGA. Le coefficient de corrélation de Spearman = -0,307. (B) Expression de CYTOR et niveau de méthylation de son promoteur dans différentes lignées mammaires.

Nous avons ensuite analysé les gènes cibles de CYTOR. Par puce à ADN, nous avons découvert 81 gènes (2 sur- et 79 sous-exprimés, taux de variation d'expression > 1.5 ; FDR < 0.05) dérégulés dans les cellules MDA-MB-231 qui sous-expriment CYTOR par rapport aux cellules contrôles (Fig. 62A). Certains de ces gènes sont impliqués dans le cancer du sein comme KIF14 (Singel et al., 2014) et GOLPH3 (Zeng et al., 2012), d'autres sont impliqués dans des voies de signalisation clés (par exemple mTOR, PRKCA et IGFR2) et dans le remodelage du cytosquelette comme la tubuline tyrosine ligase (TTL), la triphosphatase Rho guanosine (RHOBTB3) et la plakophiline-4 (PKP4). Nous n'avons identifié aucun gène dérégulé à proximité (+/- 1 méga bases) du locus de CYTOR, suggérant une action « *en trans* » de ce lncRNA. Enfin, nous avons sélectionné 5 cibles de CYTOR que nous avons validées par qPCR (Fig 62B).

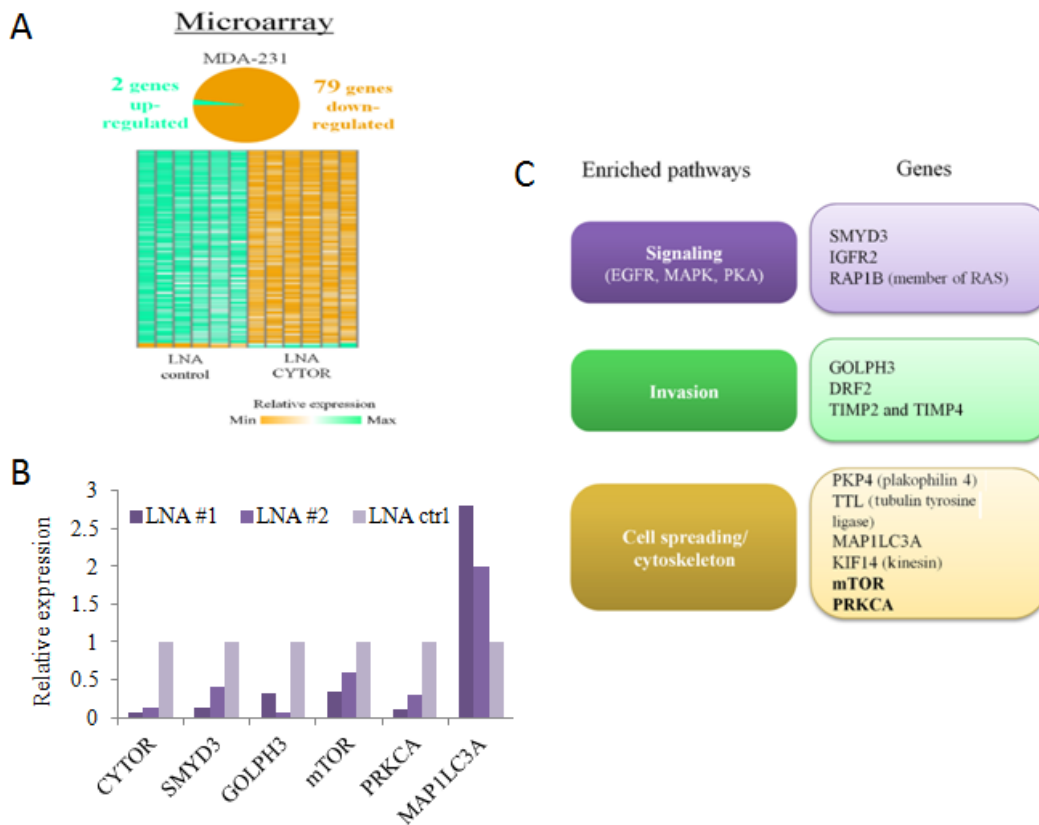


Fig. 62 : Identification des gènes régulés par CYTOR.

(A) Nombre de gènes dérégulés lorsque CYTOR est sous-exprimé en cellules MDA-MB-231. (B) Validation par qPCR de cinq cibles de CYTOR. (C) Voies biologiques et moléculaires statistiquement associées aux gènes régulés par CYTOR. Les gènes dérégulés, représentatifs de ces voies, sont représentés à droite.

Une analyse des gènes cibles de CYTOR par le logiciel « Ingenuity Pathways Analysis » a révélé les fonctions associées à ces gènes (Fig. 62C). Il en ressort que CYTOR module l'expression de gènes liés aux voies moléculaires EGFR et MAPK, comme prédit par notre analyse de culpabilité par association. La motilité cellulaire et la régulation du cytosquelette, dirigées par la polymérisation de l'actine et contrôlées par le complexe mTORC2, font partie des termes fonctionnels les plus associés aux gènes cibles de CYTOR. Dès lors, nous avons réalisé des expériences d'immunofluorescence pour marquer l'actine filamenteuse dans les cellules qui sous-expriment CYTOR. Ces cellules apparaissent plus rondes, plus petites et ont un cytosquelette réorganisé par rapport aux cellules contrôles. Les fibres longitudinales de stress d'actines polymérisées sont

moins présentes, et d'épaisses fibres d'actine se concentrent en périphérie, sur le cortex cellulaire (Fig. 63).

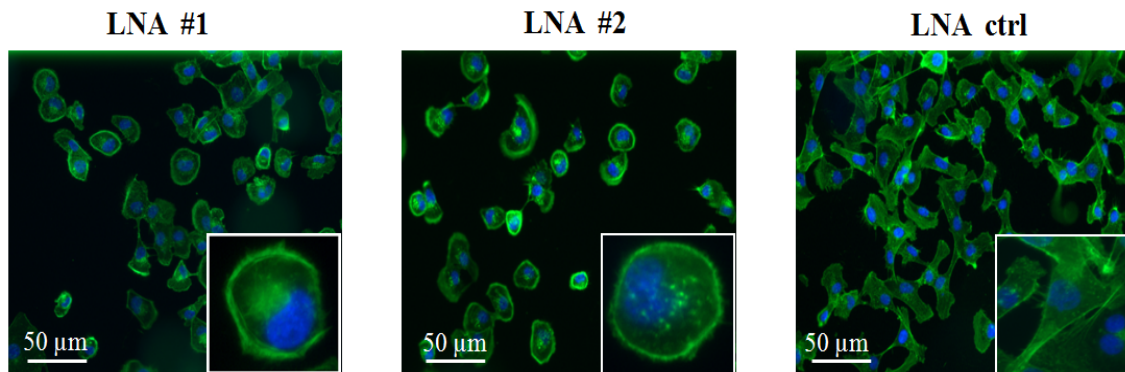


Fig. 63 : Organisation de l'actine filamenteuse dans les cellules qui sous-expriment CYTOR. Les cellules MDA-MB-231 transfectées par des LNA ciblant CYTOR (LNA #1 et LNA #2) ou un LNA contrôle sont représentées. L'actine filamenteuse est marquée en vert par l'Acti-Stain 488 phalloïdine, et le noyau en bleu par le Dapi. Les images ont été prises par un objectif 40x. Un grossissement (100x) est montré en bas à droite de chaque image.

Dans son ensemble, ces expériences révèlent que CYTOR, qui est de mauvais pronostic clinique, influence la prolifération et la migration des cellules cancéreuses mammaires. De plus, CYTOR régule la morphologie physiologique de la cellule, en agissant potentiellement sur la voie mTORC2. CYTOR pourrait donc être un bon candidat pour une nouvelle approche thérapeutique.

II.6 Matériels et méthodes

Cette section contient les procédures expérimentales qui ne sont pas décrites dans les sections Matériels et Méthodes des deux publications jointes. De plus, elle contient les méthodes plus inédites que nous avons utilisées pour élaborer nos résultats, à savoir : la méthode statistique pour identifier les lncRNAs dérégulés, la méthode de culpabilité par association et le système xCELLigence qui permet d'analyser en temps réel la prolifération et la migration cellulaires.

Préparation des ARNs de patientes

Les tissus mammaires ont été collectés à l'institut Bordet et directement congelés. L'ARN des tumeurs et des tissus normaux a été extrait au Trizol® selon le protocole recommandé par le fabricant (ThermoFisher). La qualité de l'ARN a été vérifiée par le bioanalyseur 2100® d'Agilent. La concentration a été estimée par le NanoDrop 1000®. Cette étude a été approuvée par le comité d'éthique de l'institut Bordet avec le consentement des patientes.

Quantification par PCR quantitative.

L'ARN des cellules a été extrait au Trizol® selon le protocole recommandé par le fabricant (ThermoFisher). Les ARNs totaux ont été traités à la DNase selon le protocole recommandé par le fabricant (DNA-free DNase kit Ambion). 1µg d'ARN a été rétro-transcrit en ADN complémentaire en utilisant des amorces hexamères aléatoires (Amersham/Pharma Biotech) et la Superscript II (Life Technologies). La PCR quantitative a été réalisée avec le marqueur SYBR Green sur une machine LightCycler 480 (Roche). Les résultats ont été normalisés avec les trois gènes de référence suivants : GAPDH, HPRT et ACTB.

Les amorces suivantes ont été utilisées :

lnc-KIN-2

F : TGCATCCCCTAGCGGCG

R : TGGAGTCTTCTTTCCCGCCT

GATA3

F°: CTCATTAAGCCCAAGCGAAG

R°: TCCTCCAGAGTGTGGTTGTG

ESR1

F°: GCTTACTGACCAACCTGGCAGA

R : GGATCTCTAGCCAGGCACATTC

Transfection par les LNA

Les LNA Gappers ont été transfectées (30 µM) dans les cellules MCF-7 avec 5µl de lipofectamine 2000, suivant les recommandations du fabricant (Exiqon). Les cellules ont été recueillies 24h post-transfection.

lnc-KIN-2 LNA #1 : 5' --> 3' GAGTGGATTTGGAGTC

lnc-KIN-2 LNA #2 : 5' --> 3' CCGCTCACACCGTT

Statistiques.

Pour identifier les lncRNAs dérégulés dans la partie I des résultats (puce à ADN personnalisée pour l'analyse des lncRNAs sur 45 tumeurs mammaires et 9 tissus normaux), les tests statistiques de Student et les corrections multi-tests (FDR) ont été réalisés dans R.

Dans la partie II, qui comporte plus de 900 échantillons, nous avons voulu identifier les lncRNAs qui sont dérégulés (normal vs cancer) dans une fraction des échantillons, car le cancer du sein est une maladie très hétérogène. Cette méthode nous permet d'identifier des lncRNAs qui sont dérégulés dans au moins 10% des tumeurs. L'avantage de cette méthode, adaptée de Yao *et al*, est qu'elle permet d'identifier des lncRNAs qui sont potentiellement oncogènes, même s'ils ne sont dérégulés que dans une partie relativement restreinte des tumeurs, ce que ne permet pas le test de Student classique. Pour identifier les gènes surexprimés dans les tumeurs, nous avons comparé les 10% de tumeurs et les 10% de tissus sains avec la plus haute expression, grâce à un test non paramétrique de Mann-Whitney. Nous avons étendu cette approche à une proportion croissante d'échantillons (15, 20, 25, 30, 35, 40 et 45%) et gardé les lncRNAs dérégulés dans la plus grande proportion d'échantillons. En parallèle, nous avons identifié les gènes sous-exprimés dans les tumeurs mammaires par la même approche, mais en nous focalisant sur les échantillons avec la plus basse expression. Les lncRNAs identifiés ici comme dérégulés dans les tumeurs ont une FDR $< 0,05$ et un taux de variation (fold change) $> 1,5$ dans au moins 10% des échantillons.

Sauf mentionné explicitement, les résultats expérimentaux se basent sur des triplicats biologiques, la barre d'erreur représentant la déviation standard.

Clustering.

Le clustering hiérarchique non supervisé a été réalisé dans R avec la bibliothèque Heatmap3, la fonction hclust et les paramètres suivants : méthode d'agrégation Ward et la distance Euclidienne comme mesure de dissimilarité.

Pour identifier les clusters significatifs dans la partie I des résultats, nous avons utilisé la bibliothèque pvclust, avec les paramètres par défaut. Les clusters avec un score de stabilité AU $> 0,70$ sont considérés comme statistiquement significatifs.

La bibliothèque pvclust n'étant pas adaptée à l'analyse d'une grosse cohorte, nous avons utilisé le clustering consensus et l'algorithme ConsensusClusterPlus ($n = 1000$ répétitions, 80% des échantillons à chaque répétition ($pI = 0,8$) et $pF = 1$) dans la partie II des résultats.

Analyse de culpabilité par association.

L'approche de culpabilité par association a été largement utilisée dans le contexte de recherche de fonctions des lncRNAs. Cette approche se base sur la corrélation entre l'expression des lncRNAs et de leur contrepartie codante. Dans un premier temps, nous avons divisé aléatoirement notre base de données en deux, une avec 411 tissus cancéreux, l'autre avec 412 tissus cancéreux. Pour chaque moitié de données, nous avons calculé le coefficient de corrélation de Pearson entre chaque lncRNA d'intérêt et tous les gènes codants. Ensuite, nous avons ordonné les gènes codants selon le coefficient de Pearson. Cette liste ordonnée est donnée au logiciel GSEA (paramètres : 1000 permutations, taille [15 – 500]) qui calcule un score d'enrichissement et une valeur statistique entre cette liste ordonnée et des fonctions biologiques / voies moléculaires, ce qui permet de découvrir les fonctions potentielles de chaque lncRNA d'intérêt. Afin d'obtenir des prédictions reproductibles, nous n'avons gardé que les fonctions associées à chaque lncRNA dans les deux moitiés de données, et avons calculé leur score en prenant leur moyenne arithmétique.

Les fonctions biologiques et voies moléculaires analysées proviennent de bases de données libres (KEGG, MsigDB et bdfunc.enrichment.human de la bibliothèque sRAP dans R)

Les résultats de l'analyse de culpabilité par association sont visualisés avec le logiciel Gene-E (<http://www.broadinstitute.org/cancer/software/GENE-E/index.html>).

Pour associer un cluster de lncRNAs à des fonctions, nous avons calculé un métascore d'enrichissement pour ce cluster. Le métascore est la somme pondérée des scores d'enrichissement de chaque lncRNA présent dans le cluster. La pondération vaut -1 si le lncRNA est réprimé dans la condition d'intérêt et de +1 s'il est surexprimé. Ensuite, 10 000 groupes aléatoires de lncRNAs (de même taille que le nombre de lncRNAs dans le cluster) ont été générés. Pour chaque groupe, un métascore est généré (selon la méthode précédente). Enfin, la p-valeur est définie comme la proportion de métascore aléatoire meilleure que le métascore du cluster.

Prolifération et migration par le système xCELLigence

Nous avons utilisé le système xCELLigence qui permet de suivre en temps réel la prolifération cellulaire. Ce système se base sur l'évolution de l'impédance électrique entre deux électrodes, qui va varier lorsque des cellules adhèrent et se multiplient entre ces électrodes. La valeur d'impédance est utilisée pour dériver un « index cellulaire » qui est directement proportionnel au nombre de cellules.

Pour analyser la prolifération, nous avons déposé 8000 cellules par puits, et suivi l'évolution de l'impédance pendant deux ou trois jours.

Pour analyser la migration, le système dispose de deux chambres séparées par une membrane microporeuse (qui ressemble à une chambre de Boyden). La chambre inférieure contient un chémoattractant (10% de FBS). Les cellules (40 000) sont déposées dans la chambre supérieure, dans un milieu qui ne contient pas de FBS. La migration des cellules mammaires à travers la membrane est monitorée pendant 24h, via la modification d'impédance. Les détails techniques sont fournis dans les deux publications en annexes.

DISCUSSION

I. Les micro-ARNs régulent les histones déméthylases de la famille KDM5 dans les cellules cancéreuses mammaires

A ce jour, de nombreuses études ont démontré l'impact prépondérant de miRs dans le développement du cancer, dont le cancer du sein. Nos analyses ont montré que miR-137 est sous-exprimé dans les cellules cancéreuses mammaires et pourrait donc jouer un rôle dans le développement du cancer. Notre analyse, par traitement au bisulfite suivi de pyroséquençage, révèle que dans les cellules mammaires non tumorales, le promoteur de miR-137 est peu méthylé, contrairement aux cellules cancéreuses. Un résultat qui va en accord avec une étude qui montre que miR-137 est plus méthylé et moins exprimé dans les tissus mammaires cancéreux que dans les tissus sains (Vrba et al., 2013). Ceci suggère que l'état de méthylation contrôle l'expression de miR-137. Cependant, un traitement avec un inhibiteur des ADN méthyltransférases ne permet pas une réexpression totale du miR, suggérant que d'autres mécanismes, tels que des modifications post-traductionnelles d'histones ou du complexe de remodelage de la chromatine, peuvent intervenir dans la régulation de l'expression de miR-137.

Nous avons montré que la surexpression de miR-137 en lignée cancéreuse mammaire résulte en une diminution drastique de la prolifération et de la migration et donc que miR-137 peut fonctionner comme un suppresseur de tumeur. Deux études confirment nos résultats. La première montre que la surexpression de miR-137 empêche la migration et l'invasion des cellules MCF-7, en promouvant un phénotype épithélial (augmentation d'E-cadhérine) et en inhibant les marqueurs mésenchymateux (vimentine et N-cadhérine) (Yong Han et al., 2016). Une autre étude valide nos résultats *in vivo* en démontrant que la surexpression de miR-137 en cellules MDA-MB-231 diminue la formation de tumeurs dans un modèle murin de xénogreffe (J.-M. Lee et al., 2015).

Pour mieux caractériser le rôle de miR-137, nous avons recherché ses cibles. D'une manière intéressante, nous avons découvert que l'histone déméthylase KDM5B, impliquée dans la tumorigénèse mammaire (Yamane et al., 2007),

est une cible directe de miR-137. KDM5B est surexprimée dans les tumeurs mammaires et son expression est associée à un mauvais pronostic (Yamamoto et al., 2014). De plus, il a été récemment montré que les cellules mammaires dont l'activité de KDM5B est inhibée, présentent un taux de prolifération et de migration réduit (Bamodu et al., 2016). Enfin KDM5B régule également l'état de différenciation des cellules souches embryonnaires (Bamodu et al., 2016) et la transition épithélio-mésenchymateuse (Enkhbaatar et al., 2013). Ceci pourrait expliquer le phénotype épithélial retrouvé dans les cellules MCF-7 qui surexpriment miR-137 (Yong Han et al., 2016). Nous envisageons de confirmer, *in vivo*, le rôle de miR-137 sur KDM5B et sur la transition épithélio-mésenchymateuse par des expériences de xénogreffes en modèle murin.

Nous avons également montré que miR-138, qui est également sous-exprimé dans les cellules cancéreuses mammaires, module la prolifération. Très récemment, un autre groupe a confirmé le rôle de miR-138 sur la prolifération en modèle mammaire (C. Zhang and Hua, 2016). De plus, nous avons identifié que miR-138 cible directement une autre déméthylase, moins décrite dans la littérature : KDM5C. Nous avons montré, pour la première fois, que KDM5C est surexprimée dans un modèle cellulaire de cancer mammaire. Des expériences additionnelles sont nécessaires pour déterminer les fonctions que KDM5C joue sur la tumorigénèse, par exemple en modèle murin invalidé pour KDM5C. Néanmoins, les résultats de nos expériences *in vitro* suggèrent que les histones H3K4me3 déméthylases jouent un rôle important dans le développement de tumeurs et devraient être investiguées dans le futur pour mieux comprendre leur rôle dans la tumorigénèse et l'apparition de métastases.

Les miRs sont des acteurs bien connus de la régulation de protéines épigénétiques. Par exemple, les miRs modulent l'expression des méthylases de l'ADN (DNMT3A et DNMT3B), de la protéine EZH2 du complexe Polycomb et des désacétylases d'histones HDAC1 dans les cancers. Nos résultats inédits montrent que les histones déméthylases sont également régulées par les miRs. En effet, miR-137 et miR-138 régulent directement l'expression des KDMs. Les autres cibles de miR-137 et miR-138, ainsi que les cibles des KDMs feront

l'objet d'autres investigations. D'une manière intéressante, d'autres études ont révélé l'importance clinique de ces deux miRs en montrant que miR-137 est sous-exprimé dans les tumeurs mammaires avancées (stade supérieur à 1) (Iorio et al., 2005) et que miR-138 est sous-exprimé dans les biopsies mammaires cancéreuses de patientes. De plus, son expression est pronostic (Bockhorn et al., 2014) et anti-corrélée au stade de la tumeur (J. Zhang et al., 2016).

Dans son ensemble, notre étude suggère que l'expression réduite des miRs mène à une accumulation des histones déméthylases KDMs, ce qui favorise la prolifération et la migration cellulaire et contribue à la tumorigénèse. Ces miRs sont des cibles thérapeutiques potentielles très intéressantes et devraient faire l'objet de nouvelles recherches pour l'élaboration de nouveaux traitements qui mimeraient ou réactiveraient ces miRs.

II. Description et caractérisation fonctionnelle des longs ARNs non codants dans le cancer du sein

Ces dernières années, les lncRNAs sont devenus un sujet de recherche très en vogue. En effet, l'émergence des nouvelles technologies a permis de démontrer que la majorité de notre génome est transcrit en lncRNAs. Ces lncRNAs sont fonctionnels et ils sont associés à de nombreuses maladies dont le cancer. Des études pionnières ont démontré l'implication des lncRNAs dans le cancer du sein (Gupta et al., 2010; Su et al., 2014). Cependant, leur fonction dans le développement du cancer du sein reste à élucider. Notre travail apporte un élément de réponse à cette question en associant les lncRNAs dérégulés dans le cancer du sein avec différentes voies moléculaires clés qui dysfonctionnent durant la tumorigénèse mammaire. Nous avons utilisé deux approches complémentaires pour analyser les lncRNAs dans le cancer du sein. La première a l'avantage de couvrir plus de 10 000 lncRNAs, soit la majorité des lncRNAs connus mais porte sur un nombre restreint d'échantillons, tandis que la seconde est basée sur une large cohorte de 995 tissus mammaires mais ne couvre que 3053 lncRNAs. Notre étude montre que les lncRNAs permettent de discriminer les différents sous-types de tumeurs et suggère que les lncRNAs y jouent des fonctions spécifiques. De plus, nous avons identifié 210 lncRNAs qui peuvent servir de marqueurs pronostics indépendants. Enfin, nous avons caractérisé expérimentalement deux lncRNAs (lnc-KIN-2 et CYTOR) et confirmé certaines de leurs fonctions prédites par notre méthode de culpabilité par association, telles que la régulation de la prolifération.

Analyse des lncRNAs à l'échelle du génome

Actuellement, l'analyse du transcriptome repose généralement sur deux technologies : les puces à ADN et le séquençage à haut débit. Nous avons choisi d'utiliser les puces à ADN pour analyser l'expression des lncRNAs dans le cancer du sein pour différentes raisons. (i) Les puces à ADN sont spécifiques du

brin transcrit, alors que la majorité des données de séquençage produite ne fournit pas d'information sur le brin d'origine, ce qui est particulièrement ennuyeux pour l'analyse des lncRNAs. En effet, de nombreux lncRNAs sont anti-sens des ARNm et peuvent les chevaucher. Dès lors, les fragments séquencés qui correspondent à ces régions peuvent provenir du brin sens ou anti-sens, ce qui rend impossible la quantification précise de ces lncRNAs. (ii) Par rapport au séquençage, les puces à ADN semblent avoir moins de variations techniques. Le séquençage montre plus de variations dépendantes de l'intensité, avec beaucoup de variations pour les transcrits peu exprimés tels que les lncRNAs (Raghavachari et al., 2012; McIntyre et al., 2011; Robinson, Wang, and Storey, 2015). De plus, sur une puce à ADN, la présence de transcrits très abondants n'affecte pas la quantification de transcrits moins abondants, contrairement au séquençage qui couvre peu les transcrits rares, spécialement dans les données de séquençage de faible profondeur. (iii) La puce à ADN est une technologie mature et peu onéreuse. Les méthodes d'analyses des données sont bien établies et acceptées. Il est relativement aisé de compiler différentes sources de données et de les analyser ensemble. Par contre, il n'existe pas de protocole standardisé pour l'analyse des données de séquençage, et la comparaison de données provenant de groupes différents est complexe. De plus, ces analyses requièrent un matériel informatique de pointe et peuvent être très longues. Par exemple, positionner les fragments séquencés sur le génome (mapping) de données brutes de séquençage de 1000 tumeurs mammaires prendrait plusieurs mois avec le programme d'alignement populaire TopHat (Trapnell, Pachter, and Salzberg, 2009). Ce qui explique que nos validations se basent sur des données de séquençage déjà traitées, qui consistent en une table d'expression de chaque gène (codant et non codant) dans chaque échantillon. (iv) Les données de survie associées aux données de puces à ADN sont de meilleure qualité (suivi moyen dans notre cohorte de découverte de 6,75 ans et de 9,2 ans dans notre cohorte de validation) que les données de séquençage (suivi moyen de 1,08 années dans les données du TCGA). (v) Enfin, au début de ce travail de thèse, la taille des cohortes publiquement accessibles analysées par puces à ADN (cohorte de découverte de 995

tissus et cohorte de validation de 327 patients) était beaucoup plus grande que les cohortes analysées par séquençage.

Notons néanmoins que le séquençage de l'ARN surmonte certains défauts des puces à ADN. La plage dynamique des puces à ADN est limitée (souvent aux alentours d'un taux de variation d'un facteur 1000) et le signal peut atteindre la saturation pour les gènes fortement exprimés, alors que la plage dynamique du séquençage (à haute profondeur) est théoriquement illimitée. Les puces à ADN sont dépendantes des sondes d'hybridation et ne permettent donc pas la découverte de nouveaux transcrits. Cependant, l'étude réalisée sur la grande cohorte de 995 tissus et limitée à l'analyse d'environ 3000 lncRNAs fournit des résultats nouveaux et intéressants. Nous avons identifié des signatures de lncRNAs dérégulés, de lncRNAs associés à ER et de lncRNAs spécifiques des sous-types que nous avons validées dans d'autres cohortes indépendantes. De futures études, basées sur le séquençage de l'ARN (spécifique du brin d'origine), permettront de confirmer nos résultats et de les étendre à l'ensemble des lncRNAs.

Comparaison des résultats de la partie 1 (puce à ADN spécifique) et de la partie 2 (données publiques)

Dans l'ensemble, les messages clés sont similaires entre notre première étude basée sur l'analyse de plus de 10 000 lncRNAs dans une cohorte de 45 tissus et notre seconde étude basée sur l'exploitation de données publiques de puces à ADN (> 3000 lncRNAs analysés) dans une large cohorte. Il en ressort que, à l'instar des mRNAs, on observe une dérégulation des lncRNAs dans le cancer mammaire comparé aux tissus sains. 28 lncRNAs sont communément dérégulés dans les deux études ($p\text{-valeur} = 9,7 \times 10^{-13}$). Ce nombre peut sembler limité, mais s'explique par différents facteurs : (i) il n'y a que 2216 lncRNAs présents à la fois sur la puce spécifique pour les lncRNAs et sur la puce Affymetrix. Parmi eux, 96 sont dérégulés dans la première étude et 136 sont dérégulés dans la seconde étude, (ii) les méthodes et les seuils pour identifier les lncRNAs dérégulés diffèrent entre les deux études, et (iii) les sondes utilisées pour cibler un

lncRNA varient d'une puce à l'autre et peuvent donc cibler des isoformes de lncRNAs différentes.

Nos deux études ont permis de mettre en évidence que l'expression des lncRNAs permet d'une part de discriminer les tissus ER+ et ER-, mais aussi de discriminer les différents sous-types moléculaires de cancer du sein. D'autres études confirment nos résultats (Zhao, Luo, and Jiao, 2014; Su et al., 2014). Cependant, nos analyses n'ont pas permis de découvrir de nouveaux sous-types significatifs du cancer du sein. Il pourrait toutefois être intéressant de comparer, en termes de réponses aux différentes thérapies (hormonales, anti-HER2, ...), si les sous-types moléculaires déterminés par les lncRNAs sont plus prédictifs de réponse que les sous-types déterminés par l'expression des mRNAs (classification PAM50). Cette question reste ouverte car nous ne disposons pas, actuellement, de données de réponses aux thérapies dans une large cohorte. La suite de cette discussion se focalise principalement sur les résultats obtenus dans la deuxième étude réalisée sur une large cohorte.

Fonctions associées aux lncRNAs

Pour identifier les fonctions potentielles des lncRNAs, nous avons utilisé une approche de culpabilité par association. Cette méthode a été largement utilisée dans ce contexte (Rinn and Chang, 2012; Sauvageau et al., 2013; Carnero et al., 2016) et repose sur l'idée que les ARNm et les lncRNAs corégulés sont impliqués dans les mêmes fonctions / voies moléculaires. Nous avons adapté cette méthode afin d'être plus confiant dans les résultats obtenus. En bref, nous avons divisé aléatoirement nos données en deux et réalisé les analyses statistiques sur chaque moitié des données avec le logiciel GSEA. Afin d'obtenir des prédictions robustes, nous avons uniquement considéré les enrichissements fonctionnels lorsque les résultats étaient significatifs dans les deux moitiés. Cette méthode nous a permis d'identifier les fonctions potentielles de nombreux lncRNAs. Par exemple, nous avons prédit que lnc-KIN-2 et CYTOR sont impliqués dans la prolifération et nous avons validé ces prédictions expérimentalement. D'autres prédictions fonctionnelles ont été validées dans la littérature comme les lncRNAs PVT1 et MINCR, prédits pour être associés à la voie MYC, qui ont été

impliqués expérimentalement dans le contrôle de cette voie (Tseng et al., 2014; Doose et al., 2015).

Nous avons identifié et validé différentes signatures de lncRNAs associées au statut ER et aux sous-types de cancer du sein et nous nous sommes intéressés aux fonctions de ces lncRNAs spécifiques. Nous avons montré que les 38 lncRNAs associés à ER peuvent être répartis en trois clusters (Fig. 52). La voie de signalisation mTOR est la voie la plus associée au cluster I. La voie mTOR est associée à la prolifération des cellules et promeut leur survie. Son hyperactivation est impliquée dans les cancers ER+ et dans la résistance au traitement par les inhibiteurs de l'aromatase (Ma et al., 2015; Ciruelos Gil, 2014). De nouvelles études sur les lncRNAs présents dans ce premier cluster pourraient permettre de comprendre comment ces lncRNAs contrôlent la voie mTOR et mener au développement de nouvelles thérapies. Le cluster II contient des lncRNAs, surexprimés dans les tumeurs ER-, qui sont associés à l'activation de la voie E2F1. Cette voie est également impliquée dans la résistance à l'hormonothérapie (Louie et al., 2010). De plus, le gène E2F1 est plus exprimé dans les tumeurs ER+ et son expression est de pronostic défavorable (Vuaroqueaux et al., 2007). Enfin le cluster III est associé à l'inhibition de l'immunité dans les tumeurs ER+. Le rôle de l'immunité dans le cancer reste flou mais certaines études suggèrent que la signalisation ER joue une fonction immuno-suppressive (Curran et al., 2006; Mostafa et al., 2014). Ajoutons que 9 lncRNAs sur les 17 du cluster III sont également associés à la production d'interleukine-6, un régulateur de la réponse immunitaire et du maintien des cellules souches. Les cancers du sein de types basal-like ont une expression élevée d'interleukine-6, qui est corrélée avec l'EMT et un pronostic vital défavorable (Sullivan et al., 2009; Schafer and Brugge, 2007). Nos résultats révèlent ici le lien potentiel entre certains lncRNAs, la production d'interleukine-6 et la réponse immunitaire.

Certains lncRNAs que nous avons associés avec les sous-types moléculaires du cancer du sein ont déjà été décrits dans ce contexte. Par exemple, nous avons montré que H19 et MAPT-AS1 sont surexprimés dans les tumeurs luminal A de notre cohorte de découverte et de validation. MAPT-AS1 a été décrit

comme surexprimé dans les tumeurs luminal A par Su et al (Su et al., 2014). Un autre groupe qui étudie l'expression des lncRNAs par hybridation *in situ*, montre que H19 est surexprimé dans les tumeurs ER+. Bien que H19 ait été largement étudié, ses fonctions restent floues, en effet il semble avoir à la fois des fonctions oncogènes et onco-suppressives (L. Zhang et al., 2013). En accord avec la littérature (Bergström et al., 2010; Matouk et al., 2014), nos résultats associent H19 avec EMT et la voie TGF- β . De plus, nous montrons que H19 est associé avec les protéines épigénétiques SETD7, KDM1 et EZH1 ce qui fournit un nouveau contexte d'étude de H19. Un autre exemple implique LINC00993, que nous avons identifié comme le lncRNA le plus sous-exprimé dans les tumeurs basal-like, à l'instar de Chen *et al* (C. Chen et al., 2015). D'une manière intéressante, nous avons montré que les clusters 2 et 3 des lncRNAs spécifiques des cancers basal-like sont associés à la voie EGFR (Fig 55 D). Cette voie est régulièrement activée dans ce sous-type (Gazinska et al., 2013) et peut être thérapeutiquement ciblée par des inhibiteurs spécifiques. A nouveau, nous fournissons ici une liste de lncRNAs dont l'implication dans la voie EGFR doit être confirmée expérimentalement, et qui pourraient à terme devenir des cibles thérapeutiques. Enfin, un troisième exemple implique HOTAIR. Nous avons montré que HOTAIR est le lncRNA le plus surexprimé dans le sous-type HER2+. Récemment, deux études ont également associé HOTAIR à la signalisation HER2 (Su et al., 2014; X.-H. Liu et al., 2014). Notre méthode de prédiction associe HOTAIR à la voie PI3K, cette relation a été décrite dans d'autres cancers mais les mécanismes impliqués ne sont pas bien connus (Guang Yang et al., 2014; Hui and Xianglin, 2016). Dès lors, de nouvelles études sont requises pour mieux comprendre le lien entre HOTAIR et la voie PI3K. Dans l'ensemble, nous observons que les prédictions fonctionnelles que nous avons établies par la méthode de culpabilité par association sont en adéquation avec les fonctions connues des lncRNAs, établies expérimentalement. Ceci nous conforte dans la méthodologie appliquée et dans les nouvelles fonctions identifiées.

Les lncRNAs comme marqueurs pronostics

Une analyse univariée du pouvoir pronostic des lncRNAs révèle que l'expression de 300 lncRNAs est associée au risque de rechute. En corrigeant les effets des facteurs confondants classiques, une analyse multivariée identifie que l'expression de 210 lncRNAs est pronostic, suggérant leur pouvoir prédictif indépendamment d'autres facteurs. Cependant, nous n'observons pas que valeurs pronostics des lncRNAs sont meilleures que les valeurs pronostics des mRNAs. Parmi ces 210 lncRNAs, nous avons sélectionné les 27 lncRNAs qui sont également dérégulés dans le cancer par rapport au tissu mammaire sain et qui ont donc un sens clinique plus pertinent. Nous avons validé le pouvoir prédictif de cette signature de 27 lncRNAs dans une cohorte indépendante, ce qui montre que cette signature prédictive peut être utilisée de manière reproductible et pourrait permettre de discriminer les patientes avec un cancer du sein de bon pronostic qui ne nécessiterait pas de thérapie agressive.

La majorité des lncRNAs associés au pronostic que nous avons identifiés ici sont nouveaux dans le contexte du cancer mammaire. Cependant, certains d'entre eux ont déjà été associés avec la survie. Citons par exemple CYTOR, qui est aussi pronostic dans le carcinome rénal à cellules claires (Y et al., 2015). Nous avons comparé nos signatures de lncRNAs associés au risque de rechute avec les études qui portent sur les lncRNAs pronostics dans le cancer du sein. Un groupe a identifié 9 lncRNAs associés avec l'apparition de métastases (Sun et al., 2015). Parmi ces 9, 4 sont présents sur la puce à ADN que nous avons utilisée, cependant aucun d'entre eux n'est pronostic dans notre étude. Une autre équipe a découvert que 45 lncRNAs sont associés à l'apparition de métastases dans les cancers du sein qui ne sont pas propagés aux ganglions lymphatiques (Sørensen et al., 2015), dont 24 sont présents sur notre puce à ADN. Parmi eux, seuls 5 et 2 lncRNAs se retrouvent respectivement dans nos listes uni- et multivariées. Enfin, parmi les 9 lncRNAs corrélés au risque de rechute identifiés par Liu et al (H. Liu et al., 2016), 5 sont présents sur notre puce, mais seulement un est pronostic (lnc-MRPS6-1). Notons qu'aucun lncRNAs pronostics identifiés par un de ces trois groupes ne se retrouve simultanément dans au

moins une autre étude. Ceci suggère que des études futures sont nécessaires pour identifier de manière reproductible les lncRNAs pronostics.

Ajoutons que de très nombreuses études proposent des gènes ou des signatures de gènes pronostics. Notre analyse ne fournit finalement qu'une liste de gènes lncRNAs et une signature supplémentaire du risque de rechute. En sus, nous n'observons pas que les valeurs pronostics des lncRNAs sont meilleures que les valeurs pronostics des mRNAs. Néanmoins, notre analyse est intéressante car elle montre pour la première fois que de très nombreux lncRNAs sont pronostics et qu'il existe une signature prédictive et reproductible de 27 lncRNAs. Nous pensons qu'il faudrait affiner le potentiel pronostic (au sein des sous-types) et prédictif (de réponse au traitement) des lncRNAs et le comparer aux marqueurs connus (ce qui pourrait faire l'objet d'une thèse à part entière). En effet, la grosse majorité des études pronostics/prédictives se limite aux paramètres cliniques, aux marquages histologiques et aux gènes codants. Elles n'ont pas encore permis de mettre en lumière des marqueurs de risque robustes au sein d'un sous-type (on sait qu'une fraction des tumeurs luminal A va devenir agressive et mener au décès, mais on arrive difficilement à les déceler précocement) ou des marqueurs prédictifs face aux différentes thérapies. Les lncRNAs pourraient être la pièce manquante de ces analyses et devraient être intégrés dans les futures recherches de nouveaux marqueurs pronostics et prédictifs.

Caractérisation expérimentale des lncRNAs

Nous avons sélectionné deux lncRNAs (lnc-KIN-2 et CYTOR) que nous avons analysés expérimentalement. Nos résultats démontrent leur implication dans le contrôle de la prolifération et confirment nos prédictions fonctionnelles. En outre, nous avons montré que lnc-KIN-2 régule l'expression de GATA3 et du gène ESR1 qui code pour le récepteur aux œstrogènes. Ceci est intéressant car GATA3 et ESR1 sont des gènes clés dans le développement de la glande mammaire et du cancer du sein. Ce sont des marqueurs des tumeurs de type luminal. GATA3 et ER participent à une boucle de rétroaction positive, chacun stimulant l'expression de l'autre. Dans le cancer du sein, les fonctions précises de GATA3 restent floues, il semble agir à la fois comme suppresseur de tumeur et comme

oncogène (Takaku, Grimm, and Wade, 2015). Vu leur proximité génomique, il est probable que lnc-KIN-2 et GATA3 partagent leur région promotrice et que lnc-KIN-2 régule « *en cis* » l'expression de GATA3, qui lui-même induit la prolifération par la stimulation de la voie ER. Des expériences supplémentaires, par exemple de recherche des partenaires protéiques, permettront de mieux comprendre le rôle de lnc-KIN-2 d'un point de vue mécanistique.

D'autre part, nous avons caractérisé CYTOR et montré son rôle dans la prolifération, la migration et le maintien de l'architecture normale de la cellule. Nous avons montré que l'état de méthylation de CYTOR est anti-corrélé à son expression. Une autre étude révèle que le promoteur de CYTOR est hypométhylé dans les cancers hépatiques en comparaison avec les tissus sains (J. Zhou et al., 2015). Ceci conforte nos résultats qui suggèrent que la méthylation de CYTOR joue un rôle dans la régulation de son expression. De futures analyses devront être mises en place afin de caractériser les modifications d'histones qui pourraient également influencer l'expression de CYTOR.

Nos résultats suggèrent que CYTOR module les voies EGFR et mTOR et le maintien du cytosquelette. Deux groupes ont confirmé cela en démontrant l'implication de CYTOR dans le contrôle de la voie mTOR dans un modèle de cancer du foie (Ji et al., 2015) et l'interaction entre EGFR et CYTOR dans un modèle de cancer gastrique (J. Zhou et al., 2015). Des données préliminaires du laboratoire suggèrent que CYTOR est à la fois nucléaire et cytoplasmique. Ces éléments nous permettent de proposer le modèle suivant : dans le noyau, CYTOR peut directement recruter des facteurs de transcription qui vont augmenter l'expression de gènes cibles tels que mTOR, PRKC α et RHOBTB3 qui modulent le cytosquelette. Parallèlement, CYTOR peut exercer ses fonctions dans le cytoplasme, en se liant à EGFR et permettre l'activation de certaines voies en aval telles que les voies mTORC1 (qui active la traduction et la prolifération) et mTORC2 (qui, via PKC α , module la polymérisation de l'actine et le maintien du cytosquelette). Enfin CYTOR pourrait également directement se lier à l'actine polymérisée et la stabiliser (Fig. 64).

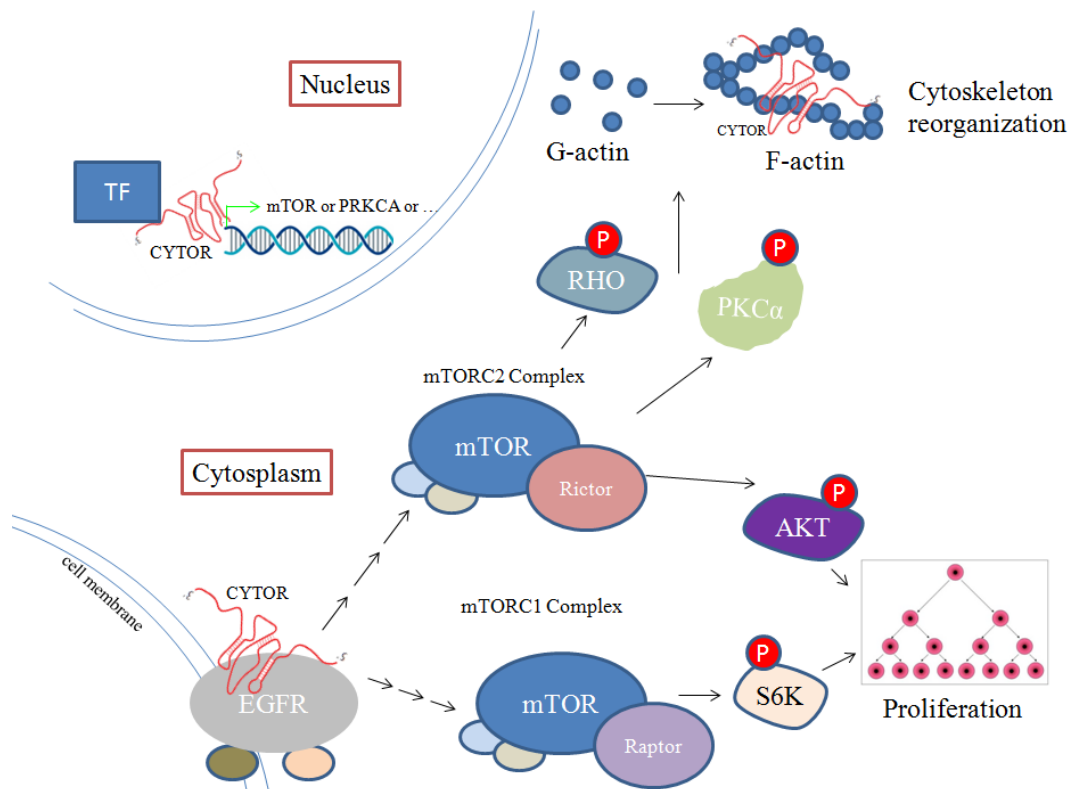


Fig. 64 : Modélisation de l'action de CYTOR.

Dans le noyau, CYTOR pourrait recruter des facteurs de transcription activateurs qui vont permettre la transcription des gènes cibles de CYTOR. Dans le cytoplasme, CYTOR pourrait interagir avec EGFR pour favoriser l'activation des voies mTORC1 et mTORC2. L'activation de la voie mTORC2 est importante pour la polymérisation de l'actine et l'organisation correcte du cytosquelette. Enfin, CYTOR pourrait également directement favoriser et stabiliser l'actine-F (polymérisée).

Ce modèle très spéculatif devra être confirmé par diverses expériences (dont certaines sont déjà en cours), telles que des western-blot en sous- et sur-expression de CYTOR, pour analyser les protéines suivantes et leur niveau de phosphorylation : EGFR, mTOR, Rictor, Raptor, AKT et PKC α (en aval de mTORC2), S6K (en aval de mTORC1). D'autre part, l'identification des partenaires protéiques de CYTOR (voir ci-dessous) devrait permettre d'affiner le modèle.

D'une manière intéressante, nos analyses, basées sur les données du TCGA, ont révélé que CYTOR est surexprimé dans de nombreux cancers, tels que le cancer du colon, du poumon et des reins (Fig. 65), ce qui pourrait en faire une cible thérapeutique intéressante dans différents cancers. Nos résultats *in vi-*

tro doivent évidemment être validés *in vivo*, par exemple en modèle murin de sur- ou sous-expression. En effet, plusieurs fonctions de lncRNAs identifiés *in vitro* n'ont pas été reproduites *in vivo*. Citons par exemple, HOTAIR qui est connu, *in vitro* pour inhiber les gènes *Hoxd* « *en trans* » en recrutant PRC2, qui peut triméthyliser les histones H3K27. Cependant, la délétion de HOTAIR en modèle murin a peu d'effet sur l'expression des gènes *Hoxd* et sur la marque H3K27me3 (Kohtz, 2014). En ce qui concerne CYTOR, deux équipes ont pu valider son rôle sur la prolifération, dans un modèle murin de cancer hépatique (J. Zhou et al., 2015) et de cancer gastrique (W. Chen et al., 2016). Ces résultats encourageants pourraient mener à de nouvelles études qui se focaliseraient sur CYTOR comme cible thérapeutique dans différentes tumeurs.

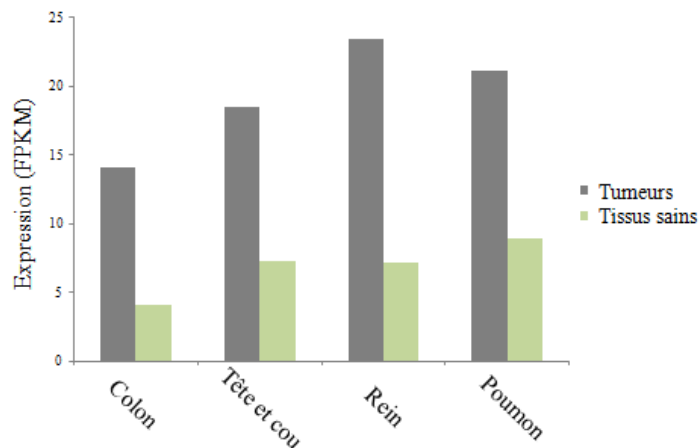


Fig. 65 : Expression de CYTOR dans différents cancers.

Enfin, discutons la stratégie expérimentale que nous avons mise en place pour étudier les lncRNAs. Nous avons utilisé des oligonucléotides antisens LNA pour diminuer spécifiquement l'expression de deux lncRNAs. Contrairement aux systèmes basés sur l'ARN interférence qui requièrent la machinerie cytoplasmique RISC pour fonctionner, les LNA agissent à la fois sur les lncRNAs nucléaires et cytoplasmiques. Les LNAs simples brins peuvent s'hybrider (par complémentarité des bases) sur leur cible lncRNA. Le duplex LNA/RNA est reconnu par la RNase H qui digère l'ARN en deux fragments qui seront ensuite digérés par des exonucléases. Au cours de ce travail, nous n'avons pas surexprimé nos lncRNAs d'intérêt car il n'est pas aisé de déterminer quelle

isoforme est majoritaire ou fonctionnelle. De plus, le début et la fin des transcrits lncRNAs ne sont pas bien définis et compliquent les étapes de clonage nécessaires à la surexpression. Notre laboratoire met en place des expériences de 3' et 5' RACE (*Rapid Amplification of cDNA Ends*) afin de répondre à ces questions et de permettre à terme la surexpression de lncRNAs d'intérêt.

Perspectives (CYTOR)

Terminons cette discussion en proposant quelques stratégies qui permettront de continuer à améliorer notre compréhension du cancer du sein, particulièrement du point de vue de l'implication des ARN non codants dans le développement de cette maladie complexe.

D'abord, de nouveaux outils bioinformatiques pourraient être mis en place afin d'affiner les prédictions fonctionnelles des lncRNAs et par exemple de prédire des liens de causalité entre les lncRNAs et d'autres lncRNAs ou ARNm comme cela a été appliqué pour les gènes codants (Haibe-Kains et al., 2012; Olsen et al., 2014).

Découvrir les partenaires protéiques des lncRNAs nous semble important pour mieux comprendre les mécanismes sous-jacents à l'action des lncRNAs. Pour ce faire, notre laboratoire met au point la technique de ChIRP (*Chromatin isolation by RNA purification*) qui permet d'isoler un lncRNA d'intérêt et d'identifier ses interacteurs protéiques par spectrométrie de masse. Nos premiers résultats de ChIRP, en utilisant une sonde contrôle positif (qui cible le petit ARN nucléaire U1), permettent de valider les interactants de U1 et suggèrent que cette technique fonctionne au sein de notre laboratoire. Cependant, jusqu'à présent, nous n'avons pas pu identifier de partenaire protéique de CYTOR. Enfin, nous avons également mis au point la technique de RIP (*RNA immunoprecipitation*). Nos premiers résultats, obtenus en modèle cellulaire mammaire, suggèrent que CYTOR interagit avec EGFR, mais pas avec mTOR. Ces données sont très intéressantes car elles indiquent que CYTOR pourrait agir en amont de vastes voies moléculaires en régulant l'activité d'EGFR. De futures études seront réalisées pour confirmer ces résultats préliminaires et mieux comprendre le rôle de CYTOR sur le fonctionnement d'EGFR.

Perspectives cliniques des lncRNAs

Comme évoqué dans l'introduction, de nombreux espoirs existent quant à l'utilisation des lncRNAs en clinique. D'une part, ils peuvent se retrouver dans les fluides corporels et servir de marqueurs diagnostics du développement précoce d'une tumeur. D'autre part, les lncRNAs oncogènes peuvent être ciblés par différentes approches thérapeutiques pour enrayer leur expression.

Nous avons largement démontré que les lncRNAs sont dérégulés au sein des tumeurs et qu'ils peuvent servir de marqueurs diagnostics et pronostics. Cependant, ces dernières années de plus en plus de recherches ont démontré que des lncRNAs « biomarqueurs » peuvent également se retrouver dans les fluides corporels facilement accessibles pour le praticien. Par exemple, la présence d'HOTAIR dans le sang permet de prédire le cancer colorectal avec une spécificité de 92% et une sensibilité de 67%. UCA1, un lncRNA présent dans les urines de patients atteints d'un cancer de la vessie permet de prédire cette maladie avec une spécificité de 92% et une sensibilité de 81% (Silva, Bullock, and Calin, 2015). Quand à CYTOR, présent dans les exosomes du plasma, semble être prédictif du cancer de l'estomac et de l'œsophage (H. Hu, Jie, and Zheng, 2016; Q. Li et al., 2015). Ces études ont démontré la valeur pronostic des lncRNAs dans les fluides corporels, il reste maintenant à démontrer que les lncRNAs sont présents dans ces fluides, dès les phases très précoces de la maladie.

Parallèlement, de nombreux essais cliniques sont en cours afin de démontrer l'efficacité thérapeutique de nouvelles drogues ciblant les lncRNAs. Différentes approches existent telles que l'utilisation d'« aptamères » (petits ARNs avec une conformation 3D particulière qui se lient sur une région spécifique d'un lncRNA pour empêcher son association avec des partenaires protéiques), de ribozymes en tête de marteau (« hammerhead ribozyme ») (ARNs simples brins qui catalysent le clivage d'un site spécifique sur un lncRNA), de siRNAs et de petits oligonucléotides antisens, tels que les LNA GapmeR utilisés dans ce travail. Cette dernière approche est celle qui a le plus avancé en recherche clinique (Silva, Bullock, and Calin, 2015; Parasramka et al., 2016). L'inhibition de CYTOR, qui est surexprimé dans différents cancers, diminue le

développement de tumeur en modèle murin (W. Chen et al., 2016) et suggère que des recherches cliniques pourraient être envisagées. Cependant, vu que ce lncRNA est exprimé à un relativement haut niveau de manière physiologique, il est probable que son inhibition entraîne des effets secondaires non négligeables.

Dans leur ensemble, les perspectives thérapeutiques des lncRNAs semblent donc intéressantes mais de nombreuses recherches sont encore à fournir pour améliorer la biodisponibilité des drogues ciblant les lncRNAs et vérifier leur innocuité.

IV. Conclusion générale

L'aboutissement de ce travail de thèse a permis de mieux comprendre l'action des ARNs non codants dans le cancer du sein. D'une part nous avons identifié de nouvelles cibles épigénétiques de miR-137 et miR-138 : les histones déméthylases de la famille KDM5. Nous avons également montré que ces miRs influencent la prolifération de cellules cancéreuses mammaires.

D'autre part, nous avons réalisé une étude à large échelle des lncRNAs dans les tumeurs mammaires. Les lncRNAs sont récemment devenus un sujet d'étude en vogue et nos résultats confirment le rôle prépondérant qu'ils jouent dans le cancer du sein. En collaboration avec l'équipe du Pr. Sotiriou de l'Institut Bordet et le groupe de bio-informaticiens du Pr. Bontempi, nous avons identifié 215 lncRNAs exprimés de manière aberrante dans les tumeurs mammaires et avons mis en évidence l'implication des lncRNAs dans différentes voies de signalisation moléculaire dérégulées dans le cancer. Nous avons montré que les lncRNAs permettent une stratification des cancers du sein en différents sous-groupes qui possèdent des propriétés cliniques variables. En outre, nous avons découvert que de nombreux lncRNAs peuvent servir de marqueurs pronostics de l'évolution du cancer et apportent des informations additionnelles et complémentaires aux marqueurs déjà connus.

D'une manière intéressante, ces travaux ont d'ores et déjà permis d'ouvrir un nouveau chapitre des connaissances sur la complexité du cancer du sein. En démontrant l'importance biologique des lncRNAs dans le développement des tumeurs mammaires, cette découverte laisse entrevoir la mise en lumière de nouveaux mécanismes par lesquels ces ARNs particuliers, qui ne codent pas des protéines, contribuent au processus de cancérogénèse.

Dans son ensemble cette étude ouvre une nouvelle approche thérapeutique en visant directement les lncRNAs dérégulés, et pourrait donc mener à des thérapies plus ciblées et plus efficaces.

BIBLIOGRAPHIE

Bibliographie

- Aebi, S, T Davidson, G Gruber, F Cardoso, and On behalf of the ESMO Guidelines Working Group. 2011. "Primary Breast Cancer: ESMO Clinical Practice Guidelines for Diagnosis, Treatment and Follow-Up." *Annals of Oncology*: Official Journal of the European Society for Medical Oncology / ESMO 22 Suppl 6 (suppl 6). Oxford University Press: vi12-24. doi:10.1093/annonc/mdr371.
- Alaimo, Salvatore, Rosalba Giugno, and Alfredo Pulvirenti. 2014. "ncPred: ncRNA-Disease Association Prediction through Tripartite Network-Based Inference." *Frontiers in Bioengineering and Biotechnology* 2: 71. doi:10.3389/fbioe.2014.00071.
- Allis, C. D., T. Jenuwein, and D. Reinberg. 2007. "Overview and Concepts." *Epigenetics, Overview and Concepts*. <http://www.cshlpress.com/pdf/sample/epigen.pdf>. "American Cancer Society. *Cancer Facts & Figures 2016*." 2016.
- Arrowsmith, Cheryl H., Chas Bountra, Paul V. Fish, Kevin Lee, and Matthieu Schapira. 2012. "Epigenetic Protein Families: A New Frontier for Drug Discovery." *Nature Reviews Drug Discovery* 11 (5). Nature Publishing Group: 384–400. doi:10.1038/nrd3674.
- Balaguer, Francesc, Alexander Link, Juan Jose Lozano, Miriam Cuatrecasas, Takeshi Nagasaka, C Richard Boland, and Ajay Goel. 2010. "Epigenetic Silencing of miR-137 Is an Early Event in Colorectal Carcinogenesis." *Cancer Research* 70 (16): 6609–18. doi:10.1158/0008-5472.CAN-10-0622.
- Bamodu, Oluwaseun Adebayo, Wen-Chien Huang, Wei-Hwa Lee, Alexander Wu, Liang Shun Wang, Michael Hsiao, Chi-Tai Yeh, et al. 2016. "Aberrant KDM5B Expression Promotes Aggressive Breast Cancer through MALAT1 Overexpression and Downregulation of Hsa-miR-448." *BMC Cancer* 16 (1). BioMed Central: 160. doi:10.1186/s12885-016-2108-5.
- Bannister, Andrew J, and Tony Kouzarides. 2011. "Regulation of Chromatin by Histone Modifications." *Cell Research* 21 (3): 381–95. doi:10.1038/cr.2011.22.
- Barrett, Angela, Bente Madsen, John Copier, Pei Juan Lu, Lucienne Cooper, Angelo G. Scibetta, Joy Burchell, and Joyce Taylor-Papadimitriou. 2002. "PLU-1 Nuclear Protein, Which Is Upregulated in Breast Cancer, Shows Restricted Expression in Normal Human Adult Tissues: A New Cancer/testis Antigen?" *International Journal of Cancer* 101 (6). Wiley Subscription Services, Inc., A Wiley Company: 581–88. doi:10.1002/ijc.10644.
- Bartel, David P. 2009. "MicroRNAs: Target Recognition and Regulatory Functions." *Cell* 136 (2). NIH Public Access: 215–33. doi:10.1016/j.cell.2009.01.002.
- Bergström, Rosita, Katia Savary, Anita Morén, Sylvain Guibert, Carl-Henrik Heldin, Rolf Ohlsson, and Aristidis Moustakas. 2010. "Transforming Growth Factor Beta Promotes Complexes between Smad Proteins and the CCCTC-Binding Factor on the H19 Imprinting Control Region Chromatin." *The Journal of Biological Chemistry* 285 (26): 19727–37. doi:10.1074/jbc.M109.088385.
- Bernstein, Bradley E, Alexander Meissner, and Eric S Lander. 2007. "The Mammalian Epigenome." *Cell* 128 (4): 669–81. doi:10.1016/j.cell.2007.01.033.
- Biterge. 2016. "A Mini Review on Post-Translational Histone Modifications.pdf." *MOJ Cell Science & Report* 3 (1).
- Blair, Lauren P, Jian Cao, Mike Ran Zou, Joyce Sayegh, and Qin Yan. 2011. "Epigenetic Regulation by Lysine Demethylase 5 (KDM5) Enzymes in Cancer." *Cancers* 3 (1): 1383–1404. doi:10.3390/cancers3011383.
- Bockhorn, Jessica, Aleix Prat, Ya-Fang Chang, Xia Liu, Simo Huang, Meng Shang, Chika Nwachukwu, et al. 2014. "Differentiation and Loss of Malignant Character of Spontaneous Pulmonary Metastases in Patient-Derived Breast Cancer Models." *Cancer Research* 74 (24). American Association for Cancer Research: 7406–17. doi:10.1158/0008-5472.CAN-14-1188.
- Breiling, Achim, and Frank Lyko. 2015. "Epigenetic Regulatory Functions of DNA Modifications: 5-Methylcytosine and beyond." *Epigenetics & Chromatin* 8 (1). BioMed Central: 24.

- doi:10.1186/s13072-015-0016-6.
- Brookes, Emily, and Yang Shi. 2014. Diverse Epigenetic Mechanisms of Human Disease. Annual Review of Genetics. Vol. 48. doi:10.1146/annurev-genet-120213-092518.
- Cabili, Moran N, Margaret C Dunagin, Patrick D McClanahan, Andrew Biaesch, Olivia Padovan-Merhar, Aviv Regev, John L Rinn, et al. 2015. "Localization and Abundance Analysis of Human lncRNAs at Single-Cell and Single-Molecule Resolution." Genome Biology 16 (1). BioMed Central: 20. doi:10.1186/s13059-015-0586-4.
- Carnero, Elena, Marina Barriocanal, Celia Prior, Juan Pablo Unfried, Victor Segura, Elizabeth Guruceaga, Mónica Enguita, et al. 2016. "Long Noncoding RNA EGOT Negatively Affects the Antiviral Response and Favors HCV Replication." EMBO Reports 17 (7). EMBO Press: 1013–28. doi:10.15252/embr.201541763.
- Chakravarty, Dimple, Andrea Sboner, Sujit S Nair, Eugenia Giannopoulou, Ruohan Li, Sven Hennig, Juan Miguel Mosquera, et al. 2014. "The Oestrogen Receptor Alpha-Regulated lncRNA NEAT1 Is a Critical Modulator of Prostate Cancer." Nature Communications 5. Nature Publishing Group: 5383. doi:10.1038/ncomms6383.
- Chen, Chen, Zhilu Li, Yuan Yang, Tingxiu Xiang, Weihong Song, and Shengchun Liu. 2015. "Microarray Expression Profiling of Dysregulated Long Non-Coding RNAs in Triple-Negative Breast Cancer." Cancer Biology & Therapy 16 (6): 856–65. doi:10.1080/15384047.2015.1040957.
- Chen, Geng, Ziyun Wang, Dongqing Wang, Chengxiang Qiu, Mingxi Liu, Xing Chen, Qipeng Zhang, Guiying Yan, and Qinghua Cui. 2013. "LncRNADisease: A Database for Long-Non-Coding RNA-Associated Diseases." Nucleic Acids Research 41 (D1). Oxford University Press: D983-6. doi:10.1093/nar/gks1099.
- Chen, Wen-ming, Ming-de Huang, Dao-ping Sun, Rong Kong, Tong-peng Xu, Rui Xia, Er-bao Zhang, and Yong-qian Shu. 2016. "Long Intergenic Non-Coding RNA 00152 Promotes Tumor Cell Cycle Progression by Binding to EZH2 and Repressing p15 and p21 in Gastric Cancer." Oncotarget 7 (9): 9773–87. doi:10.18632/oncotarget.6949.
- Chen, Xing, P. Bertone, E. Birney, P. Carninci, J. M. Claverie, L. J. Core, J. J. Waterfall, et al. 2015. "KATZLDA: KATZ Measure for the lncRNA-Disease Association Prediction." Scientific Reports 5 (November). Nature Publishing Group: 16840. doi:10.1038/srep16840.
- Chen, Xing, Chenggang Clarence Yan, Xu Zhang, and Zhu-Hong You. 2016. "Long Non-Coding RNAs and Complex Diseases: From Experimental Results to Computational Models." Briefings in Bioinformatics, June. Oxford University Press, bbw060. doi:10.1093/bib/bbw060.
- Christensen, Jesper, Karl Agger, Paul A C Cloos, Diego Pasini, Simon Rose, Lau Sennels, Juri Rappsilber, Klaus H Hansen, Anna Elisabetta Salcini, and Kristian Helin. 2007. "RBP2 Belongs to a Family of Demethylases, Specific for Tri- and Dimethylated Lysine 4 on Histone 3." Cell 128 (6): 1063–76. doi:10.1016/j.cell.2007.02.003.
- Chu, Ci, Robert C Spitale, and Howard Y Chang. 2015. "Technologies to Probe Functions and Mechanisms of Long Noncoding RNAs." Nature Structural & Molecular Biology. doi:10.1038/nsmb.2921.
- Ciruelos Gil, Eva Maria. 2014. "Targeting the PI3K/AKT/mTOR Pathway in Estrogen Receptor-Positive Breast Cancer." Cancer Treatment Reviews 40 (7): 862–71. doi:10.1016/j.ctrv.2014.03.004.
- Croce, Carlo M. 2010. "LINCing Chromatin Remodeling to Metastasis." Nature Biotechnology 28 (9). Nature Research: 931–32. doi:10.1038/nbt0910-931.
- Curran, Edward M, Barbara M Judy, Ngozi A Duru, Hui-Qun Wang, Leoncio A Vergara, Dennis B Lubahn, and D Mark Estes. 2006. "Estrogenic Regulation of Host Immunity against an Estrogen Receptor-Negative Human Breast Cancer." Clinical Cancer Research: An Official Journal of the American Association for Cancer Research 12 (19): 5641–47. doi:10.1158/1078-0432.CCR-05-1117.
- Dai, Xiaofeng, Liangjian Xiang, Ting Li, and Zhonghu Bai. 2016. "Cancer Hallmarks, Biomarkers and Breast Cancer Molecular Subtypes." Journal of Cancer 7 (10): 1281–94. doi:10.7150/jca.13141.

- Dawson, Mark A, and Tony Kouzarides. 2012. "Cancer Epigenetics: From Mechanism to Therapy." *Cell* 150 (1). Elsevier: 12–27. doi:10.1016/j.cell.2012.06.013.
- Deans, Carrie, and Keith A Maggert. 2015. "What Do You Mean, 'epigenetic'?" *Genetics* 199 (4): 887–96. doi:10.1534/genetics.114.173492.
- Dedeurwaerder, Sarah, Christine Desmedt, Emilie Calonne, Sandeep K Singhal, Benjamin Haibe-Kains, Matthieu Defrance, Stefan Michiels, et al. 2011. "DNA Methylation Profiling Reveals a Predominant Immune Component in Breast Cancers." *EMBO Molecular Medicine* 3 (12). EMBO Press: 726–41. doi:10.1002/emmm.201100801.
- Delatte, Benjamin, Rachel Deplus, and François Fuks. 2014. "Playing TETris with DNA Modifications." *The EMBO Journal* 33 (11). EMBO Press: 1198–1211. doi:10.15252/embj.201488290.
- Delatte, Benjamin, and François Fuks. 2013. "TET Proteins: On the Frenetic Hunt for New Cytosine Modifications." *Briefings in Functional Genomics* 12 (3): 191–204. doi:10.1093/bfpg/elt010.
- Denis, H., Olivier Van Grembergen, Benjamin Delatte, Sarah Dedeurwaerder, Pascale Putmans, Emilie Calonne, Françoise Rothé, Christos Sotiriou, François Fuks, and Rachel Deplus. 2015. "MicroRNAs Regulate KDM5 Histone Demethylases in Breast Cancer Cells." *Molecular bioSystems* 12 (2): 404–13. doi:10.1039/c5mb00513b.
- Denis, H, M N Ndlovu, and F Fuks. 2011. "Regulation of Mammalian DNA Methyltransferases: A Route to New Mechanisms." *EMBO Rep* 12 (7). Nature Publishing Group: 647–56. doi:10.1038/embor.2011.110.
- Denis, Hélène, Olivier Van Grembergen, Maria Agresta, Magdalena Stugiewicz, Valeria Spadotto, and Tiziana Bonaldi. 2015. "Molecular BioSystems," 2231–47. doi:10.1039/c5mb00513b.
- Derrien, Thomas, Rory Johnson, Giovanni Bussotti, Andrea Tanzer, Sarah Djebali, Hagen Tilgner, Gregory Guernec, et al. 2012. "The GENCODE v7 Catalog of Human Long Noncoding RNAs: Analysis of Their Gene Structure, Evolution, and Expression." *Genome Research* 22 (9): 1775–89. doi:10.1101/gr.132159.111.
- Di Leva, Gianpiero, Michela Garofalo, and Carlo M Croce. 2014. "MicroRNAs in Cancer." *Annual Review of Pathology* 9 (January). *Annual Reviews*: 287–314. doi:10.1146/annurev-pathol-012513-104715.
- Diederichs, Sven. 2014. "The Four Dimensions of Noncoding RNA Conservation." *Trends in Genetics* 30 (4): 121–23. doi:10.1016/j.tig.2014.01.004.
- Dimitrova, Emilia, Anne H Turberfield, and Robert J Klose. 2015. "Histone Demethylases in Chromatin Biology and Beyond." *EMBO Reports* 16 (12). EMBO Press: 1620–39. doi:10.15252/embr.201541113.
- DiTacchio, Luciano, Hiep D Le, Christopher Vollmers, Megumi Hatori, Michael Witcher, Julie Secombe, and Satchidananda Panda. 2011. "Histone Lysine Demethylase JARID1a Activates CLOCK-BMAL1 and Influences the Circadian Clock." *Science (New York, N.Y.)* 333 (6051): 1881–85. doi:10.1126/science.1206022.
- Djebali, Sarah, Carrie A Davis, Angelika Merkel, Alex Dobin, Timo Lassmann, Ali Mortazavi, Andrea Tanzer, et al. 2012. "Landscape of Transcription in Human Cells." *Nature* 489 (7414). Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.: 101–8. doi:10.1038/nature11233.
- Doose, Gero, Andrea Haake, Stephan H Bernhart, Cristina López, Sujitha Duggimpudi, Franziska Wojciech, Anke K Bergmann, et al. 2015. "MINCR Is a MYC-Induced lncRNA Able to Modulate MYC's Transcriptional Network in Burkitt Lymphoma Cells." *Proceedings of the National Academy of Sciences of the United States of America* 112 (38): E5261–5270. doi:10.1073/pnas.1505753112.
- Dozmorov, Mikhail G, Cory B Giles, Kristi A Koelsch, and Jonathan D Wren. 2013. "Systematic Classification of Non-Coding RNAs by Epigenomic Similarity." *BMC Bioinformatics* 14 Suppl 1 (14). BioMed Central: S2. doi:10.1186/1471-2105-14-S14-S2.
- Du, Zhou, Teng Fei, Roel G W Verhaak, Zhen Su, Yong Zhang, Myles Brown, Yiwen Chen, and X Shirley Liu. 2013. "Integrative Genomic Analyses Reveal Clinically Relevant Long Noncoding RNAs in Human Cancer." *Nature Structural & Molecular Biology* 20 (7). *Nature*

- Publishing Group: 908–13. doi:10.1038/nsmb.2591.
- Egger, Gerda, and Paola Arimondo. 2015. Drug Discovery in Cancer Epigenetics.
- Enkhbaatar, Zanabazar, Minoru Terashima, Dulamsuren Oktyabri, Shoichiro Tange, Akihiko Ishimura, Seiji Yano, and Takeshi Suzuki. 2013. “KDM5B Histone Demethylase Controls Epithelial-Mesenchymal Transition of Cancer Cells by Regulating the Expression of the microRNA-200 Family.” <http://dx.doi.org/10.4161/cc.25142>. Taylor & Francis. doi:10.4161/CC.25142.
- Ercan, C, P J van Diest, and M Vooijs. 2011. “Mammary Development and Breast Cancer: The Role of Stem Cells.” *Current Molecular Medicine* 11 (4). Europe PMC Funders: 270–85.
- “ESMO Breast Cancer: A Guide for Patients.” 2013. <http://www.esmo.org/Patients/Patient-Guides/Breast-Cancer>.
- Fatemi, Roya Pedram, Dmitry Velmeshev, and Mohammad Ali Faghihi. 2014. “De-Repressing lncRNA-Targeted Genes to Upregulate Gene Expression: Focus on Small Molecule Therapeutics.” *Molecular Therapy. Nucleic Acids* 3 (January). American Society of Gene & Cell Therapy: e196. doi:10.1038/mtna.2014.45.
- Fatima, Roshan, Vijay Suresh Akhade, Debosree Pal, Satyanarayana MR Rao, W Gilbert, ES Lander, LM Linton, et al. 2015. “Long Noncoding RNAs in Development and Cancer: Molecular Biomarkers and Therapeutic Targets.” *Molecular and Cellular Therapies* 3 (1). BioMed Central: 5. doi:10.1186/s40591-015-0042-6.
- Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray, F. 2012. “GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11.” Lyon, France: International Agency for Research on Cancer. <http://globocan.iarc.fr>.
- Fuks, François, Paul J Hurd, Rachel Deplus, and Tony Kouzarides. 2003. “The DNA Methyltransferases Associate with HP1 and the SUV39H1 Histone Methyltransferase.” *Nucleic Acids Research* 31 (9): 2305–12.
- Gajria, Devika, and Sarat Chandralapaty. 2011. “HER2-Amplified Breast Cancer: Mechanisms of Trastuzumab Resistance and Novel Targeted Therapies.” *Expert Review of Anticancer Therapy* 11 (2). NIH Public Access: 263–75. doi:10.1586/era.10.226.
- Gazinska, Patrycja, Anita Grigoriadis, John P Brown, Rosemary R Millis, Anca Mera, Cheryl E Gillett, Lars H Holmberg, Andrew N Tutt, and Sarah E Pinder. 2013. “Comparison of Basal-like Triple-Negative Breast Cancer Defined by Morphology, Immunohistochemistry and Transcriptional Profiles.” *Modern Pathology* 26 (7). Nature Publishing Group: 955–66. doi:10.1038/modpathol.2012.244.
- Gibb, Ewan A, Carolyn J Brown, and Wan L Lam. 2011. “The Functional Role of Long Non-Coding RNA in Human Carcinomas.” *Molecular Cancer* 10. BioMed Central: 38. doi:10.1186/1476-4598-10-38.
- Goldberg, Aaron D, C David Allis, and Emily Bernstein. 2007. “Epigenetics: A Landscape Takes Shape.” *Cell* 128 (4): 635–38. doi:10.1016/j.cell.2007.02.006.
- Gong, Jing, Wei Liu, Jiayou Zhang, Xiaoping Miao, and An-Yuan Guo. 2015. “lncRNASNP: A Database of SNPs in lncRNAs and Their Potential Functions in Human and Mouse.” *Nucleic Acids Research* 43 (Database issue). Oxford University Press: D181-6. doi:10.1093/nar/gku1000.
- Gray, Steven G. 2015. Epigenetic Cancer Therapy.
- Gupta, Rajnish A, Nilay Shah, Kevin C Wang, Jeewon Kim, Hugo M Horlings, J David, Miao-chih Tsai, et al. 2010. “Long Noncoding RNA HOTAIR Reprograms Chromatin State to Promote Cancer Metastasis” 464 (7291): 1071–76. doi:10.1038/nature08975.Long.
- Gutschner, Tony, and Sven Diederichs. 2012. “The Hallmarks of Cancer: A Long Non-Coding RNA Point of View.” *RNA Biology* 9 (6). Taylor & Francis: 703–19. doi:10.4161/rna.20481.
- Gutschner, Tony, Monika Hämmerle, Moritz Eissmann, Jeff Hsu, Youngsoo Kim, Gene Hung, Alexey Revenko, et al. 2013. “The Noncoding RNA MALAT1 Is a Critical Regulator of the Metastasis Phenotype of Lung Cancer Cells.” *Cancer Research* 73 (3). American Association for Cancer Research: 1180–89. doi:10.1158/0008-5472.CAN-12-2850.

- Ha, Minju, and V Narry Kim. 2014. "Regulation of microRNA Biogenesis." *Nature Reviews. Molecular Cell Biology* 15 (8). Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.: 509–24. doi:10.1038/nrm3838.
- Haibe-Kains, Benjamin, Catharina Olsen, Amira Djebbari, Gianluca Bontempi, Mick Correll, Christopher Bouton, and John Quackenbush. 2012. "Predictive Networks: A Flexible, Open Source, Web Application for Integration and Analysis of Human Gene Networks." *Nucleic Acids Research* 40 (Database issue): D866-75. doi:10.1093/nar/gkr1050.
- Han, Y, Z Wu, T Wu, Y Huang, Z Cheng, X Li, T Sun, X Xie, Y Zhou, and Z Du. 2016. "Tumor-Suppressive Function of Long Noncoding RNA MALAT1 in Glioma Cells by Downregulation of MMP2 and Inactivation of ERK/MAPK Signaling." *Cell Death & Disease* 7 (3). Nature Publishing Group: e2123. doi:10.1038/cddis.2015.407.
- Han, Yong, Yueyang Bi, Haiyang Bi, Guoqiang Zhang, Kai Cheng, and Zhenlin Yang. 2016. "miR-137 Suppresses the Invasion and Procedure of EMT of Human Breast Cancer Cell Line MCF-7 through Targeting CtBP1." *Human Cell* 29 (1). Springer Japan: 30–36. doi:10.1007/S13577-015-0124-4.
- Hangauer, Matthew J, Ian W Vaughn, and Michael T McManus. 2013. "Pervasive Transcription of the Human Genome Produces Thousands of Previously Unidentified Long Intergenic Noncoding RNAs." *PLoS Genetics* 9 (6). Public Library of Science: e1003569. doi:10.1371/journal.pgen.1003569.
- Hansji, Herah, Euphemia Y. Leung, Bruce C. Baguley, Graeme J. Finlay, and Marjan E. Askarian-Amiri. 2014. "Keeping Abreast with Long Non-Coding RNAs in Mammary Gland Development and Breast Cancer." *Frontiers in Genetics* 5 (October): 379. doi:10.3389/fgene.2014.00379.
- Hasanali, Zainul S, Bikramajit Singh Saroya, August Stuart, Sara Shimko, Juanita Evans, Mithun Vinod Shah, Kamal Sharma, et al. 2015. "Epigenetic Therapy Overcomes Treatment Resistance in T Cell Prolymphocytic Leukemia." *Science Translational Medicine* 7 (293): 293ra102. doi:10.1126/scitranslmed.aaa5079.
- Hoffmann-Cucuz, Pascale. 2010. "Le Sein : De L'embryologie À L'allaitement." *Université Joseph Fourier de Grenoble*.
- Holliday, R. 1994. "Epigenetics: An Overview." *Developmental Genetics* 15 (6): 453–57. doi:10.1002/dvg.1020150602.
- Holm, Karolina, Dorthe Grabau, Kristina Lövgren, Steina Aradottir, Sofia Gruvberger-Saal, Jillian Howlin, Lao H. Saal, et al. 2012. "Global H3K27 Trimethylation and EZH2 Abundance in Breast Tumor Subtypes." *Molecular Oncology* 6 (5): 494–506. doi:10.1016/j.molonc.2012.06.002.
- Horton, John R, Amanda Engstrom, Elizabeth L Zoeller, Xu Liu, John R Shanks, Xing Zhang, Margaret A Johns, Paula M Vertino, Haiyan Fu, and Xiaodong Cheng. 2016. "Characterization of a Linked Jumonji Domain of the KDM5/JARID1 Family of Histone H3 Lysine 4 Demethylases." *The Journal of Biological Chemistry* 291 (6). American Society for Biochemistry and Molecular Biology: 2631–46. doi:10.1074/jbc.M115.698449.
- Hu, Hai-bo, Hong-Ying Jie, and Xiang-Xiang Zheng. 2016. "Three Circulating LncRNA Predict Early Progress of Esophageal Squamous Cell Carcinoma." *Cellular Physiology and Biochemistry* 40 (1–2): 117–25. doi:10.1159/000452529.
- Hu, Xiaowen, Jinyi Guo, Lan Zheng, Chunsheng Li, Tim M Zheng, Janos L Tanyi, Shun Liang, et al. 2013. "The Heterochronic microRNA Let-7 Inhibits Cell Motility by Regulating the Genes in the Actin Cytoskeleton Pathway in Breast Cancer." *Molecular Cancer Research* : MCR 11 (3): 240–50. doi:10.1158/1541-7786.MCR-12-0432.
- Huang, He, Benjamin R. Sabari, Benjamin A. Garcia, C. David Allis, and Yingming Zhao. 2014. "SnapShot: Histone Modifications." *Cell* 159 (2). Elsevier: 458–458.e1. doi:10.1016/j.cell.2014.09.037.
- Huang, Yi, Shweta Nayak, Rachel Jankowitz, Nancy E Davidson, Steffi Oesterreich, T Jenuwein, CD Allis, et al. 2011. "Epigenetics in Breast Cancer: What's New?" *Breast Cancer Research* 13 (6). BioMed Central: 225. doi:10.1186/bcr2925.
- Huarte, Maite. 2015. "The Emerging Role of lncRNAs in Cancer." *Nature Medicine* 21 (11).

- Nature Research*: 1253–61. doi:10.1038/nm.3981.
- Huffman, Kenneth, and Elisabeth D Martinez. 2013. “Pre-Clinical Studies of Epigenetic Therapies Targeting Histone Modifiers in Lung Cancer.” *Frontiers in Oncology* 3. *Frontiers Media SA*: 235. doi:10.3389/fonc.2013.00235.
- Hui, Zhang, and Meng Xianglin. 2016. “Association of HOTAIR Expression with PI3K/Akt Pathway Activation in Adenocarcinoma of Esophagogastric Junction.” *Open Medicine* 11 (1): 36–40. doi:10.1515/MED-2016-0008.
- Inserm Expertise collective. 2008. “Inserm (Dir.). *Cancer et Environnement*.” Les éditions Inserm.
- Iorio, Marilena V, Manuela Ferracin, Chang-Gong Liu, Angelo Veronese, Riccardo Spizzo, Silvia Sabbioni, Eros Magri, et al. 2005. “MicroRNA Gene Expression Deregulation in Human Breast Cancer.” *Cancer Research* 65 (16). American Association for Cancer Research: 7065–70. doi:10.1158/0008-5472.CAN-05-1783.
- Jacob, François, and Jacques Monod. 1961. “Genetic Regulatory Mechanisms in the Synthesis of Proteins.” *Journal of Molecular Biology* 3 (3): 318–56. doi:10.1016/S0022-2836(61)80072-7.
- Janaki Ramaiah, M, A Lavanya, Mohsen Honarpisheh, Mojtaba Zarea, Utpal Bhadra, and Manika Pal Bhadra. 2014. “MiR-15/16 Complex Targets p70S6 Kinase 1 and Controls Cell Proliferation in MDA-MB-231 Breast Cancer Cells.” *Gene* 552 (2): 255–64. doi:10.1016/j.gene.2014.09.052.
- Ji, Jie, Junwei Tang, Lei Deng, Yu Xie, Runqiu Jiang, Guoqiang Li, and Beicheng Sun. 2015. “LINC00152 Promotes Proliferation in Hepatocellular Carcinoma by Targeting EpCAM via the mTOR Signaling Pathway.” *Oncotarget* 5 (November). *Impact Journals*.
- Jin, Yi, Dan Chen, Robert J Cabay, Anxun Wang, David L Crowe, and Xiaofeng Zhou. 2013. “Role of microRNA-138 as a Potential Tumor Suppressor in Head and Neck Squamous Cell Carcinoma.” *International Review of Cell and Molecular Biology* 303 (January): 357–85. doi:10.1016/B978-0-12-407697-6.00009-X.
- Jones, Peter A. 2012. “Functions of DNA Methylation: Islands, Start Sites, Gene Bodies and Beyond.” *Nature Reviews. Genetics* 13 (7). *Nature Publishing Group*: 484–92. doi:10.1038/nrg3230.
- Kim, KiBem, Andrew D Skora, Zhaobo Li, Qiang Liu, Ada J Tam, Richard L Blosser, Luis A Diaz, et al. 2014. “Eradication of Metastatic Mouse Cancers Resistant to Immune Checkpoint Blockade by Suppression of Myeloid-Derived Cells.” *Proceedings of the National Academy of Sciences of the United States of America* 111 (32). *National Academy of Sciences*: 11774–79. doi:10.1073/pnas.1410626111.
- Klein, Brianna J., Lianhua Piao, Yuanxin Xi, Hector Rincon-Arano, Scott B. Rothbart, Danni Peng, Hong Wen, et al. 2014. “The Histone-H3K4-Specific Demethylase KDM5B Binds to Its Substrate and Product through Distinct PHD Fingers.” *Cell Reports* 6 (2): 325–35. doi:10.1016/j.celrep.2013.12.021.
- Kohtz, Jhumku D. 2014. “Long Non-Coding RNAs Learn the Importance of Being in Vivo.” *Frontiers in Genetics* 5 (March). *Frontiers*: 45. doi:10.3389/fgene.2014.00045.
- Kooistra, Susanne Marije, and Kristian Helin. 2012. “Molecular Mechanisms and Potential Functions of Histone Demethylases.” *Nature Publishing Group* 13 (5). *Nature Publishing Group*: 297–311. doi:10.1038/nrm3327.
- Kouzarides, Tony. 2007. “Chromatin Modifications and Their Function.” *Cell* 128 (4): 693–705. doi:10.1016/j.cell.2007.02.005.
- Kozaki, Ken-ichi, Issei Imoto, Seiki Mogi, Ken Omura, and Johji Inazawa. 2008. “Exploration of Tumor-Suppressive microRNAs Silenced by DNA Hypermethylation in Oral Cancer.” *Cancer Research* 68 (7): 2094–2105. doi:10.1158/0008-5472.CAN-07-5194.
- Koziol, Magdalena J, Charles R Bradshaw, George E Allen, Ana S H Costa, Christian Frezza, and John B Gurdon. 2015. “Identification of Methylated Deoxyadenosines in Vertebrates Reveals Diversity in DNA Modifications.” *Nature Structural & Molecular Biology* 23 (1). *Nature Publishing Group*: 24–30. doi:10.1038/nsmb.3145.
- Kozomara, Ana, and Sam Griffiths-Jones. 2014. “miRBase: Annotating High Confidence

- microRNAs Using Deep Sequencing Data.*” *Nucleic Acids Research* 42 (Database issue): D68-73. doi:10.1093/nar/gkt1181.
- Kriaucionis, Skirmantas, and Nathaniel Heintz. 2009. “The Nuclear DNA Base 5-Hydroxymethylcytosine Is Present in Purkinje Neurons and the Brain.” *Science* (New York, N.Y.) 324 (5929). American Association for the Advancement of Science: 929–30. doi:10.1126/science.1169786.
- Kumar, Madhu S, Jun Lu, Kim L Mercer, Todd R Golub, and Tyler Jacks. 2007. “Impaired microRNA Processing Enhances Cellular Transformation and Tumorigenesis.” *Nature Genetics* 39 (5): 673–77. doi:10.1038/ng2003.
- La Fondation Registre du Cancer. 2013. “Les Chiffres Du Cancer.” 2013. http://www.kankerregister.org/Les_chiffres_du_cancer.
- Lao, Victoria Valinluck, and William M Grady. 2011. “Epigenetics and Colorectal Cancer.” *Nature Reviews. Gastroenterology & Hepatology* 8 (12). Nature Publishing Group: 686–700. doi:10.1038/nrgastro.2011.173.
- Lee, Jong-Min, Kyoung-Won Cho, Eun-Jung Kim, Qinghuang Tang, Kye-Seong Kim, Cheryll Tickle, Han-Sung Jung, et al. 2015. “A Contrasting Function for miR-137 in Embryonic Mammogenesis and Adult Breast Carcinogenesis.” *Oncotarget* 6 (26). Impact Journals: 22048–59.
- Lee, Min Gyu, Jessica Norman, Ali Shilatifard, and Ramin Shiekhattar. 2007. “Physical and Functional Association of a Trimethyl H3K4 Demethylase and Ring6a/MBLR, a Polycomb-like Protein.” *Cell* 128 (5): 877–87. doi:10.1016/j.cell.2007.02.004.
- Lee, R C, R L Feinbaum, and V Ambros. 1993. “The *C. Elegans* Heterochronic Gene *Lin-4* Encodes Small RNAs with Antisense Complementarity to *Lin-14*.” *Cell* 75 (5): 843–54.
- Leucci, Eleonora, Roberto Vendramin, Marco Spinazzi, Patrick Laurette, Mark Fiers, Jasper Wouters, Enrico Radaelli, et al. 2016. “Melanoma Addiction to the Long Non-Coding RNA SAMMSON.” *Nature* 531 (7595). Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.: 518–22. doi:10.1038/nature17161.
- Li, Jiefang, Qinrong Wang, Ruiling Wen, Jieman Liang, Xiaoling Zhong, Wei Yang, Dongxiang Su, and Jun Tang. 2015. “MiR-138 Inhibits Cell Proliferation and Reverses Epithelial-Mesenchymal Transition in Non-Small Cell Lung Cancer Cells by Targeting *GIT1* and *SEMA4C*.” *Journal of Cellular and Molecular Medicine* 19 (12): 2793–2805. doi:10.1111/jcmm.12666.
- Li, Jing, Zhenyu Xuan, and Changning Liu. 2013. “Long Non-Coding RNAs and Complex Human Diseases.” *International Journal of Molecular Sciences* 14 (9). Multidisciplinary Digital Publishing Institute: 18790–808. doi:10.3390/ijms140918790.
- Li, Juan, Hui Meng, Yun Bai, and Kai Wang. 2016. “Regulation of lncRNA and Its Role in Cancer Metastasis.” *Oncology Research* 23 (5): 205–17. doi:10.3727/096504016X14549667334007.
- Li, Jun-Hao, Shun Liu, Hui Zhou, Liang-Hu Qu, and Jian-Hua Yang. 2014. “starBase v2.0: Decoding miRNA-ceRNA, miRNA-ncRNA and Protein-RNA Interaction Networks from Large-Scale CLIP-Seq Data.” *Nucleic Acids Research* 42 (Database issue). Oxford University Press: D92-7. doi:10.1093/nar/gkt1248.
- Li, Qier, Yongfu Shao, Xinjun Zhang, Tuo Zheng, Min Miao, Lijun Qin, Bojun Wang, Guoliang Ye, Bingxiu Xiao, and Junming Guo. 2015. “Plasma Long Noncoding RNA Protected by Exosomes as a Potential Stable Biomarker for Gastric Cancer.” *Tumor Biology* 36 (3): 2007–12. doi:10.1007/s13277-014-2807-y.
- Lin, Shuibin, and Richard I. Gregory. 2015. “MicroRNA Biogenesis Pathways in Cancer.” *Nature Reviews Cancer* 15 (6). Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.: 321–33. doi:10.1038/nrc3932.
- Liu, Hairong, Juan Li, Pratirodh Koirala, Xianfeng Ding, Binghai Chen, Yiheng Wang, Zheng Wang, Chuanxin Wang, Xu Zhang, and Yin-Yuan Mo. 2016. “Long Non-Coding RNAs as Prognostic Markers in Human Breast Cancer.” *Oncotarget* 7 (15): 20584–96. doi:10.18632/oncotarget.7828.
- Liu, Xiang-Hua, Ming Sun, Feng-Qi Nie, Ying-Bin Ge, Er-Bao Zhang, Dan-Dan Yin, Rong Kong,

- et al. 2014. “Lnc RNA HOTAIR Functions as a Competing Endogenous RNA to Regulate HER2 Expression by Sponging miR-331-3p in Gastric Cancer.” *Molecular Cancer* 13 (January): 92. doi:10.1186/1476-4598-13-92.
- Liu, Xingyin, and Julie Secombe. 2015. “The Histone Demethylase KDM5 Activates Gene Expression by Recognizing Chromatin Context through Its PHD Reader Motif.” *Cell Reports* 13 (10). Elsevier: 2219–31. doi:10.1016/j.celrep.2015.11.007.
- Louie, Maggie C, Ashley McClellan, Christina Siewit, and Lauren Kawabata. 2010. “Estrogen Receptor Regulates E2F1 Expression to Mediate Tamoxifen Resistance.” *Molecular Cancer Research : MCR* 8 (3): 343–52. doi:10.1158/1541-7786.MCR-09-0395.
- Ma, Cynthia X., Tomás Reinert, Izabela Chmielewska, and Matthew J. Ellis. 2015. “Mechanisms of Aromatase Inhibitor Resistance.” *Nature Reviews Cancer* 15 (5): 261–75. doi:10.1038/nrc3920.
- Maio, Michele, Alessia Covre, Elisabetta Fratta, Anna Maria Di Giacomo, Pietro Taverna, Pier Giorgio Natali, Sandra Coral, and Luca Sigalotti. 2015. “Molecular Pathways: At the Crossroads of Cancer Epigenetics and Immunotherapy.” *Clinical Cancer Research* 21 (18). *Clinical Cancer Research*: 4040–47. doi:10.1158/1078-0432.CCR-14-2914.
- Maleszewska, Marta, and Bożena Kaminska. 2013. “Is Glioblastoma an Epigenetic Malignancy?” *Cancers* 5 (3). *Multidisciplinary Digital Publishing Institute*: 1120–39. doi:10.3390/cancers5031120.
- Malumbres, Marcos. 2013. “miRNAs and Cancer: An Epigenetics View.” *Molecular Aspects of Medicine* 34 (4): 863–74. doi:10.1016/j.mam.2012.06.005.
- Marques, Ana C, Jim Hughes, Bryony Graham, Monika S Kowalczyk, Doug R Higgs, Chris P Ponting, P Carninci, et al. 2013. “Chromatin Signatures at Transcriptional Start Sites Separate Two Equally Populated yet Distinct Classes of Intergenic Long Noncoding RNAs.” *Genome Biology* 14 (11). *BioMed Central*: R131. doi:10.1186/gb-2013-14-11-r131.
- Matouk, Imad J, Eli Raveh, Rasha Abu-lail, Shaul Mezan, Michal Gilon, Eitan Gershtain, Tatiana Birman, et al. 2014. “Oncofetal H19 RNA Promotes Tumor Metastasis.” *Biochimica et Biophysica Acta* 1843 (7): 1414–26. doi:10.1016/j.bbamcr.2014.03.023.
- McIntyre, Lauren M, Kenneth K Lopiano, Alison M Morse, Victor Amin, Ann L Oberg, Linda J Young, and Sergey V Nuzhdin. 2011. “RNA-Seq: Technical Variability and Sampling.” *BMC Genomics* 12 (1). *BioMed Central*: 293. doi:10.1186/1471-2164-12-293.
- Medina-Franco, José L. 2016. *Epi-Informatics : Discovery and Development of Small Molecule Epigenetic Drugs and Probes*.
- Mercer, Tim R, and John S Mattick. 2013. “Structure and Function of Long Noncoding RNAs in Epigenetic Regulation.” *Nature Structural & Molecular Biology* 20 (3). *Nature Publishing Group*: 300–307. doi:10.1038/nsmb.2480.
- Michalik, K. M., X. You, Y. Manavski, A. Doddaballapur, M. Zornig, T. Braun, D. John, et al. 2014. “Long Noncoding RNA MALAT1 Regulates Endothelial Cell Function and Vessel Growth.” *Circulation Research* 114 (9). *Lippincott Williams & Wilkins*: 1389–97. doi:10.1161/CIRCRESAHA.114.303265.
- Mondal, Tanmoy, Santhilal Subhash, Roshan Vaid, Stefan Enroth, Sireesha Uday, Björn Reinius, Sanhita Mitra, et al. 2015. “MEG3 Long Noncoding RNA Regulates the TGF- β Pathway Genes through Formation of RNA–DNA Triplex Structures.” *Nature Communications* 6 (July). *Nature Research*: 7743. doi:10.1038/ncomms8743.
- Morris, Kevin V., ed. 2016. *Long Non-Coding RNAs in Human Disease. Vol. 394. Current Topics in Microbiology and Immunology. Cham: Springer International Publishing*. doi:10.1007/978-3-319-23907-1.
- Morton, Sarah U, Paul J Scherz, Kimberly R Cordes, Kathryn N Ivey, Didier Y R Stainier, and Deepak Srivastava. 2008. “microRNA-138 Modulates Cardiac Patterning during Embryonic Development.” *Proceedings of the National Academy of Sciences of the United States of America* 105 (46): 17830–35. doi:10.1073/pnas.0804673105.
- Mostafa, Ahmed A, Dianne Codner, Kensuke Hirasawa, Yumiko Komatsu, Matthew N Young, Viktor Steimle, and Sheila Drover. 2014. “Activation of ER α Signaling Differentially Modulates IFN- γ Induced HLA-Class II Expression in Breast Cancer Cells.” Edited by

- Susan Kovats. PLoS One 9 (1). Public Library of Science: e87377.
doi:10.1371/journal.pone.0087377.
- Neault, Mathieu, Frédérick A Mallette, and Stéphane Richard. 2016. "miR-137 Modulates a Tumor Suppressor Network-Inducing Senescence in Pancreatic Cancer Cells." *Cell Reports* 14 (8): 1966–78. doi:10.1016/j.celrep.2016.01.068.
- NIH. 2016. "Fact Sheets: Female Breast Cancer."
<http://seer.cancer.gov/statfacts/html/breast.html>.
- Olsen, Catharina, Gianluca Bontempi, Frank Emmert-Streib, John Quackenbush, and Benjamin Haibe-Kains. 2014. "Relevance of Different Prior Knowledge Sources for Inferring Gene Interaction Networks." *Frontiers in Genetics* 5 (June). *Frontiers*: 177.
doi:10.3389/fgene.2014.00177.
- Orang, Ayla Valinezhad, Reza Safaralizadeh, and Mina Kazemzadeh-Bavili. 2014. "Mechanisms of miRNA-Mediated Gene Regulation from Common Downregulation to mRNA-Specific Upregulation." *International Journal of Genomics* 2014 (June). doi:10.1155/2014/970607.
- Outchkourov, Nikolay S., Jose M. Muiño, Kerstin Kaufmann, Wilfred F.J. van IJcken, Marian J. Groot Koerkamp, Dik van Leenen, Petra de Graaf, Frank C.P. Holstege, Frank G. Grosveld, and H.T. Marc Timmers. 2013. "Balancing of Histone H3K4 Methylation States by the Kdm5c/SMCX Histone Demethylase Modulates Promoter and Enhancer Function." *Cell Reports* 3 (4): 1071–79. doi:10.1016/j.celrep.2013.02.030.
- Parasramka, Mansi A., Sayantan Maji, Akiko Matsuda, Irene K. Yan, and Tushar Patel. 2016. "Long Non-Coding RNAs as Novel Targets for Therapy in Hepatocellular Carcinoma." *Pharmacology & Therapeutics* 161: 67–78. doi:10.1016/j.pharmthera.2016.03.004.
- Parker, Joel S, Michael Mullins, Maggie C U Cheang, Samuel Leung, David Voduc, Tammi Vickery, Sherri Davies, et al. 2009. "Supervised Risk Predictor of Breast Cancer Based on Intrinsic Subtypes." *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology* 27 (8): 1160–67. doi:10.1200/JCO.2008.18.1370.
- Pasquinelli, A E, B J Reinhart, F Slack, M Q Martindale, M I Kuroda, B Maller, D C Hayward, et al. 2000. "Conservation of the Sequence and Temporal Expression of Let-7 Heterochronic Regulatory RNA." *Nature* 408 (6808): 86–89. doi:10.1038/35040556.
- Peng, Weizhao, Shuang Si, Qingxia Zhang, Chaofeng Li, Fang Zhao, Fang Wang, Jia Yu, and Ren Ma. 2015. "Long Non-Coding RNA MEG3 Functions as a Competing Endogenous RNA to Regulate Gastric Cancer Progression." *Journal of Experimental & Clinical Cancer Research : CR* 34 (1). *BioMed Central*: 79. doi:10.1186/s13046-015-0197-7.
- Perou, C M, T Sørlie, M B Eisen, M van de Rijn, S S Jeffrey, C A Rees, J R Pollack, et al. 2000. "Molecular Portraits of Human Breast Tumours." *Nature* 406 (6797): 747–52.
doi:10.1038/35021093.
- Portela, A, and M Esteller. 2010. "Epigenetic Modifications and Human Disease." *Nat Biotechnol* 28 (10). *Nature Publishing Group*: 1057–68. doi:10.1038/nbt.1685.
- Prensner, John R, Matthew K Iyer, Anirban Sahu, Irfan a Asangani, Qi Cao, Lalit Patel, Ismael a Vergara, et al. 2013. "The Long Noncoding RNA SCHLAPI Promotes Aggressive Prostate Cancer and Antagonizes the SWI/SNF Complex." *Nature Genetics* 45 (11). *Nature Publishing Group*: 1392–98. doi:10.1038/ng.2771.
- Pusztai, Lajos, Roman Rouzier, and W Fraser Symmans. 2015. "CCR 20th Anniversary Commentary: Divide and Conquer-Breast Cancer Subtypes and Response to Therapy." *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research* 21 (16): 3575–77. doi:10.1158/1078-0432.CCR-14-3121.
- Quinn, Jeffrey J., and Howard Y. Chang. 2015. "Unique Features of Long Non-Coding RNA Biogenesis and Function." *Nature Reviews Genetics* 17 (1). *Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.*: 47–62.
doi:10.1038/nrg.2015.10.
- Raghavachari, Nalini, Jennifer Barb, Yanqin Yang, Poching Liu, Kimberly Woodhouse, Daniel Levy, Christopher J O'Donnell, Peter J Munson, and Gregory J Kato. 2012. "A Systematic Comparison and Evaluation of High Density Exon Arrays and RNA-Seq Technology Used to Unravel the Peripheral Blood Transcriptome of Sickle Cell Disease." *BMC Medical*

- Genomics 5 (1). *BioMed Central*: 28. doi:10.1186/1755-8794-5-28.
- Rahman, Mumtahena, Laurie K Jackson, W Evan Johnson, Dean Y Li, Andrea H Bild, and Stephen R Piccolo. 2015. "Alternative Preprocessing of RNA-Sequencing Data in The Cancer Genome Atlas Leads to Improved Analysis Results." *Bioinformatics* (Oxford, England) 31 (22): 3666–72. doi:10.1093/bioinformatics/btv377.
- Reddy, Kaladhar B. 2015. "MicroRNA (miRNA) in Cancer." *Cancer Cell International* 15 (1). *BioMed Central*: 38. doi:10.1186/s12935-015-0185-1.
- Rinn, John L, and Howard Y Chang. 2012. "Genome Regulation by Long Noncoding RNAs." *Annual Review of Biochemistry* 81 (January): 145–66. doi:10.1146/annurev-biochem-051410-092902.
- Robinson, David G, Jean Y Wang, and John D Storey. 2015. "A Nested Parallel Experiment Demonstrates Differences in Intensity-Dependence between RNA-Seq and Microarrays." *Nucleic Acids Research* 43 (20): e131. doi:10.1093/nar/gkv636.
- Rotili, Dante, and Antonello Mai. 2011. "Targeting Histone Demethylases: A New Avenue for the Fight against Cancer." *Genes & Cancer* 2 (6): 663–79. doi:10.1177/1947601911417976.
- Rottach, Andrea, Heinrich Leonhardt, and Fabio Spada. 2009. "DNA Methylation-Mediated Epigenetic Control." *Journal of Cellular Biochemistry* 108 (1): 43–51. doi:10.1002/jcb.22253.
- Sauvageau, Martin, Loyal A Goff, Simona Lodato, Boyan Bonev, Abigail F Groff, Chiara Gerhardinger, Diana B Sanchez-Gomez, et al. 2013. "Multiple Knockout Mouse Models Reveal lincRNAs Are Required for Life and Brain Development." *eLife* 2 (December). doi:10.7554/eLife.01749.
- Schafer, Zachary T, and Joan S Brugge. 2007. "IL-6 Involvement in Epithelial Cancers." *The Journal of Clinical Investigation* 117 (12): 3660–63. doi:10.1172/JCI34237.
- Scheen, A.J., and C. Junien. 2012. "Epigenetique, Interface Entre Environnement et Genes: Rôle Dans Les Maladies Complexes." *Rev Med Liège*, 67 5-6. https://orbi.ulg.ac.be/bitstream/2268/130382/1/20120506_06_epigenetique.pdf.
- Schmitt, Adam M., and Howard Y. Chang. 2016. "Long Noncoding RNAs in Cancer Pathways." *Cancer Cell* 29 (4): 452–63. doi:10.1016/j.ccell.2016.03.010.
- Schmitz, Sandra U., Phillip Grote, and Bernhard G. Herrmann. 2016. "Mechanisms of Long Noncoding RNA Function in Development and Disease." *Cellular and Molecular Life Sciences* 73 (13): 2491–2509. doi:10.1007/s00018-016-2174-5.
- Senkus, E, S Kyriakides, S Ohno, F Penault-Llorca, P Poortmans, E Rutgers, S Zackrisson, and F Cardoso. 2015. "Primary Breast Cancer: ESMO Clinical Practice Guidelines for Diagnosis, Treatment and Follow-Up†." *Annals of Oncology : Official Journal of the European Society for Medical Oncology / ESMO* 26 Suppl 5 (suppl_5): v8–30. doi:10.1093/annonc/mdv298.
- Shan, L, X Li, L Liu, X Ding, Q Wang, Y Zheng, Y Duan, et al. 2014. "GATA3 Cooperates with PARP1 to Regulate CCND1 Transcription through Modulating Histone H1 Incorporation." *Oncogene* 33 (24). *Nature Publishing Group*: 3205–16. doi:10.1038/onc.2013.270.
- Shore, Amy N, and Jeffrey M Rosen. 2014. "Regulation of Mammary Epithelial Cell Homeostasis by lincRNAs." *The International Journal of Biochemistry & Cell Biology, March*. doi:10.1016/j.biocel.2014.03.012.
- Siegel, R, and D Naishadham. 2013. "Cancer Statistics, 2013." *CA: A Cancer Journal for ...* 63 (1). *Wiley Subscription Services, Inc., A Wiley Company*: 11–30. doi:10.3322/caac.20073.
- Silva, Andreia, Marc Bullock, and George Calin. 2015. "The Clinical Relevance of Long Non-Coding RNAs in Cancer." *Cancers* 7 (4). *Multidisciplinary Digital Publishing Institute*: 2169–82. doi:10.3390/cancers7040884.
- Singel, Stina M, Crystal Cornelius, Elma Zaganjor, Kimberly Batten, Venetia R Sarode, Dennis L Buckley, Yan Peng, et al. 2014. "KIF14 Promotes AKT Phosphorylation and Contributes to Chemoresistance in Triple-Negative Breast Cancer." *Neoplasia* (New York, N.Y.) 16 (3): 247–56, 256.e2. doi:10.1016/j.neo.2014.03.008.
- Smith, Zachary D, and Alexander Meissner. 2013. "DNA Methylation: Roles in Mammalian Development." *Nature Reviews. Genetics* 14 (3). *Nature Publishing Group*: 204–20.

- doi:10.1038/nrg3354.
- Sørensen, Kristina P, Mads Thomassen, Qihua Tan, Martin Bak, Søren Cold, Mark Burton, Martin J Larsen, and Torben A Kruse. 2015. "Long Non-Coding RNA Expression Profiles Predict Metastasis in Lymph Node-Negative Breast Cancer Independently of Traditional Prognostic Markers." *Breast Cancer Research : BCR* 17: 55. doi:10.1186/s13058-015-0557-4.
- Sorlie, Therese, Robert Tibshirani, Joel Parker, Trevor Hastie, J S Marron, Andrew Nobel, Shihong Deng, et al. 2003. "Repeated Observation of Breast Tumor Subtypes in Independent Gene Expression Data Sets." *Proceedings of the National Academy of Sciences of the United States of America* 100 (14). *National Academy of Sciences*: 8418–23. doi:10.1073/pnas.0932692100.
- Sotiriou, Christos, Soek-Ying Neo, Lisa M McShane, Edward L Korn, Philip M Long, Amir Jazaeri, Philippe Martiat, Steve B Fox, Adrian L Harris, and Edison T Liu. 2003. "Breast Cancer Classification and Prognosis Based on Gene Expression Profiles from a Population-Based Study." *Proceedings of the National Academy of Sciences of the United States of America* 100 (18): 10393–98. doi:10.1073/pnas.1732912100.
- Sotiriou, Christos, and Lajos Pusztai. 2009. "Gene-Expression Signatures in Breast Cancer." <http://dx.doi.org/10.1056/NEJMra0801289>. *Massachusetts Medical Society*.
- Steponaitiene, Ruta, Juozas Kupcinskis, Cosima Langner, Francesc Balaguer, Linas Venclauskas, Henrikas Pauzas, Algimantas Tamelis, et al. 2016. "Epigenetic Silencing of miR-137 Is a Frequent Event in Gastric Carcinogenesis." *Molecular Carcinogenesis* 55 (4): 376–86. doi:10.1002/mc.22287.
- Stirzaker, Clare, Elena Zotenko, Jenny Z. Song, Wenjia Qu, Shalima S. Nair, Warwick J. Locke, Andrew Stone, et al. 2015. "Methylome Sequencing in Triple-Negative Breast Cancer Reveals Distinct Methylation Clusters with Prognostic Value." *Nature Communications* 6 (February). *Nature Publishing Group*: 5899. doi:10.1038/ncomms6899.
- Stresemann, Carlo, and Frank Lyko. 2008. "Modes of Action of the DNA Methyltransferase Inhibitors Azacytidine and Decitabine." *International Journal of Cancer* 123 (1). *Wiley Subscription Services, Inc., A Wiley Company*: 8–13. doi:10.1002/ijc.23607.
- Su, Xiaoping, Gabriel G. Malouf, Yunxin Chen, Jianping Zhang, Hui Yao, Vicente Valero, John N. Weinstein, et al. 2014. "Comprehensive Analysis of Long Non-Coding RNAs in Human Breast Cancer Clinical Subtypes." *Oncotarget* 5 (20): 9864–76.
- Subramanian, A., P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, et al. 2005. "Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles." *Proceedings of the National Academy of Sciences* 102 (43). *National Academy of Sciences*: 15545–50. doi:10.1073/pnas.0506580102.
- Sullivan, N J, A K Sasser, A E Axel, F Vesuna, V Raman, N Ramirez, T M Oberyzyzn, and B M Hall. 2009. "Interleukin-6 Induces an Epithelial-Mesenchymal Transition Phenotype in Human Breast Cancer Cells." *Oncogene* 28 (33): 2940–47. doi:10.1038/onc.2009.180.
- Sun, Jie, Xihai Chen, Zhenzhen Wang, Maoni Guo, Hongbo Shi, Xiaojun Wang, Liang Cheng, et al. 2015. "A Potential Prognostic Long Non-Coding RNA Signature to Predict Metastasis-Free Survival of Breast Cancer Patients." *Scientific Reports* 5 (November): 16553. doi:10.1038/srep16553.
- Taft, Ryan J, Michael Pheasant, and John S Mattick. 2007. "The Relationship between Non-Protein-Coding DNA and Eukaryotic Complexity." *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology* 29 (3): 288–99. doi:10.1002/bies.20544.
- Tahiliani, Mamta, Kian Peng Koh, Yinghua Shen, William A Pastor, Hozefa Bandukwala, Yevgeny Brudno, Suneet Agarwal, et al. 2009. "Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1." *Science (New York, N.Y.)* 324 (5929). *American Association for the Advancement of Science*: 930–35. doi:10.1126/science.1170116.
- Takahashi, Ryou-u, Hiroaki Miyazaki, and Takahiro Ochiya. 2015. "The Roles of MicroRNAs in Breast Cancer." *Cancers* 7 (2). *Multidisciplinary Digital Publishing Institute*: 598–616. doi:10.3390/cancers7020598.

- Takaku, Motoki, Sara A Grimm, and Paul A Wade. 2015. "GATA3 in Breast Cancer: Tumor Suppressor or Oncogene?" *Gene Expression* 16 (4). NIH Public Access: 163–68. doi:10.3727/105221615X14399878166113.
- Tang, Jen-Yang, Jin-Ching Lee, Yung-Ting Chang, Ming-Feng Hou, Hurng-Wern Huang, Chih-Chuang Liaw, Hsueh-Wei Chang, et al. 2013. "Long Noncoding RNAs-Related Diseases, Cancers, and Drugs." *TheScientificWorldJournal* 2013. Hindawi Publishing Corporation: 943539. doi:10.1155/2013/943539.
- Tarantino, C., G. Paoletta, L. Cozzuto, G. Minopoli, L. Pastore, S. Parisi, and T. Russo. 2010. "miRNA 34a, 100, and 137 Modulate Differentiation of Mouse Embryonic Stem Cells." *The FASEB Journal* 24 (9). Federation of American Societies for Experimental Biology: 3255–63. doi:10.1096/fj.09-152207.
- Tessarz, Peter, and Tony Kouzarides. 2014. "Histone Core Modifications Regulating Nucleosome Structure and Dynamics." *Nature Reviews. Molecular Cell Biology* 15 (11). Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.: 703–8. doi:10.1038/nrm3890.
- The Cancer Genome Atlas. 2012. "Comprehensive Molecular Portraits of Human Breast Tumours." *Nature* 490 (7418): 61–70. doi:10.1038/nature11412.
- Tian, Xiaoling, and Guoxiong Xu. 2015. "Clinical Value of lncRNA MALAT1 as a Prognostic Marker in Human Cancer: Systematic Review and Meta-Analysis." *BMJ Open* 5 (9). British Medical Journal Publishing Group: e008653. doi:10.1136/bmjopen-2015-008653.
- Tokunaga, Eriko, Yuichi Hisamatsu, Kimihiro Tanaka, Nami Yamashita, Hiroshi Saeki, Eiji Oki, Hiroyuki Kitao, and Yoshihiko Maehara. 2014. "Molecular Mechanisms Regulating the Hormone Sensitivity of Breast Cancer." *Cancer Science* 105 (11): 1377–83. doi:10.1111/cas.12521.
- Trapnell, Cole, Lior Pachter, and Steven L Salzberg. 2009. "TopHat: Discovering Splice Junctions with RNA-Seq." *Bioinformatics (Oxford, England)* 25 (9). Oxford University Press: 1105–11. doi:10.1093/bioinformatics/btp120.
- Tripathi, Vidisha, Jonathan D Ellis, Zhen Shen, David Y Song, Qun Pan, Andrew T Watt, Susan M Freier, et al. 2010. "The Nuclear-Retained Noncoding RNA MALAT1 Regulates Alternative Splicing by Modulating SR Splicing Factor Phosphorylation." *Molecular Cell* 39 (6). NIH Public Access: 925–38. doi:10.1016/j.molcel.2010.08.011.
- Tseng, Yuen-Yi, Branden S. Moriarity, Wuming Gong, Ryutarō Akiyama, Ashutosh Tiwari, Hiroko Kawakami, Peter Ronning, et al. 2014. "PVT1 Dependence in Cancer with MYC Copy-Number Increase." *Nature* 512 (7512): 82–86. doi:10.1038/nature13311.
- Turner, N C, J S Reis-Filho, A M Russell, R J Springall, K Ryder, D Steele, K Savage, et al. 2007. "BRCA1 Dysfunction in Sporadic Basal-like Breast Cancer." *Oncogene* 26 (14). Nature Publishing Group: 2126–32. doi:10.1038/sj.onc.1210014.
- Van Keymeulen, Alexandra, Ana Sofia Rocha, Marielle Ousset, Benjamin Beck, Gaëlle Bouvencourt, Jason Rock, Neha Sharma, Sophie Dekoninck, and Cédric Blanpain. 2011. "Distinct Stem Cells Contribute to Mammary Gland Development and Maintenance." *Nature* 479 (7372). Nature Publishing Group: 189–93. doi:10.1038/nature10573.
- Viré, Emmanuelle, Carmen Brenner, Rachel Deplus, Loïc Blanchon, Mario Fraga, Céline Didelot, Lluís Morey, et al. 2006. "The Polycomb Group Protein EZH2 Directly Controls DNA Methylation." *Nature* 439 (7078): 871–74. doi:10.1038/nature04431.
- Volders, Pieter-Jan, Kenny Helsens, Xiaowei Wang, Björn Menten, Lennart Martens, Kris Gevaert, Jo Vandesompele, and Pieter Mestdagh. 2013. "LNCipedia: A Database for Annotated Human lncRNA Transcript Sequences and Structures." *Nucleic Acids Research* 41 (Database issue): D246-51. doi:10.1093/nar/gks915.
- Vrba, Lukas, José L. Muñoz-Rodríguez, Martha R. Stampfer, Bernard W. Futscher, H Guo, NT Ingolia, JS Weissman, et al. 2013. "miRNA Gene Promoters Are Frequent Targets of Aberrant DNA Methylation in Human Breast Cancer." Edited by J.Christopher States. *PLoS ONE* 8 (1). Public Library of Science: e54398. doi:10.1371/journal.pone.0054398.
- Vuaroqueaux, Vincent, Patrick Urban, Martin Labuhn, Mauro Delorenzi, Pratyaksha Wirapati, Christopher C Benz, Renata Flury, et al. 2007. "Low E2F1 Transcript Levels Are a Strong

- Determinant of Favorable Breast Cancer Outcome.* Breast Cancer Research : BCR 9 (3): R33. doi:10.1186/bcr1681.
- Waddington, C.H. 1942. "The Epigenotype." *Endeavour* 1: 18–20.
- Wang, Junyi, Ruixia Ma, Wei Ma, Ji Chen, Jichun Yang, Yaguang Xi, and Qinghua Cui. 2016. "LncDisease: A Sequence Based Bioinformatics Tool for Predicting lncRNA-Disease Associations." *Nucleic Acids Research* 44 (9). Oxford University Press: e90–e90. doi:10.1093/nar/gkw093.
- Wei, J. T., Z. Feng, A. W. Partin, E. Brown, I. Thompson, L. Sokoll, D. W. Chan, et al. 2014. "Can Urinary PCA3 Supplement PSA in the Early Detection of Prostate Cancer?" *Journal of Clinical Oncology* 32 (36). American Society of Clinical Oncology: 4066–72. doi:10.1200/JCO.2013.52.8505.
- Wei, Jun, Edjah K Nduom, Ling-Yuan Kong, Yuuri Hashimoto, Shuo Xu, Konrad Gabrusiewicz, Xiaoyang Ling, et al. 2015. "MiR-138 Exerts Anti-Glioma Efficacy by Targeting Immune Checkpoints." *Neuro-Oncology*, December. doi:10.1093/neuonc/nov292.
- Weintraub, Karen. 2016. "Take Two: Combining Immunotherapy with Epigenetic Drugs to Tackle Cancer." *Nature Medicine* 22 (1). Nature Publishing Group: 8–10. doi:10.1038/nm0116-8.
- Woodson, Karen, Keith J. O'Reilly, Jeffrey C. Hanson, Dayne Nelson, Elyse L. Walk, and Joseph A. Tangrea. 2008. "The Usefulness of the Detection of GSTP1 Methylation in Urine as a Biomarker in the Diagnosis of Prostate Cancer." *The Journal of Urology* 179 (2): 508–12. doi:10.1016/j.juro.2007.09.073.
- Wright, Mathew W. 2014. "A Short Guide to Long Non-Coding RNA Gene Nomenclature." *Human Genomics* 8: 7. doi:10.1186/1479-7364-8-7.
- Wu, Zhuomin, Xiaoxia Liu, Li Liu, Houliang Deng, Jingjing Zhang, Qian Xu, Bohong Cen, and Aimin Ji. 2014. "Regulation of lncRNA Expression." *Cellular and Molecular Biology Letters* 19 (4): 561–75. doi:10.2478/s11658-014-0212-6.
- Xu, C, H Fu, L Gao, L Wang, W Wang, J Li, Y Li, et al. 2014. "BCR-ABL/GATA1/miR-138 Mini Circuitry Contributes to the Leukemogenesis of Chronic Myeloid Leukemia." *Oncogene* 33 (1). Macmillan Publishers Limited: 44–54. doi:10.1038/ncr.2012.557.
- Y, Wu, Tan C, Weng WW, Deng Y, Zhang QY, Yang XQ, Gan HL, et al. 2015. "Long Non-Coding RNA Linc00152 Is a Positive Prognostic Factor for and Demonstrates Malignant Biological Behavior in Clear Cell Renal Cell Carcinoma." *American Journal of Cancer Research* 6 (2): 285–99.
- Yamamoto, Shoji, Zhenhua Wu, Hege G. Russnes, Shinji Takagi, Guillermo Peluffo, Charles Vaske, Xi Zhao, et al. 2014. "JARID1B Is a Luminal Lineage-Driving Oncogene in Breast Cancer." *Cancer Cell* 25 (6). Elsevier: 762–77. doi:10.1016/j.ccr.2014.04.024.
- Yamane, Kenichi, Keisuke Tateishi, Robert J Klose, Jia Fang, Laura A Fabrizio, Hediye Erdjument-Bromage, Joyce Taylor-Papadimitriou, Paul Tempst, and Yi Zhang. 2007. "PLU-1 Is an H3K4 Demethylase Involved in Transcriptional Repression and Breast Cancer Cell Proliferation." *Molecular Cell* 25 (6): 801–12. doi:10.1016/j.molcel.2007.03.001.
- Yang, Guang, Shenghong Zhang, Fei Gao, Zhenyin Liu, Mingjian Lu, Sheng Peng, Tao Zhang, and Fujun Zhang. 2014. "Osteopontin Enhances the Expression of HOTAIR in Cancer Cells via IRF1." *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1839 (9): 837–48. doi:10.1016/j.bbagr.2014.06.020.
- Yang, Guodong, Xiaozhao Lu, and Lijun Yuan. 2014. "LncRNA: A Link between RNA and Cancer." *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1839 (11): 1097–1109. doi:10.1016/j.bbagr.2014.08.012.
- Yang, Tian, Hong Li, Asmitananda Thakur, Tianjun Chen, Jing Xue, Dan Li, and Mingwei Chen. 2015. "FOXP4 Modulates Tumor Growth and Independently Associates with miR-138 in Non-Small Cell Lung Cancer Cells." *Tumour Biology : The Journal of the International Society for Oncodevelopmental Biology and Medicine* 36 (10): 8185–91. doi:10.1007/s13277-015-3498-8.
- Yang, Yongfang, Liwei Wen, Hongliang Zhu, JM Johnson, S Edwards, D Shoemaker, EE Schadt, et al. 2015. "Unveiling the Hidden Function of Long Non-Coding RNA by Identifying Its Major Partner-Protein." *Cell & Bioscience* 5 (1). BioMed Central: 59. doi:10.1186/s13578-

015-0050-x.

- Yao, Lijing, Hui Shen, Peter W Laird, Peggy J Farnham, and Benjamin P Berman. 2015. "Inferring Regulatory Element Landscapes and Transcription Factor Networks from Cancer Methylomes." *Genome Biology* 16 (1): 105. doi:10.1186/s13059-015-0668-3.
- Yin, Jingwen, Juda Lin, Xudong Luo, Yanyan Chen, Zheng Li, Guoda Ma, and Keshen Li. 2014. "miR-137: A New Player in Schizophrenia." *International Journal of Molecular Sciences* 15 (2). Multidisciplinary Digital Publishing Institute: 3262–71. doi:10.3390/ijms15023262.
- Yoshimoto, Rei, Akila Mayeda, and Minoru Yoshida. 2016. "MALAT1 Long Non-Coding RNA in Cancer." *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1859 (1): 192–99. doi:10.1016/j.bbagr.2015.09.012.
- Zeng, Zhaolei, Huanxin Lin, Xiaohui Zhao, Guanglin Liu, Xi Wang, Ruihua Xu, Kun Chen, Jun Li, and Libing Song. 2012. "Overexpression of GOLPH3 Promotes Proliferation and Tumorigenicity in Breast Cancer via Suppression of the FOXO1 Transcription Factor." *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research* 18 (15): 4059–69. doi:10.1158/1078-0432.CCR-11-3156.
- Zhang, B., L. Gunawardane, F. Niazi, F. Jahanbani, X. Chen, and S. Valadkhan. 2014. "A Novel RNA Motif Mediates the Strict Nuclear Localization of a Long Noncoding RNA." *Molecular and Cellular Biology* 34 (12). American Society for Microbiology: 2318–29. doi:10.1128/MCB.01673-13.
- Zhang, Chao, and Yanmei Hua. 2016. "MicroRNA-138 Inhibits Proliferation and Migration of Breast Cancer Cells by Targeting c-Met." *Int J Clin Exp Me* 9 (2): 1290–97.
- Zhang, Jun, Dan Liu, Zhuo Feng, Jun Mao, Chunying Zhang, Ying Lu, Jiazhi Li, Qingqing Zhang, Qing Li, and Lianhong Li. 2016. "MicroRNA-138 Modulates Metastasis and EMT in Breast Cancer Cells by Targeting Vimentin." *Biomedicine & Pharmacotherapy* 77: 135–41. doi:10.1016/j.biopha.2015.12.018.
- Zhang, Ling, Fu Yang, Ji-hang Yuan, Sheng-xian Yuan, Wei-ping Zhou, Xi-song Huo, Dan Xu, Hai-shan Bi, Fang Wang, and Shu-han Sun. 2013. "Epigenetic Activation of the MiR-200 Family Contributes to H19-Mediated Metastasis Suppression in Hepatocellular Carcinoma." *Carcinogenesis* 34 (3): 577–86. doi:10.1093/carcin/bgs381.
- Zhang, Xun, Kimberley Rice, Yingying Wang, Wendy Chen, Ying Zhong, Yuki Nakayama, Yunli Zhou, and Anne Klibanski. 2010. "Maternally Expressed Gene 3 (MEG3) Noncoding Ribonucleic Acid: Isoform Structure, Expression, and Functions." *Endocrinology* 151 (3). The Endocrine Society: 939–47. doi:10.1210/en.2009-0657.
- Zhao, Weihong, Jiancheng Luo, and Shunchang Jiao. 2014. "Comprehensive Characterization of Cancer Subtype Associated Long Non-Coding RNAs and Their Clinical Implications." *Scientific Reports* 4 (January). Nature Publishing Group: 6591. doi:10.1038/srep06591.
- Zhou, Jianping, Xiaofei Zhi, Linjun Wang, Weizhi Wang, Zheng Li, Jie Tang, Jiwei Wang, Qun Zhang, and Zekuan Xu. 2015. "Linc00152 Promotes Proliferation in Gastric Cancer through the EGFR-Dependent Pathway." *Journal of Experimental & Clinical Cancer Research : CR* 34 (1). BioMed Central: 135. doi:10.1186/s13046-015-0250-6.
- Zhou, Yunli, Xun Zhang, and Anne Klibanski. 2012. "MEG3 Noncoding RNA: A Tumor Suppressor." *Journal of Molecular Endocrinology* 48 (3): R45-53. doi:10.1530/JME-12-0008.
- Zhuang, Joanna, Allison Jones, Shih-Han Lee, Esther Ng, Heidi Fiegl, Michal Zikan, David Cibula, et al. 2012. "The Dynamics and Prognostic Potential of DNA Methylation Changes at Stem Cell Gene Loci in Women's Cancer." Edited by Devin Absher. *PLoS Genetics* 8 (2). Public Library of Science: e1002517. doi:10.1371/journal.pgen.1002517.

ANNEXE I

MicroRNAs regulate KDM5 histone demethylases in breast cancer cells,

Molecular Biosystems, (2015)

Hélène Denis, **Olivier Van Grembergen**, Benjamin Delatte, Sarah Dedeurwaerder, Pascale Putmans, Emilie Calonne, Françoise Rothé, Christos Sotiriou, François Fuksa and Rachel Deplus

CrossMark
click for updates

MicroRNAs regulate KDM5 histone demethylases in breast cancer cells†

Cite this: *Mol. BioSyst.*, 2016, 12, 404Hélène Denis,^a Olivier Van Grembergen,^a Benjamin Delatte,^a Sarah Dedeurwaerder,^a Pascale Putmans,^a Emilie Calonne,^a Françoise Rothé,^b Christos Sotiriou,^b François Fuks^a and Rachel Deplus*^a

MicroRNAs (miRNAs) are small non-coding RNAs that post-transcriptionally regulate gene expression. Alteration of miRNA levels is common in tumors and contributes to the pathogenesis of human malignancies. In the present study we examined the role played by miR-137 in breast tumorigenesis. We found miR-137 levels to be lower in breast cancer cells than in their non-tumorigenic counterparts and observed reduced proliferation and migration of breast cancer cells overexpressing miR-137. We further identified KDM5B, a histone demethylase known to be involved in breast cancer tumorigenesis, as a target of miR-137. As the involvement of histone demethylases in cancer is still poorly understood and as the role of miRNAs in controlling epigenetic mechanisms in cancer is emerging, we broadened our study to the whole KDM5 histone demethylase family to see if the genes coding for these epigenetic enzymes might be regulated by miRNAs in cancer cells. We discovered that KDM5C is overexpressed in breast cancer cells, providing evidence that miR-138 regulates its expression. We found miR-138 overexpression to affect breast cancer cell proliferation. Altogether, our findings suggest that miRNAs may regulate KDM5 histone demethylase levels in breast cancer and thereby control breast cancer cell proliferation and migration.

Received 30th July 2015,
Accepted 20th November 2015

DOI: 10.1039/c5mb00513b

www.rsc.org/molecularbiosystems

Introduction

Epigenetic mechanisms include DNA methylation, post-translational histone modifications, and chromatin remodelling. They are involved in modulating the structure of chromatin to make it suitable for gene transcription or silencing.¹ An altered epigenomic pattern is a major feature of cancer, and it is now clear that misregulation of histone modifications contributes to human cancer.² Histone modification patterns are dynamically regulated by enzymes that add, remove, or read covalent modifications. For instance, histone lysine methyltransferases (KMTs) add methyl groups onto lysine residues of histones, while histone lysine demethylases (KDMs) remove these methyl groups. The contribution of KMTs to cancer has been extensively investigated.^{3–6} Although less studied, KDM misregulation appears also to be involved in cancer progression.^{3,7} For instance, KDM1A (also known as LSD1) is highly expressed in poorly differentiated neuroblastoma, and pharmacological inhibition of LSD1 reduces neuroblastoma xenograft growth.^{8,9} Among the KDMs, we focus

here on the KDM5 family and on the possible role of its members in cancer. The KDM5 proteins contain a JmjC domain and are the only KDMs capable of removing from lysine 4 of histone H3 the tri- and di-methyl marks for transcriptionally active chromatin.¹⁰ This enzyme family comprises four members: KDM5A (also known as JARID1A or RBP2); KDM5B (also known as JARID1B or PLU-1); KDM5C, located on the X chromosome (also known as JARID1C or SMCX), and KDM5D, located on the Y chromosome (also known as JARID1D or SMCY).¹⁰ Studies on human malignancies have pointed to a role of two members of the KDM5 family, KDM5A and KDM5B, in cancer development. KDM5A is reported to be translocated in leukemia^{11,12} and overexpressed in gastric cancer,¹³ and its loss has been shown to inhibit tumorigenesis in a mouse cancer model.¹⁴ KDM5B is overexpressed in various types of malignancies,^{15–18} and its down-regulation suppresses breast cancer growth in a syngenic mouse cancer model.¹⁹ The involvement of KDM5C and KDM5D in carcinogenesis is still less clear.

The fairly recent discovery of the crucial role played by miRNAs has broadened our view of the gene expression – epigenetic regulation network. These approximately 22-nucleotides-long noncoding RNAs negatively regulate the expression of genes, usually through direct imperfect alignment with the 3' untranslated region (3'UTR) of target mRNAs. This binding causes mRNA translational repression, deadenylation, or degradation.^{20,21}

^a Laboratory of Cancer Epigenetics, Faculty of Medicine, Université Libre de Bruxelles, Brussels, Belgium. E-mail: rdeplus@ulb.ac.be^b Breast Cancer Translational Research Laboratory, Jules Bordet Institute, Université Libre de Bruxelles, Brussels, Belgium

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c5mb00513b

More than 1000 different miRNAs have been described in the human genome, and efforts to delineate their specific gene targets are still in progress. MicroRNAs influence multiple biological processes. Importantly, a growing number of reports indicate that they affect cell transformation and tumor progression, including the lethal metastatic phase of cancer.^{22–25} Global expression of miRNAs is deregulated in most cancer types, being widely downregulated in tumors as compared to normal tissues.²⁶ MicroRNAs are crucial regulators of tumorigenesis, where they can function as tumor suppressors or oncogenic miRNAs depending on the targets they regulate.^{27,28} For instance, miR-137 expression is decreased in oral squamous cell carcinoma²⁹ and colorectal cancer,^{30,31} and its overexpression in the corresponding cancer cells inhibits cell growth, suggesting tumor-suppressive activity.^{29,30} Both the biological role of miR-137 and its specific downstream mRNA targets in breast carcinogenesis remain unknown.

Breast cancer is a heterogeneous disease, and with around 500 000 deaths yearly, it is the most frequent cause of cancer death in women.^{32,33} Advances in prevention, diagnosis, and treatment are responsible for an increased survival rate,³⁴ but the molecular pathogenesis of breast cancer is not yet fully understood. In the present paper we show that miR-137 levels are lower in breast cancer cells than in their non-tumorigenic counterparts and that miR-137 overexpression in breast cancer cells causes inhibition of growth and migration. We identify the histone demethylase KDM5B as a target of miR-137 and show that KDM5C is also overexpressed in breast cancer cells. We provide evidence that miR-138 negatively regulates the expression of KDM5C and that miR-138 overexpression inhibits breast cancer cell proliferation. Altogether, our results suggest that miR-137 and miR-138 may act as tumor suppressor miRNAs in breast cancer.

Results and discussion

miR-137 is silenced in human breast cancer cell lines

Global miRNA expression is deregulated in most cancer types, being widely downregulated in tumors as compared to normal tissues.²⁶ In particular, miR-137 expression is decreased in oral squamous cell carcinoma²⁹ and colorectal cancer,^{30,31} where miR-137 seems to affect tumor cell growth. To see if miR-137 expression is also reduced in breast cancer, we used TaqMan real-time PCR to measure miR-137 expression in breast cancer cell lines (MCF7, MDA-MB-231, MDA-MB-361, BT20, SKBR3, and ZR-75-1) and in non-tumorigenic human mammary epithelial cell line MCF10A. In all breast cancer cell lines tested, miR-137 expression was found to be lower than in non-tumorigenic cells (Fig. 1A). RNU6B was used control to normalize the expression of target microRNAs, with the copy number being shown at the bottom part of the Fig. 1A. Comparison with a second normal cell line, MCF12A, was performed and the expression of miR-137 was higher in MCF12A as compared to breast cancer cell lines (Fig. S1, ESI[†]). We also examined publicly accessible data from ENCODE, which showed the higher expression of miR-137 in normal HMEpC as compared to MCF7 (Fig. S2, ESI[†]).

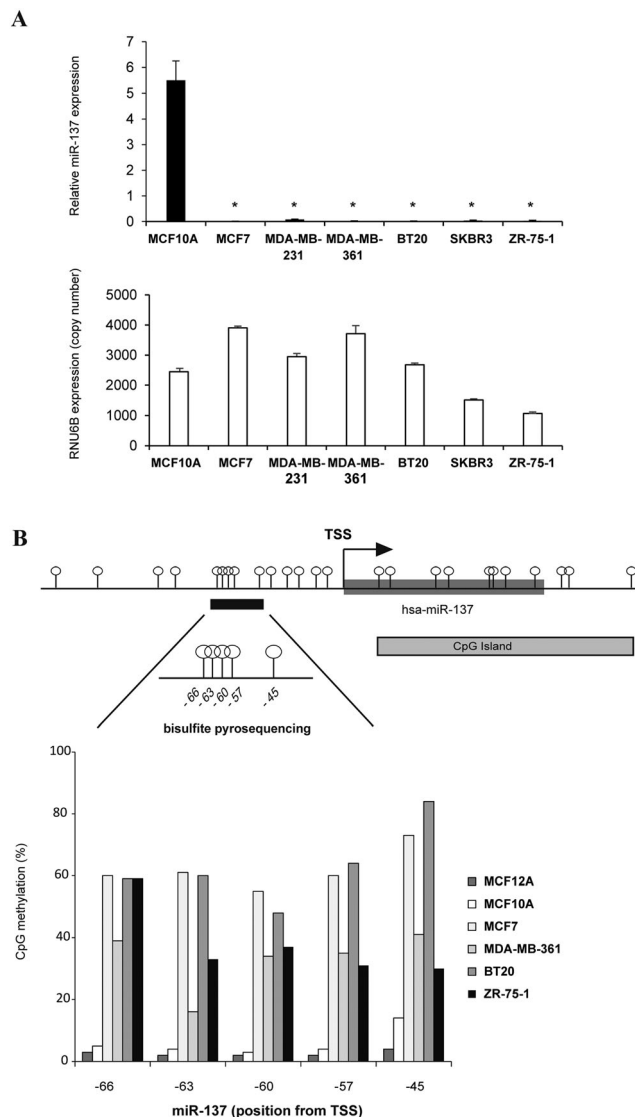


Fig. 1 miR-137 is downregulated in breast cancer. (A) miR-137 expression was analysed by TaqMan RT-qPCR in non-tumorigenic human mammary epithelial cell line (MCF10A) and in five breast cancer cell lines (MCF7, MDA-MB-361, BT20, SKBR3, and ZR-75-1). Results are normalized with respect to the endogenous control RNU6B presented in the bottom panel (copy number) for direct comparison. (* = $p < 0.05$) ($n = 3$). (B) Upper part: map of the miR-137 promoter, with the position of the CpG island and the region used for methylation analysis indicated. The UCSC genome browser was used to identify the upstream CpG island. Lollipops represent all the CG dinucleotides. Lower part: methylation levels of the miR-137 promoter in the MCF12A, MCF10A, MCF7, MDA-MB-361, BT20, and ZR-75-1 breast cell lines. A representative experiment is shown.

Epigenetic mechanisms such as promoter methylation or histone acetylation can modulate microRNA expression, and aberrant regulation at this level is found in cancer. To investigate whether miR-137 downregulation in breast cancer might result from promoter methylation, we analysed the genomic sequence of the miR-137 gene with the UCSC Genome Browser (<http://genome.ucsc.edu>). We found that the miR-137 gene is associated with a CpG island (Fig. 1B, upper part). As cancer cells are characterised by hypermethylation of specific CpG island

promoters,³⁵ we hypothesised that miR-137 gene might be regulated by DNA methylation in cancer. To examine experimentally the methylation status of the miR-137 promoter, we applied bisulfite pyrosequencing to five CpG sites located respectively 66 bp, 63 bp, 60 bp, 57 bp, and 45 bp upstream of the transcription start site (TSS) of the miR-137 gene. Fig. 1B depicts the level of methylation of each CpG site in non-tumorigenic MCF10A and MCF12A cells and in the MCF7, MDA-MB-361, BT20, and ZR-75-1 breast cancer cell lines. For a given cell type, all five CpG sites analyzed tended to show a similar degree of methylation (Fig. 1B). The miR-137 promoter appeared weakly methylated in non-tumorigenic breast cells (MCF10A and MCF12A) but more highly methylated in all breast cancer cell lines (Fig. 1B). Previous reports showed that miR-137 undergoes promoter methylation in oral squamous cell carcinoma,²⁹ gastric cancer cell lines,^{36,37} colorectal cancer^{30,31} and lung cancer³⁸ suggesting that miR-137 promoter methylation could be a common and general feature of tumor cells. Previous studies have shown treatment with the demethylating agent 5-aza-2-deoxycytidine (5-aza) to restore miR-137 expression in oral squamous carcinoma cells and some colorectal cancer cells.^{29,30} Here, we show that this agent induced significant miR-137 reexpression in breast cancer cells (Fig. S3, ESI†). It thus appears that miR-137 is silenced in breast tumor cells, but that promoter methylation alone cannot explain the lower levels of miR-137 expression observed. Given the known interplay between DNA methylation and histone deacetylation, we tested if a synergism between these two repressive machineries might explain silencing of miR-137 in breast cancer cells. For this, MCF7 cells were treated with 5-aza and/or TSA. As presented in Fig. S3 (ESI†), histone deacetylase inhibitor or DNA methyltransferase inhibitor led to substantial reactivation of miR-137 expression in MCF7 cells. However, combination of 5-aza and TSA didn't show elevated miR-137 expression, suggesting independent regulation by histone deacetylation and DNA methylation of miR-137 expression.

miR-137 overexpression inhibits breast cancer cell proliferation and migration

Having shown that miR-137 is silenced in breast cancer cells, we next performed proliferation assays to determine the functional effects of miR-137 overexpression in these cells. Cell proliferation was evaluated in the xCELLigence system. This instrument monitors cellular events in real time without exogenous labeling, through impedance-based technology.³⁹ It provides a kinetic cell-response profile throughout the experiment, detailing the onset and rate of proliferation. After transfection of MCF7 cells with miR-137 precursors, cell growth was measured every 30 min for 3 days (see Materials and methods for details). TaqMan real-time PCR analysis (Fig. S4, ESI†) confirmed the overexpression of mature miR-137 in the transfected cells. As compared to control cells (ctrl pre-miR), pre-miR-137-transfected cells showed strongly reduced proliferation (Fig. 2A). Fig. 2B, where the rate of cell proliferation is estimated as the mean slope of the cell growth curves for three independent experiments, shows that this rate was reduced by more than 70% in the pre-miR-137-transfected breast cancer cells.

We then evaluated the effect of miR-137 on the migration behavior of breast cancer cells. Cell migration kinetics was recorded in real time with the xCELLigence system (see scheme of the system used in Fig. S5, ESI†). As is common,⁴⁰ we performed the migration assays with MDA-MB-231 cells, using fetal bovine serum (FBS) as a chemoattractant. The cells were transfected with miRNA precursors and overexpression of the mature miR-137 was confirmed (Fig. S6, ESI†). Cell migration was then measured every 30 min for 24 h (Fig. 2C). Without any chemoattractant in the medium of the lower chamber, the cells did not migrate (ctrl pre-miR and pre-miR-137, grey and yellow curves respectively in Fig. 2C). When FBS was used as a chemotaxis inducer, the cells migrated to the lower chamber. We found the control cells (ctrl pre-miR, black curve in Fig. 2C) to migrate faster than the pre-miR-137-transfected cells (pre-miR-137, red curve in Fig. 2C). We focused on the first 4 hours of the migration kinetics (*cf.* Material and Methods) to calculate the rate of cell migration, represented in Fig. 2D as the mean slope of the cell migration curves from three independent experiments. Cell migration was found to decrease by 40% in response to miR-137 overexpression (Fig. 2D). Our findings thus suggest that miR-137 reduces both the proliferation and migration of breast cancer cells.

KDM5B is a direct target of miR-137

Although miRNAs appear to regulate the expression of many different genes, their function in tumorigenesis might be due to regulation of a few specific ones.⁴¹ To determine the potential target of miR-137 in breast cancer, we looked at the different gene targets described in the literature.^{42–45} Of particular interest was a study on mouse ES cells, where miR-137 was shown to repress the histone demethylase KDM5B (also known as PLU-1 or JARID1B) and to be required for proper ES cell differentiation.⁴⁴ Interestingly, KDM5B is frequently overexpressed in breast cancer, and KDM5B knockdown impairs the proliferative capacity of MCF7 breast cancer cells.^{15,17,19} On the basis of these studies, we hypothesized that miR-137 might affect the proliferation of breast cancer cells by targeting KDM5B. To test this hypothesis, we first performed western blotting using an anti-KDM5B in breast cancer cells (BT20, MCF7, MDA-MB-231, MDA-MB-361, SKBR3, and ZR-75-1) and non-tumorigenic cells (MCF10A). As shown in Fig. 3A, breast cancer cell lines harbored higher levels of KDM5B compared to non-tumorigenic mammary cells, suggesting a potential link between miR-137 expression and decreased KDM5B protein level. To confirm that miR-137 regulates KDM5B in breast cancer cells, we transiently transfected MCF7 cells with pre-miR-137 and evaluated the KDM5B mRNA and protein levels. We used MCF7 cells for this experiment because they have a high level of endogenous KDM5B expression (Fig. 3A)^{15,17} and undetected levels of miR-137 (Fig. 1A). The KDM5B transcript level (assessed by RT-qPCR) was found to significantly reduced (by ~30%) in pre-miR-137-transfected MCF7 cells (Fig. 3B). Similar results were also obtained in two other cancer cell lines, MDA-MB-231 and MDA-MB-361 (Fig. S7, ESI†), suggesting a common feature in breast cancer cells. The fact that this transcript-level effect is

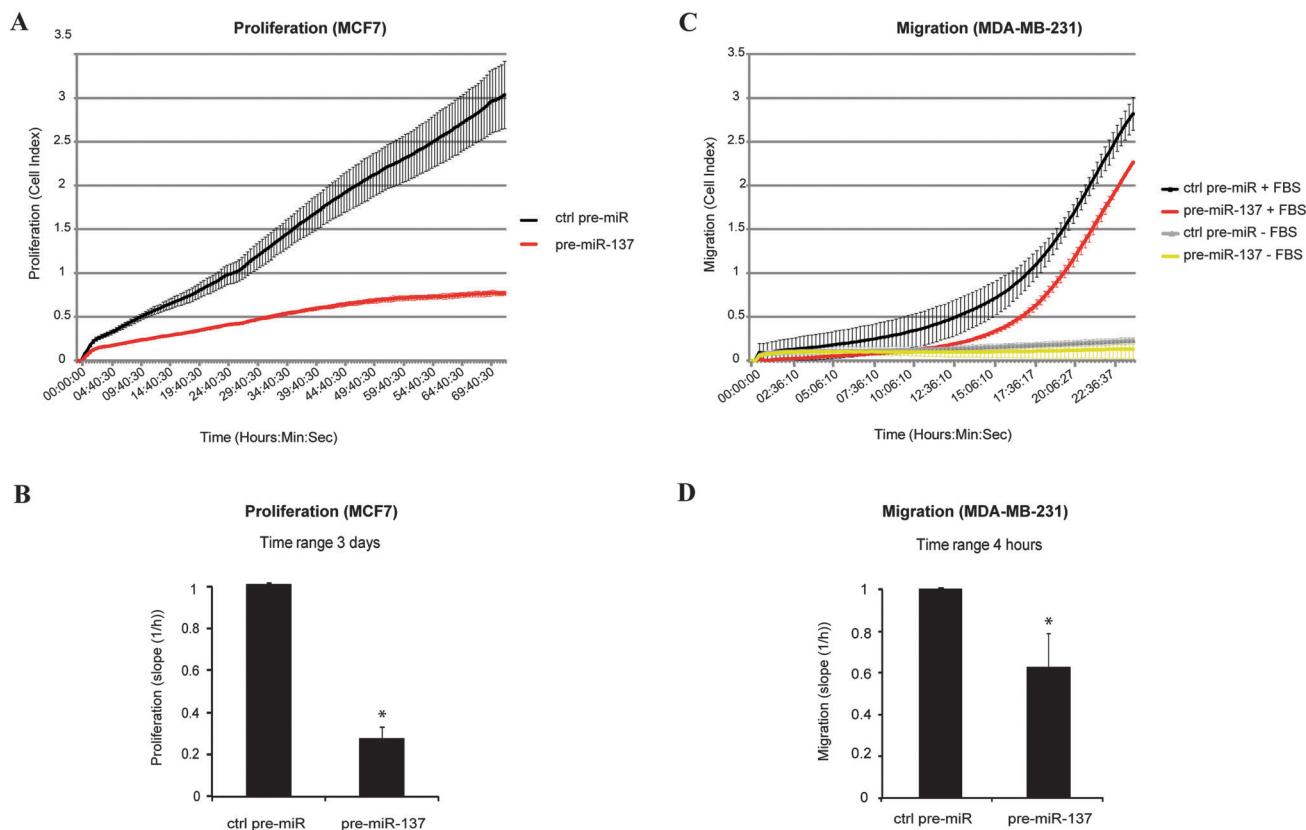


Fig. 2 miR-137 regulates breast cancer cell proliferation and migration. (A) Cell index (mean \pm standard deviation) as a measure of MCF7 cell proliferation. Measurements were automatically collected by the RTCA DP analyser every 30 min for up to 3 days ($n = 4$). Black: MCF-7 cells transfected with control pre-miR (ctrl pre-miR); Red: MCF7 cells transfected with precursors of miR-137 (pre-miR-137). (B) Results of the cell proliferation assays represented as slopes (changes in cell index per hour) ($n = 3$) (* = $p < 0.05$). (C) Migration kinetics of MDA-MB-231 cells assessed by continuous monitoring of live cell migration for approximately 24 hours. Black: MDA-MB-231 cells transfected with control pre-miR (ctrl pre-miR + FBS), 10% FBS in lower chamber (LC), $n = 4$; Red: MDA-MB-231 cells transfected with miR-137 precursors (pre-miR-137 + FBS), 10% FBS in LC, $n = 4$. Negative controls: grey: MDA-MB-231 cells transfected with control pre-miR (ctrl pre-miR - FBS), serum-free medium in LC, $n = 2$; Yellow: MDA-MB-231 cells transfected with miR-137 precursors (pre-miR-137 - FBS), SFM in LC, $n = 2$. (D) Results of the cell migration assays represented as slopes (changes in cell index per hour) ($n = 3$) (* = $p < 0.05$).

moderate is not surprising: miR-137 might predominantly affect the translational efficiency, as shown in the literature for other miRNAs.⁴⁶ Immunoblot analyses with an antibody targeting KDM5B revealed significantly reduced levels of this protein in miR-137-expressing cancer cells (Fig. 3C). To see if miR-137 regulates KDM5B directly, we cloned into a luciferase reporter vector 25 bp of the KDM5B 3'UTR, corresponding to the miR-137 target sequence in this UTR (Fig. 3D, upper part). Overexpression of pre-miR-137, but not of control precursors, was found to decrease slightly but reproducibly the activity of the luciferase reporter (Fig. 3D). Taken together, our results show that miR-137 regulates KDM5B in breast cancer cells by directly targeting the KDM5B 3'UTR.

miR-138 regulates the histone demethylase KDM5C in breast cancer

While there are previous reports of KDM5B overexpression in various types of malignancies,^{15–17} it is not known if other members of the KDM5 histone demethylase family are involved in carcinogenesis generally or in breast tumorigenesis in

particular. We therefore found it interesting to assess the role of KDM5C in the context of breast cancer. Immunoblot analysis revealed that KDM5C, like KDM5B (ref. 15 and 17 and Fig. 3A), is overexpressed in BT20, MCF7, MDA-MB-231, MDA-MB-361, SKBR3, and ZR-75-1 breast cancer cells as compared to non-tumorigenic MCF10A (Fig. 4A). We then examined whether KDM5C might, like KDM5B (Fig. 3), be regulated by miRNAs in breast tumor cells. We used online prediction tools⁴⁷ to identify putative KDM5C-targeting miRNA candidates (data not shown). Individual TaqMan assays showed reduced levels of one candidate miRNA, miR-138, in all the cancer cell lines tested except MDA-MB-231, where it was highly overexpressed (Fig. 4B and Fig. S8, ESI[†]). We also examined publicly available data from ENCODE, which showed higher expression of miR-138 in HMEpC versus MCF7 (Fig. S9, ESI[†]).

Interestingly, MDA-MB-231 was the cancer cell line where the level of KDM5C protein was lowest (Fig. 4A). Nevertheless, despite a stronger expression of miR-138, KDM5C protein is higher in MDA-MB-231 than in non tumorigenic cells, suggesting an additional mechanism of regulation. This supplemental

(MCF7)

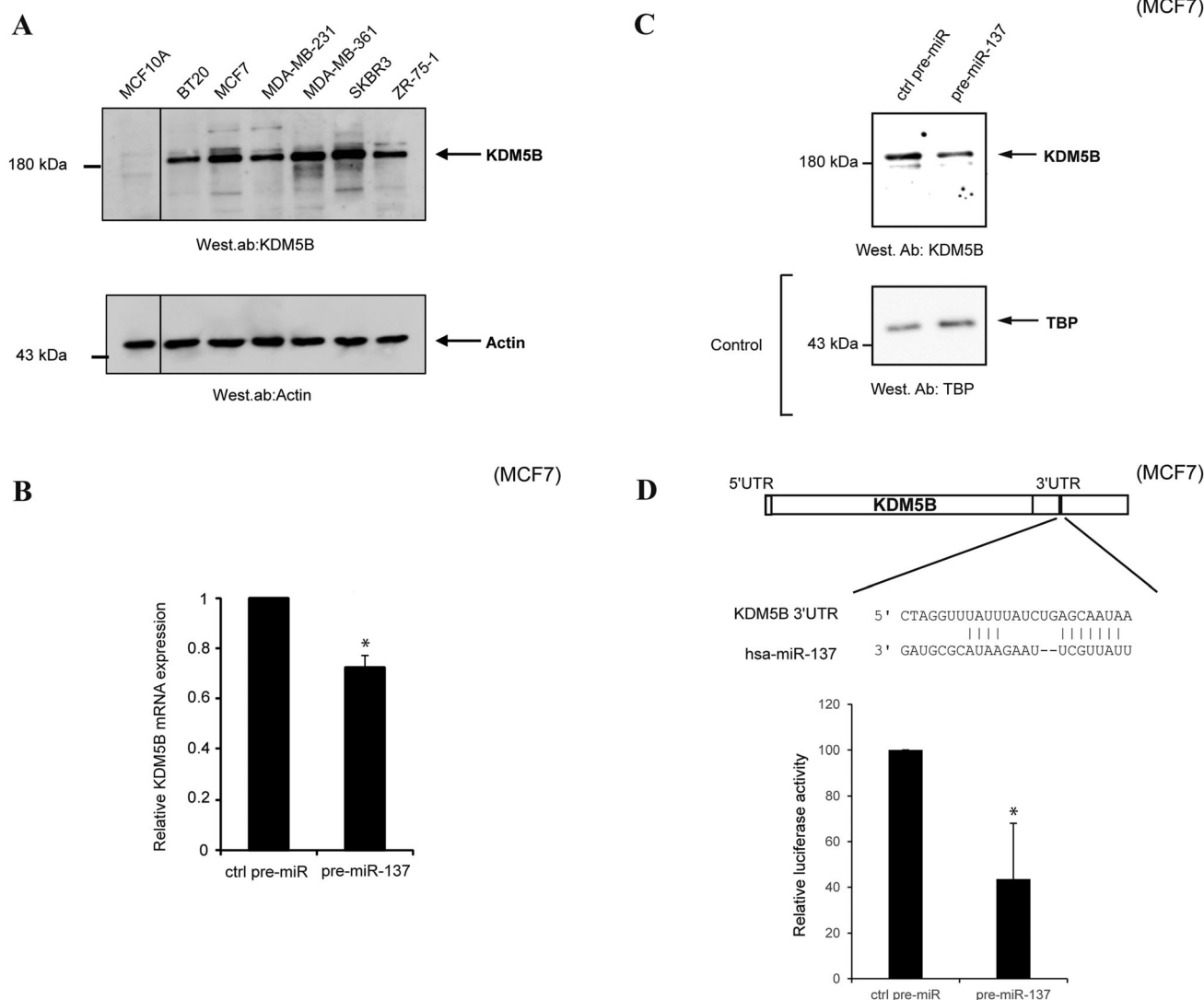


Fig. 3 KDM5B is a direct target of miR-137 in breast cancer cells. (A) Immunoblot analysis of KDM5B protein levels in MCF10A, BT20, MCF7, MDA-MB-231, MDA-MB-361, SKBR3, and ZR-75-1 cell lines. Vertical line indicates juxtaposition of lanes non-adjacent within the same blot, exposed for the same time. (B) Following transfection of MCF7 cells with miR-137 precursors (pre-miR-137) or with a scrambled pre-miR used as a negative control (ctrl pre-miR), KDM5B mRNA levels were analysed by real-time qPCR and standardized with respect to the endogenous control GAPDH. Relative expression levels were determined for control-pre-miR- and pre-miR-137-transfected cells. ($n = 3$) ($* = p < 0.05$) (C) KDM5B protein levels in MCF7 cells were assessed by western blotting following transfection of MCF7 cells with control pre-miR or pre-miR-137. TBP was used as a loading control. (D) A 25-bp sequence from the KDM5B 3'UTR was cloned into the 3'UTR of a luciferase reporter vector (upper part). MCF7 cells were transfected with ctrl pre-miR or pre-miR-137 together with the pMIR-3'UTR construct and pMIR-REPORT Beta-Gal. Luciferase activity was assayed 48 h post-transfection and normalized with respect to beta-galactosidase activity. Relative luciferase activity was determined on control-pre-miR- and pre-miR-137-transfected samples ($n = 3$) ($* = p < 0.05$).

mechanism need to be identified and could regulated KDM5C transcription through transcription factor or epigenetic regulation, or could also influence translation or protein stability.

We then transiently transfected MCF7 cells, characterized by high levels of endogenous KDM5C and an undetectable level of miR-138 (Fig. 4A and B), with pre-miR-138. The cells transfected with pre-miR-138 showed respectively $\sim 20\%$ lower levels of KDM5C transcripts than cells transfected with a control pre-miR (Fig. 4C), and they also showed a substantially decreased KDM5C protein level on western blots (Fig. 4D). In order to examine whether miR-138 could regulate KDM5C directly, the miR-138 target sequence of the KDM5C 3'UTR was cloned into a

luciferase reporter vector (Fig. 4E, upper part). Significant decrease of activity of the luciferase reporter was detected after overexpression of pre-miR-138. This was not the case for the control precursors (Fig. 3D). Thus, miR-138 regulates KDM5C in breast cancer cells by directly targeting the KDM5B 3'UTR. Taken together, our data indicate that miR-138 causes downregulation of endogenous KDM5C in human breast tumor cells.

High expression of miR-138 decreases breast cancer cell proliferation

We then examined whether enhanced expression of miR-138 might affect breast cancer cell proliferation. We transfected

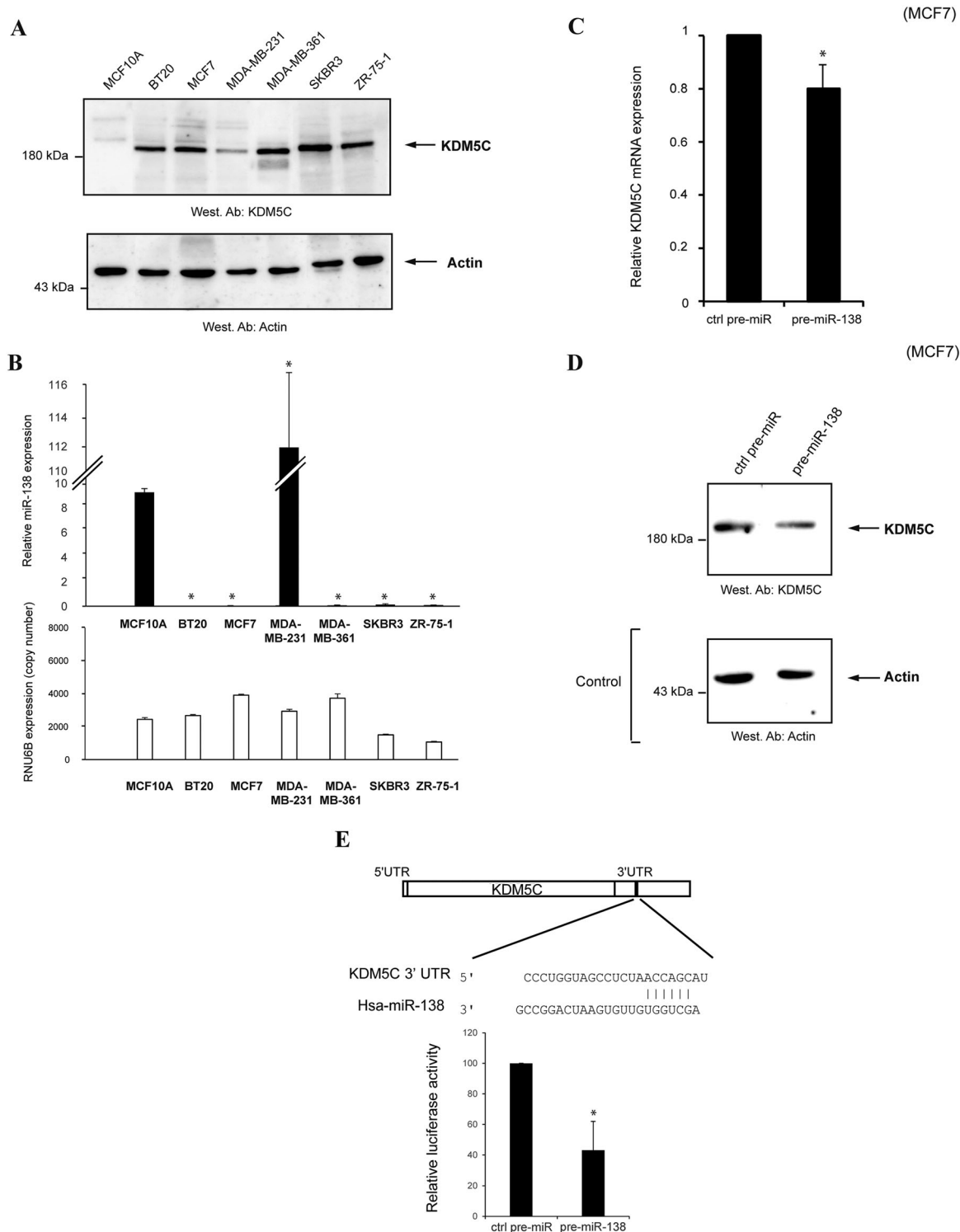


Fig. 4 miR-138 regulates the histone demethylase KDM5C in breast cancer cells. (A) Immunoblot analysis of KDM5C protein levels in the MCF10A, BT20, MCF7, MDA-MB-231, MDA-MB-361, SKBR3, and ZR-75-1 cell lines. (B) miR-138 levels were analysed by TaqMan RT-qPCR in the non-tumorigenic human mammary epithelial cell line (MCF10A) and in several breast cancer cell lines (MCF7, MDA-MB-361, BT20, SKBR3, and ZR-75-1). Results are normalized with respect to the endogenous control RNU6B shown in the bottom part (copy number) (* = $p < 0.05$) ($n = 3$). The bottom panel shown the level of RNU6B expression (copy number) for direct comparison. (C) Following transfection of MCF7 cells with miRNA precursors (pre-miR-138) or nonspecific control (ctrl pre-miR), KDM5C mRNA levels were analysed by qPCR and normalized with respect to the endogenous control GAPDH. ($n = 3$) (* = $p < 0.05$) (D) MCF7 cells were transfected with pre-miR-138, or the negative control and then western blot analysis was performed to detect KDM5C. (E) A 25-bp sequence from the KDM5C 3'UTR was cloned into the 3'UTR of a luciferase reporter vector (upper part). MCF7 cells were transfected with ctrl pre-miR or pre-miR-138 together with the pMIR-3'UTR construct and pMIR-REPORT Beta-Gal. Luciferase activity was assayed 48 h post-transfection and normalized with respect to beta-galactosidase activity. Relative luciferase activity was determined on control-pre-miR- and pre-miR-138-transfected samples ($n = 3$) (* = $p < 0.05$).

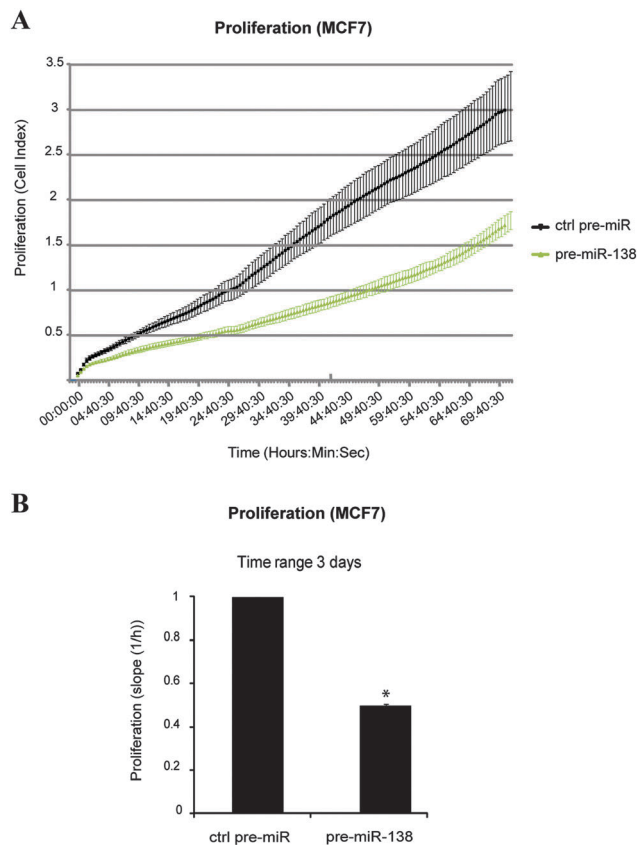


Fig. 5 miR-138 regulates breast cancer cell proliferation. (A) Cell index (mean \pm standard deviation) as a measure of the growth of MCF7 cells. Measurements were automatically collected by the RTCA DP analyzer every 30 min for up to 3 days ($n = 4$). Black: MCF-7 cells transfected with control pre-miR; Green: MCF7 cells transfected with precursors of miR-138 (pre-miR-138). (B) Results of the cell proliferation assays are represented as slopes (changes in cell index/hour) ($n = 3$) (* = $p < 0.05$).

MCF7 breast cancer cells with pre-miR-138 and monitored cell growth in real time in the xCELLigence system (Fig. 5A). Overexpression of mature miR-138 was confirmed by real-time PCR analysis (Fig. S10, ESI†). In this experiment, cells transfected with pre-miR-138 showed substantially slower proliferation than control cells (ctrl pre-miR) (Fig. 5A). Fig. 5B shows that the proliferation rate, estimated here as the mean slope of the cell growth curves from three independent experiments, was halved when miR-138 was overexpressed. Taken together, our findings show that miR-138 can regulate KDM5C expression in breast cancer cells, and that this miR can affect their proliferation.

Conclusion

In recent years, several reports have highlighted deregulation of miRNA expression in human cancers, including breast cancer.^{26,48} For example, downregulation of miR-137 has been observed in oral squamous cell carcinoma and in colorectal cancer.^{30,31} We find that miR-137 is also downregulated in breast cancer cells, and this suggests that miR-137 might be a central event in tumorigenesis. Our bisulfite pyrosequencing data show that CpG sites located in

the miR-137 promoter are weakly methylated in non-tumorigenic breast cells (MCF10A), but more strongly methylated in breast cancer cell lines. Yet treatment with a DNA methyltransferase inhibitor alone or in combination with histone deacetylase inhibitors did not fully restore miR-137 expression. This indicates that other mechanisms must regulate the miR-137 level. These mechanisms deserve further investigation.

We also provide evidence that miR-137 might play a role in breast tumorigenesis: overexpression of miR-137 in breast cancer cells reduces significantly both the proliferation and the migration of these cells. Identifying miR-137-specific targets is crucial to understanding its possible role in tumorigenesis. Interestingly, we have identified KDM5B, a key histone demethylase involved in breast cancer tumorigenesis,^{17,19} as a novel miR-137 target. Altogether, both our data and other published results concerning miR-137 in oral squamous cell carcinoma²⁹ and colorectal cancer^{30,31} suggest that this microRNA may function as a tumor suppressor gene and thus be a good candidate for use in anticancer therapy.

The contribution of histone lysine methyltransferases to cancer has been extensively explored.^{3–5} Although less studied, overexpression of histone demethylases has also been implicated in cancer progression.³ One piece of evidence suggesting that KDM5 H3K4me3 demethylases might participate in tumorigenesis is the observation that KDM5B is overexpressed in several cancers, including breast cancer.^{15–17,19} Here we show for the first time that the histone demethylase KDM5C is also overexpressed in breast tumor cells. Additional studies will also be necessary to determine what functional role KDM5C might play in breast carcinogenesis *in vivo*. The relevance of KDM5C to tumor formation and/or metastasis must be investigated in detail.

KDM5 family members might present functional overlap and KDM5 protein might exert their function together. Nevertheless, different tissues tropism⁴⁹ or divergent phenotypes of knockout mice^{50,51} point also to independent mode of actions. Common and divergent functional properties shared or not by KDM5B and KDM5C and involved in tumorigenesis should be investigated in future.

Over the past few years, miRNAs have been emerging as agents controlling epigenetics in cancer. The importance of this control is underscored by the finding that a number of crucial epigenetic regulators, including key enzymes regulating DNA methylation (DNMT3A and DNMT3B),^{52,53} histone methylation (EZH2),⁵⁴ and histone deacetylation (HDAC1),⁵⁵ are targeted by miRNAs in human cancers. Until our present study, nothing was known regarding the regulation of histone demethylases by miRNAs in carcinogenesis. We demonstrate for the first time that miRNAs can also control KDMs in cancer cells: miR-137 controls the histone demethylase KDM5B, and overexpression of miR-138 affects KDM5C expression in breast cancer cells. We also show that miR-138 overexpression, like miR-137, affects breast cancer cell proliferation. We thus add two new examples illustrating how miRNAs can regulate epigenetic enzymes, here KDMs, in cancer cells.

Altogether, our findings suggest that reduced miRNA levels may contribute to breast tumorigenesis by allowing KDM5

histone demethylases to accumulate to an abnormally high level and thus favoring unchecked cell proliferation and migration. As cancer cells depend on dysregulation of specific miRNAs for proliferation and survival, miRNAs might be prove useful as therapeutic tools favoring tumor regression.

Materials and methods

Cell lines

MCF10A cells were grown in a 1:1 mixture of DMEM and F12 medium, supplemented with 10% fetal calf serum. MCF12A cells were grown in a 1:1 mixture of Dulbecco's modified Eagle's medium and Ham's F12 medium supplemented with 20 ng ml⁻¹ Human epidermal growth factor, 100 ng ml⁻¹ cholera toxin, 0.01 mg ml⁻¹ bovine insulin, 500 ng ml⁻¹ hydrocortisone and 5% horse serum. MCF7, SKBR3, and MDA-MB-231 cells were grown in DMEM supplemented with 10% fetal calf serum. ZR-75-1 and MDA-MB-361 cells were grown in RPMI supplemented with 10% fetal calf serum. The BT-20 cell line was grown in Eagle's Minimum Essential Medium (MEM) supplemented with 10% fetal calf serum. MCF7 cells were treated with 5-aza (2.5 μM, SIGMA) for 72 h and/or TSA (200 mM, SIGMA) during 48 h.

Transfection with pre-miRs

The individual precursors of miRNAs (pre-miRs) used in this study were purchased from Applied Biosystems. Cells were transfected with pre-miRs at 100 nM working concentration, using siPORT™ (Ambion) according to the manufacturer's protocol. Pre-miR negative control #1 (Ambion) was used as a scrambled control. Treatment proceeded for 72 h before RNA extraction and/or western blot analysis. It proceeded for 24 h before cell-based assays.

RNA purification and real-time quantitative PCR

Extraction of total RNA was carried out with the TriPure reagent (Roche) according to the manufacturer's instructions. DNase treatment was performed with a DNA-free DNase kit (Ambion) according to the manufacturer's protocol.

Quantitative PCRs were performed with SYBR Green dye (Eurogentec) on a LightCycler 480 (Roche). Briefly, cDNA was reverse transcribed from 1 μg RNA with random hexamers (Amersham/Pharmacia Biotech) and Superscript II reverse transcriptase (Life Technologies, Inc.). Real-time PCRs were performed to measure gene expression levels under different conditions. The primer sequences for the KDM5B and KDM5C amplifications were as follows: forward KDM5B, 5'-ACGTCGA ATGGGTTGTCC-3'; reverse KDM5B, 5'-TCTTGGGCTTTTCCTT CTCA-3'; forward KDM5C, 5'-CAGATGAACAACCTGCCTTG-3'; reverse KDM5C, 5'-CCAGAACACCCTTGACATCC-3'. GAPDH was amplified as an internal control.

For microRNA quantitative PCRs, total RNA (10 ng) was reverse transcribed with the TaqMan MicroRNA Reverse Transcription Kit (Applied Biosystem) and miRNA-specific primers. For miRNA quantitation, the 7500 Fast Real-Time System (Applied Biosystem) was used in conjunction with gene-specific

TaqMan assay kits for miR-137, miR-138, miR-329, and RNU6B. RNU6B was used as an endogenous control to normalize the expression of target microRNAs.

Cell extracts and western blot analyses

Whole-cell extracts were prepared with IPH lysis buffer (50 mM Tris-HCl pH 8, 150 mM NaCl, 5 mM EDTA, 0.5% NP40). Nuclear extracts were prepared as follow: cells were first lysed in buffer A (10 mM HEPES pH 7.9, 10 mM KCl, 0.1 mM EDTA, 1 mM DTT, protease inhibitors) and then centrifuged at 10 000g for 5 min. Buffer B (20 mM HEPES pH 7.9, 0.4 M NaCl, 1 mM EDTA, 10% glycerol, 1 mM DTT, protease inhibitors) was then added to the nuclear pellet fraction, then the mixture was shaken vigorously for 2 h and centrifuged at 12 000g for 10 min. The supernatant (nuclear fraction) was stored at -80 °C until used. All procedures were performed at 4 °C.

Standard procedures were used for western blotting. The primary antibodies used in these experiments were against the following: KDM5B (Abnova, pAb-053-050), KDM5C (Abcam, ab34718), TBP (Abcam, ab818) and actin (Abcam, ab8226). Anti-TBP and anti-actin were used as loading control antibodies.

miR reporter luciferase assay

Complementary oligonucleotides containing the miR-137 target site of the KDM5B UTR (a 25-bp sequence from the KDM5B 3'UTR) were synthesized by Sigma. Oligonucleotides were annealed and ligated into the pMIR-REPORT luciferase reporter vector (Applied Biosystems). pMIR-REPORT Beta-Gal (Applied Biosystems) was used as a control to monitor transfection efficiency. Constructs were used to co-transfect MCF7 cells in a 24-well plate using siPORT. Briefly, cells were transfected with 100 nM miRNA precursor, 0.6 μg pMIR-3'UTR clone, and 0.4 μg pMIR-REPORT Beta-Gal for 48 h. The Dual-Light System Chemiluminescent Reporter Gene Assay was used according to the manufacturer's protocol to quantify luciferase and beta-galactosidase activity (Applied Biosystems).

The following oligonucleotides were used for cloning:

KDM5B_miR-137_F: AATGCACTAGTCTAGGTTTATTATCT
GAGCAATAAGCTCAGCAAGCTTAATGC

KDM5B_miR-137_R: GCATTAAGCTTGCTGAGCTTATTGCTC
AGATAAATAAACCTAGACTAGTGCATT

Bisulfite pyrosequencing

750 ng genomic DNA was subjected to bisulfite conversion performed with the Epitect® Bisulfite Kit (Qiagen) according to the manufacturer's protocol. The converted DNA was generally eluted with 26 μl elution buffer, and 8 μl of the eluted DNA was used as template in subsequent PCRs. PCR and sequencing primers for bisulfite pyrosequencing were deduced with the PyroMark® Assay Design 2.0 software (Qiagen). Primers were obtained from Sigma. The following PCR and sequencing primers were used:

miR-137 Fwd: AGGGAAATAGAGTTATGGATTATGGT;

miR-137 Rev: [Btm]CACCCAAAAAATCAAAAAACCAACTAC
and miR-137 sequencing primer: GGGTTTAGAGTAGTAAG.

PCRs were performed with the HotStarTaq PCR kit (Qiagen) according to the manufacturer's recommendations. Sample preparation and pyrosequencing reactions were performed with the PyroMark™ Q24 system (Qiagen).

Cell proliferation and migration assays

To evaluate breast cancer cell proliferation, MCF7 cells transfected with miRNA precursors were seeded 24 h post-transfection into the xCELLigence E-plate 16 (Roche) (8000 cells per well) according to the manufacturer's instructions. In this system, the electrical impedance is used to derive a cell index which, when continually monitored, gives a real-time representation of the growth characteristics of the cells. Measurements were automatically collected by the RTCA DP analyser every 30 min for up to 3 days. Four replicate measurements per condition were obtained. The data were analyzed with the provided RTCA software.

To examine breast cancer cell migration, MDA-MB-231 cells transfected with miRNA precursors were seeded 24 h post-transfection into the xCELLigence CIM-plate 16 (Roche). Briefly, a 165 µl volume of fresh medium containing 10% FBS (chemoattractant) or with serum-free medium (control) was added to the lower chambers of the CIM-plate 16. The upper chambers were filled with serum-free medium (30 µl per well) and the plate was incubated at 37 °C in 5% CO₂ for 1 h. Cells (40 000 cells per well) were then added to each well of the upper chamber. After 30 min, the CIM plate was assembled onto the RTCA DP analyser and cell migration was assessed at 30 min intervals for 4 h at 37 °C in 5% CO₂. Upon migration, cells adhere to the surface of the filter electrode and increase the impedance (Fig. S5, ESI[†]). Four replicate measurements per condition were obtained. The data were analyzed with the provided RTCA software. When analyzing serum-induced migration data, it is important to choose a migration time well below the doubling time of the cells in order to capture signals derived from cell migration only, rather than both cell migration and proliferation (recommended by the xCELLigence supplier). We therefore focused on the first 4 hours of the migration kinetics to calculate the rate of MDA-MB-231 migration.

Statistics

Student tests were used to perform statistical analysis. * is indicated if $p < 0.05$.

Funding

H. D. was supported by the Belgian Fonds National de la Recherche Scientifique and the Brussels Region, O. V. G. by the Télévie, S. D. by the Télévie and the Innoviris/Brubrest grant and R. D. by the Interuniversity Attraction Poles [IUAP P6/28] and the Télévie. B. D. is a Research Fellow from the Fonds National de la Recherche Scientifique. F. F. is a Professor of the "Université Libre de Bruxelles". This work was funded by grants from the Fonds National de la Recherche Scientifique and Télévie, the "Plan National Cancer", the IAP P7/03, by the Brussels Region, by the Belgian Foundation against Cancer

and by the "Action de Recherche Concertée". Funding for open access charge: Belgian ARC.

Conflicts of interest

The authors declare that they have no conflicts of interest.

Acknowledgements

We thank Michael Volkmar and 'Matladi N. Ndlovu for their helpful advices.

References

- 1 G. Felsenfeld and M. Groudine, *Nature*, 2003, **421**, 448–453.
- 2 P. Chi, C. D. Allis and G. G. Wang, *Nat. Rev. Cancer*, 2010, **10**, 457–469.
- 3 R. A. Varier and H. T. M. Timmers, *Biochim. Biophys. Acta*, 2011, **1815**, 75–89.
- 4 K. Collett, G. E. Eide, J. Arnes, I. M. Stefansson, J. Eide, A. Braaten, T. Aas, A. P. Otte and L. A. Akslen, *Clin. Cancer Res.*, 2006, **12**, 1168–1174.
- 5 C. G. Kleer, Q. Cao, S. Varambally, R. Shen, I. Ota, S. A. Tomlins, D. Ghosh, R. G. Sewalt, A. P. Otte, D. F. Hayes, M. S. Sabel, D. Livant, S. J. Weiss, M. A. Rubin and A. M. Chinnaiyan, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 11606–11611.
- 6 T. Suzuki, M. Terashima, S. Tange and A. Ishimura, *Cancer Sci.*, 2013, **104**, 795–800.
- 7 S. Lim, E. Metzger, R. Schule, J. Kirfel and R. Buettner, *Int. J. Cancer*, 2010, **127**, 1991–1998.
- 8 J. H. Schulte, S. Lim, A. Schramm, N. Friedrichs, J. Koster, R. Versteeg, I. Ora, K. Pajtler, L. Klein-Hitpass, S. Kuhfittig-Kulle, E. Metzger, R. Schule, A. Eggert, R. Buettner and J. Kirfel, *Cancer Res.*, 2009, **69**, 2065–2071.
- 9 S. Amente, G. Milazzo, M. C. Sorrentino, S. Ambrosio, G. Di Palo, L. Lania, G. Perini and B. Majello, *Oncotargets Ther.*, 2015, **6**, 14572–14583.
- 10 R. J. Klose, E. M. Kallin and Y. Zhang, *Nat. Rev. Genet.*, 2006, **7**, 715–727.
- 11 L. J. van Zutven, E. Onen, S. C. Velthuisen, E. van Drunen, A. R. von Bergh, M. M. van den Heuvel-Eibrink, A. Veronese, C. Mecucci, M. Negrini, G. E. de Greef and H. B. Beverloo, *Genes, Chromosomes Cancer*, 2006, **45**, 437–446.
- 12 G. G. Wang, J. Song, Z. Wang, H. L. Dormann, F. Casadio, H. Li, J. L. Luo, D. J. Patel and C. D. Allis, *Nature*, 2009, **459**, 847–851.
- 13 J. Zeng, Z. Ge, L. Wang, Q. Li, N. Wang, M. Bjorkholm, J. Jia and D. Xu, *Gastroenterology*, 2009, **138**, 981–992.
- 14 W. Lin, J. Cao, J. Liu, M. L. Beshiri, Y. Fujiwara, J. Francis, A. D. Cherniack, C. Geisen, L. P. Blair, M. R. Zou, X. Shen, D. Kawamori, Z. Liu, C. Grisanzio, H. Watanabe, Y. A. Minamishima, Q. Zhang, R. N. Kulkarni, S. Signoretti, S. J. Rodig, R. T. Bronson, S. H. Orkin, D. P. Tuck, E. V. Benevolenskaya, M. Meyerson, W. G. Kaelin Jr. and Q. Yan, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 13379–13386.

- 15 A. Barrett, B. Madsen, J. Copier, P. J. Lu, L. Cooper, A. G. Scibetta, J. Burchell and J. Taylor-Papadimitriou, *Int. J. Cancer*, 2002, **101**, 581–588.
- 16 S. Hayami, M. Yoshimatsu, A. Veerakumarasivam, M. Unoki, Y. Iwai, T. Tsunoda, H. I. Field, J. D. Kelly, D. E. Neal, H. Yamaue, B. A. Ponder, Y. Nakamura and R. Hamamoto, *Mol. Cancer*, 2010, **9**, 59.
- 17 P. J. Lu, K. Sundquist, D. Baekstrom, R. Poulosom, A. Hanby, S. Meier-Ewert, T. Jones, M. Mitchell, P. Pitha-Rowe, P. Freemont and J. Taylor-Papadimitriou, *J. Biol. Chem.*, 1999, **274**, 15633–15645.
- 18 Z. Wang, F. Tang, G. Qi, S. Yuan, G. Zhang, B. Tang and S. He, *Am. J. Cancer Res.*, 2015, **5**, 87–100.
- 19 K. Yamane, K. Tateishi, R. J. Klose, J. Fang, L. A. Fabrizio, H. Erdjument-Bromage, J. Taylor-Papadimitriou, P. Tempst and Y. Zhang, *Mol. Cell*, 2007, **25**, 801–812.
- 20 W. Filipowicz, S. N. Bhattacharyya and N. Sonenberg, *Nat. Rev. Genet.*, 2008, **9**, 102–114.
- 21 E. Huntzinger and E. Izaurralde, *Nat. Rev. Genet.*, 2011, **12**, 99–110.
- 22 A. Ventura and T. Jacks, *Cell*, 2009, **136**, 586–591.
- 23 M. S. Nicoloso, R. Spizzo, M. Shimizu, S. Rossi and G. A. Calin, *Nat. Rev. Cancer*, 2009, **9**, 293–302.
- 24 Q. Huang, K. Gumireddy, M. Schrier, C. le Sage, R. Nagel, S. Nair, D. A. Egan, A. Li, G. Huang, A. J. Klein-Szanto, P. A. Gimotty, D. Katsaros, G. Coukos, L. Zhang, E. Pure and R. Agami, *Nat. Cell Biol.*, 2008, **10**, 202–210.
- 25 L. Ma, J. Teruya-Feldstein and R. A. Weinberg, *Nature*, 2007, **449**, 682–688.
- 26 J. Lu, G. Getz, E. A. Miska, E. Alvarez-Saavedra, J. Lamb, D. Peck, A. Sweet-Cordero, B. L. Ebert, R. H. Mak, A. A. Ferrando, J. R. Downing, T. Jacks, H. R. Horvitz and T. R. Golub, *Nature*, 2005, **435**, 834–838.
- 27 J. Le Quesne and C. Caldas, *Mol. Oncol.*, 2010, **4**, 230–241.
- 28 E. O'Day and A. Lal, *Breast Cancer Res.*, 2010, **12**, 201.
- 29 K. Kozaki, I. Imoto, S. Mogi, K. Omura and J. Inazawa, *Cancer Res.*, 2008, **68**, 2094–2105.
- 30 F. Balaguer, A. Link, J. J. Lozano, M. Cuatrecasas, T. Nagasaka, C. R. Boland and A. Goel, *Cancer Res.*, 2010, **70**, 6609–6618.
- 31 E. Bandres, X. Agirre, N. Bitarte, N. Ramirez, R. Zarate, J. Roman-Gomez, F. Prosper and J. Garcia-Foncillas, *Int. J. Cancer*, 2009, **125**, 2737–2743.
- 32 J. Ferlay, D. M. Parkin and E. Steliarova-Foucher, *Eur. J. Cancer*, 2010, **46**, 765–781.
- 33 J. Ferlay, H. R. Shin, F. Bray, D. Forman, C. Mathers and D. M. Parkin, *Int. J. Cancer*, 2011, **127**, 2893–2917.
- 34 B. T. Bilynskij, *Exp. Oncol.*, 2011, **32**, 190–194.
- 35 P. A. Jones and S. B. Baylin, *Cell*, 2007, **128**, 683–692.
- 36 T. Ando, T. Yoshida, S. Enomoto, K. Asada, M. Tatematsu, M. Ichinose, T. Sugiyama and T. Ushijima, *Int. J. Cancer*, 2009, **124**, 2367–2374.
- 37 R. Steponaitiene, J. Kupcinskas, C. Langner, F. Balaguer, L. Venclauskas, H. Pauzas, A. Tamelis, J. Skieceviciene, L. Kupcinskas, P. Malfetheriner and A. Link, *Mol. Carcinog.*, 2015, DOI: 10.1002/mc.22287.
- 38 N. Kang, S. Y. Choi, Y. K. Kim, I. R. Yoo, D. H. Han, D. S. Lee, Y. S. Kim, S. H. Hong, J. H. Kang, K. Y. Lee, J. K. Park, S. W. Sung, M. S. Park, H. W. Yim, S. J. Kim and J. Y. Park, *Lung Cancer*, 2015, DOI: 10.1016/j.lungcan.2015.05.011.
- 39 P. Mestdagh, A. K. Bostrom, F. Impens, E. Fredlund, G. Van Peer, P. De Antonellis, K. von Stedingk, B. Ghesquiere, S. Schulte, M. Dewes, A. Thomas-Tikhonenko, J. H. Schulte, M. Zollo, A. Schramm, K. Gevaert, H. Axelson, F. Speleman and J. Vandesompele, *Mol. Cell*, 2011, **40**, 762–773.
- 40 G. Martin-Manso, M. J. Calzada, Y. Chuman, J. M. Sipes, C. P. Xavier, V. Wolf, S. A. Kuznetsova, J. S. Rubin and D. D. Roberts, *Arch. Biochem. Biophys.*, 2011, **509**, 147–156.
- 41 C. M. Croce, *Nat. Rev. Genet.*, 2009, **10**, 704–714.
- 42 L. T. Bemis, R. Chen, C. M. Amato, E. H. Classen, S. E. Robinson, D. G. Coffey, P. F. Erickson, Y. G. Shellman and W. A. Robinson, *Cancer Res.*, 2008, **68**, 1362–1368.
- 43 M. Liu, N. Lang, M. Qiu, F. Xu, Q. Li, Q. Tang, J. Chen, X. Chen, S. Zhang, Z. Liu, J. Zhou, Y. Zhu, Y. Deng, Y. Zheng and F. Bi, *Int. J. Cancer*, 2010, **128**, 1269–1279.
- 44 C. Tarantino, G. Paoletta, L. Cozzuto, G. Minopoli, L. Pastore, S. Parisi and T. Russo, *FASEB J.*, 2010, **24**, 3255–3263.
- 45 G. Sun, P. Ye, K. Murai, M. F. Lang, S. Li, H. Zhang, W. Li, C. Fu, J. Yin, A. Wang, X. Ma and Y. Shi, *Nat. Commun.*, 2011, **2**, 529.
- 46 A. Derfoul, A. H. Juan, M. J. Difilippantonio, N. Palanisamy, T. Ried and V. Sartorelli, *Carcinogenesis*, 2011, **32**, 1607–1614.
- 47 D. P. Bartel, *Cell*, 2009, **136**, 215–233.
- 48 C. A. Andorfer, B. M. Necela, E. A. Thompson and E. A. Perez, *Trends Mol. Med.*, 2011, **17**, 313–319.
- 49 S. Iwase, F. Lan, P. Bayliss, L. de la Torre-Ubieta, M. Huarte, H. H. Qi, J. R. Whetstone, A. Bonni, T. M. Roberts and Y. Shi, *Cell*, 2007, **128**, 1077–1088.
- 50 S. Catchpole, B. Spencer-Dene, D. Hall, S. Santangelo, I. Rosewell, M. Guenatri, R. Beatson, A. G. Scibetta, J. M. Burchell and J. Taylor-Papadimitriou, *Int. J. Oncol.*, 2011, **38**, 1267–1277.
- 51 M. R. Zou, J. Cao, Z. Liu, S. J. Huh, K. Polyak and Q. Yan, *J. Biol. Chem.*, 2014, **289**, 17620–17633.
- 52 M. Fabbri, R. Garzon, A. Cimmino, Z. Liu, N. Zanesi, E. Callegari, S. Liu, H. Alder, S. Costinean, C. Fernandez-Cymering, S. Volinia, G. Guler, C. D. Morrison, K. K. Chan, G. Marcucci, G. A. Calin, K. Huebner and C. M. Croce, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 15805–15810.
- 53 R. Garzon, S. Liu, M. Fabbri, Z. Liu, C. E. Heaphy, E. Callegari, S. Schwind, J. Pang, J. Yu, N. Muthusamy, V. Havelange, S. Volinia, W. Blum, L. J. Rush, D. Perrotti, M. Andreeff, C. D. Bloomfield, J. C. Byrd, K. Chan, L. C. Wu, C. M. Croce and G. Marcucci, *Blood*, 2009, **113**, 6411–6418.
- 54 S. Varambally, Q. Cao, R. S. Mani, S. Shankar, X. Wang, B. Ateeq, B. Laxman, X. Cao, X. Jing, K. Ramnarayanan, J. C. Brenner, J. Yu, J. H. Kim, B. Han, P. Tan, C. Kumar-Sinha, R. J. Lonigro, N. Palanisamy, C. A. Maher and A. M. Chinnaiyan, *Science*, 2008, **322**, 1695–1699.
- 55 E. J. Noonan, R. F. Place, D. Pookot, S. Basak, J. M. Whitson, H. Hirata, C. Giardina and R. Dahiya, *Oncogene*, 2009, **28**, 1714–1724.

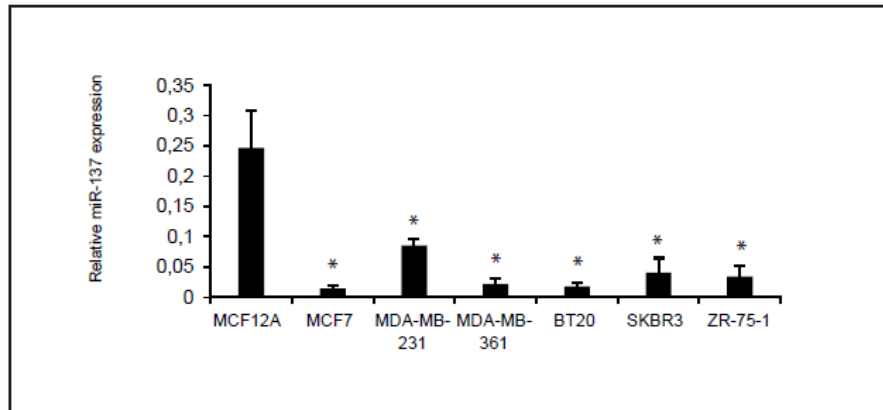


Fig S1 (related to Fig. 1A). MiR-137 is downregulated in breast cancer. MiR-137 expression was analyzed by TaqMan RT-qPCR in non-tumorigenic human mammary epithelial cell line (MCF12A) and in six breast cancer cell lines (MCF7, MDA-MB-231, MDA-MB-361, BT20, SKBR3, and ZR-75-1). Results are normalized with respect to the endogenous control RNU6B (*= $p < 0,05$) (n=3).

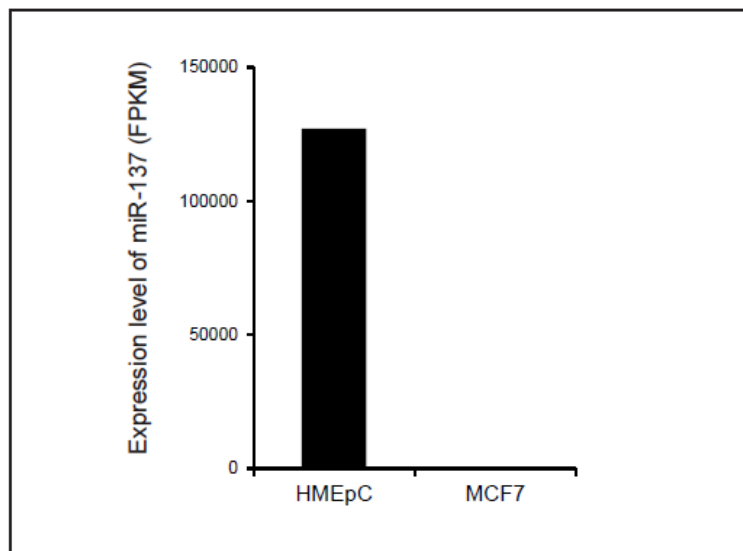


Fig. S2 (related to Fig. 1): MiR-137 is less expressed in MCF7 breast cancer cells than in HMEpC normal cells. Levels of miR-137 from RNA-seq data of ENCODE consortium are given as numbers of RNA-seq fragments per kilobase of transcript per million fragments mapped (FPKM) for the breast cancer cell line MCF-7 and human mammary epithelial cell, HMEpC.

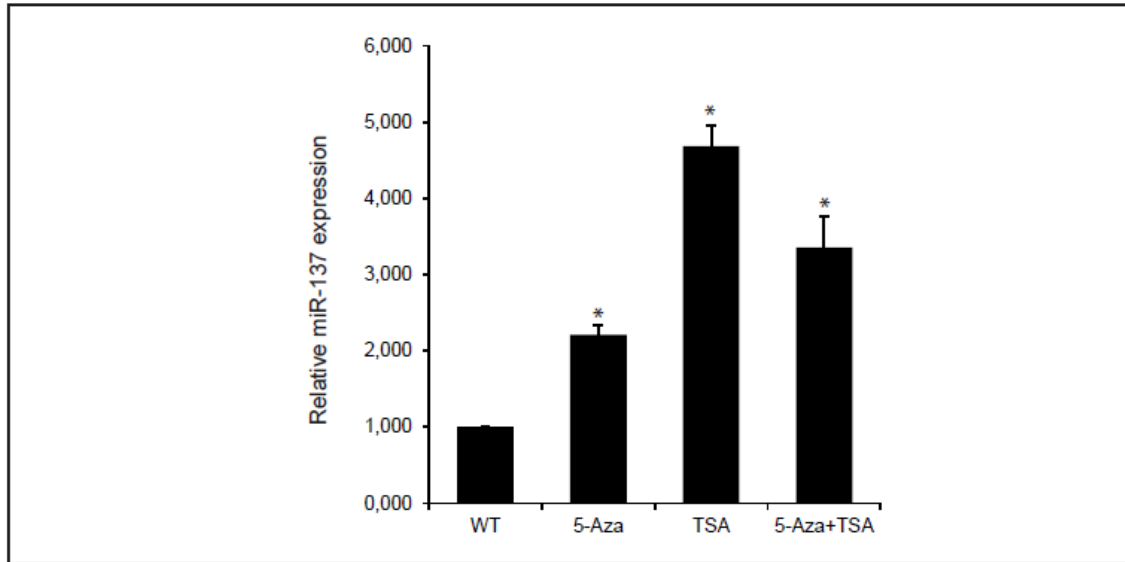


Fig. S3 (related to Fig. 1B): MiR-137 is epigenetically regulated in breast cancer cells. MiR-137 expression was evaluated by RT-qPCR in MCF7 cells, treated or not with 5-Aza (2,5µM) and/or TSA (200mM). Results are standardized with respect to the endogenous control RUN6B (n=2) (*=p>0,05).

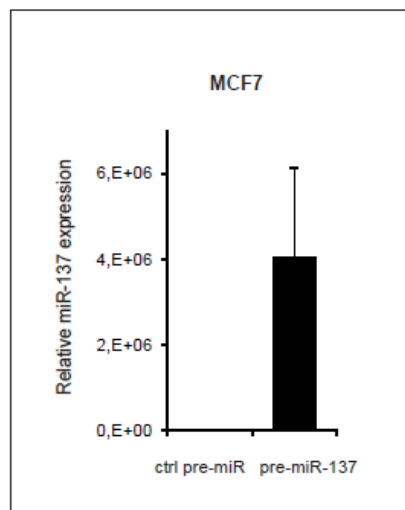


Fig. S4 (related to Figs. 2A, B): Mature miR-137 expression in proliferation experiments. Mature miR-137 expression was evaluated by RT-qPCR analysis. Transcript levels were measured following transfection of MCF-7 cells with miR-137 precursor (pre-miR-137). Results are normalized with respect to the endogenous control RNU6B (n=3).

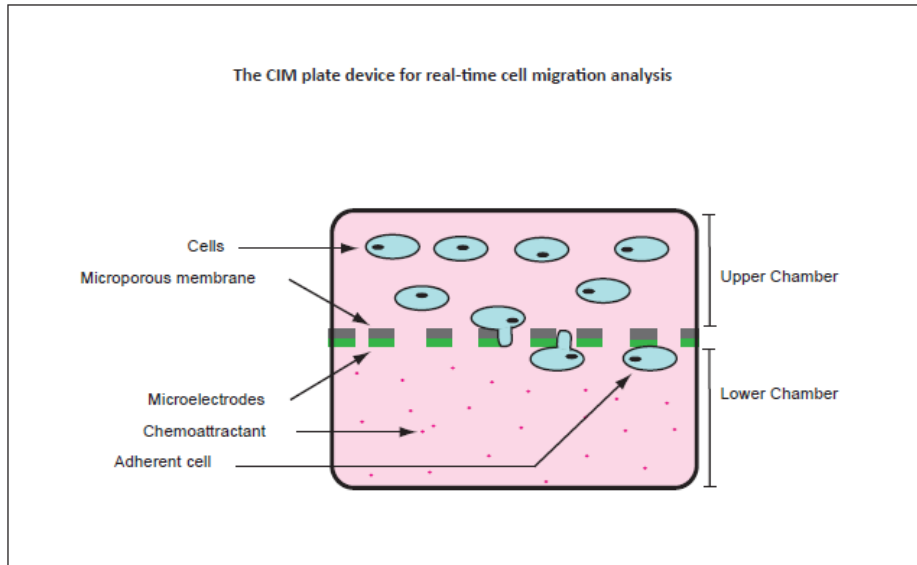


Fig. S5 (related to Fig. 2): Graphical representation of the CIM-Plate 16 device for real-time cell migration analysis. The xCELLigence instrument uses the CIM-Plate 16 which features microelectronic sensors integrated onto the underside of microporous polyethylene terephthalate membrane of a boyden-like chamber. As cells migrate from the upper chamber through the membrane into the bottom chamber in response to chemoattractant, they contact and adhere to the electronic sensor on the impedance changes are automatically and continuously by the RTCA instrument. Therefore, cell migration can be monitored via the cell index profile.

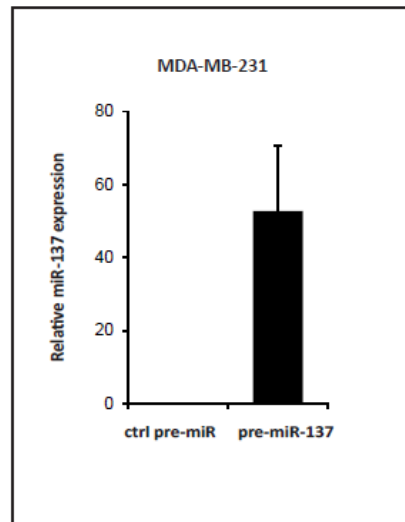


Fig. S6 (related to Figs. 2C, D): Mature miR-137 expression in migration experiments. Mature miR-137 expression was evaluated by RT-qPCR analysis. Transcript levels were measured following transfection of MDA-MB-231 cells with miR-137 precursor (pre-miR-137). Results are normalized with respect to the endogenous control RNU6B (n=3).

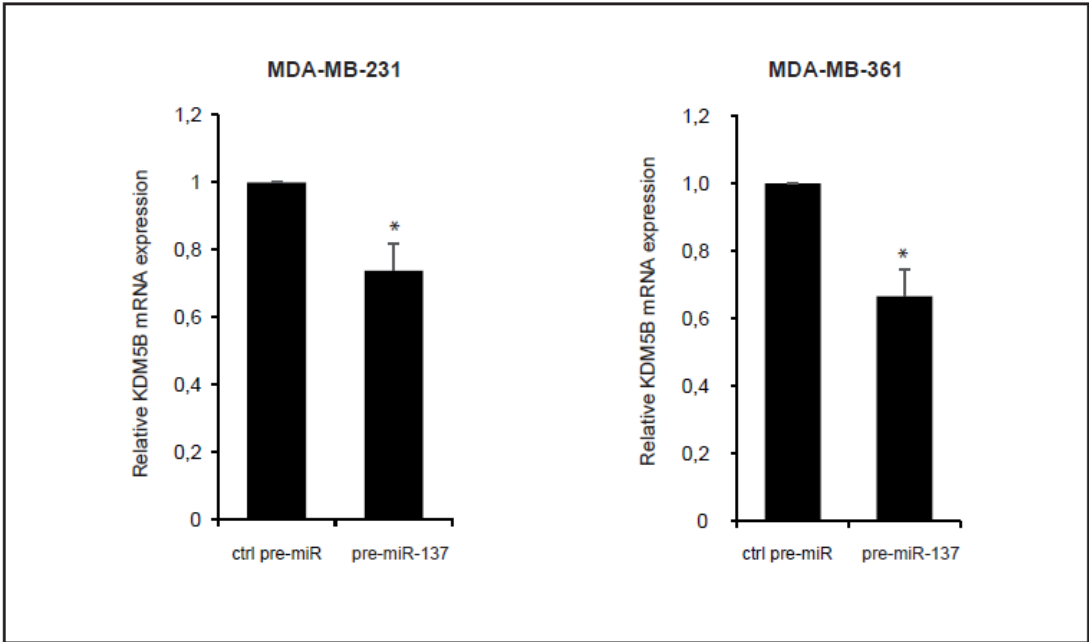


Fig. S7 (related to Fig. 3B): KDM5B is a direct target of miR-137 in breast cancer cells. Following transfection of MDA-MB-231 or MDA-MB-361 cells with miR-137 precursors (pre-miR-137) or with a scrambled pre-miR used as a negative control (ctrl pre-miR), KDM5B mRNA levels were analysed by real-time qPCR and standardized with respect to the endogenous control GAPDH. Relative expression levels were determined for control-pre-miR- and pre-miR-137-transfected cells (n=2) (*=p<0,05).

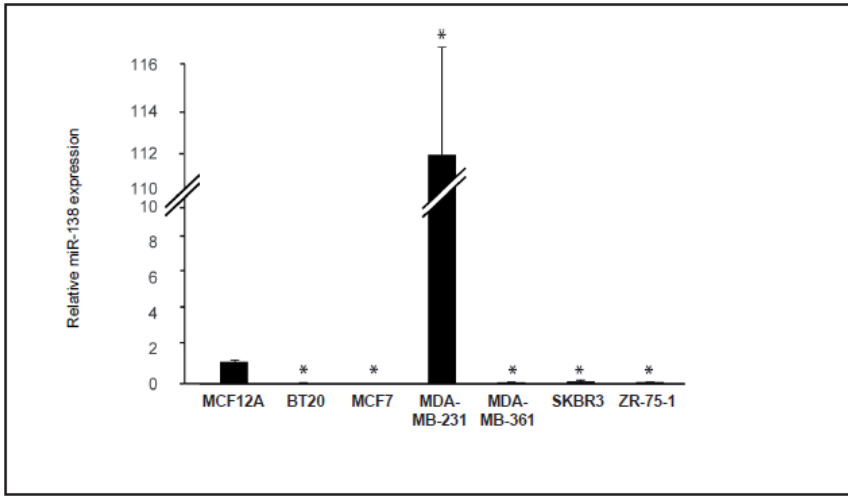


Fig S8 (related to Fig. 4B). MiR-138 is downregulated in breast cancer. MiR-138 expression was analyzed by TaqMan RT-qPCR in non-tumorigenic human mammary epithelial cell line (MCF12A) and in six breast cancer cell lines (BT20, MCF7, MDA-MB-231, MDA-MB-361, SKBR3, and ZR-75-1). Results are normalized with respect to the endogenous control RNU6B (*=p<0,05) (n=3).

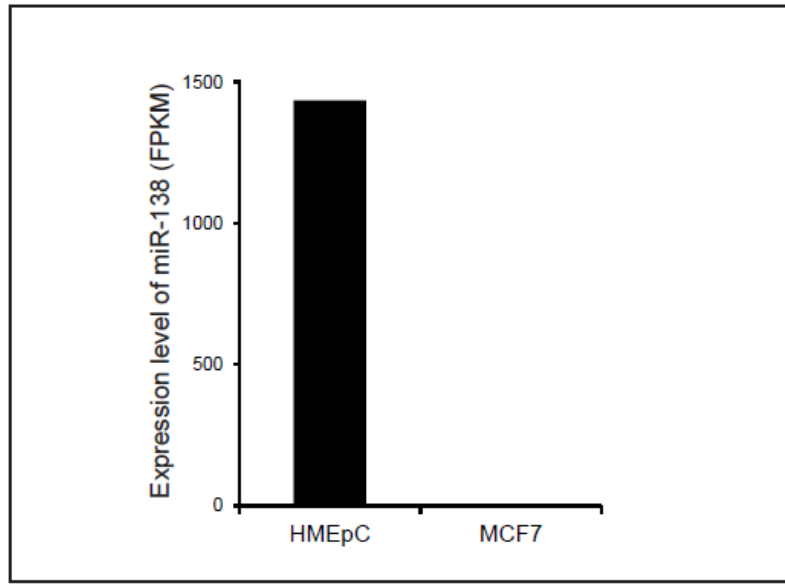


Fig. S9. (related to Fig. 4B): MiR-138 is less expressed in MCF7 breast cancer cells than in HMEpC normal cells. Levels of miR-138 from RNA-seq data of ENCODE consortium are given as numbers of RNA-seq fragments per kilobase of transcript per million fragments mapped (FPKM) for breast cancer cell line MCF-7 and human mammary epithelial cell, HMEpC.

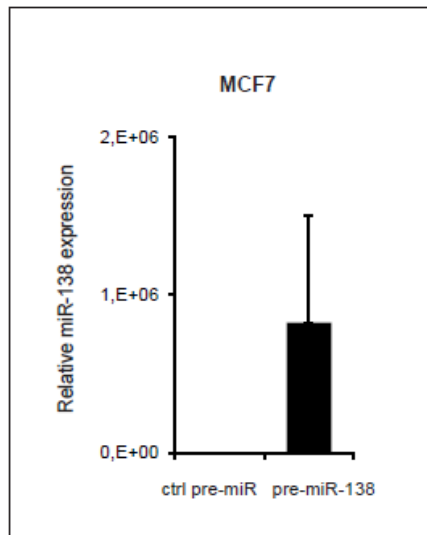


Fig. S10 (related to Fig. 5): Mature miR-138 expression in proliferation experiments. Mature miR-138 expression was evaluated by RT-qPCR analysis. Transcript levels were measured following transfection of MCF-7 cells with miR-138 precursor (pre-miR-138). Results are normalized with respect to the endogenous control RNU6B (n=3).

ANNEXE II

Portraying breast cancers with long
noncoding RNAs

Science Advances, (2016)

Olivier Van Grembergen, Martin Bizet, Eric J. de Bony, Emilie Calonne, Pascale Putmans, Sylvain Brohée, Catharina Olsen, Mingzhou Guo, Gianluca Bontempi, Christos Sotiriou, Matthieu Defrance and François Fuks

Portraying breast cancers with long noncoding RNAs

Olivier Van Grembergen,¹ Martin Bizet,^{1,2,3} Eric J. de Bony,¹ Emilie Calonne,¹ Pascale Putmans,¹ Sylvain Brohée,⁴ Catharina Olsen,² Mingzhou Guo,⁵ Gianluca Bontempi,^{2,3} Christos Sotiriou,⁴ Matthieu Defrance,^{1,3} François Fuks^{1*}

2016 © The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC). 10.1126/sciadv.1600220

Evidence is emerging that long noncoding RNAs (lncRNAs) may play a role in cancer development, but this role is not yet clear. We performed a genome-wide transcriptional survey to explore the lncRNA landscape across 995 breast tissue samples. We identified 215 lncRNAs whose genes are aberrantly expressed in breast tumors, as compared to normal samples. Unsupervised hierarchical clustering of breast tumors on the basis of their lncRNAs revealed four breast cancer subgroups that correlate tightly with PAM50-defined mRNA-based subtypes. Using multivariate analysis, we identified no less than 210 lncRNAs prognostic of clinical outcome. By analyzing the coexpression of lncRNA genes and protein-coding genes, we inferred potential functions of the 215 dysregulated lncRNAs. We then associated subtype-specific lncRNAs with key molecular processes involved in cancer. A correlation was observed, on the one hand, between luminal A-specific lncRNAs and the activation of phosphatidylinositol 3-kinase, fibroblast growth factor, and transforming growth factor- β pathways and, on the other hand, between basal-like-specific lncRNAs and the activation of epidermal growth factor receptor (EGFR)-dependent pathways and of the epithelial-to-mesenchymal transition. Finally, we showed that a specific lncRNA, which we called CYTOR, plays a role in breast cancer. We confirmed its predicted functions, showing that it regulates genes involved in the EGFR/mammalian target of rapamycin pathway and is required for cell proliferation, cell migration, and cytoskeleton organization. Overall, our work provides the most comprehensive analyses for lncRNA in breast cancers. Our findings suggest a wide range of biological functions associated with lncRNAs in breast cancer and provide a foundation for functional investigations that could lead to new therapeutic approaches.

INTRODUCTION

Breast cancer is a major public health issue. According to the most recent worldwide cancer statistics, more than 1,675,000 women are diagnosed with this disease each year and more than 500,000 die of it (1). Breast cancer is a heterogeneous disease, and different subtypes have been described (2). Beyond the classic grading system (based on tumor cell differentiation status) and the TNM (tumor size, lymph node involvement, and metastasis) classification, breast tumors are also classified on the basis of protein and gene status. Clinically, breast tumors are subclassified into three main subgroups on the basis of estrogen receptor (ER) and human epidermal growth factor receptor 2 (HER2) gene expression: ER-positive tumors (the most frequently diagnosed subtype), HER2-positive tumors (harboring an *ERBB2* amplification), and triple-negative breast cancers lacking ER, HER2, and the progesterone receptor (3). Additionally, microarray-based coding mRNA expression profiling has identified five “intrinsic” subtypes (2). Luminal A (low-grade) and luminal B (high-grade) tumors are predominantly ER⁺. The HER2⁺ subtype mostly comprises tumors overexpressing the HER2⁺ gene. The basal-like subtype is particularly frequent among triple-negative cancers. The least frequent subtype, called “normal-like,” comprises tumors that have an expression pattern similar to that of normal breast tissues. Recently, because of whole-transcriptome sequencing [RNA sequencing (RNA-seq)], this classification has been refined, with the identification of 12 breast tumor subgroups (4). Despite these advances, we are far from completely

understanding breast cancer heterogeneity because tumors of the same subtype can respond differently to therapy and can have different outcomes (5, 6). Understanding the molecular mechanisms that underlie breast cancer heterogeneity thus remains a major challenge improving diagnosis and therapy, and new approaches are needed to meet this challenge. Recent work has suggested that long noncoding RNAs (lncRNAs) and small noncoding RNAs (such as microRNAs) might play key roles in mammary tumor development (7, 8) and provide new biomarkers and potential targets for future therapies.

lncRNAs are transcripts more than 200 base pairs long that lack an extended open reading frame and thus do not code for proteins (9). According to recent studies, the human transcriptome contains up to 16,000 lncRNAs, frequently spliced and polyadenylated, whose genes are mainly transcribed by RNA polymerase II (9). Expression of lncRNA genes is lower than that of protein-coding genes, but is tissue-specific (9). In recent years, studies have linked lncRNAs to a wide variety of physiological and pathological mechanisms, including pluripotency regulation and cancer development (10). Like proteins, lncRNAs may mediate oncogenic or tumor-suppressive effects (10, 11). They can exert various functions in the cytoplasm (for example, as scaffolds between proteins or as microRNA sponges) and the nucleus (10). They have emerged as key players in the transcriptional regulation of protein-coding genes, in which case they can act either distally (in trans) or proximally (in cis) (10). For instance, the lncRNA HOTAIR (HOX transcript antisense RNA), which is up-regulated in some breast cancer tissues and whose expression is associated with poor prognosis and tumor metastasis, is suggested to silence tumor suppressor genes (7, 10). Other lncRNAs appear as key regulators of pathways underlying carcinogenesis; an example is lincRNA-p21, which mediates global gene repression in the p53 response (12). Genome-wide association studies on cancer have

¹Laboratory of Cancer Epigenetics, Faculty of Medicine, ULB-Cancer Research Center (U-CRC), Université Libre de Bruxelles (ULB), 1070 Brussels, Belgium. ²Machine Learning Group, Computer Science Department, Université Libre de Bruxelles, 1050 Brussels, Belgium. ³Inter-university Institute of Bioinformatics Brussels, Université Libre de Bruxelles-Vrije Universiteit Brussel, 1050 Brussels, Belgium. ⁴Breast Cancer Translational Research Laboratory, Jules Bordet Institute, Université Libre de Bruxelles, 1000 Brussels, Belgium. ⁵Department of Gastroenterology and Hepatology, Chinese People's Liberation Army General Hospital, Beijing 100853, China. *Corresponding author. Email: ffuks@ulb.ac.be

revealed that more than 80% of cancer-associated single-nucleotide polymorphisms occur in noncoding regions. This suggests that a significant fraction of the genetic etiology of cancer is related to lncRNAs (13). Previous studies have documented aberrantly expressed lncRNA genes in breast tumors and have notably established associations between certain lncRNAs and known gene expression-based breast cancer subtypes (14, 15). However, these lncRNAs have not been precisely related to molecular pathways, and their functions have not been investigated.

The goal of the present study was to explore lncRNA landscape in breast cancers and to extract novel biological and clinical information. On the basis of an array-based transcriptional survey of more than 3000 lncRNA genes, we have identified lncRNAs aberrantly expressed across 823 breast tumors, as compared to 172 normal samples. Although gene chips contain less lncRNAs than other technologies for which data are publicly available (for example, RNA-seq), we used microarray data for the following reasons: (i) low technical variation (16–18), (ii) strand specificity, (iii) larger number of publicly available data, and (iv) long follow-up clinical annotation.

From our in-depth analyses, we have inferred potential functions of dysregulated lncRNAs and demonstrated their relevance to breast cancer classification. We have investigated lncRNAs as survival markers, associating several of them with prognosis. Finally, we have experimentally characterized the functions of one breast cancer-related lncRNA—CYTOR (cytoskeleton regulator). Overall, this work provides the most comprehensive data sets so far for lncRNA in breast cancers. It highlights the influence of lncRNAs in numerous pathways that are dysregulated in tumors and may provide novel approaches to cancer prognosis and treatment.

RESULTS

Breast tumors display lncRNA gene expression profiles that are distinct from those of normal breast tissues

Seven data sets from the Gene Expression Omnibus (GEO) data repository (19) were selected and compiled to generate a large cohort of 823 breast tumors and 172 normal breast tissues (Fig. 1A and table S1). The data sets were selected on the basis of their size (more than 50 samples), the presence of breast tumors from each subtype, and extensive clinical annotation, including relapse information. To increase the number of normal samples, we included the GSE10780 series that profiled 143 normal breasts and 42 tumors.

First, we reannotated the entire collection of probe sets of the Affymetrix Human Genome U133 Plus 2.0 Array to analyze levels of 16,951 mRNAs and 3053 lncRNAs (table S2). When the distribution of lncRNA gene expression levels across all breast tissue samples was compared with that of coding gene expression levels, it was confirmed that the former generally show lower level expression than do the latter (fig. S1A) (9). On the basis of the observed lncRNA gene expression profiles in tumors and normal breast tissues, we then performed unsupervised consensus clustering (20, 21) and identified three robust clusters, one of which (cluster II) contained almost all of the normal samples (95%) (P value of the association = 2.3×10^{-164}). Cancer and healthy tissues thus appear to have different lncRNA gene expression profiles (Fig. 1B). We therefore sought to identify lncRNAs whose genes are differentially expressed in breast tumors versus normal breast tissue. Because of the heterogeneity of breast cancer, we did not apply the classical t test but adapted a method that allows detection of differentially expressed

lncRNAs in a fraction of cancer samples (22). We identified 215 lncRNAs whose genes appeared aberrantly expressed in at least 10% of the breast tumors (see Materials and Methods, fig. S1B, and table S3). Among these 215 lncRNAs, 124 appeared up-regulated and 91 appeared down-regulated in breast cancer (Fig. 1, C and D, top). Some of these identified lncRNAs have already been described as involved in breast cancer development, such as HOTAIR, MALAT1, H19, and GAS5 (10, 14). The genes that encode the 215 identified lncRNAs appeared scattered across the genome, without any focal hot spot (fig. S1C).

We next examined whether lncRNA levels, like the levels of protein-coding transcripts, can serve as biomarkers for breast cancer diagnosis. For this, we generated an lncRNA metagene and assessed its ability to discriminate tumor samples from healthy ones. In the discovery data set, significant discrimination of breast tumors from healthy samples was achieved ($P < 0.0001$) with high specificity (0.96) and high sensitivity (0.95). To validate our findings on an independent data set, we used RNA-seq data from The Cancer Genome Atlas (TCGA) consortium reanalyzed by Rahman *et al.* (23) that we called “validation set 1A.” This processed data set allows assessment of the expression of 1161 lncRNAs from 1052 breast tumors and 113 normal samples. On the basis of lncRNA expression, we performed unsupervised hierarchical clustering and observed three stable clusters (Fig. 1E), one of which contained almost all of the normal samples, suggesting that our results could be extended to independent cohorts profiled by RNA-seq. Of the 215 lncRNAs identified as dysregulated in the discovery cohort, we could assess the expression of 87 lncRNAs present in the RNA-seq data set. Of them, 82 were dysregulated in the RNA-seq data (see examples in Fig. 1D, bottom). In addition, the expression pattern of a metagene that represents the expression of these 87 lncRNAs allowed highly accurate classification of the cancer versus normal samples (specificity, 0.98; sensitivity, 0.93; $P < 0.0001$) (fig. S1D). We also re-annotated the custom Agilent 244K microarray used by the TCGA consortium (4). Of the 215 dysregulated lncRNAs identified from the discovery cohort, 167 were detectable on the TCGA microarray, which we called “validation set 1B.” The expression patterns of the corresponding genes allowed a highly predictive classification of breast cancer and normal breast tissues in the TCGA validation set composed of 524 breast tumors and 63 normal tissues (specificity, 0.97; sensitivity, 0.98; $P < 0.0001$) (fig. S1E). Together, these analyses provide a validated set of lncRNAs that are dysregulated in breast tumors, as compared to normal human samples.

lncRNA expression is associated with ER signaling

On the basis of hierarchical clustering (Fig. 1B), we noticed that breast tumors are separated into two different clusters, and we further searched for their relationship with clinical properties. We observed a significant association with the immunohistochemistry (IHC)-based ER status, that is, marked differential expression of lncRNAs between ER-positive and ER-negative tumors. Cluster I was found to contain 91% ER-positive tumors (P value of the association = 6.7×10^{-61}), whereas cluster III contained 76% ER-negative tumors (P value of the association = 3.6×10^{-63}) (Fig. 2A and table S4). Given this finding, we next used a supervised approach to identify lncRNAs specifically associated with the ER status. From the whole set of lncRNAs (that is, 3053 lncRNAs), our analysis revealed 38 lncRNA genes differentially expressed between ER-positive and ER-negative breast tumors, determined by IHC [false discovery rate (FDR) < 0.05 ; fold change > 1.5 , t test] (see examples in Fig. 2B). Whereas an association between

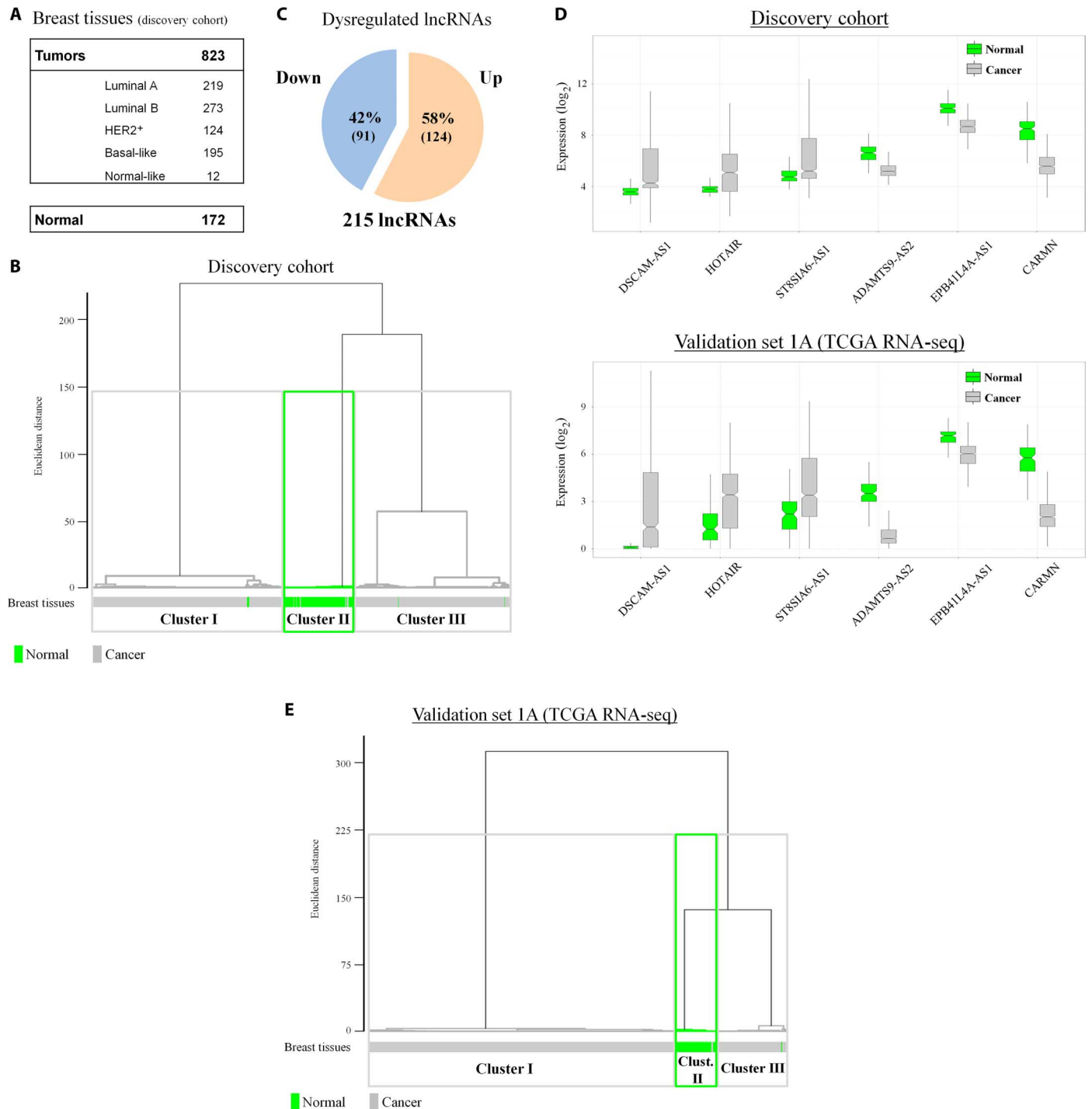


Fig. 1. lncRNA gene expression profiling in breast tissues reveals 215 dysregulated lncRNAs. (A) Description of the human breast tissues analyzed in this study. (B) Unsupervised consensus clustering of the samples on the basis of lncRNA gene expression. Primary tumors (823) and normal samples (172) were used for hierarchical clustering on the basis of the 500 most variant lncRNAs (based on SD). (C) Pie chart showing, among the lncRNAs that are dysregulated in breast tumors versus normal tissues, the numbers of down- and up-regulated ones. (D) Box plot for expression levels of the top dysregulated lncRNAs in the discovery cohort (top) and their expression in the TCGA RNA-seq cohort (bottom). Notches are used to compare groups; if the notches of two boxes do not overlap, the medians differ significantly. The whiskers extend to the most extreme data point, which is no more than 1.5 times the interquartile range of the box. (E) Same as in (B) for the TCGA RNA-seq cohort composed of 1052 breast tumors and 113 normal samples.

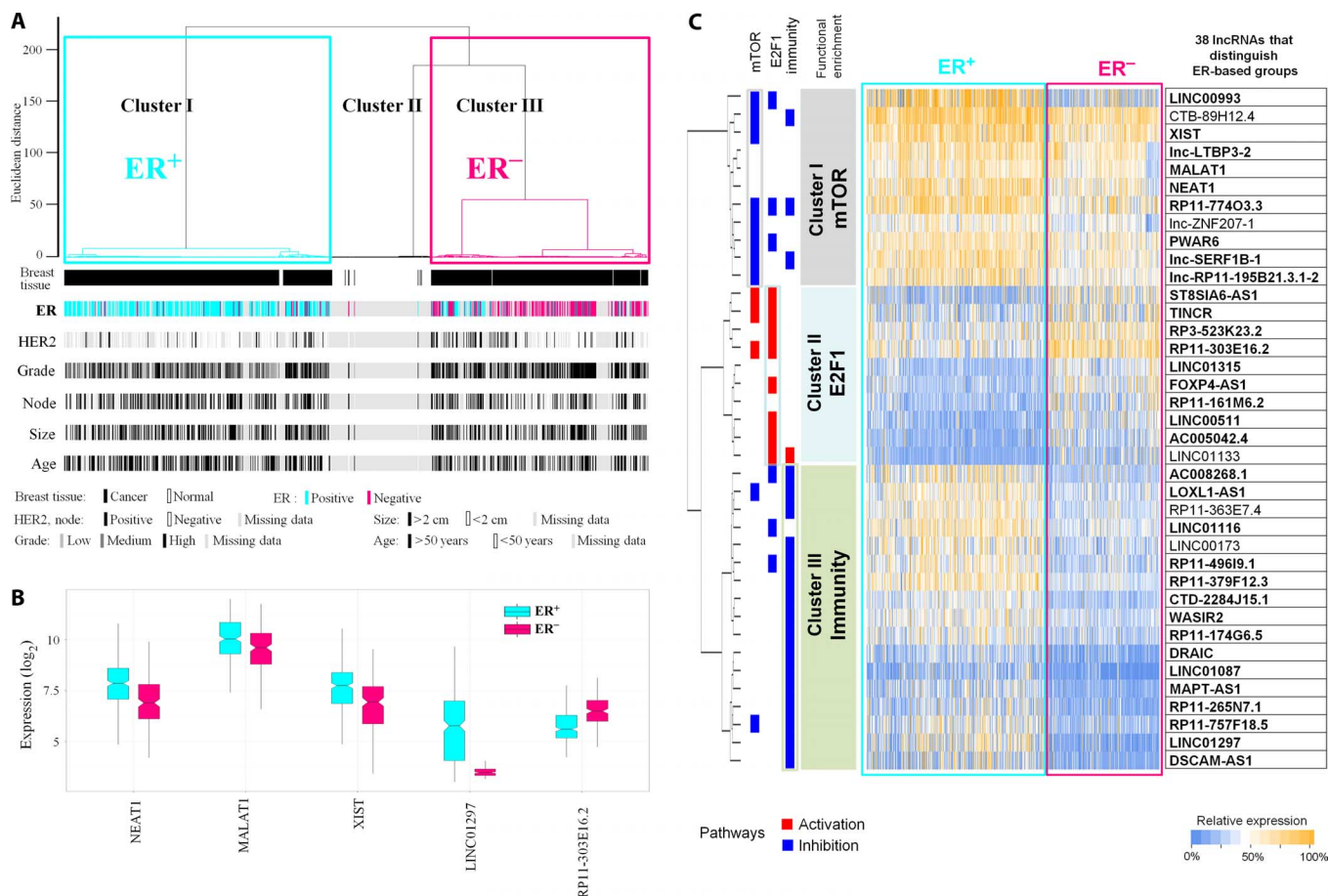


Fig. 2. IncRNA gene expression profiling identifies two main breast tumor categories differing with regard to ER status. (A) Dendrogram of 823 primary tumors and 172 normal samples obtained by consensus hierarchical clustering of the samples on the basis of expression of the top 500 most variant lncRNAs. Clusters I and III, encompassing almost all tumors, are related to the ER status. (B) Box plot illustrating the expression levels of five lncRNA genes differentially expressed between ER⁺ and ER⁻ tissues. Notches are used to compare groups; if the notches of two boxes do not overlap, the medians differ significantly. The whiskers extend to the most extreme data point, which is no more than 1.5 times the interquartile range of the box. (C) Heat map illustrating the expression of the 38 lncRNA genes (rows) of the ER signature across the breast tumors (columns). The lncRNAs in bold represent lncRNAs dysregulated between breast cancer and normal samples. The color scale of the heat map indicates the relative expression of each lncRNA gene. Hierarchical clustering reveals three clusters of lncRNAs. For each cluster, the most significant functional enrichment term from the guilt-by-association analysis is shown.

the ER status and some of the identified lncRNAs [for example, NEAT1 (24), MALAT1 (25), and Xist (26)] had previously been evidenced, most lncRNAs highlighted here are novel in the context of breast cancer. This is the case for LINC01297, the most significantly up-regulated lncRNA in ER-positive tumors, and RP11-303E16.2, the most significantly down-regulated lncRNA in these tumors (Fig. 2B). Notably, 32 of the 38 ER-associated lncRNAs were also dysregulated in breast tumors, as compared to normal samples.

We next used different data sets [validation set 1A (RNA-seq data from TCGA), validation set 1B (microarray data from TCGA), and validation set 2 (GSE20685)] to assess to what extent one could predict the ER status on the basis of expression levels of the lncRNA genes identified above. In validation set 1A (TCGA RNA-seq), we assessed the expression of 12 of the 38 lncRNAs from the ER-associated lncRNAs and found that their expression allowed highly accurate prediction of the ER status (specificity, 0.93; sensitivity, 0.83; $P < 0.0001$). Validation set 1B (TCGA microarray) was generated on a different microarray platform, which allowed assessment of the expression of 29 of the 38 lncRNA

genes identified. The expression patterns of these 29 lncRNA were again highly predictive of the ER status (specificity, 0.89; sensitivity, 0.78; $P < 0.0001$). Finally, in the GSE20685 data set, the expression of the 38 genes can also predict the ER status, with a specificity of 0.86 and a sensitivity of 0.91 ($P < 0.0001$). Prediction of the ER status on the basis of these patterns was thus highly reproducible (fig. S2).

To gain insight into the biological relevance of the 38 lncRNAs, we used the “guilt-by-association” approach (27) to investigate their relationship to different pathways (see Materials and Methods). Briefly, we computed the matrix of correlation between lncRNA levels and expression levels of protein-coding genes across all breast cancer samples and then generated hypotheses regarding the function(s) of each dysregulated lncRNA on the basis of the known biological functions and the molecular pathways of protein-coding genes that show a good correlation (table S5). We used hierarchical clustering to subdivide the 38 lncRNAs that constitute the ER signature into three groups on the basis of their levels in the various tumors (Fig. 2C). This analysis revealed enrichment in key breast cancer-related pathways for each group

of lncRNAs. Levels of group 1 lncRNAs, which tended to be up-regulated in ER-positive tumors, appeared to correlate most strongly with mammalian target of rapamycin (mTOR) pathway inhibition. Group 2 lncRNAs appeared most strongly related to E2F1 pathway activation. Group 3 contained lncRNAs that are potentially involved in inhibition of immunity. Overall, our findings suggest that lncRNA levels can distinguish ER⁺ from ER⁻ tumors and that lncRNAs are involved in various processes beyond ER biology. This sheds new light on pathways related to the ER status of breast cancers, such as the mTOR pathway and immunity (4, 28, 29).

Identification and inferred functions of lncRNAs that are associated with molecular subtypes of breast cancers

We next examined the relationship between the expression of lncRNA genes and the intrinsic breast tumor subtypes. lncRNA level-based consensus clustering of the 823 tumors identified four robust clusters (assessed by the ConsensusClusterPlus algorithm). There was also good agreement between lncRNA-based clusters and PAM50-defined mRNA-based subtypes: clusters 1, 2, 3, and 4 were found to contain 96% basal-like (P value of the association = 1.9×10^{-142}), 52% HER2⁺ (P value of the association = 2.6×10^{-60}), 84% luminal A (P value of the association = 1.6×10^{-50}), and 64% luminal B (P value of the association = 1.8×10^{-47}) samples, respectively ($P = 1.2 \times 10^{-242}$, χ^2 test) (table S6). We thus used

the PAM50 classification as a reference to identify molecular subtype-specific lncRNAs, which defines a specific lncRNA signature for each breast cancer subtype (see Materials and Methods). These signatures consisted of 42 lncRNAs (36 up-regulated and 6 down-regulated) for luminal A, 9 (8 up-regulated and 1 down-regulated) for luminal B, 14 (8 up-regulated and 6 down-regulated) for HER2⁺, and 74 (28 up-regulated and 46 down-regulated) for the basal-like subtype (Fig. 3B and table S7). The large size of the basal-like signature distinguishes the basal-like subtype as particularly perturbed at the level of lncRNA gene expression.

Again, we looked at the predictive values of these signatures in the validation data sets used previously [validation set 1A (RNA-seq data from TCGA), validation set 1B (microarray data from TCGA), and validation set 2 (GSE20685)]. We found each subtype-specific lncRNA signature to identify with high-efficiency samples belonging to the considered subtype. As in the case of subtype mRNA signatures (30), the best prediction scores were obtained for the basal-like subtype, thus confirming that this subtype has a more distinct profile than the others. Overall, these analyses performed on cohorts independent of our initial 823 tumor samples validated the four subtype-specific signatures (fig. S3).

We next wondered whether these subtype-specific lncRNAs might be globally related to the alteration of different biological functions and molecular pathways. To investigate this, we again used the guilt-by-association approach, relating the functions predicted for subtype-specific

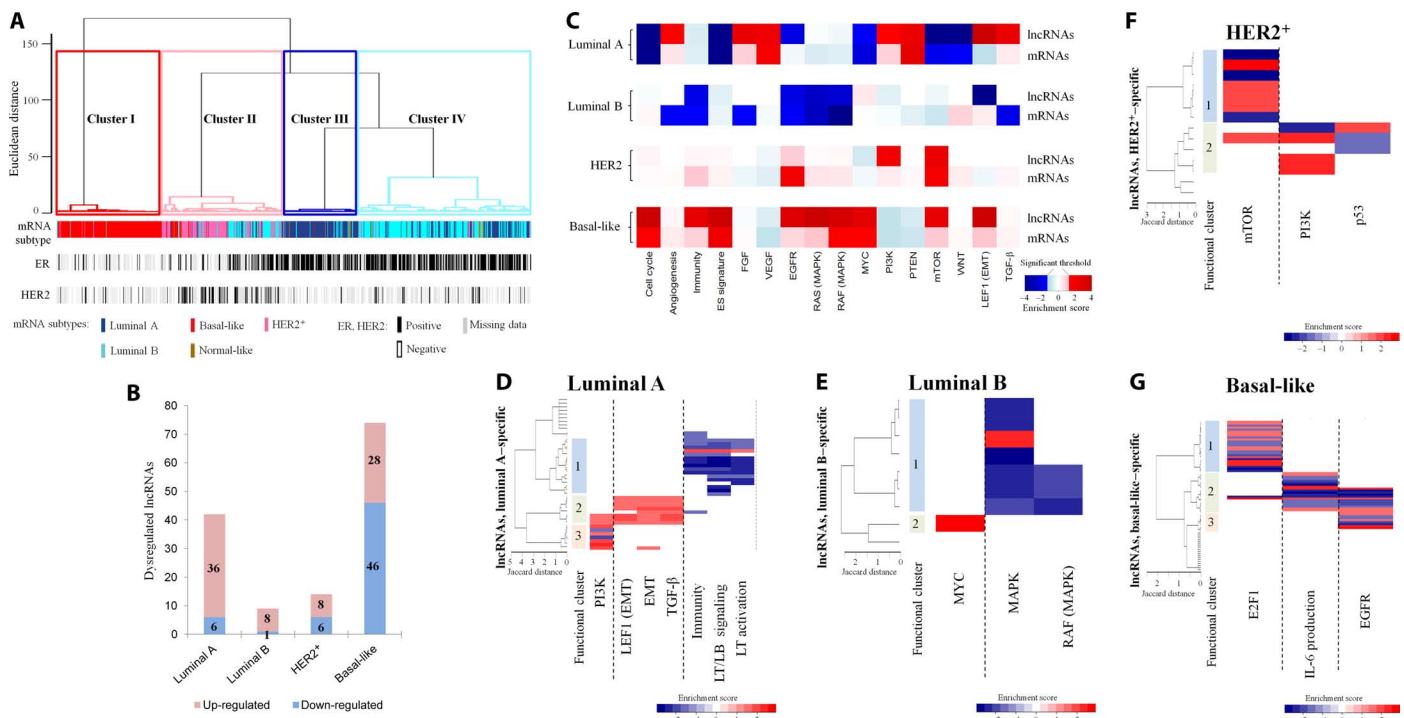


Fig. 3. Identification of four lncRNA-related clusters correlating with the known molecular subtypes and enriched in specific functional terms.

(A) Dendrogram of the 823 breast tumors obtained by consensus hierarchical clustering according to the levels of the 500 most variant dysregulated lncRNAs, revealing four groups of tumors (clusters I to IV). How these clusters relate to the mRNA-based breast cancer subtype (based on PAM50) is also shown. (B) Histogram illustrating the number of specific lncRNAs in each molecular subtype of breast cancer. (C) Heat map illustrating the pathways whose activation or inhibition correlates with levels of subtype-specific lncRNAs and mRNAs. To relate subtype-specific lncRNAs to gene sets, an enrichment metascore was computed for each gene set. The P value of the metascore was defined as the proportion of random metascores being at least as high (low) as the metascore of the positively (negatively) subtype-specific lncRNAs. Conventional GSEA analysis was used to analyze specific enrichment in mRNAs, comparing one subtype to the three others. The significance score was defined as the log of the P value, adjusted by the sign of the enrichment metascore. (D to G) Heat maps illustrating the enrichment scores of subtype-specific lncRNAs for representative gene sets in the (D) luminal A, (E) luminal B, (F) HER2⁺, and (G) basal-like subtypes. A positive (negative) score is associated with the activation (repression) of the gene set.

lncRNAs to those predicted for subtype-specific mRNAs (Fig. 3C). We observed good agreement between the two sets of predicted functions. In the basal-like subtype, for example, we found both predicted lncRNAs and mRNAs to be associated with cell cycle activation and the RAF [mitogen-activated protein kinase (MAPK)] and MYC pathways; yet, there were cases where subtype-specific lncRNAs appeared more strongly associated with a function or pathway than subtype-specific mRNAs. This was true for angiogenesis stimulation and activation of phosphatidylinositol 3-kinase (PI3K), fibroblast growth factor (FGF), and transforming growth factor- β (TGF- β) pathways in the luminal A subtype and for activation of epidermal growth factor receptor (EGFR) and lymphoid enhancer binding factor 1 (LEF1) pathways in the basal-like subtype.

We then took a closer look at predicted functions of subtype-specific lncRNAs. Within each subtype-specific lncRNA signature, we first performed hierarchical clustering of the lncRNAs on the basis of their coexpression with coding mRNAs and then examined which biological functions or molecular pathways appeared overrepresented in each group (Fig. 3, D to G). We observed three functional groups for luminal A-specific lncRNAs: lncRNAs related to the PI3K pathway, lncRNAs associated with the epithelial-to-mesenchymal transition (EMT), and lncRNAs related to immunity (Fig. 3D). Hierarchical clustering of the nine lncRNAs that compose the luminal B signature identified one main lncRNA group related to the MAPK pathway, including a group of lncRNAs enriched for the RAF pathway and one additional lncRNA associated with the activation of the MYC pathway (Fig. 3E). We identified two functional groups of HER2⁺-specific lncRNAs: one related to the mTOR pathway and one related to the PI3K pathway. This second functional cluster encompasses lncRNAs related to the p53 pathway, known to be influenced by the PI3K pathway in the context of HER2⁺ tumors (Fig. 3F) (31). Three groups of basal-like-specific lncRNAs were identified: group 1, which contains lncRNAs whose genes are coexpressed with E2F1 target genes; group 2, which is related to interleukin-6 (IL-6) production and partly to the EGFR pathway [two pathways known to be interconnected (32)]; and group 3, which is more specifically related to the EGFR pathway (Fig. 3G).

Together, our analyses highlight important putative functions for subgroups of subtype-specific lncRNA genes whose expression correlates tightly with that of cancer-related coding genes. Moreover, some groups of subtype-specific lncRNAs are involved in common pathways (that is, the PI3K pathway), whereas others appear more specialized (that is, group 3 in the basal-like subtype, which is related to E2F1). Our results show that some lncRNAs seem to be more significantly regulated in many pathways than mRNAs. This illustrates the importance of lncRNAs in breast cancer.

Dysregulated lncRNAs are markers of breast cancer clinical outcome

Like protein-coding RNAs, several lncRNAs have been linked to clinical outcome in different diseases (33). We thus used univariate Cox regression models to assess whether expression levels of lncRNAs might correlate with relapse-free survival, using our discovery data set with long follow-up (median follow-up, 6.75 years). No less than 300 lncRNAs emerged as significantly prognostic markers of risk of relapse (table S8), including 41 lncRNAs dysregulated in breast cancer, as compared to healthy tissues. Next, we performed multivariate Cox analysis to examine the possible impact of confounding factors known to affect prognosis (size, node, grade, ER, and HER2). The expression of 210 lncRNAs appeared to be significant risk-of-relapse predictors in this analysis, sug-

gesting that these lncRNAs could be independent prognostic factors (Fig. 4A and table S9). Most of them are novel survival markers in breast cancer, including the newly identified RP11-863K10.2, which has the highest hazard ratio (HR) in our study (HR, 8.5; $P = 0.007$) (Fig. 4, B and C), and LINC00152, which we called CYTOR [following the HUGO Gene Nomenclature Committee guidelines (34)] (HR, 1.42; $P = 0.012$) (Fig. 4, D and E), an lncRNA up-regulated in all subtypes of breast cancer and recently revealed as a marker of gastric cancer (35). As shown in Fig. 4, Kaplan-Meier curves highlighted significant differences in relapse-free survival between patients whose tumors showed high and low levels of certain lncRNAs.

Twenty-seven of the 210 prognostic lncRNAs were also dysregulated in breast cancer, as compared to normal tissues, suggesting their importance in terms of biological relevance. We sought to validate this set of 27 dysregulated and prognostic lncRNAs using a metagene-based approach. As shown in fig. S4, we observed similar Cox HR values for the prognostic lncRNAs in the discovery cohort (HR, 1.23; $P = 1.9 \times 10^{-6}$) and in an independent validation cohort composed of 327 samples with a median follow-up of 9.2 years (HR, 1.22; $P = 2.2 \times 10^{-5}$; validation P value = 0.0029). Together, the above results suggest that a set of 27 dysregulated lncRNAs might be used as novel independent markers reliably predicting the risk of relapse in breast cancer.

CYTOR is required for cell proliferation, cell migration, and cytoskeletal organization

We selected CYTOR (previously known as LINC00152) for further functional characterization and experimental validation of the guilt-by-association prediction. CYTOR is an intergenic lncRNA located more than 100 kb away from the nearest protein coding (PLGLB2) in the chromosome 2p11.2. We chose this lncRNA for the following reasons: (i) CYTOR may be a general tumor marker because it is up-regulated in all subtypes of breast cancer (Fig. 5A) and in other types of cancer such as thyroid, stomach, lung, renal, and liver cancer (36); (ii) CYTOR is prognostic of relapse in both our discovery and validation cohorts; (iii) our guilt-by-association analysis highlighted an association between lncRNA and key cancer-related pathways such as cell proliferation, cell migration, and EMT and the EGFR, mTOR, and MAPK pathways (Table 1).

In addition to this, the transcription start site of CYTOR appears associated with H3K27 acetylation, H3K4 trimethylation, and weak H3K4 monomethylation marks in different breast cancer cell lines [MDA-MB-231 (Fig. 5B), HMEC, and MCF-7 (fig. S5A)], suggesting that this intergenic lncRNA is transcribed from promoter-like elements (37). The promoter DNA methylation profile of CYTOR in tumors and various breast cell lines suggests that it may be regulated by DNA methylation in breast tumors (fig. S5, B and C).

To assess the function of CYTOR, we used locked nucleic acid (LNA) gapmers to efficiently knock down CYTOR in MDA-MB-231 breast cancer cells (fig. S6A), wherein the basal level of CYTOR is high. Silencing of CYTOR resulted in a significant decrease in cell proliferation, as assessed with xCELLigence technology (Fig. 5C). In 5-bromo-2'-deoxyuridine (BrdU)/7-aminoactinomycin D (7-AAD) flow cytometry experiments, CYTOR knockdown cells accumulated in the G₂/M phase, at the expense of the S phase, confirming cell cycle inhibition (fig. S6, B and C). Thus, CYTOR is required for normal proliferation and cell cycle progression. We then evaluated the effect of CYTOR knockdown on cell migration, another process predicted to be affected by the guilt-by-association analysis. Cell migration kinetics was recorded with the xCELLigence system, using fetal bovine serum (FBS) as a chemoattractant.

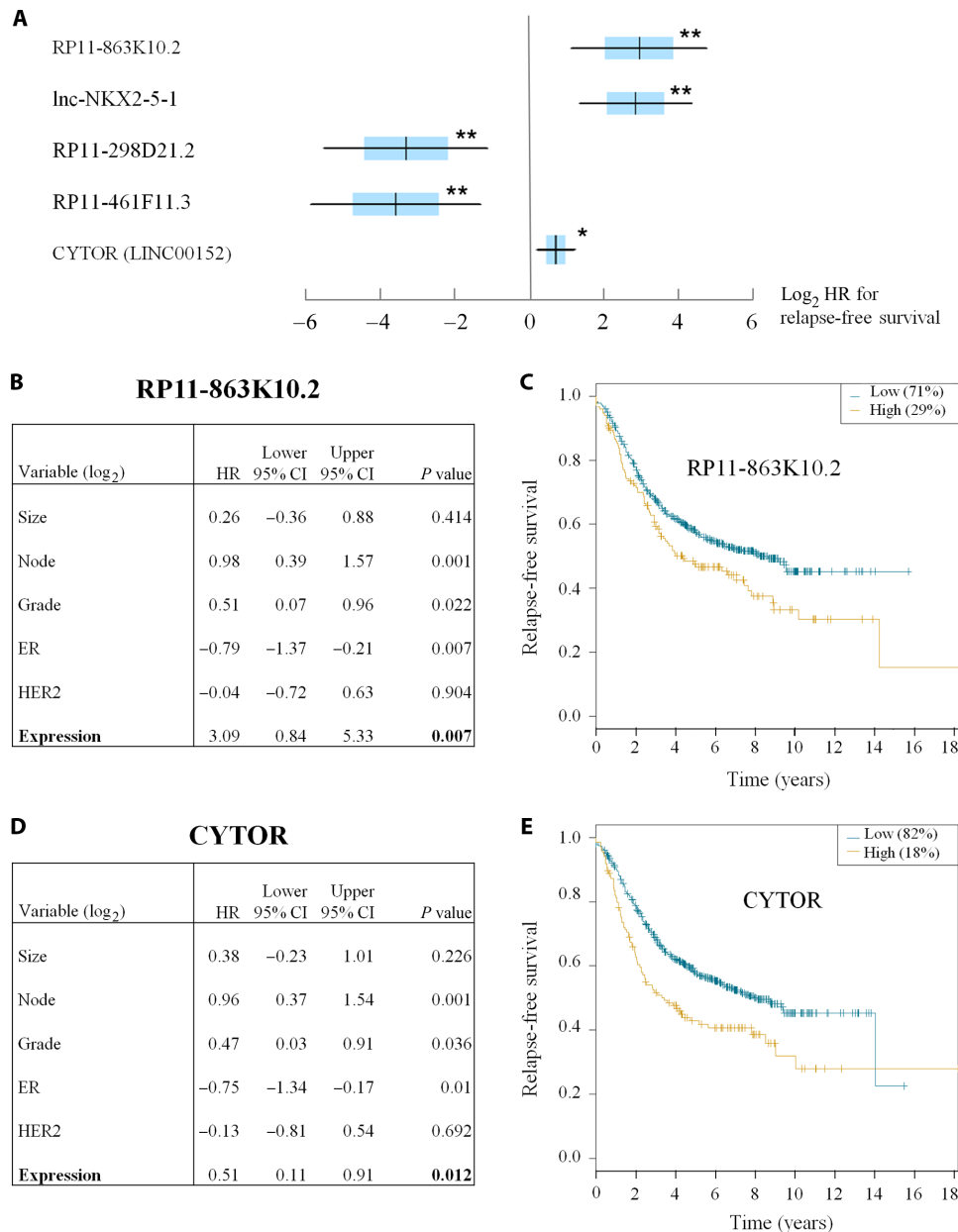


Fig. 4. lncRNAs in breast cancer have prognostic value. (A) Forest plot showing the log₂ HR with the SD (blue boxes) and the 95% confidence interval (bars) of the relapse-free survival analysis (multivariate Cox analysis). A negative HR reveals that a high lncRNA level is associated with a good outcome, and conversely. For example, five lncRNAs significantly related to relapse are shown. **P* < 0.05, ***P* < 0.01. (B) Multivariate analysis with all the classical markers used clinically. RP11-863K10.2 is used as an example (see table S9 for the complete analysis). CI, confidence interval. (C) Exemplative Kaplan-Meier curves for RP11-863K10.2 (*P* = 0.01, log-rank test). (D) Same as in (B) for CYTOR. (E) Same as in (C) for CYTOR (*P* = 0.000776, log-rank test).

Without FBS in the lower chamber medium, cells did not migrate. With FBS, control cells migrated, whereas cells depleted for CYTOR did not (Fig. 5D).

We next examined the possible role of CYTOR in controlling gene expression. In gene microarray experiments on control and CYTOR-knockdown MDA-MB-231 cells, we found 2 genes that are up-regulated and 79 genes that are down-regulated in the latter (fold change > 1.5; FDR < 0.05) (Fig. 5E), suggesting a rather positive role for this lncRNA in gene regulation. Some of the identified targets are involved in breast

cancer [*KIF14* (38) and *GOLPH3* (39)], some are involved in key signaling pathways (for example, *mTOR*, *PRKCA*, and *IGFR2*), and some are involved in cytoskeleton remodeling [for example, tubulin tyrosine ligase (*TTL*), Rho guanosine triphosphatase (GTPase) (*Rhobtb3*), and plakophilin-4 (*PKP4*) (40)]. Note that none of the genes in the large neighborhood (±1 Mb) of *CYTOR* was affected by the knockdown, suggesting that this lncRNA acts in trans. We selected five CYTOR target genes (*mTOR*, *GOLPH3*, *KIF14*, *PRKCA*, and *SMYD3*) for validation by reverse transcription quantitative polymerase chain reaction (RT-qPCR)

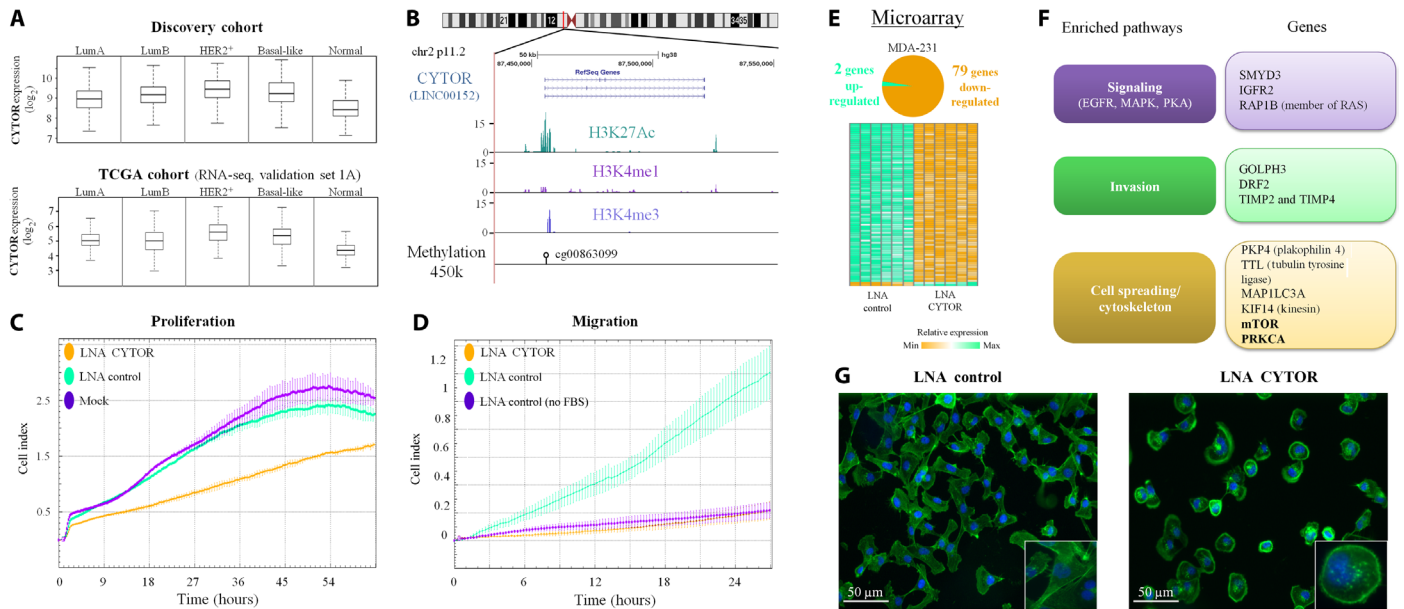


Fig. 5. CYTOR controls cell proliferation, cell migration, and cytoskeleton organization. (A) Expression of the gene encoding CYTOR in the discovery data set (823 tumors and 172 normal breast samples) and the TCGA RNA-seq data set GSE62944 (971 tumors and 103 normal breast samples). (B) UCSC (University of California, Santa Cruz) genome browser view of chromosome locus 2p11.2, which contains CYTOR. The magnified view depicts CYTOR with its associated histone marks (H3K27 acetylation, H3K4 trimethylation, and H3K4 monomethylation, from GSE49651) and methylation marks in MDA-MB-231 cells. The unfilled lollipop represents unmethylated CG dinucleotides from the Infinium HumanMethylation450. (C) Proliferation curve of MDA-MB-231 cells with and without LNA gapmer-mediated knockdown of CYTOR. A real-time cell analyzer (RTCA) software representative trace of a triplicate experiment is shown. (D) Migration kinetics of MDA-MB-231 cells assessed by continuous monitoring for approximately 24 hours. FBS (10%) in the lower chamber was used as chemoattractant, except for the control curve (purple), which represents untransfected cells with serum-free medium in the lower chamber. An RTCA software representative trace of a triplicate experiment is shown. (E) Pie chart and heat map showing the distribution of differentially expressed genes after CYTOR knockdown in MDA-MB-231 cells. (F) Enriched pathways (left) from Ingenuity Pathway Analysis in CYTOR knockdown cells. Representative differentially expressed genes of enriched gene sets are shown on the right (see also fig. S6E). (G) Staining of F-actin with Acti-stain 488 fluorescent phalloidin (green) and of DNA with 4',6-diamidino-2-phenylindole (blue) reveals reorganization of the actin cytoskeleton in MDA-MB-231 cells transfected with an LNA gapmer against CYTOR (right), as compared to cells transfected with a control LNA gapmer (left). Images captured with a 40x objective are shown.

after CYTOR knockdown. The results confirmed changes in their transcript-level expression (fig. S6D). Furthermore, Ingenuity Pathway Analysis of the targets revealed significant overrepresentation of pathways related to EGFR and MAPK signaling (Fig. 5F and fig. S6E), as predicted by our guilt-by-association analysis (Table 1). The most overrepresented term was cell spreading, a process driven by actin polymerization and promoted by the Rho family of GTPase and the mTORC2 complex (41, 42). Therefore, we stained the filamentous actin (F-actin) cytoskeleton with Acti-stain 488 fluorescent phalloidin in cells treated with the LNA gapmer targeting CYTOR, or with LNA control cells. The former cells appeared smaller and rounder than the latter, and global reorganization of actin fibers was observed in these cells, with fewer stress fibers and thick actin fibers present mainly on the cell cortex (Fig. 5G).

Overall, these results indicate that CYTOR, which correlates with poor outcome, is required for breast cancer cell growth, migration, and normal morphology. They also show that CYTOR may act in trans to control genes involved in the mTOR pathway. It may thus be a good candidate target for new therapeutic approaches.

DISCUSSION

Although previous studies have demonstrated the involvement of lncRNAs in breast cancer pathology (14, 15), the functions they exert

in breast cancer development remain poorly understood. The microarray approach we used here is limited to detecting lncRNAs that are known and present on the Affymetrix U133 Plus design. However, while representing a fraction of lncRNAs encoded by the human genome, our integrative analysis also brings significant insights and advances over previous studies because it provides the most comprehensive data sets so far for breast cancers, a resource of clinically relevant lncRNAs, and a potential lncRNA function in the breast cancer context and uncovers their utility in prognosis. It is worth adding that the clinical information available from the array-based expression profiles are more extensive than publicly available data from the TCGA RNA-seq experiment, especially in terms of median follow-up [median follow-up of 6.75 years for the discovery cohort and 9.2 years for the validation cohort, in comparison to 1.08 years for the biggest RNA-seq cohort publicly available from the TCGA consortium (4)]. Our work provides an important foundation for the potential function and clinical relevance of lncRNA in breast cancers, and future work would certainly be valuable to extend our results to the entire set of lncRNAs, notably by RNA-seq.

Here, we show that 215 lncRNAs are dysregulated in breast cancers. Using (i) already processed TCGA RNA-seq data, hence easily usable, and (ii) two independent microarray cohorts, we could validate our discovery signature on data profiled by two different technologies. We also demonstrated an association between lncRNAs and clinical features and relapse. In terms of function, these lncRNAs appear to be associated with

Table 1. Predicted functions of CYTOR. A selected gene set significantly associated with CYTOR by the guilt-by-association analysis (see also table S5).

Gene set	Score
Epithelial-mesenchymal transition (HALLMARK_EMT)	2.75
Proliferation (BENPORATH_PROLIFERATION)	2.44
EGFR signaling (EGFR_UP.V1)	2.33
MAPK signaling (MAPK_CASCADE.POS)	2.25
RAS (MAPK) signaling (RAS.POS)	2.14
mTORC signaling (HALLMARK_MTORC_SIGNAL.)	2.12
RAF (MAPK) signaling (RAF_UP.V1.POS)	2.12
mTOR signaling (MTOR_UP.N4.V1.POS)	1.99
Migration (CELL_MIGRATION.POS)	1.92

numerous key molecular processes, including the EGFR, PI3K, MAPK, and E2F1 pathways. In addition, we find that CYTOR, an lncRNA that is dysregulated in breast cancer and is associated with bad outcome, is essential to the proliferation, migration, and normal morphology of breast cancer cells.

Our in-depth transcriptomic analysis has revealed that lncRNA profiling of breast tumors distinguishes ER⁺ from ER⁻ tumors and allows stratification into different molecular subtypes, in agreement with findings of recent studies (14, 15). By RNA-seq in breast cancer cell lines, two recent studies identified lncRNAs regulated by the ER in the presence or absence of an estrogen agonist (43, 44). However, we found only two lncRNAs (DSCAM-AS1 and RP11-161M6.2) that overlap with our ER-related signature and also between these two studies themselves, suggesting differential mechanism between cell lines and tissues and/or specificities related to the method/platform used.

In addition, we provide validated ER status- and subtype-specific lncRNA signatures and highlight possible roles of lncRNAs in ER- and subtype-related pathways. We report that many dysregulated lncRNAs are related to processes or pathways that play key roles in breast cancer development, such as immunity and the MAPK, PI3K, and mTOR pathways. A few of the functions predicted here have been validated by previous studies, suggesting that our guilt-by-association approach is effective. For example, *PVT1* and *MINCR*, two lncRNAs predicted here to play a role in the MYC pathway, were recently found to control MYC activity in mammary tumor and in Burkitt's lymphoma, respectively (45, 46).

Notably, we show that the lncRNAs that compose our ER-related signature can be subdivided into three clusters, which predicted to influence various biological functions previously described in ER signaling. For instance, the cluster I lncRNAs appear associated with the mTOR pathway, which drives cell growth and promotes survival. Hyperactivation of this pathway is involved in the development of ER⁺ breast cancer and in resistance to endocrine therapy (47). Further characterization of the lncRNAs involved in this pathway may be of interest in the search for potential new therapies. Here, we demonstrate an association of ER-related cluster II lncRNAs, most of which are up-regulated in ER-negative tumors, with activation of E2F1 signaling. Regulation of E2F1 by ER signaling is reported to mediate resistance to hormone therapy

(48). Moreover, the *E2F1* gene appears more highly expressed in ER tissues, and high levels of E2F1 transcript correlate with an unfavorable outcome (49). In cluster III, we highlight lncRNAs that may play roles related to the immune response. Although the link between ER and the immune response in breast cancer remains unclear, some studies suggest that ER plays an immunosuppressive role (50, 51). Furthermore, 9 of the 17 lncRNAs in cluster III are associated with the production of IL-6, a potential regulator of normal and tumor stem cell self-renewal. Abnormally high IL-6 levels seen in basal-like breast tumors (ER-negative) are associated with EMT and with poor clinical outcome (52, 53). Our present results reveal a potential link between lncRNAs, IL-6 production, and immunity in ER-negative tumors.

We have taken a closer look at the identified molecular subtype-specific lncRNAs, some of which have already been described in breast cancer. For example, H19, the first identified imprinted lncRNA (54), is up-regulated in luminal A samples in both our discovery and validation cohorts. These results agree with *in situ* hybridization data that show overexpression of the H19 gene in ER-positive tumors and with the finding that estradiol transcriptionally regulates H19 (55). Although H19 and its involvement in breast cancer have been extensively studied, its function remains unclear because it is reported to both promote and suppress metastasis (56). The results of our guilt-by-association study highlight an association of this lncRNA with the TGF- β pathway and with the EMT, as already described in several studies (57, 58). Our analysis also links H19 to epigenetic proteins such as SETD7, KDM1, and EZH1 and provides a new context for further characterization of H19. In conjunction with Chen *et al.* (59), we observed that most of our basal-like-specific lncRNAs are down-regulated and also identified LINC00993 as the most down-regulated lncRNA in the basal-like subtype. Our last example involved HOTAIR; in agreement with a recent study that suggests that HOTAIR may be related to the HER2⁺ subtype (14), we find it to be the most up-regulated lncRNA of the HER2⁺ signature. In support of a link between HER2 and HOTAIR, the latter has recently been reported as an HER2⁺ regulator in gastric cancer (60). We also predicted a positive association between HOTAIR and the PI3K pathway; this relation was recently reported in other cancers (61, 62), but the mechanisms involved are still poorly understood. Therefore, further studies are needed to better understand HOTAIR/PI3K signaling transduction. Note that, in addition to linking HOTAIR to the HER2⁺ and PI3K pathways, our analysis also links HOTAIR to endocrine therapy resistance and to other processes, such as histone modification by EHMT1 (euchromatic histone lysine *N*-methyltransferase 1) or CREBP (cAMP response element-binding protein). These links should be further explored.

Also noteworthy is our observation that subtype-specific lncRNAs and subtype-specific mRNAs are globally predicted to activate or inhibit the same pathways. However, some specificities appear. For example, lncRNAs specific to the basal-like subtype seem particularly involved in the activation of the EGFR pathway. Amplification of the EGFR gene is common in basal-like tumors (4). Here, we provide evidence for an interesting correlation between EGFR gene amplification and lncRNAs associated with the EGFR pathway. For instance, the gene *TPT1-AS1*, which is associated with EGFR pathway activation, is more expressed in EGFR-amplified tumors, whereas *DRAIC*, which is associated with EGFR pathway inhibition, is less expressed in such tumors. This suggests that these EGFR signaling-associated lncRNAs may be regulated by somatic copy number alteration of the EGFR. Overall, we show that our prediction method provides results in keeping with

known functions of lncRNAs. This lends weight to the hypotheses it has enabled us to generate regarding new functions of dysregulated lncRNAs in breast cancer. Confirming these hypotheses may provide a strong basis for further functional studies.

To test the reliability of the guilt-by-association approach, we selected CYTOR for validation. The results of our functional assays confirm the predicted involvement of CYTOR in cell proliferation and migration and its link to the EGFR and mTOR pathways. These findings support recent studies that show that CYTOR promotes proliferation (i) in gastric cancer, through the EGFR-dependent pathway (63), and (ii) in hepatocellular carcinoma, through the mTOR signaling pathway (64). Moreover, we reveal a drastic reorganization of F-actin upon CYTOR knockdown. We propose a model wherein CYTOR ultimately influences F-actin organization by regulating *GOLPH3* expression, which in turn regulates cell size (65), and affects the mTORC2 complex by controlling *mTOR* and *PRKCA* expression. Alternatively, CYTOR could bind directly to the mTORC2 complex or to modulators of the cytoskeleton. Overall, our data indicate that CYTOR may be involved in breast cancer development by playing an essential role in cell proliferation, migration, and morphology. It thus appears as a potential target for future therapies. Further research is required to better understand the mechanisms underlying the involvement of CYTOR in breast cancer.

Besides providing a basis for studying functions of dysregulated lncRNAs in breast cancer, we also identified and validated a set of 27 lncRNAs predictive of relapse, using multivariate Cox analysis. Whereas most of these markers are novel, some have already been shown in other studies to have prognostic value. Regarding HOTAIR, which was reported to be a significant predictor of metastasis and death (7), this lncRNA does not reach statistical relevance in our data set, suggesting that other lncRNAs identified here could be more informative with regard to the probability of relapse. Furthermore, a recent paper demonstrates that HOTAIR has prognostic value in ER-negative patients only (66). This may explain why HOTAIR does not emerge as related to relapse from our analysis. Globally, these 27 lncRNAs might have clinical use as molecular diagnostic markers for identifying patients with low risk of relapse and who do not need aggressive therapy. We further compared our list of relapse-associated lncRNAs with recently published survey of prognostic lncRNAs in breast tumor. Sun *et al.* (67) identified nine lncRNAs associated with metastasis in breast cancer, of which four are present in the Affymetrix U133 Plus 2.0 array. However, none of them were significantly associated with relapse in our uni- or multivariate Cox analysis. Another study identified a set of 45 lncRNAs prognostic of metastasis in lymph node-negative breast cancer (68); this said, we found only a small overlap with our results (five and two, respectively, in common with our uni- and multivariate analysis). This suggests that additional efforts should be done to robustly identify sets of prognostic lncRNAs.

Overall, our study provides an in-depth analysis of the lncRNA transcriptome in breast cancer and provides numerous new lncRNA markers associated with ER status, tumor subtype, and clinical outcome. We have inferred functions of these dysregulated lncRNAs, and we demonstrate for the first time that lncRNAs might contribute to the dysregulation of nearly every known breast cancer pathway. These data lay the ground for future studies that address the biological mechanisms involving these lncRNAs and their use as diagnostic markers and therapeutic targets. We have experimentally confirmed the predicted function of one such dysregulated lncRNA, concluding that our integrative ap-

proach is effective. These findings should contribute to a better understanding of the mechanisms of action of lncRNAs in breast cancer.

MATERIALS AND METHODS

Experimental design

The goal of this study was to analyze the transcriptome of breast cancer in order to identify aberrantly expressed lncRNAs and to infer their functions. We began by downloading publicly available microarray data and extracted information on lncRNA expression. We used a specific method to identify lncRNAs that were dysregulated in small subset of breast tumors and infer their functions through a guilt-by-association approach. We have investigated lncRNAs as survival markers. Finally, we have characterized the functions of one breast cancer-related lncRNA, CYTOR, by means of in vitro experiment. The detailed procedure is described below.

Breast cancer gene expression data and reannotation of the Affymetrix microarray

To obtain a genome-wide view of lncRNA expression in breast cancer, we reannotated the entire collection of probe sets of the Affymetrix Human Genome U133 Plus 2.0 array. We downloaded the microarray (U133 Plus 2.0 Affymetrix) gene expression data sets GSE9195, GSE10780, GSE10810, GSE12276, GSE19615, GSE20711, and GSE21653 from GEO (<http://ncbi.nlm.nih.gov/geo/>). The raw CEL files were frozen robust multiarray analysis (fRMA)-normalized in the R environment using the limma and fRMA libraries to obtain log₂-normalized expression signals for each probe set. We then applied the ComBat algorithm of the sva library with default parameters to adjust the data for batch effects, using tumor versus normal tissue as covariate. Hybridization probe sets were locally mapped by sequence alignment (National Center for Biotechnology Information BLAST 2.2.29+) against a reference transcriptome in LNCipedia database version 2.1 (69), a database dedicated to lncRNAs and Ensembl 84 transcriptome. We required that at least 80% of a probe set should hit a given transcript sequence. For the lncRNAs, we kept probes that target lncRNAs present in the LNCipedia database. We then discarded probes that were discordant between LNCipedia and Ensembl, in terms of transcript biotype. We also excluded probe sets that target multiple genes, except if the target was a duplicated lncRNA (that is, corresponding to a duplicated region of the genome). To identify these duplicated lncRNAs, we blasted all lncRNA transcripts against the LNCipedia database, where transcripts were defined as duplicated if the smallest transcript shared at least 95% of its sequence with the other. Because these duplicated lncRNAs could not be distinguished from each other, they were analyzed as arising from a single gene and the tag “multi” was added to the name of one of them. Alternative transcripts were considered to be from the same gene. When multiple probe sets mapped to the same gene (corresponding to an lncRNA or an mRNA), the one with the highest variance was selected. We computed a full annotation table for the 3053 lncRNAs, comprising their corresponding names in the Ensembl database (if available), their genomic location, their category, and the probe sets that matched each lncRNA (table S2).

Normal breast tissues are composed of a mixture of different cell types, mainly including epithelial cells and adipocytes, whereas breast tumors are composed mostly of epithelial cells. Because we focused on lncRNA genes differentially expressed between normal and cancerous

breast epithelial cells, we excluded six lncRNA genes whose expression correlated strongly (Pearson $R > 0.6$) with adipose markers (70) in normal breast tissues, suggesting that they might be expressed in adipocytes rather than in epithelial cells.

To assess potential remaining batch effects in the breast cancer data sets, we performed unsupervised hierarchical clustering of the top 500 most variable genes (coding and noncoding). None of the breast samples clustered according to the data set they came from, indicating that no strong batch effect was present.

For the validation steps, the fRMA-normalized data set GSE20685 was downloaded from InSilico DB (www.insilicodb.com). We also acquired raw data from the TCGA consortium. GSE20685 contains no normal samples and was therefore not used to validate the dysregulated lncRNAs in breast cancer.

Clinical data and molecular subtype prediction

The clinical data were downloaded from GEO and were merged (table S1). In the original studies, the ER and HER2 statuses were determined by IHC. Intrinsic mRNA-based breast cancer subtypes were determined with the 50-gene PAM50 predictor (71).

TCGA RNA-seq analysis

Because the raw RNA-seq data (level 1) from TCGA are challenging to analyze, we chose to use the TCGA data reprocessed by Rahman *et al.* (23) (GSE62944) that allowed to assess the expression of 1241 lncRNAs, of which 804 were present on the Affymetrix microarray. We first selected data from breast tissues and then applied the voom transformation from the limma package (72, 73) on the transcripts per million expression values to obtain \log_2 expression data.

TCGA microarray reannotation

TCGA raw data were processed as previously described (4). Probes of the TCGA microarray were mapped to the LNCipedia database using the TCGA annotation file. Briefly, coordinates targeted by TCGA microarray probes were first extracted from the annotation file “AnotAgilentG4502A_07_3.adf” available on the TCGA Web site and converted to hg19 genome build, using liftOver (UCSC). We selected probes where at least 58 base pairs (bp) of the targeted region overlapped, in a strand-specific way, with exons of lncRNA transcripts of the LNCipedia version 2.1 database. Because the boundaries of exons are not always clearly defined, we added 5 bp on both sides of each lncRNA exon. In parallel, probes targeting protein-coding regions were extracted similarly with Ensembl version 72 and RefSeq version 58. Finally, probes targeting lncRNAs were only kept for further analysis.

Clustering

For hierarchical clustering, we used the ward.D aggregation method of the hclust algorithm in R, with the Euclidean distance as the dissimilarity measure. To maximize the robustness of the clustering toward overfitting, we ran the ConsensusClusterPlus algorithm using 1000 repetitions with subsets obtained by sampling 80% of the samples [proportion of items (pI), 0.8] and keeping all the features [proportion of features (pF), 1]. Both inner and final clustering included in this method were realized with ward.D linkage and Euclidean distance.

Identification of dysregulated lncRNAs

Because breast tumors are highly heterogeneous, classical *t* tests cannot identify dysregulated lncRNAs in a small subset of breast tumors.

Therefore, the method was adapted to identify lncRNAs that were dysregulated in at least 10% of the breast tumors (22). This ensured selection of a reasonable number of dysregulated lncRNAs with a potentially informative variance related to the heterogeneity of breast cancer. To find significantly up-regulated genes, we explored the upper tail of the expression distribution by selecting the 10% of samples (normal or cancerous) with the highest expression. We then used a nonparametric Mann-Whitney test to compare expression levels between the normal samples and tumor samples. We extended this approach to a growing proportion of samples (15, 20, 25, 30, 35, 40, and 45%), and all lncRNA genes additionally found to be differentially expressed were retained. In parallel, we identified down-regulated lncRNAs by applying the same methodology to the lower tail of the expression distribution. The 215 lncRNA genes identified as dysregulated had an FDR of < 0.05 and a fold change of > 1.5 in at least 10% of the samples.

Prediction of lncRNA functions by the guilt-by-association approach

This approach is based on establishing correlations between the expression of lncRNA genes and that of protein-coding genes known to be involved in particular functions (gene sets). It enabled us to generate hypotheses regarding the function(s) of a given lncRNA. Publicly available gene sets were selected from both the KEGG (Kyoto Encyclopedia of Genes and Genomes) gene set [Molecular Signatures Database (MsigDB)] and the bdfunc.enrichment.human database of the sRAP library. Following the guidelines of the GSEA (Gene Set Enrichment Analysis) software, we grouped gene sets that contain redundant genes as follows: (i) we computed a between-gene-set distance matrix using the overlap distance (defined as the number of common genes/number of genes that compose the smallest gene set); (ii) we performed hierarchical clustering on the basis of this matrix (complete linkage); and (iii) we used a threshold level of 0.5 to cut the tree and group gene sets that belong to the same cluster.

We chose to focus only on the potential functions of dysregulated, ER- and subtype-specific lncRNAs because the guilt-by-association approach is computation-consuming. First, we randomly divided our breast tumor expression data into two data sets, one with 411 samples and one with 412 samples. For each data set, we computed a Pearson matrix of correlation between each lncRNA and each coding gene to produce two matrices of 236 lncRNAs \times 16,951 mRNAs. In every matrix, the protein-coding genes were ranked for each lncRNA on the basis of the correlation coefficient. The GSEA software (parameters, 1000 permutations on gene sets; minimum size, 15; maximum size, 500) was used to calculate a running sum statistic, corresponding to the enrichment score, on the basis of the ranks of the investigated gene set members, relative to those of nonmembers. We thus obtained two matrices containing an enrichment score and a statistical family-wise error rate (FWER) value for each lncRNA/gene set pair (236 lncRNAs \times 422 gene sets). To obtain high-confidence associations of lncRNAs with functions, we finally selected gene sets that were statistically (FWER < 0.05) associated with an lncRNA in both matrices and computed the mean of their enrichment scores.

To relate clusters of lncRNAs to gene sets, we started by computing, for each cluster, a metascore of enrichment for a gene set. This metascore was defined as the weighted sum of the enrichment scores obtained for the members of the lncRNA cluster, with the weight defined as -1 if the lncRNA was repressed in the condition of interest and as 1 if otherwise. Then, 10,000 random groups of lncRNAs of the same

size as the lncRNA cluster of interest were generated by random selection. For each group, a metascore was generated using the same weighted sum approach. Finally, the P value of the metascore was defined as the proportion of random metascores being at least as high (low) as the metascore of the positively (negatively) associated cluster.

Signature validation

The different signatures were validated on three cohorts using the following protocol. First, one should note that for the validation with the TCGA data, the analysis was restricted to lncRNAs common to the two platforms. This could lead to a reduction of signature size. In this case, the size of this “between-platform signature” is specified in the main text. The data were scaled to make the two data sets more comparable. For each signature, a metagene was then defined for the discovery cohort, as the first component of a principal components analysis (PCA). A receiver operating characteristic (ROC) curve was generated from that metagene, and the threshold that maximizes the Youden index was selected. A metagene was then generated from the validation data using the first component eigenvector from the discovery cohort. For visualization purposes, a ROC curve was generated using this metagene. The threshold selected for class prediction was the one obtained for the discovery cohort to avoid any bias. The prediction accuracy of the metagene was assessed using the balanced error rate (BER) metric. Finally, to evaluate the significance of signature performance, we generated 10,000 random signatures of the same size as the real between-platform signature by randomly sampling lncRNAs. For each random signature, the same process that was used for the real signature was applied. A P value was defined as the proportion of random signatures that show a BER lower than or equal to the real signature, and signatures with a P value lower than 0.05 were considered significant.

Determination of subtype-specific lncRNA signatures

For each subtype, we first selected lncRNA genes that show significant differential expression ($FDR < 0.05$; fold change > 1.5 , t test) in one particular subtype versus the three others. We then filtered out lncRNAs that were up-regulated (or down-regulated) in more than one subtype to obtain a list of lncRNAs characteristic of each subtype.

Relapse-free survival analysis

The analysis was computed in R using the “survival” library. The prognostic value of individual lncRNAs was estimated by univariate Cox regression. In parallel, multivariate Cox regression was used to test the independent prognostic values of lncRNAs, using clinical properties as covariates. Univariate analysis was used to select covariates that were prognostic. The proportional hazard assumption was verified with the “cox.zph” function (threshold, 0.01). For all analyses, $P < 0.05$ was the criterion of statistical significance. For visualization purposes, we also generated Kaplan-Meier curves using the “survfit” function. To define a high- and low-level group, we assessed all possible thresholds that lead to groups that represent at least 10% of the data set using the log-rank test. The threshold that leads to the lowest P value was selected as the final cutoff for group definition.

A set of 27 lncRNAs associated to relapse-free survival has been validated using a metagene-based approach: a metagene was defined for the discovery cohort as the first component of a PCA. The metagene was then generated from the validation data, using the PCA eigenvector from the discovery cohort. The association of this metagene with relapse-free survival was assessed on the validation cohort using a Cox model.

The significance of this association can be visualized using forest plot representation. Because random signatures can be associated to survival (74), we assessed whether the association of our metagene on the validation cohort is significantly better than metagenes obtained from randomly selected lncRNAs. Therefore, we generated 10,000 random signatures of the same size as the real relapse-free associated signature by randomly sampling lncRNAs. For each random signature, the same process of computing a Cox model from the first eigenvector from the discovery cohort was applied. Finally, we called “validation P value” the proportion of random signatures that show a Cox P value lower than or equal to the real signature. The signature was considered as significant if the validation P value was lower than 0.05.

Culture of breast cell lines and silencing of target lncRNAs

MDA-MB-231 cells were grown in Dulbecco’s modified Eagle’s medium (Gibco) supplemented with 10% FBS (Gibco). They were maintained at 37°C in 5% CO₂. To silence target lncRNAs, cells were treated with LNA GapmeRs (Exiqon), according to the manufacturer’s instructions. Briefly, 300,000 cells were transfected in six-well plates with 30 nM LNA GapmeRs and 5 µl of Lipofectamine 2000 in 2-ml total volume and were incubated for 24 hours. Staining of F-actin was performed following the manufacturer’s protocol (<http://cytoskeleton.com>).

RNA purification and RT-qPCR

RNA purification was performed with the RNeasy kit (Qiagen) according to the manufacturer’s instructions. Deoxyribonuclease (DNase) treatment was performed with a DNA-free DNase kit (Ambion) according to the manufacturer’s protocol. qPCRs were performed with SYBR Green dye (Eurogentec) in LightCycler 480 (Roche). Briefly, complementary DNA was obtained by reverse transcription of 1 µg of RNA, with random hexamers (Amersham/Pharmacia Biotech) and SuperScript II Reverse Transcriptase (Life Technologies Inc.). The results were normalized against the following housekeeping genes: SDHA, GAPDH, and ACTIN.

The following LNA gapmer sequences were used: CYTOR LNA, 5'-TCATAGAGCTTCCTGT-3'.

The following qPCR assay primer sequences were used: CYTOR, 5'-CTGGATGGTTCGCTGCTTTTT-3' (forward) and 5'-GATCTGAA-GACAGGCACGGG-3' (reverse); SMYD3, 5'-TACTGCGAGCAGTCC-GAGACA-3' (forward) and 5'-TTGTCCTGGGTTTGCAACGGA-3' (reverse); GOLPH3, 5'-CTAGAGGCTTGTGGAATGAGACG-3' (forward) and 5'-GACCGTTTCTGGAGGCTGAGTT-3' (reverse); KIF14, 5'-GCACCTTCGGAACAAGCAAACCA-3' (forward) and 5'-ATGT-TGCTGGCAGCGGGACTAA-3' (reverse); mTOR, 5'-AGCATCG-GATGCTTAGGAGTGG-3' (forward) and 5'-CAGCCAGTCATCTTG-GAGACC-3' (reverse); PRKCA, 5'-GCCTATGGCGTCCTGTTGTATG-3' (forward) and 5'-GAAACAGCCTCCTTGGACAAGG-3' (reverse).

Cell proliferation/migration

To evaluate breast cancer cell proliferation, MDA-MB-231 cells (10,000 per well) transfected with LNA gapmers were seeded into the xCELLigence E-Plate 16 (Roche) 24 hours after transfection, according to the manufacturer’s instructions. In this system, the electrical impedance is used to derive a cell index, which gives a real-time representation of cell growth when continuously monitored. Measurements were automatically collected by the RTCA Dual Plate analyzer every 30 min for up to 3 days. The data were analyzed with the RTCA software.

To examine breast cancer cell migration, MDA-MB-231 cells transfected with LNA gapmers were seeded into the xCELLigence CIM-Plate 16 (Roche) 24 hours after transfection. Briefly, a 165- μ l volume of fresh medium containing 10% FBS (chemoattractant) or serum-free medium (control) was added to the lower chambers of the CIM-Plate 16. The upper chambers were filled with serum-free medium (30 μ l per well), and the plate was incubated in 5% CO₂ for 1 hour at 37°C. Cells (20,000 per well) were then added to each well of the upper chamber, and cell migration was assessed at 30-min intervals for 18 hours at 37°C in 5% CO₂. Upon migration, cells adhere to the surface of the filter electrode and increase the impedance.

Cell cycle analysis by flow cytometry

The distribution of cells through the various phases of the cell cycle was determined by measuring the DNA content with BD BrdU Flow kits (catalog no. 552598) according to the manufacturer's instructions. Briefly, cells were incubated with 10 μ M BrdU for 2 hours. They were then fixed with paraformaldehyde, permeabilized with saponin, and treated with DNase to expose the BrdU epitope. BrdU was stained with fluorescent anti-BrdU antibodies. The total DNA level was assessed by staining with 7-AAD. Data were analyzed with the Kaluza Analysis Software (Beckman Coulter).

Illumina Expression HT12 arrays

Total RNA (200 ng) was amplified with the Illumina TotalPrep RNA Amplification Kit (Ambion) and then hybridized with the array according to the Whole-Genome Gene Expression Direct Hybridization Assay (Illumina). Chips were scanned with the HiScan Reader (Illumina).

The raw data were normalized using the quantile normalization method from the lumi package (75), and batch effect was corrected with the ComBat algorithm of the sva library with default parameters, using the CYTOR status (LNA-silenced versus control) as covariate. Then, to identify probes that are differentially expressed, a *t* test was applied, and the *P* values obtained were corrected for multitesting using the Benjamini-Hochberg method. The probes that simultaneously show an FDR of <0.05 and an absolute fold change between the median of each group higher than 1.5 were reported as differentially expressed if the median expression value in each group was significantly higher than the background level (detection *P* value < 0.05). The raw data have been uploaded in GEO database and are accessible under accession no. GSE77491.

Statistical analysis

Data values were expressed as means \pm SD of at least two independent experiments and evaluated using Student's *t* test for unpaired samples, or otherwise specified. Mean differences were considered significant when *P* < 0.05, or otherwise specified.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/2/9/e1600220/DC1>

fig. S1. Dysregulated lncRNAs in breast tumors.

fig. S2. Validation of the ER-related lncRNA signature in three data sets.

fig. S3. Validation of the subtype-specific lncRNA signature in three data sets.

fig. S4. Validation of the set of 27 lncRNAs predictive of relapse and dysregulated in breast cancer.

fig. S5. Methylation of the CYTOR gene in relation to its expression.

fig. S6. Impact of CYTOR depletion on cell cycle and gene expression.

table S1. Clinical annotation of the 823 primary tumors and the 172 normal tissue samples from the discovery cohort, including age, size, lymph node status, ER status, HER2 status, grade, PAM50-associated subtype, and relapse information.

table S2. Reannotation of the Affymetrix Human Genome U133 Plus 2.0 array.

table S3. Description of the 215 dysregulated lncRNAs in breast cancer.

table S4. Contingency table of the lncRNA-related clusters that correlate with ER status.

table S5. Significant enrichment scores from the guilt-by-association analysis.

table S6. Contingency table of the lncRNA-related clusters that correlate with the known molecular subtypes.

table S7. lncRNAs signatures of the known molecular subtypes of breast cancer.

table S8. Survival analysis: Univariate results.

table S9. Survival analysis: Multivariate results.

REFERENCES AND NOTES

- J. Ferlay, I. Soerjomataram, M. Ervik, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, D. M. Parkin, D. Forman, F. Bray, *GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11* (International Agency for Research on Cancer, Lyon, 2012).
- C. M. Perou, T. Sørli, M. B. Eisen, M. van de Rijn, S. S. Jeffrey, C. A. Rees, J. R. Pollack, D. T. Ross, H. Johnsen, L. A. Akslen, Ø. Fluge, A. Pergamenschikov, C. Williams, S. X. Zhu, P. E. Lønning, A.-L. Børresen-Dale, P. O. Brown, D. Botstein, Molecular portraits of human breast tumours. *Nature* **406**, 747–752 (2000).
- E. Senkus, S. Kyriakides, S. Ohno, F. Penault-Llorca, P. Poortmans, E. Rutgers, S. Zackrisson, F. Cardoso, Primary breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann. Oncol.* **26** (Suppl. 5), v8–v30 (2015).
- The Cancer Genome Atlas Network, Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70 (2012).
- J. S. Reis-Filho, L. Pusztai, Gene expression profiling in breast cancer: Classification, prognostication, and prediction. *Lancet* **378**, 1812–1823 (2011).
- L. Pusztai, R. Rouzier, W. F. Symmans, *CCR 20th anniversary commentary: Divide and conquer—Breast cancer subtypes and response to therapy.* *Clin. Cancer Res.* **21**, 3575–3577 (2015).
- R. A. Gupta, N. Shah, K. C. Wang, J. Kim, H. M. Horlings, D. J. Wong, M.-C. Tsai, T. Hung, P. Argani, J. L. Rinn, Y. Wang, P. Brzoska, B. Kong, R. Li, R. B. West, M. J. van de Vijver, S. Sukumar, H. Y. Chang, Long non-coding RNA *HOTAIR* reprograms chromatin state to promote cancer metastasis. *Nature* **464**, 1071–1076 (2010).
- J. M. Silva, N. J. Boczek, M. W. Berres, X. Ma, D. I. Smith, *LSINCT5* is over expressed in breast and ovarian cancer and affects cellular proliferation. *RNA Biol.* **8**, 496–505 (2011).
- T. Derrien, R. Johnson, G. Bussotti, A. Tanzer, S. Djebali, H. Tilgner, G. Guernec, D. Martin, A. Merkel, D. G. Knowles, J. Lagarde, L. Veeravalli, X. Ruan, Y. Ruan, T. Lassmann, P. Carninci, J. B. Brown, L. Lipovich, J. M. Gonzalez, M. Thomas, C. A. Davis, R. Shiekhattar, T. R. Gingeras, T. J. Hubbard, C. Notredame, J. Harrow, R. Guigó, The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Res.* **22**, 1775–1789 (2012).
- T. Gutschner, S. Diederichs, The hallmarks of cancer: A long non-coding RNA point of view. *RNA Biol.* **9**, 703–719 (2012).
- M. Esteller, Non-coding RNAs in human disease. *Nat. Rev. Genet.* **12**, 861–874 (2011).
- M. Huarte, M. Guttman, D. Feldser, M. Garber, M. J. Koziol, D. Kenzelmann-Broz, A. M. Khalil, O. Zuk, I. Amit, M. Rabani, L. D. Attardi, A. Regev, E. S. Lander, T. Jacks, J. L. Rinn, A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* **142**, 409–419 (2010).
- S. W. Cheetham, F. Gruhl, J. S. Mattick, M. E. Dinger, Long noncoding RNAs and the genetics of cancer. *Br. J. Cancer* **108**, 2419–2425 (2013).
- X. Su, G. G. Malouf, Y. Chen, J. Zhang, H. Yao, V. Valero, J. N. Weinstein, J.-P. Spano, F. Meric-Bernstam, D. Khayat, F. J. Esteva, Comprehensive analysis of long non-coding RNAs in human breast cancer clinical subtypes. *Oncotarget* **5**, 9864–9876 (2014).
- W. Zhao, J. Luo, S. Jiao, Comprehensive characterization of cancer subtype associated long non-coding RNAs and their clinical implications. *Sci. Rep.* **4**, 6591 (2014).
- L. M. McIntyre, K. K. Lopiano, A. M. Morse, V. Amin, A. L. Oberg, L. J. Young, S. V. Nuzhdin, RNA-seq: Technical variability and sampling. *BMC Genomics* **12**, 293 (2011).
- N. Raghavachari, J. Barb, Y. Yang, P. Liu, K. Woodhouse, D. Levy, C. J. O'Donnell, P. J. Munson, G. J. Kato, A systematic comparison and evaluation of high density exon arrays and RNA-seq technology used to unravel the peripheral blood transcriptome of sickle cell disease. *BMC Med. Genomics* **5**, 28 (2012).
- L. Jiang, F. Schlesinger, C. A. Davis, Y. Zhang, R. Li, M. Salit, T. R. Gingeras, B. Oliver, Synthetic spike-in standards for RNA-seq experiments. *Genome Res.* **21**, 1543–1551 (2011).
- R. Edgar, M. Domrachev, A. E. Lash, Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207–210 (2002).
- S. Monti, P. Tamayo, J. Mesirov, T. Golub, Consensus clustering: A resampling-based method for class discovery and visualization of gene expression microarray data. *Mach. Learn.* **52**, 91–118 (2003).
- M. D. Wilkerson, D. N. Hayes, ConsensusClusterPlus: A class discovery tool with confidence assessments and item tracking. *Bioinformatics* **26**, 1572–1573 (2010).

22. L. Yao, H. Shen, P. W. Laird, P. J. Farnham, B. P. Berman, Inferring regulatory element landscapes and transcription factor networks from cancer methylomes. *Genome Biol.* **16**, 105 (2015).
23. M. Rahman, L. K. Jackson, W. E. Johnson, D. Y. Li, A. H. Bild, S. R. Piccolo, Alternative pre-processing of RNA-sequencing data in The Cancer Genome Atlas leads to improved analysis results. *Bioinformatics* **31**, 3666–3672 (2015).
24. D. Chakravarty, A. Sboner, S. S. Nair, E. Giannopoulou, R. Li, S. Hennig, J. M. Mosquera, J. Pauwels, K. Park, M. Kossai, T. Y. MacDonald, J. Fontugne, N. Erho, I. A. Vergara, M. Ghadessi, E. Davicioni, R. B. Jenkins, N. Palanisamy, Z. Chen, S. Nakagawa, T. Hirose, N. H. Bander, H. Beltran, A. H. Fox, O. Elemento, M. A. Rubin, The oestrogen receptor α -regulated lncRNA NEAT1 is a critical modulator of prostate cancer. *Nat. Commun.* **5**, 5383 (2014).
25. H. Hansji, E. Y. Leung, B. C. Baguley, G. J. Finlay, M. E. Askarian-Amiri, Keeping abreast with long non-coding RNAs in mammary gland development and breast cancer. *Front. Genet.* **5**, 379 (2014).
26. N. C. Turner, J. S. Reis-Filho, A. M. Russell, R. J. Springall, K. Ryder, D. Steele, K. Savage, C. E. Gillett, F. C. Schmitt, A. Ashworth, A. N. Tutt, BRCA1 dysfunction in sporadic basal-like breast cancer. *Oncogene* **26**, 2126–2132 (2007).
27. J. L. Rinn, H. Y. Chang, Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* **81**, 145–166 (2012).
28. M. F. Montenegro, M. d. M. Collado-González, M. P. Fernández-Pérez, M. B. Hammouda, L. Tolordava, M. Gamkrelidze, J. N. Rodríguez-López, Promoting E2F1-mediated apoptosis in oestrogen receptor- α -negative breast cancer cells. *BMC Cancer* **14**, 539 (2014).
29. X. Jiang, D. J. Shapiro, The immune system and inflammation in breast cancer. *Mol. Cell. Endocrinol.* **382**, 673–682 (2014).
30. C. Sweeney, P. S. Bernard, R. E. Factor, M. L. Kwan, L. A. Habel, C. P. Quesenberry, K. Shakespear, E. K. Weltzien, I. J. Stijleman, C. A. Davis, M. T. W. Ebbert, A. Castillo, L. H. Kushi, B. J. Caan, Intrinsic subtypes from PAM50 gene expression assay in a population-based breast cancer cohort: Differences by age, race, and tumor characteristics. *Cancer Epidemiol. Biomarkers Prev.* **23**, 714–724 (2014).
31. L. Zheng, J. Q. Ren, H. Li, Z. L. Kong, H. G. Zhu, Downregulation of wild-type p53 protein by HER-2/neu mediated PI3K pathway activation in human breast cancer cells: Its effect on cell proliferation and implication for therapy. *Cell Res.* **14**, 497–506 (2004).
32. S. Grivnikov, M. Karin, Autocrine IL-6 signaling: A key event in tumorigenesis? *Cancer Cell* **13**, 7–9 (2008).
33. Z. Du, T. Fei, R. G. W. Verhaak, Z. Su, Y. Zhang, M. Brown, Y. Chen, X. S. Liu, Integrative genomic analyses reveal clinically relevant long noncoding RNAs in human cancer. *Nat. Struct. Mol. Biol.* **20**, 908–913 (2013).
34. M. W. Wright, A short guide to long non-coding RNA gene nomenclature. *Hum. Genomics* **8**, 7 (2014).
35. Q. Pang, J. Ge, Y. Shao, W. Sun, H. Song, T. Xia, B. Xiao, J. Guo, Increased expression of long intergenic non-coding RNA LINC00152 in gastric cancer and its clinical significance. *Tumour Biol.* **35**, 5441–5447 (2014).
36. M. K. Iyer, Y. S. Niknafs, R. Malik, U. Singhal, A. Sahu, Y. Hosono, T. R. Barrette, J. R. Prensner, J. R. Evans, S. Zhao, A. Poliakov, X. Cao, S. M. Dhanasekaran, Y.-M. Wu, D. R. Robinson, D. G. Beer, F. Y. Feng, H. K. Iyer, A. M. Chinnaiyan, The landscape of long noncoding RNAs in the human transcriptome. *Nat. Genet.* **47**, 199–208 (2015).
37. A. C. Marques, J. Hughes, B. Graham, M. S. Kowalczyk, D. R. Higgs, C. P. Ponting, Chromatin signatures at transcriptional start sites separate two equally populated yet distinct classes of intergenic long noncoding RNAs. *Genome Biol.* **14**, R131 (2013).
38. S. M. Singel, C. Cornelius, E. Zaganjor, K. Batten, V. R. Sarode, D. L. Buckley, Y. Peng, G. B. John, H. C. Li, N. Sadeghi, W. E. Wright, L. Lum, T. W. Corson, J. W. Shay, KIF14 promotes AKT phosphorylation and contributes to chemoresistance in triple-negative breast cancer. *Neoplasia* **16**, 247–256.e2 (2014).
39. Z. Zeng, H. Lin, X. Zhao, G. Liu, X. Wang, R. Xu, K. Chen, J. Li, L. Song, Overexpression of GOLPH3 promotes proliferation and tumorigenicity in breast cancer via suppression of the FOXO1 transcription factor. *Clin. Cancer Res.* **18**, 4059–4069 (2012).
40. R. Keil, J. Schulz, M. Hatzfeld, p0071/PP4, a multifunctional protein coordinating cell adhesion with cytoskeletal organization. *Biol. Chem.* **394**, 1005–1017 (2013).
41. T. Wakatsuki, R. B. Wysolmerski, E. L. Elson, Mechanics of cell spreading: Role of myosin II. *J. Cell Sci.* **116**, 1617–1625 (2003).
42. D. D. Sarbassov, S. M. Ali, D.-H. Kim, D. A. Guertin, R. R. Latek, H. Erdjument-Bromage, P. Tempst, D. M. Sabatini, Rictor, a novel binding partner of mTOR, defines a rapamycin-insensitive and raptor-independent pathway that regulates the cytoskeleton. *Curr. Biol.* **14**, 1296–1302 (2004).
43. P. Jonsson, C. Coarfa, F. Mesmar, T. Raz, K. Rajapakshe, J. F. Thompson, P. H. Gunaratne, C. Williams, Single-molecule sequencing reveals estrogen-regulated clinically relevant lncRNAs in breast cancer. *Mol. Endocrinol.* **29**, 1634–1645 (2015).
44. V. Miano, G. Ferrero, S. Reineri, L. Caizzi, L. Annaratone, L. Ricci, S. Cutrupi, I. Castellano, F. Cordero, M. De Bortoli, Luminal long non-coding RNAs regulated by estrogen receptor α in a ligand-independent manner show functional roles in breast cancer. *Oncotarget* **7**, 3201–3216 (2016).
45. Y.-Y. Tseng, B. S. Moriarity, W. Gong, R. Akiyama, A. Tiwari, H. Kawakami, P. Ronning, B. Reuland, K. Guenther, T. C. Beadnell, J. Essig, G. M. Otto, M. G. O'Sullivan, D. A. Largaespada, K. L. Schwertfeger, Y. Marahrens, Y. Kawakami, A. Bagchi, PVT1 dependence in cancer with MYC copy-number increase. *Nature* **512**, 82–86 (2014).
46. G. Doose, A. Haake, S. H. Bernhart, C. López, S. Duggimpudi, F. Wojciech, A. K. Bergmann, A. Borkhardt, B. Burkhardt, A. Claviez, L. Dimitrova, S. Haas, J. I. Hoell, M. Hummel, D. Karsch, W. Klapper, K. Kleo, H. Kretzmer, M. Kreuz, R. Küppers, C. Lawerenz, D. Lenze, M. Loeffler, L. Mantovani-Löffler, P. Möller, G. Ott, J. Richter, M. Rohde, P. Rosenstiel, A. Rosenwald, M. Schilhabel, M. Schneider, I. Scholz, S. Stilgenbauer, H. G. Stunnenberg, M. Szczepanowski, L. Trümper, M. A. Weniger; ICGC MMML-Seq Consortium, S. Hoffmann, R. Siebert, I. Iaccarino, MINCR is a MYC-induced lncRNA able to modulate MYC's transcriptional network in Burkitt lymphoma cells. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E5261–E5270 (2015).
47. E. M. Ciruelos Gil, Targeting the PI3K/AKT/mTOR pathway in estrogen receptor-positive breast cancer. *Cancer Treat. Rev.* **40**, 862–871 (2014).
48. M. C. Louie, A. McClellan, S. Siewit, L. Kawabata, Estrogen receptor regulates E2F1 expression to mediate tamoxifen resistance. *Mol. Cancer Res.* **8**, 343–352 (2010).
49. V. Vuaroqueaux, P. Urban, M. Labuhn, M. Delorenzi, P. Wirapati, C. C. Benz, R. Flury, H. Dieterich, F. Spyrtatos, U. Eppenberger, S. Eppenberger-Castori, Low E2F1 transcript levels are a strong determinant of favorable breast cancer outcome. *Breast Cancer Res.* **9**, R33 (2007).
50. E. M. Curran, B. M. Judy, N. A. Duru, H.-Q. Wang, L. A. Vergara, D. B. Lubahn, D. M. Estes, Estrogenic regulation of host immunity against an estrogen receptor-negative human breast cancer. *Clin. Cancer Res.* **12**, 5641–5647 (2006).
51. A. A. Mostafa, D. Codner, K. Hirasawa, Y. Komatsu, M. N. Young, V. Steimle, S. Drover, Activation of ER α signaling differentially modulates IFN- γ induced HLA-class II expression in breast cancer cells. *PLoS One* **9**, e87377 (2014).
52. N. J. Sullivan, A. K. Sasser, A. E. Axel, F. Vesuna, V. Raman, N. Ramirez, T. M. Oberyszyn, B. M. Hall, Interleukin-6 induces an epithelial-mesenchymal transition phenotype in human breast cancer cells. *Oncogene* **28**, 2940–2947 (2009).
53. Z. T. Schafer, J. S. Brugge, IL-6 involvement in epithelial cancers. *J. Clin. Invest.* **117**, 3660–3663 (2007).
54. N. Berteaux, N. Aptel, G. Cathala, C. Genton, J. Coll, A. Daccache, N. Spruyt, H. Hondemarcq, T. Dugimont, J.-J. Cury, T. Forné, E. Adriaenssens, A novel H19 antisense RNA overexpressed in breast cancer contributes to paternal IGF2 expression. *Mol. Cell. Biol.* **28**, 6731–6745 (2008).
55. E. Adriaenssens, L. Dumont, S. Lottin, D. Bolle, A. Leprêtre, A. Delobelle, F. Bouali, T. Dugimont, J. Coll, J.-J. Cury, H19 overexpression in breast adenocarcinoma stromal cells is associated with tumor values and steroid receptor status but independent of p53 and Ki-67 expression. *Am. J. Pathol.* **153**, 1597–1607 (1998).
56. L. Zhang, F. Yang, J.-h. Yuan, S.-x. Yuan, W.-p. Zhou, X.-s. Huo, D. Xu, H.-s. Bi, F. Wang, S.-h. Sun, Epigenetic activation of the miR-200 family contributes to H19-mediated metastasis suppression in hepatocellular carcinoma. *Carcinogenesis* **34**, 577–586 (2013).
57. R. Bergström, K. Savary, A. Morén, S. Guibert, C.-H. Heldin, R. Ohlsson, A. Moustakas, Transforming growth factor β promotes complexes between Smad proteins and the CCCTC-binding factor on the H19 imprinting control region chromatin. *J. Biol. Chem.* **285**, 19727–19737 (2010).
58. I. J. Matouk, E. Raveh, R. Abu-lail, S. Mezan, M. Gilon, E. Gershtain, T. Birman, J. Gallula, T. Schneider, M. Barkali, C. Richler, Y. Fellig, V. Sorin, A. Hubert, A. Hochberg, A. Czerniak, Oncofetal H19 RNA promotes tumor metastasis. *Biochim. Biophys. Acta* **1843**, 1414–1426 (2014).
59. C. Chen, Z. Li, Y. Yang, T. Xiang, W. Song, S. Liu, Microarray expression profiling of dysregulated long non-coding RNAs in triple-negative breast cancer. *Cancer Biol. Ther.* **16**, 856–865 (2015).
60. X.-H. Liu, M. Sun, F.-q. Nie, Y.-b. Ge, E.-b. Zhang, D.-d. Yin, R. Kong, R. Xia, K.-h. Lu, J.-h. Li, W. De, K.-m. Wang, Z.-x. Wang, Lnc RNA HOTAIR functions as a competing endogenous RNA to regulate HER2 expression by sponging miR-331-3p in gastric cancer. *Mol. Cancer* **13**, 92 (2014).
61. G. Yang, S. Zhang, F. Gao, Z. Liu, M. Lu, S. Peng, T. Zhang, F. Zhang, Osteopontin enhances the expression of HOTAIR in cancer cells via IRF1. *Biochim. Biophys. Acta* **1839**, 837–848 (2014).
62. Z. Hui, M. Xianglin, Association of HOTAIR expression with PI3K/Akt pathway activation in adenocarcinoma of esophagogastric junction. *Open Med.* **11**, 36–40 (2016).
63. J. Zhou, X. Zhi, L. Wang, W. Wang, Z. Li, J. Tang, J. Wang, Q. Zhang, Z. Xu, Linc00152 promotes proliferation in gastric cancer through the EGFR-dependent pathway. *J. Exp. Clin. Cancer Res.* **34**, 135 (2015).
64. J. Ji, J. Tang, L. Deng, Y. Xie, R. Jiang, G. Li, B. Sun, LINC00152 promotes proliferation in hepatocellular carcinoma by targeting EpCAM via the mTOR signaling pathway. *Oncotarget* **6**, 42813–42824 (2015).
65. K. L. Scott, O. Kabbarah, M.-C. Liang, E. Ivanova, V. Anagnostou, J. Wu, S. Dhakal, M. Wu, S. Chen, T. Feinberg, J. Huang, A. Sacci, H. R. Widlund, D. E. Fisher, Y. Xiao, D. L. Rimm,

- A. Protopopov, K.-K. Wong, L. Chin, GOLPH3 modulates mTOR signalling and rapamycin sensitivity in cancer. *Nature* **459**, 1085–1090 (2009).
66. Y. Gökmen-Polar, I. T. Vladislav, Y. Neelamraju, S. C. Janga, S. Badve, Prognostic impact of HOTAIR expression is restricted to ER-negative breast cancers. *Sci. Rep.* **5**, 8765 (2015).
67. J. Sun, X. Chen, Z. Wang, M. Guo, H. Shi, X. Wang, L. Cheng, M. Zhou, A potential prognostic long non-coding RNA signature to predict metastasis-free survival of breast cancer patients. *Sci. Rep.* **5**, 16553 (2015).
68. K. P. Sørensen, M. Thomassen, Q. Tan, M. Bak, S. Cold, M. Burton, M. J. Larsen, T. A. Kruse, Long non-coding RNA expression profiles predict metastasis in lymph node-negative breast cancer independently of traditional prognostic markers. *Breast Cancer Res.* **17**, 55 (2015).
69. P.-J. Volders, K. Helsens, X. Wang, B. Menten, L. Martens, K. Gevaert, J. Vandesompele, P. Mestdag, LNCipedia: A database for annotated human lncRNA transcript sequences and structures. *Nucleic Acids Res.* **41**, D246–D251 (2013).
70. P. Stepniak, M. Maycock, K. Wojdan, M. Markowska, S. Perun, A. Srivastava, L. S. Wyrwicz, K. Świrski, Microarray Inspector: Tissue cross contamination detection tool for microarray data. *Acta Biochim. Pol.* **60**, 647–655 (2013).
71. J. S. Parker, M. Mullins, M. C. U. Cheang, S. Leung, D. Voduc, T. Vickery, S. Davies, C. Fauron, X. He, Z. Hu, J. F. Quackenbush, I. J. Stijleman, J. Palazzo, J. S. Marron, A. B. Nobel, E. Mardis, T. O. Nielsen, M. J. Ellis, C. M. Perou, P. S. Bernard, Supervised risk predictor of breast cancer based on intrinsic subtypes. *J. Clin. Oncol.* **27**, 1160–1167 (2009).
72. C. W. Law, Y. Chen, W. Shi, G. K. Smyth, voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).
73. M. E. Ritchie, B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, G. K. Smyth, *limma* powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
74. D. Venet, J. E. Dumont, V. Detours, Most random gene expression signatures are significantly associated with breast cancer outcome. *PLOS Comput. Biol.* **7**, e1002240 (2011).
75. P. Du, W. A. Kibbe, S. M. Lin, *lumi*: A pipeline for processing Illumina microarray. *Bioinformatics* **24**, 1547–1548 (2008).

Acknowledgments: We thank J. Vandesompele, P. Mestdag, H. Duvillier, and S. Garaud for their helpful advice. **Funding:** O.V.G. was supported by the Télévie. M.B. was supported by the Télévie. M.D. was supported by the Walloon Region (WB Health grant, CanDx). E.J.d.B. was supported by the Belgian F.R.I.A. This work was funded by grants from the Fonds de la Recherche Scientifique and Télévie, as well as by grants from the IUAP P7/03 program, the Action de Recherche Concertée (AUWB-2010-2015 ULB-No 7), the Belgian “Foundation against Cancer,” the WB Health program, and the Fonds Gaston Ithier. **Author contributions:** O.V.G. and F.F. designed the experiments and interpreted the data. O.V.G., M.B., E.J.d.B., C.O., S.B., and M.D. performed the bioinformatics analysis. O.V.G., E.C., and P.P. performed the qPCR and proliferation/migration experiments. M.G., G.B., and C.S. provided technical support and advices. F.F. directed the study, and O.V.G. and F.F. wrote the article. **Competing interests:** G.B., C.S., and F.F. are Université Libre de Bruxelles professors. The other authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Submitted 4 February 2016

Accepted 5 August 2016

Published 2 September 2016

10.1126/sciadv.1600220

Citation: O. Van Grembergen, M. Bizet, E. J. de Bony, E. Calonne, P. Putmans, S. Brohée, C. Olsen, M. Guo, G. Bontempi, C. Sotiriou, M. Defrance, F. Fuks, Portraying breast cancers with long noncoding RNAs. *Sci. Adv.* **2**, e1600220 (2016).

This article is published under a Creative Commons license. The specific license under which this article is published is noted on the first page.

For articles published under [CC BY](#) licenses, you may freely distribute, adapt, or reuse the article, including for commercial purposes, provided you give proper attribution.

For articles published under [CC BY-NC](#) licenses, you may distribute, adapt, or reuse the article for non-commercial purposes. Commercial use requires prior permission from the American Association for the Advancement of Science (AAAS). You may request permission by clicking [here](#).

The following resources related to this article are available online at <http://advances.sciencemag.org>. (This information is current as of September 29, 2016):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://advances.sciencemag.org/content/2/9/e1600220.full>

Supporting Online Material can be found at:

<http://advances.sciencemag.org/content/suppl/2016/08/29/2.9.e1600220.DC1>

This article **cites 74 articles**, 20 of which you can access for free at:

<http://advances.sciencemag.org/content/2/9/e1600220#BIBL>

Science Advances (ISSN 2375-2548) publishes new articles weekly. The journal is published by the American Association for the Advancement of Science (AAAS), 1200 New York Avenue NW, Washington, DC 20005. Copyright is held by the Authors unless stated otherwise. AAAS is the exclusive licensee. The title *Science Advances* is a registered trademark of AAAS

Supplementary Materials for

Portraying breast cancers with long noncoding RNAs

Olivier Van Grembergen, Martin Bizet, Eric J. de Bony, Emilie Calonne, Pascale Putmans, Sylvain Brohé, Catharina Olsen, Mingzhou Guo, Gianluca Bontempi, Christos Sotiriou, Matthieu Defrance, François Fuks

Published 2 September 2016, *Sci. Adv.* **2**, e1600220 (2016)

DOI: 10.1126/sciadv.1600220

The PDF file includes:

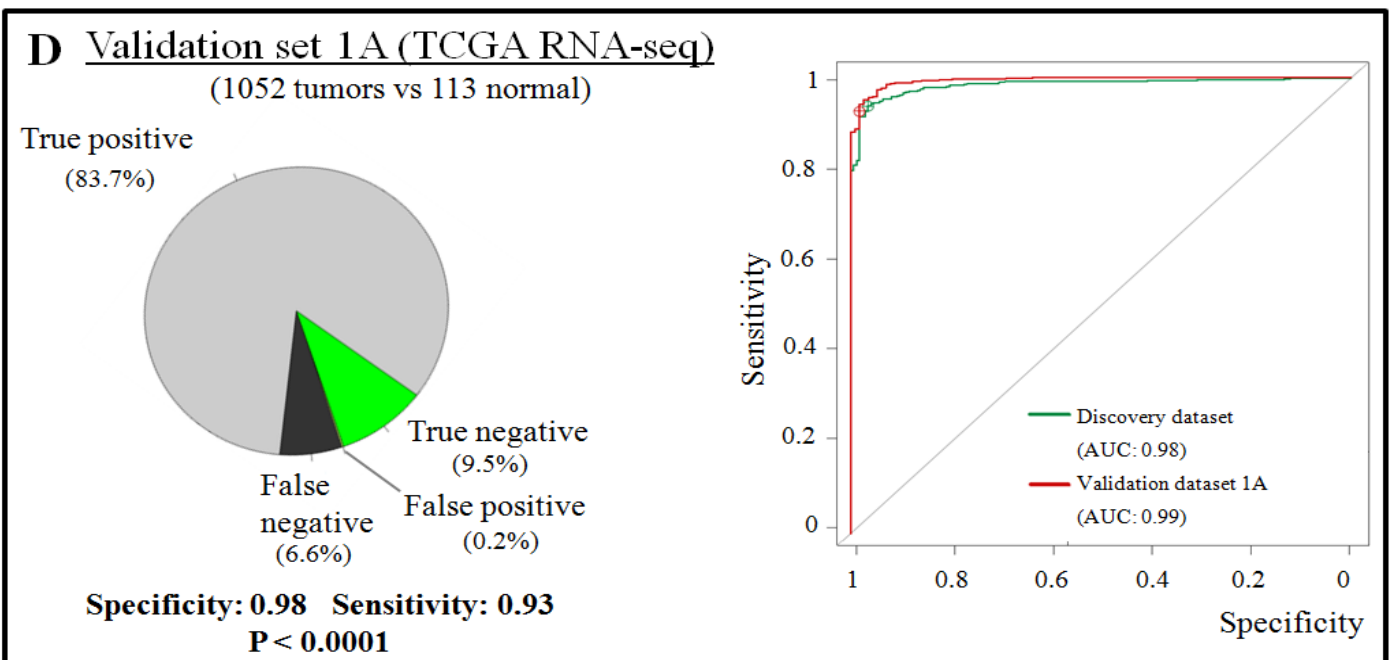
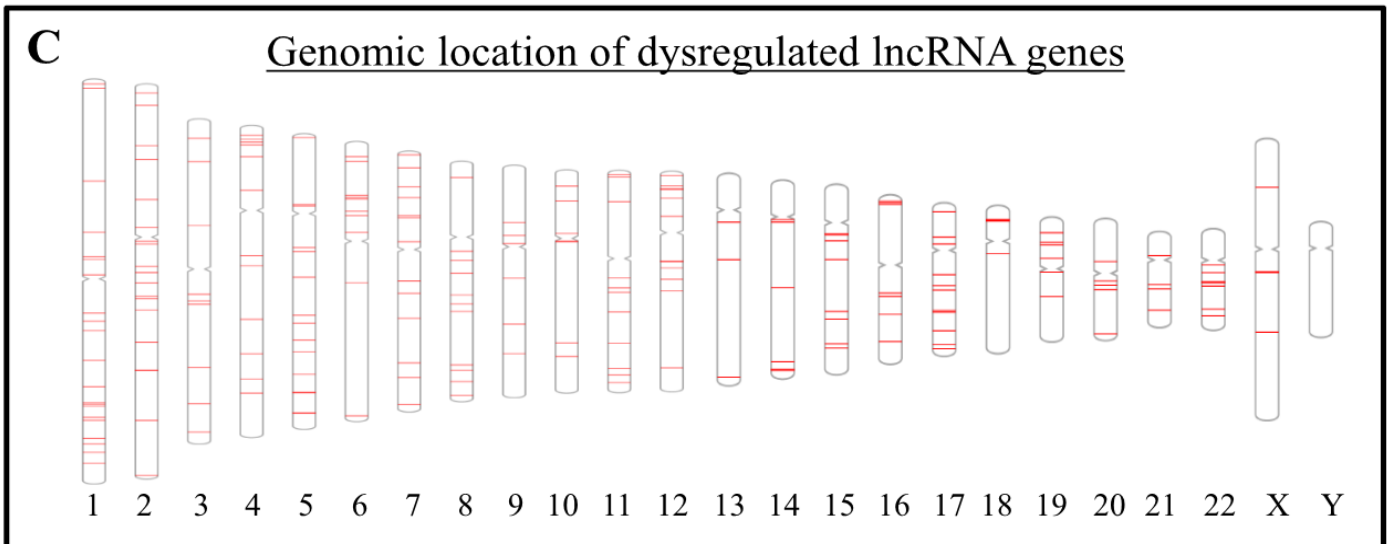
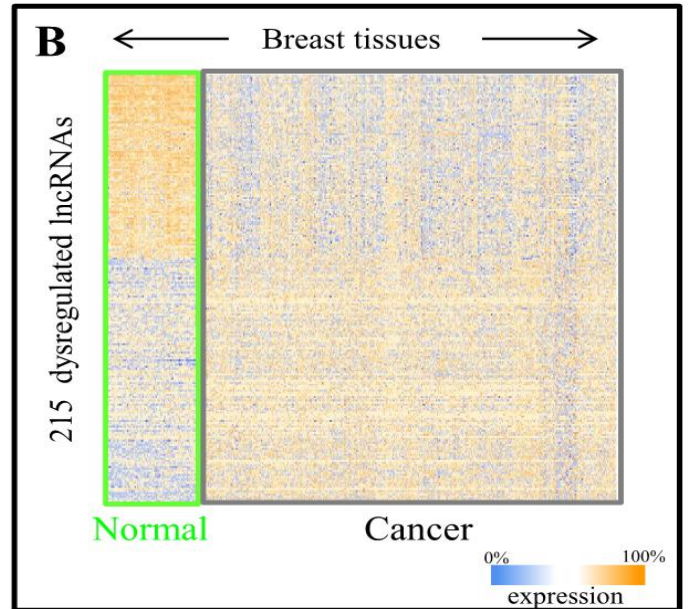
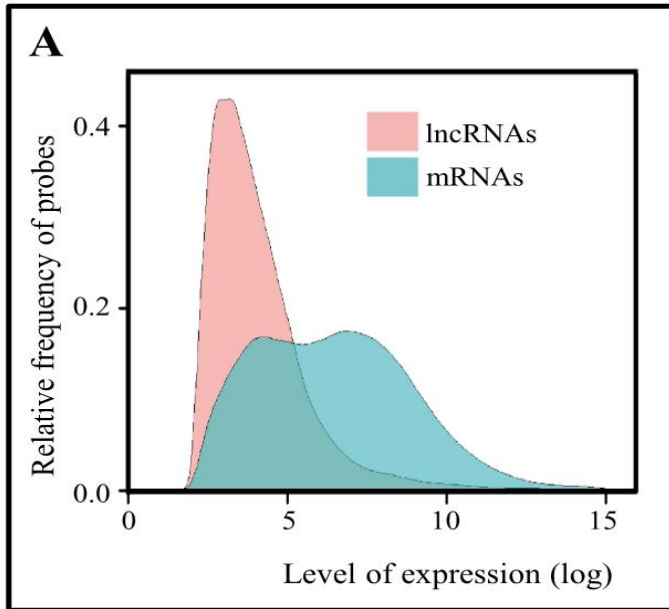
- fig. S1. Dysregulated lncRNAs in breast tumors.
- fig. S2. Validation of the ER-related lncRNA signature in three data sets.
- fig. S3. Validation of the subtype-specific lncRNA signature in three data sets.
- fig. S4. Validation of the set of 27 lncRNAs predictive of relapse and dysregulated in breast cancer.
- fig. S5. Methylation of the CYTOR gene in relation to its expression.
- fig. S6. Impact of CYTOR depletion on cell cycle and gene expression.
- table S4. Contingency table of the lncRNA-related clusters that correlate with ER status.
- table S6. Contingency tables of the lncRNA-related clusters that correlate with the known molecular subtypes.
- table S7. lncRNAs signatures of the known molecular subtypes of breast cancer.
- Legends for tables S1 to S3, S5, S8, and S9

Other Supplementary Material for this manuscript includes the following:

(available at advances.sciencemag.org/cgi/content/full/2/9/e1600220/DC1)

- table S1 (Microsoft Excel format). Clinical annotation of the 823 primary tumors and the 172 normal tissue samples from the discovery cohort, including age, size, lymph node status, ER status, HER2 status, grade, PAM50-associated subtype, and relapse information.
- table S2 (Microsoft Excel format). Reannotation of the Affymetrix Human Genome U133 Plus 2.0 array.
- table S3 (Microsoft Excel format). Description of the 215 dysregulated lncRNAs in breast cancer.

- table S5 (Microsoft Excel format). Significant enrichment scores from the guilt-by-association analysis.
- table S8 (Microsoft Excel format). Survival analysis: Univariate results.
- table S9 (Microsoft Excel format). Survival analysis: Multivariate results.



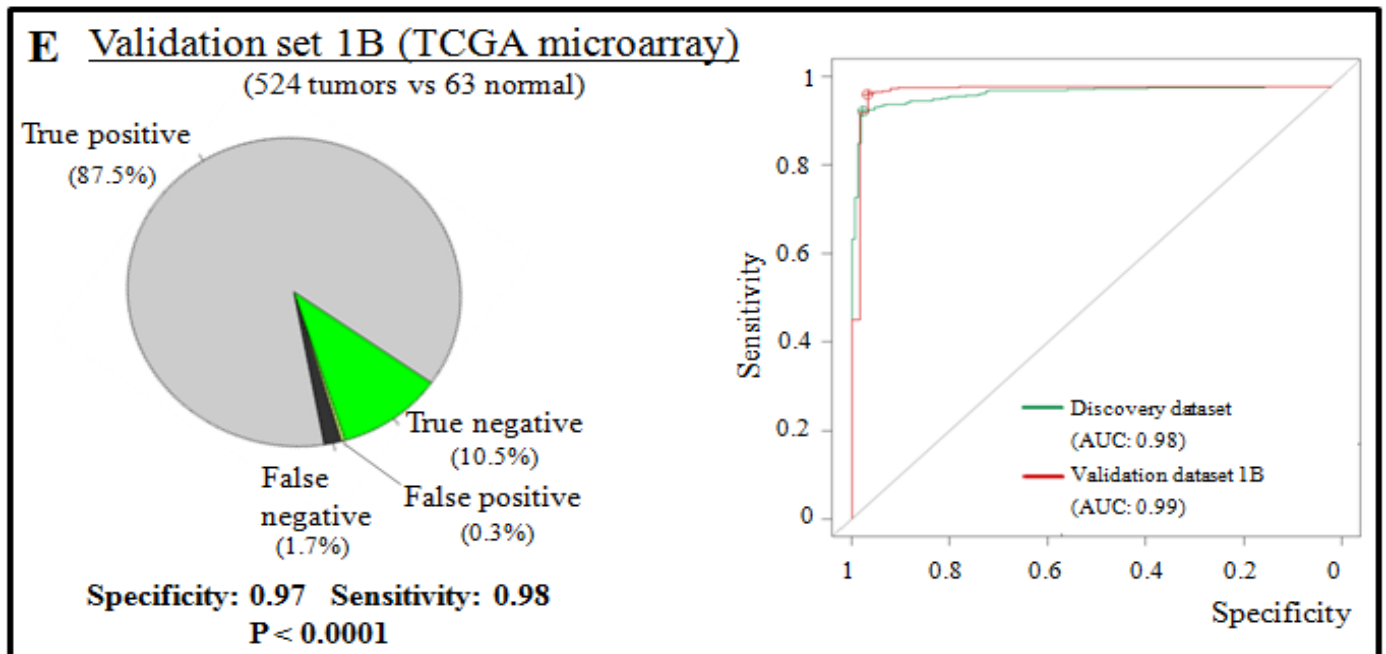
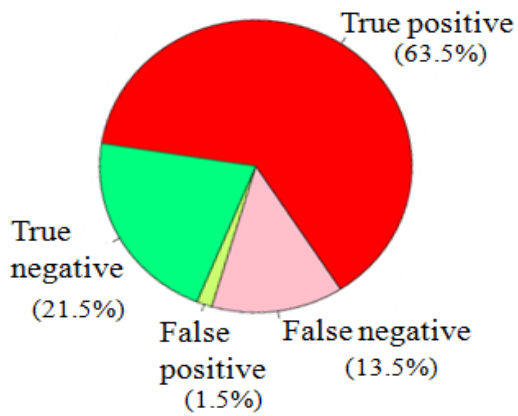


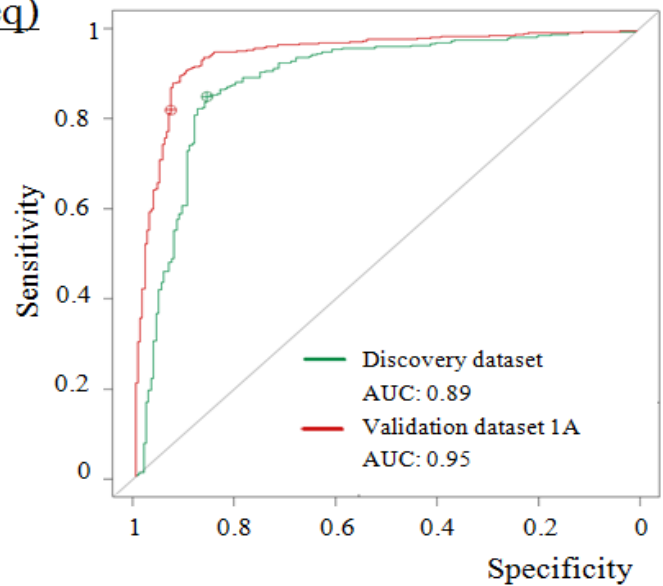
fig. S1. Dysregulated lncRNAs in breast tumors. (A) Distribution of gene expression levels, comparing lncRNA genes and protein-coding genes across all breast tissues. (B) Heat map of expression levels of the 230 lncRNA genes showing significantly different expression between 823 breast tumors and 172 normal breast tissue samples. (C) Chromosomal positions of the 215 dysregulated lncRNA genes. (D) Validation on the validation set 1A (TCGA RNA-seq). Left: pie chart of the predicted classification of breast tumors versus normal breast tissues, based on the 87 dysregulated lncRNAs that are present in the TCGA RNAseq dataset (see Materials and Methods for details of the prediction method). The P value is defined as the proportion of random signatures showing a balanced error rate lower than or equal to that of the real signature. Right : Receiver operating characteristic (ROC) analysis showing the sensitivity and specificity of the dysregulated lncRNA signature, based on the TCGA RNAseq data. The threshold (green and red crosses on the ROC curve) was determined on the discovery dataset and used to determine the sensitivity and specificity during the validation step. (E) Same as (D) based on the 167 dysregulated lncRNAs that are present in the validation set 1B (TCGA microarray). AUC: Area under the curve.

A Validation set 1A (TCGA RNA-seq)

(796 ER+ vs 238 ER- samples)

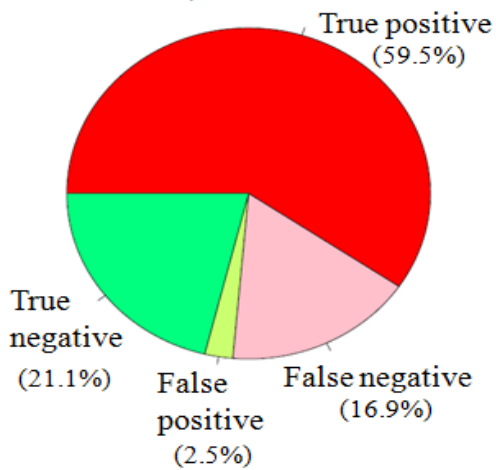


Specificity: 0.93 Sensitivity: 0.83
 $P < 0.0001$

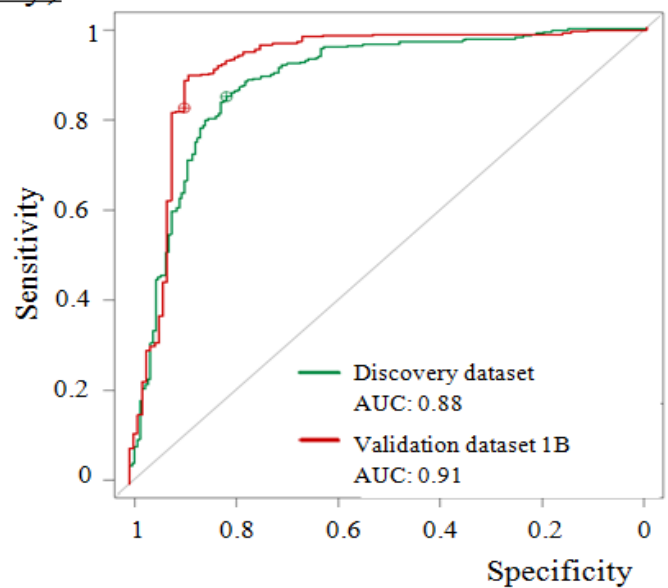


B Validation set 1B (TCGA microarray)

(398 ER+ vs 123 ER- samples)



Specificity: 0.89 Sensitivity: 0.78
 $P < 0.0001$



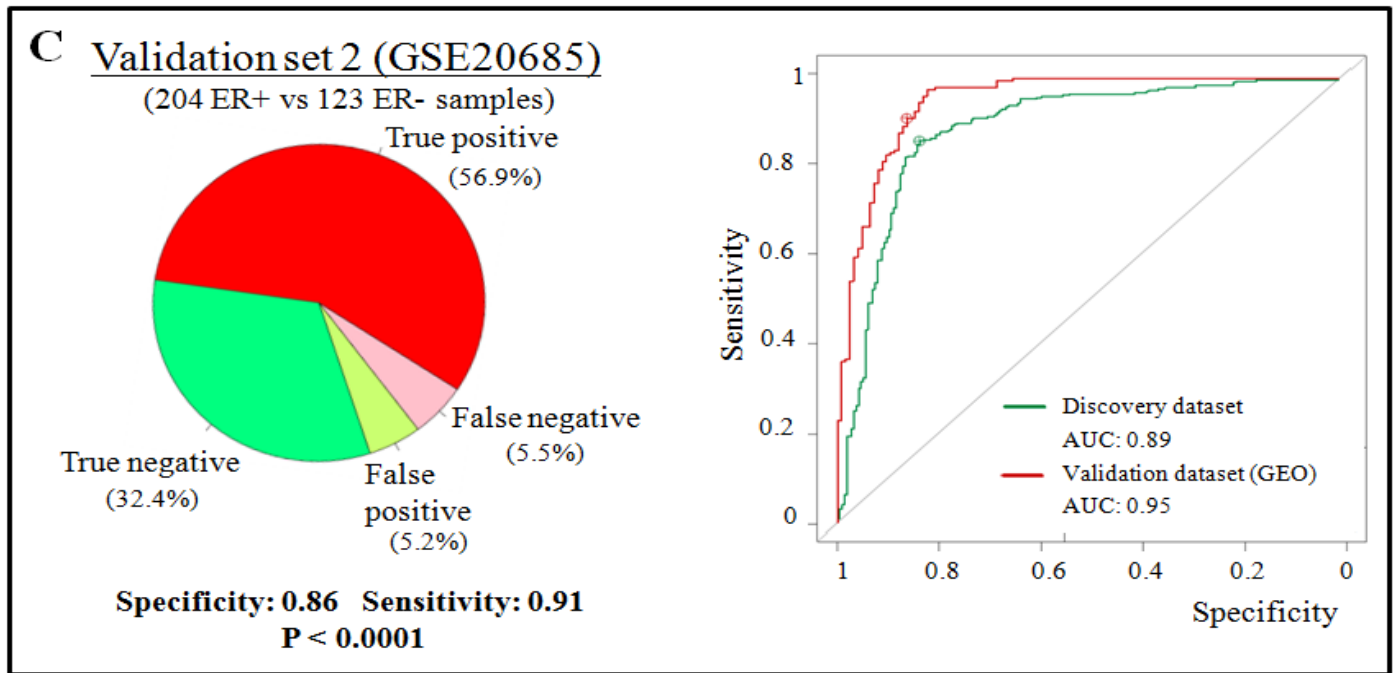
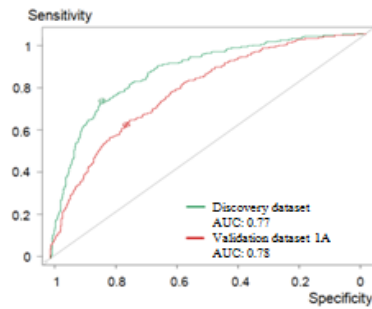
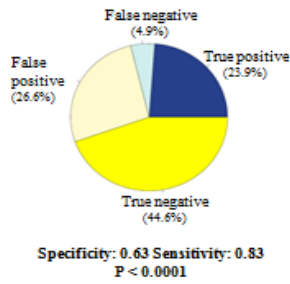
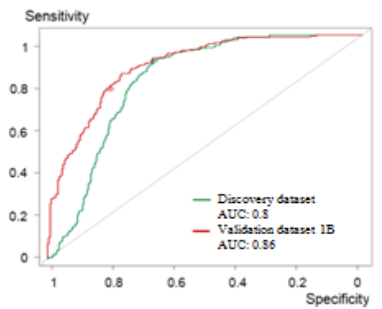
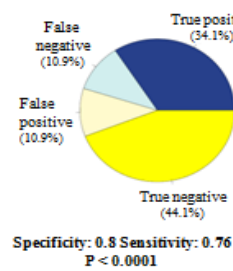
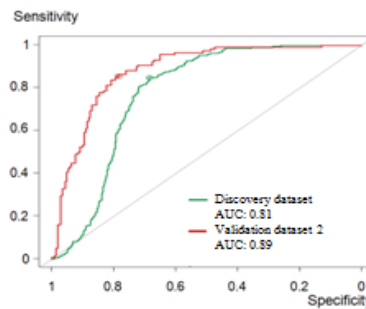
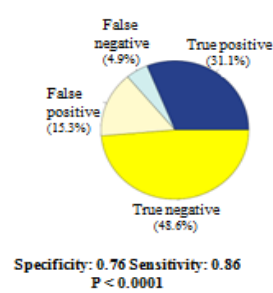
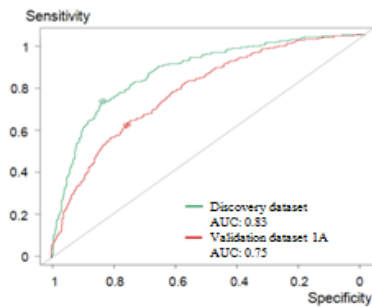
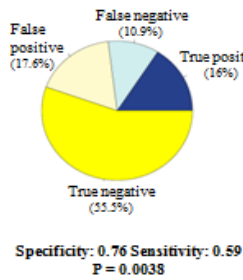
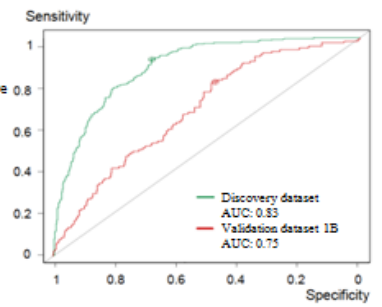
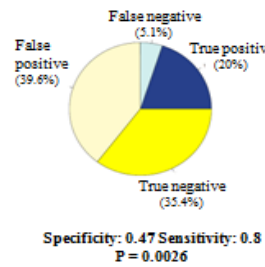
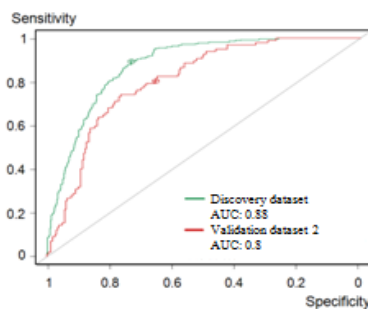
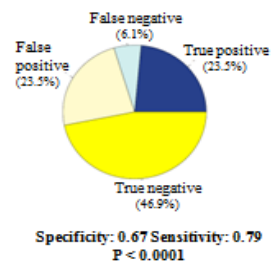


fig. S2. Validation of the ER-related lncRNA signature in three data sets. (A) Validation on the TCGA RNA-seq dataset, based on the 12 lncRNAs of the ER-related signature that are represented on the RNA-seq data. Left: Pie chart of the predicted classification of breast tumors into ER-positive and ER-negative tumors. The P value is defined as the proportion of random signatures showing a balanced error-rate lower than or equal to the real signature. Right: ROC curve analysis showing the sensitivity and specificity of the ER-related signature. The threshold (green and red crosses) was determined on the discovery dataset and was used to determine the sensitivity and specificity during the validation step. (B) Validation on the TCGA microarray dataset, based on the 29 lncRNAs of the ER-related signature that are represented on the TCGA microarray (C) Validation on validation data set 2 (GSE20685), based on the 38 lncRNAs that compose the ER-related lncRNA signature. AUC: Area under the curve.

A**Validation of the luminal A lncRNA signature****Validation set 1A (TCGA RNA-seq)**
(303 lum A vs 659 other breast cancers)**Validation set 1B (TCGA microarray)**
(232 lum A vs 283 other breast cancers)**Validation set 2 (GSE20685)**
(118 lum A vs 209 other breast cancers)**B****Validation of the luminal B lncRNA signature****Validation set 1A (TCGA RNA-seq)**
(283 lum A vs 679 other breast cancers)**Validation set 2 (TCGA microarray)**
(129 lum B vs 386 other breast cancers)**Validation set 2 (GSE20685)**
(97 lum B vs 230 other breast cancers)

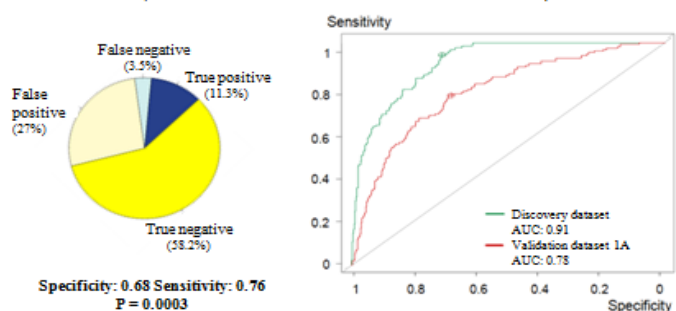
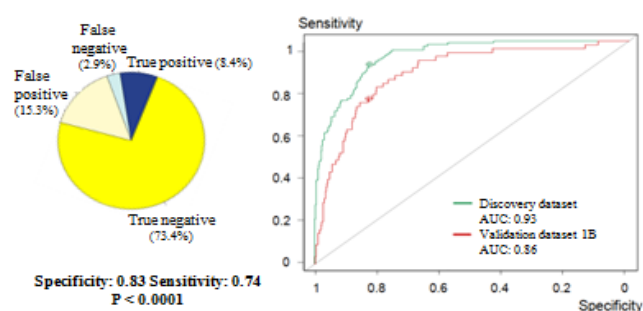
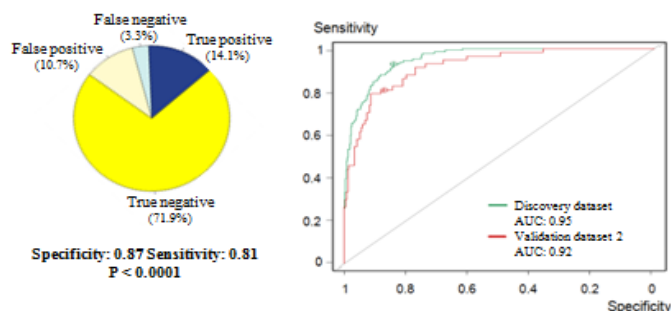
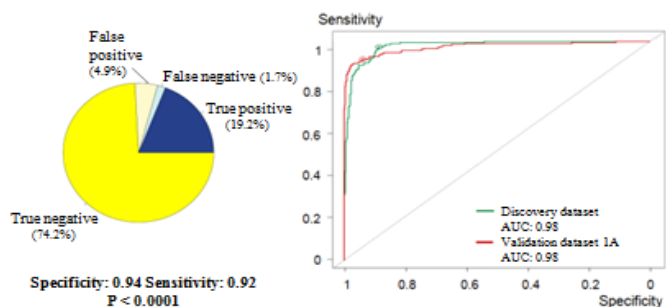
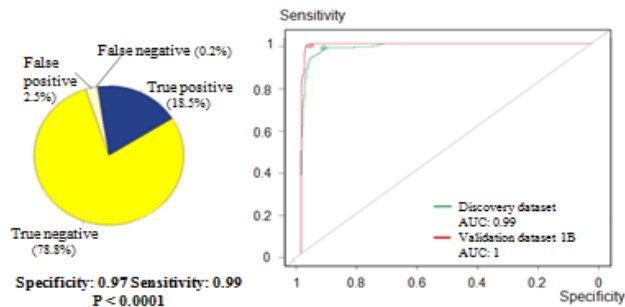
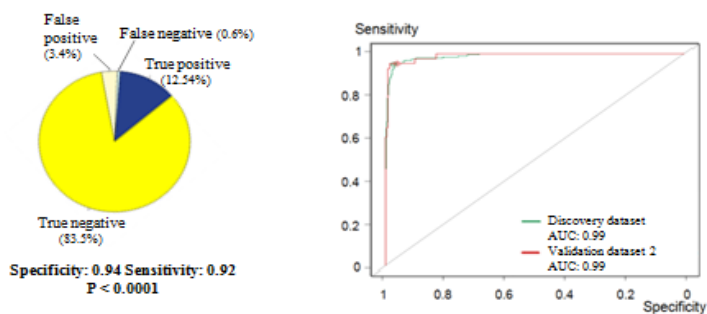
C**Validation of the HER2+ lncRNA signature****Validation set 1A (TCGA RNA-seq)**
(156 HER2+ vs 806 other breast cancers)**Validation set 1B (TCGA microarray)**
(58 HER2+ vs 457 other breast cancers)**Validation set 2 (GSE20685)**
(57 HER2+ vs 270 other breast cancers)**D****Validation of the basal-like lncRNA signature****Validation set 1A (TCGA RNA-seq)**
(220 Basal-like vs 742 other breast cancers)**Validation set 1B (TCGA microarray)**
(96 Basal-like vs 419 other breast cancers)**Validation set 2 (GSE20685)**
(43 Basal-like vs 284 other breast cancers)

fig. S3. Validation of the subtype-specific lncRNA signature in three data sets. Validation of the (A) luminal A, (B) luminal B, (C) HER2+, and (D) basal-like lncRNA signatures. For each breast cancer subtype, validation is shown for three datasets (validation set 1A based on TCGA RNA-seq, validation set 1B based on TCGA microarray and validation set 2 based on GSE20685). Pie charts depict the predicted classification of breast tumors as the subtype of interest versus other subtypes. The P value is defined as the proportion of random signatures showing a balanced error rate lower than or equal to that of the real signature. ROC curves depict the sensitivity and specificity of the signatures. The threshold (green and red crosses) was determined on the discovery dataset and was used to determine the sensitivity and specificity during the validation step. For validation set 1A, 15 of the 42 lncRNAs composing the luminal A signature, 3 of the 9 lncRNAs composing the luminal B signature, 7 of the 14 lncRNAs composing the HER2+ signature, and 27 of the 74 lncRNAs composing the basal-like signature were represented on the TCGA RNA-seq data and used for validation. For validation set 1B, 35 of the 42 lncRNAs composing the luminal A signature, 6 of the 9 lncRNAs composing the luminal B signature, 10 of the 14 lncRNAs composing the HER2+ signature, and 59 of the 74 lncRNAs composing the basal-like signature were represented on the TCGA array and used for validation. AUC: Area under the curve.

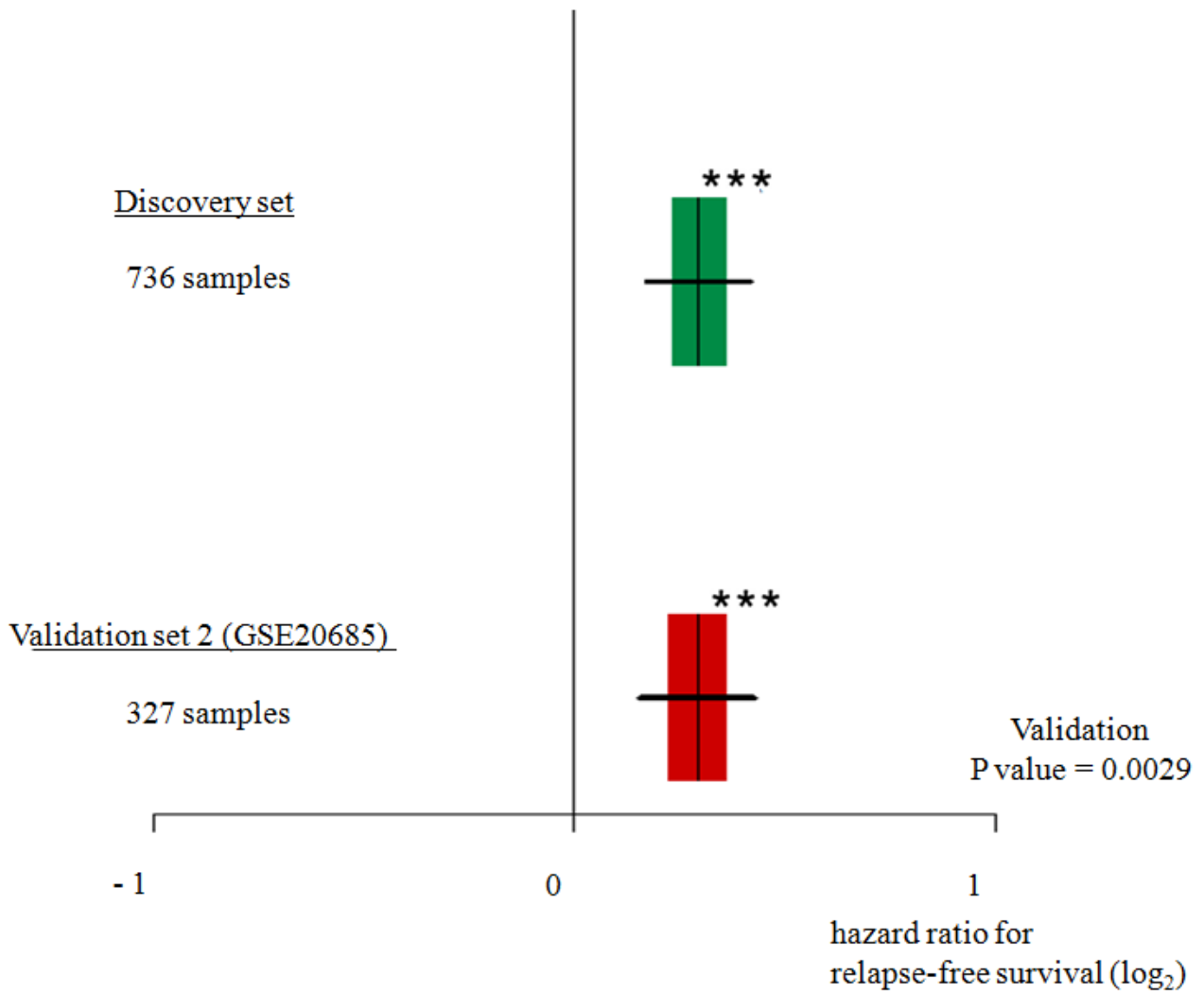


fig. S4. Validation of the set of 27 lncRNAs predictive of relapse and dysregulated in breast cancer. Forest plot showing the log₂ hazard ratio with the standard error of the mean (boxes) and the 95% confidence interval (bars) of the relapse free survival analysis in the discovery set and in the validation set 2.

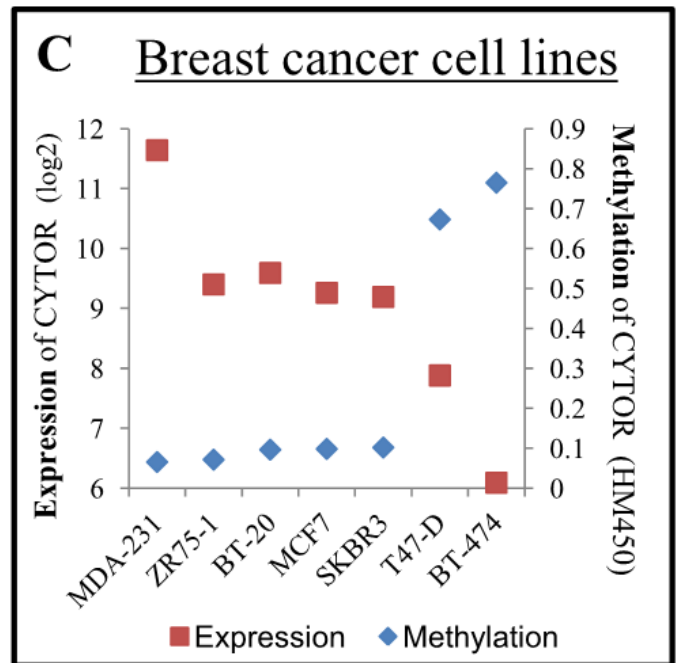
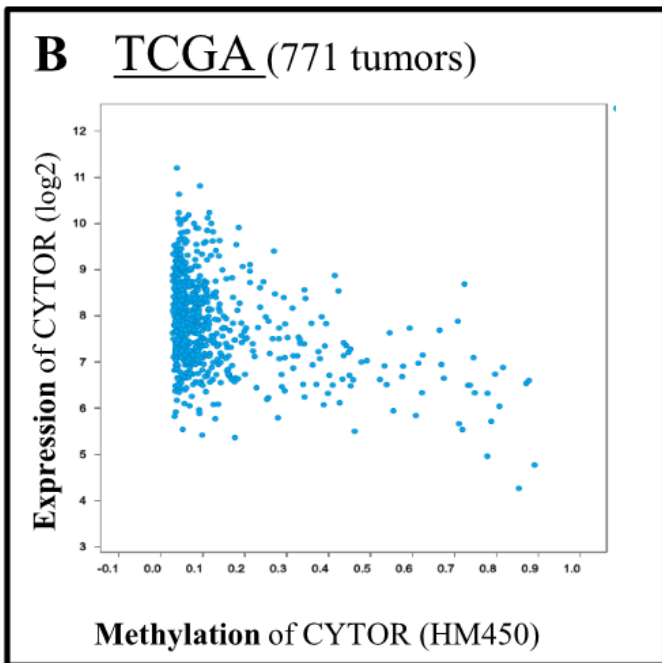
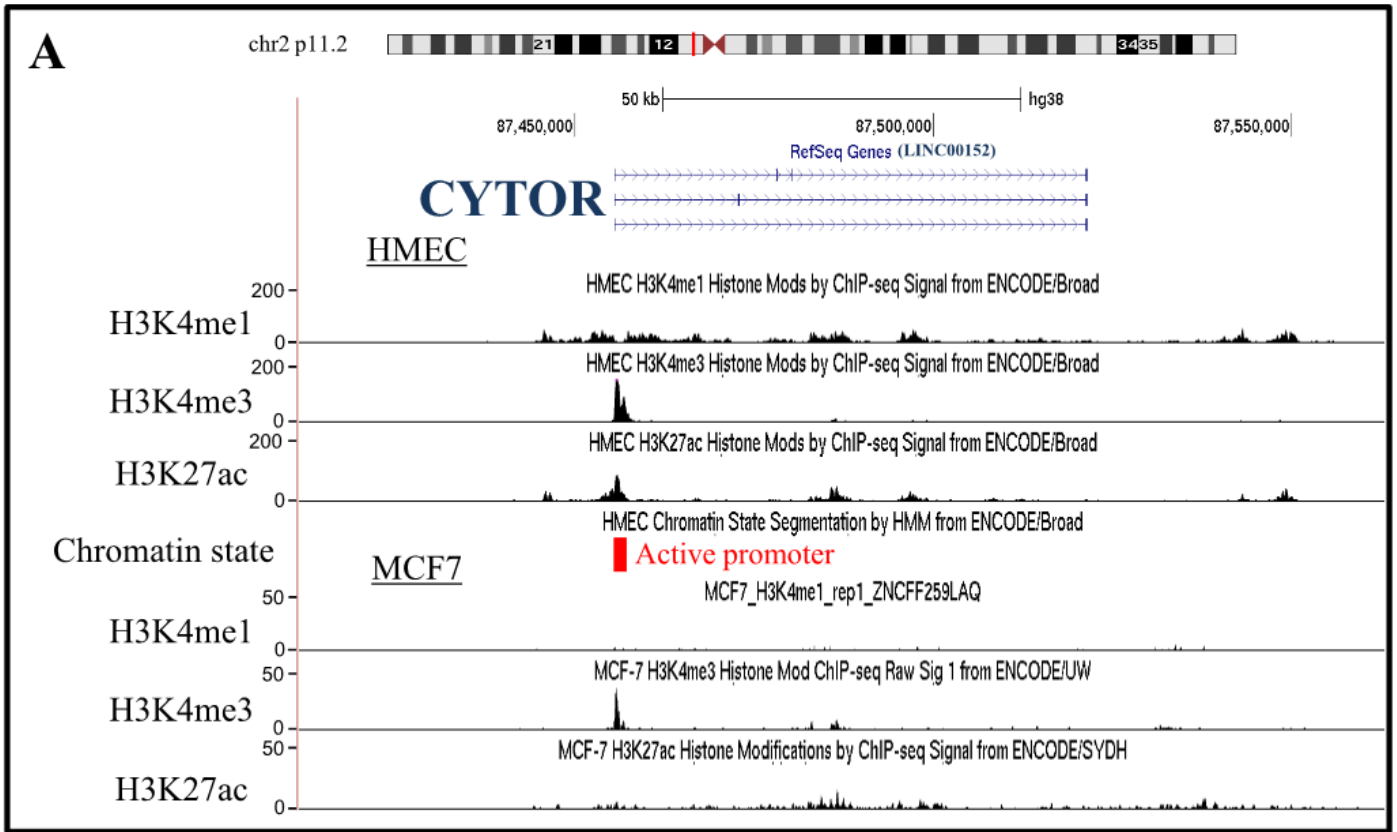
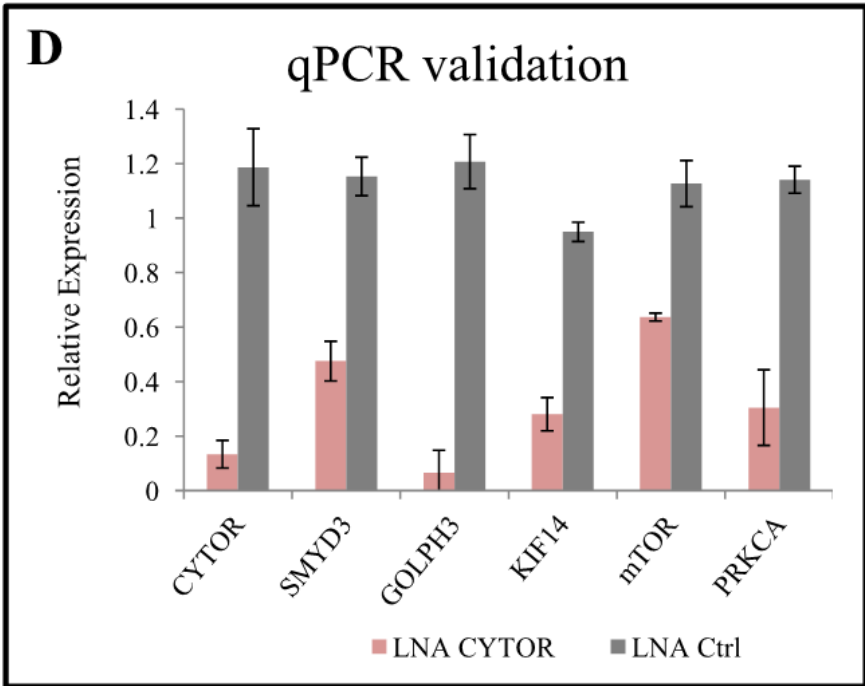
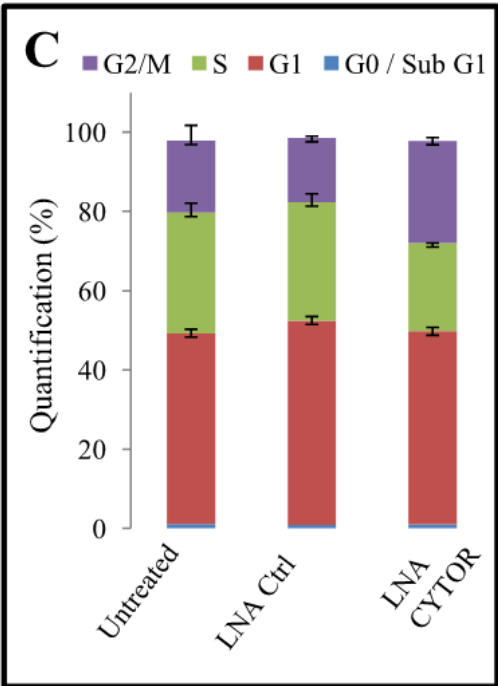
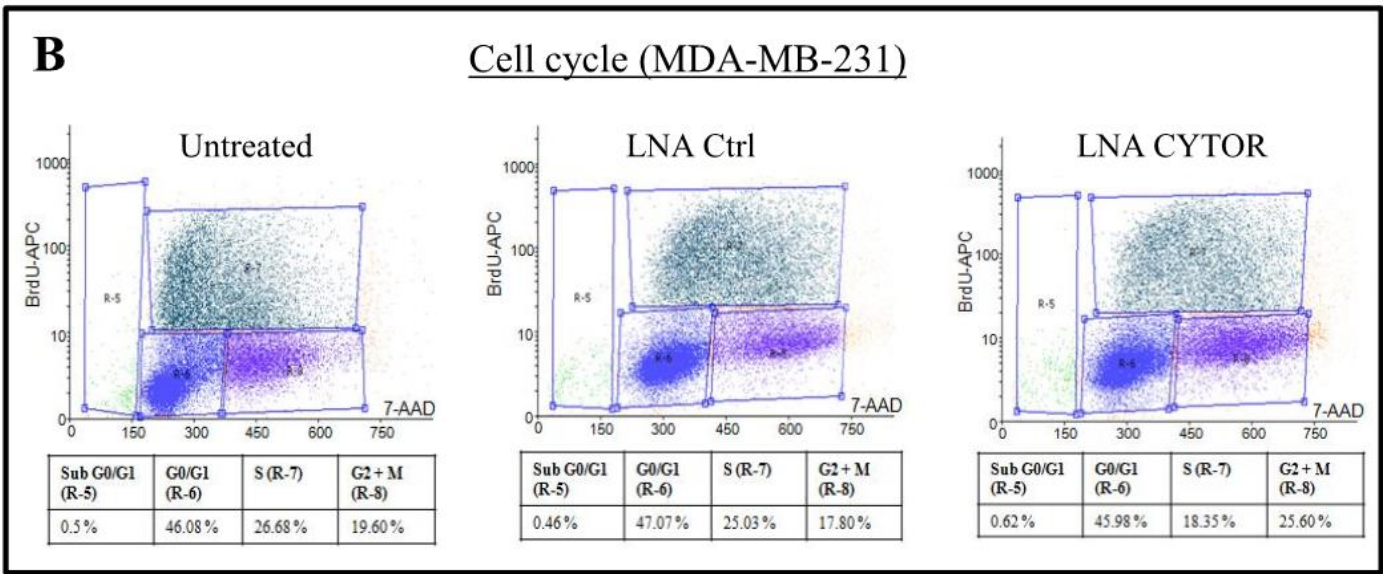
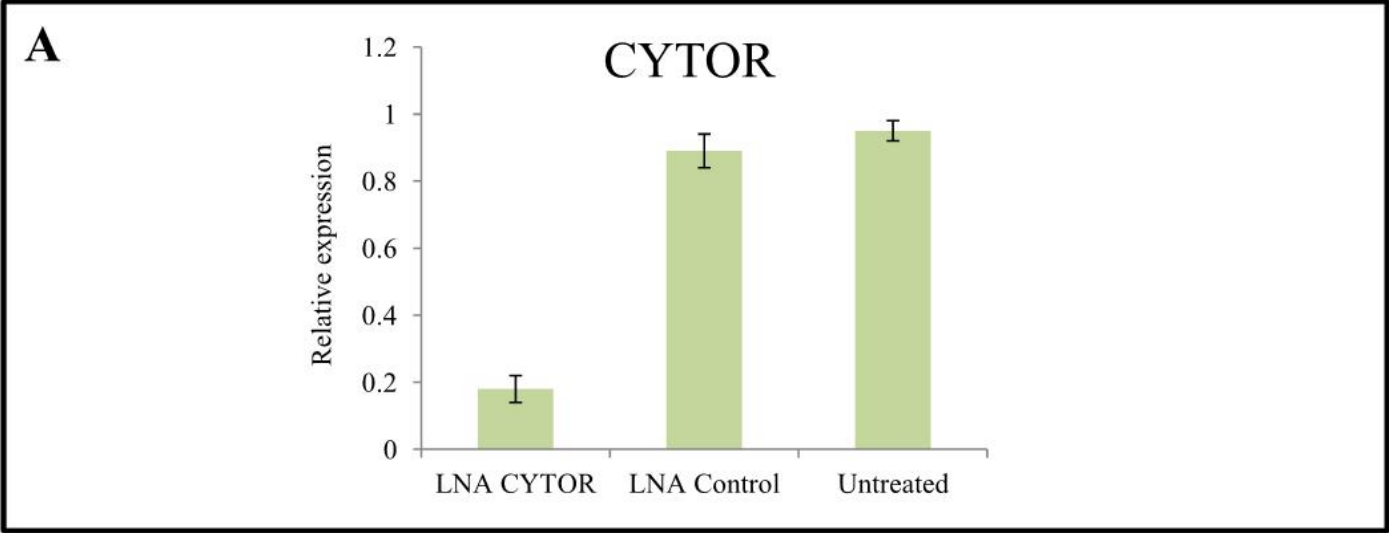


fig. S5. Methylation of the CYTOR gene in relation to its expression. (A) UCSC genome browser view of chromosome locus 2p11.2, which contains CYTOR. The magnified view depicts CYTOR with its associated histone marks (H3K27 acetylation, H3K4 trimethylation, H3K4 monomethylation and the ChromHMM from ENCODE) in HMEC and MCF-7 breast cell lines (B) Expression level of the CYTOR gene (RNA seq V2 RSEM log2) versus the methylation level of its promoter in 772 tumors of the TCGA cohort (result from cbiportal.org with the TCGA provisional dataset). Spearman correlation = -0.307. (C) Expression of the CYTOR gene and methylation level of its promoter in different breast cancer cell lines.



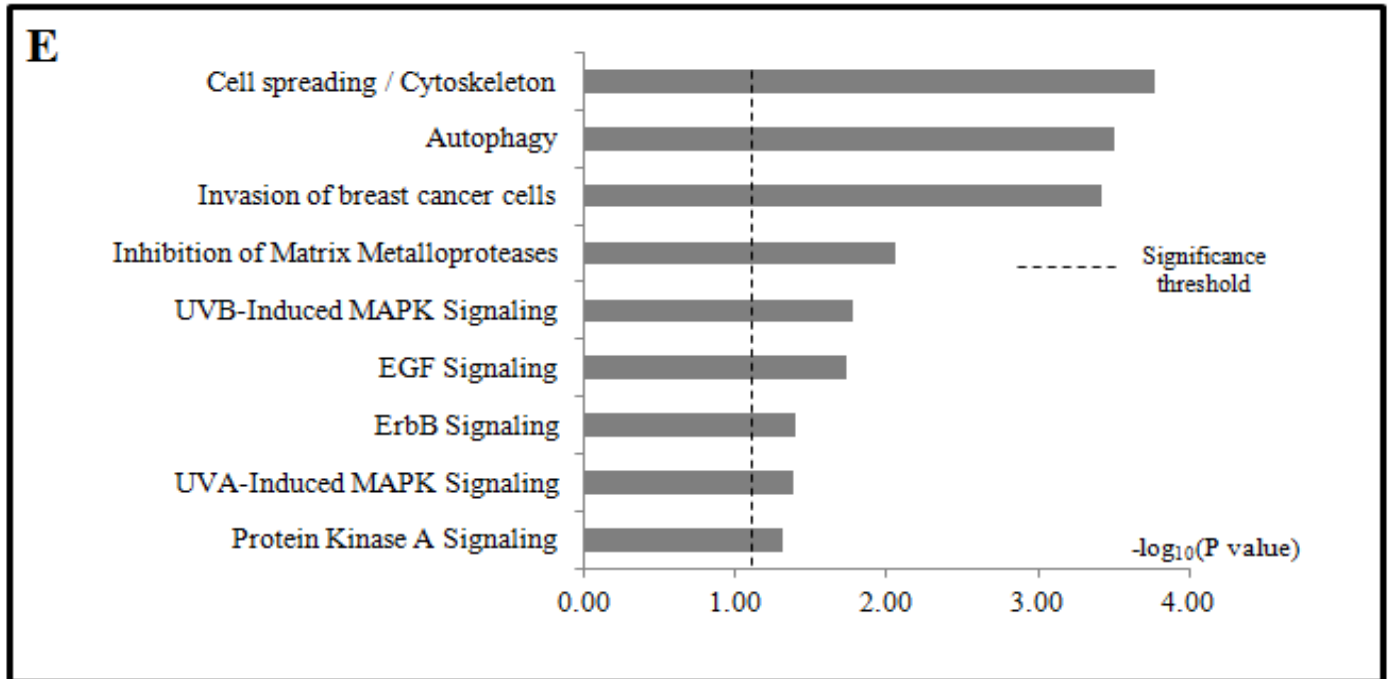


fig. S6. Impact of CYTOR depletion on cell cycle and gene expression. (A) Relative expression of CYTOR after knockdown with LNA gapmers (Exiquon) in MDA-MB-231 cells. Bars represent the standard deviation of triplicates. (B-C) Quantification (%), by flow cytometry, of the cell cycle distribution of MDA-MB-231 cells stained with BrdU and 7-AAD, after depletion of CYTOR by transfection with LNA gapmers. The results of all experiments were analyzed 48 h after treatment (n=2). (D) qPCR validation of 6 genes regulated by CYTOR (n = 3). (E) Ingenuity Pathway Analysis of dysregulated genes (Fold change > 1.5, FDR < 0.05) in CYTOR-knockdown MDA-MB-231 cells (n=6) compared to wild-type cells (n=6). The x-axis represents the $-\log_{10}(P \text{ value})$ and the dashed line shows the significance threshold of pathway over-representation.

table S4. Contingency table of the lncRNA-related clusters that correlates with ER status.

Observed number of patients in each cluster described in main Fig. 2A and their ER status are provided. This table only includes samples with ER status information based on IHC. $P = 2.5 \times 10^{-58}$, χ^2 test.

Observed	cluster 1 (ER+)	cluster 2 (Normal)	cluster 3 (ER-)	Total
ER+	291	1	58	350
ER-	28	3	183	214
Total	319	4	241	564

table S6. Contingency table of the lncRNA-related clusters that correlates with the known molecular subtypes. Observed number of patients in each cluster of the main Fig. 3A and their associated molecular subtype are provided. Molecular subtypes were defined by the PAM-50 predictor. $P = 1.2 \times 10^{-242}$, χ^2 test.

Observed	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Total
LumA	0	20	107	92	219
LumB	3	60	16	194	273
HER2	5	110	0	9	124
BLBC	174	19	1	1	195
Normal-like	0	3	4	5	12
Total	182	212	128	301	823

table S7. LncRNA signatures of known molecular subtypes of breast cancer (A) lncRNAs specific of the luminal-A subtype (B) lncRNAs specific of the luminal-B subtype (C) lncRNAs specific of the HER2+ subtype (D) lncRNAs specific of the basal-like subtype.

table S7A. Luminal A specific lncRNAs			
lncRNA name	HUGO Name	Fold Change	FDR
lnc-ATMIN-1	RP11-303E16.2	0.46	6.31E-77
lnc-SLC39A11-1	LINC00511	0.57	2.34E-36
lnc-TFEB-1	FOXP4-AS1	0.58	1.55E-30
lnc-SAFB2-1	TINCR	0.64	3.71E-12
lnc-FAM83B-1	RP3-523K23.2	0.64	2.02E-13
lnc-JAM2-2	MIR155HG	0.66	5.34E-16
lnc-ABT1-3	RP11-457M11.5	1.50	7.12E-21
lnc-NKIRAS1-2	THRBI-IT1	1.51	6.82E-23
lnc-multi-AL358813.2-2	#N/A	1.51	7.56E-12
lnc-RP11-809H16.2.1-2	EWSAT1	1.51	2.13E-19
lnc-ZNF506-1	#N/A	1.52	1.28E-34
lnc-ZBTB49-2	#N/A	1.53	1.48E-21
lnc-IL17B-2	CTB-89H12.4	1.53	1.70E-18
lnc-ATXN7-8	ADAMTS9-AS2	1.54	5.58E-29
lnc-ACER2-1	RP11-363E7.4	1.55	1.44E-22
lnc-KIAA1267-3	MAPT-AS1	1.55	6.86E-17
lnc-HOXC4-3	RP11-834C11.4	1.55	8.52E-29
lnc-TMED5-1	CCDC18-AS1	1.56	4.82E-27
lnc-HNMT-2	AC092620.2	1.56	4.18E-24
lnc-C14orf37-1	RP11-999E24.3	1.57	3.81E-21
lnc-PPP1R14D-1	RP11-532F12.5	1.57	6.01E-31
lnc-MAP1LC3B2-2	LINC00173	1.57	5.11E-12
lnc-LEMD3-1	RP11-305O6.3	1.62	2.68E-24
lnc-NAP1L2-1	XIST	1.64	3.78E-13
lnc-CCDC54-2	DUBR	1.64	2.01E-22
lnc-CLGN-2	#N/A	1.64	1.72E-27
lnc-ZNF207-1	#N/A	1.65	1.66E-19
lnc-TAF9-2	#N/A	1.65	2.34E-21
lnc-SERF1B-1	#N/A	1.69	4.80E-22
lnc-DLK1-4	MEG3	1.70	5.89E-14
lnc-multi-AP000769.1-2	NEAT1	1.70	7.94E-13
lnc-multi-C11orf89-2	H19	1.72	5.28E-12
lnc-RP11-195B21.3.1-2	#N/A	1.74	6.81E-11
lnc-TC2N-1	RP11-747H7.3	1.74	1.70E-18
lnc-BLID-1	RP11-166D19.1	1.76	1.20E-20
lnc-SNURF-1	PWAR6	1.80	4.95E-24
lnc-INADL-2	#N/A	1.82	1.28E-24
lnc-PCYOX1L-2	CARMN	1.84	6.35E-23
lnc-ZBTB20-1	RP11-553L6.5	1.85	4.30E-42
lnc-RASSF7-1	RP11-496I9.1	1.93	3.41E-26
lnc-B3GAT2-3	#N/A	1.98	1.45E-30
lnc-ACOX3-1	RP11-774O3.3	2.16	3.46E-44

table S7B. Luminal B specific lncRNAs			
lncRNA name	HUGO Name	Fold Change	FDR
lnc-PCP4-1	DSCAM-AS1	2.98	3.24E-15
lnc-FSIP1-7	RP11-265N7.1	1.65	9.04E-18
lnc-SLC7A13-1	CTD-2284J15.1	1.64	9.91E-23
lnc-MYC-2	PVT1	1.57	2.89E-17
lnc-multi-SNRNP25-1	WASIR2; WASIR1	1.56	5.00E-11
lnc-HIST4H4-1	RP11-174G6.5	1.56	2.76E-13
lnc-C2orf27A-1	LINC01087	1.52	1.15E-10
lnc-IRX3-4	CRNDE	1.51	3.32E-17
lnc-C1QTNF8-3	RP11-161M6.2	0.60	3.57E-34

table S7C. HER2+ specific lncRNAs			
lncRNA name	HUGO Name	Fold Change	FDR
lnc-MAP1LC3B2-2	LINC00173	0.45	8.21E-38
lnc-B3GAT2-3	#N/A	0.54	7.74E-52
lnc-MYC-2	PVT1	0.62	5.09E-11
lnc-ZBTB49-2	#N/A	0.64	4.26E-14
lnc-B4GALNT3-2	RP11-218M22.1	0.65	3.75E-15
lnc-AC007401.2.1-1	RP11-490M8.1	0.65	1.24E-10
lnc-ZIC5-3	RP11-12G12.7	1.50	4.59E-10
lnc-C4orf39-2	LINC01207	1.55	2.93E-05
lnc-ATP6V1H-1	RP11-400K9.4	1.64	9.68E-10
lnc-ATP10B-1	RP11-109J4.1	1.71	5.63E-07
lnc-AASDHPPT-1	RP11-680E19.1	1.72	2.48E-13
lnc-SAFB2-1	TINCR	1.72	1.27E-09
lnc-CDK12-1	RP11-390P24.1	2.00	1.24E-08
lnc-multi-SMUG1-7	HOTAIR	2.22	3.75E-15

table S7D. Basal-like specific lncRNAs

lncRNA name	HUGO Name	Fold Change	FDR
lnc-ANKRD30A-2	LINC00993	0.07	1.50E-144
lnc-RPLP1-1	DRAIC	0.37	1.19E-105
lnc-SLC7A13-1	CTD-2284J15.1	0.50	1.75E-100
lnc-KIN-2	RP11-379F12.3	0.47	6.76E-85
lnc-PCP4-1	DSCAM-AS1	0.22	1.85E-73
lnc-AC009336.1-5	LINC01116	0.36	3.07E-73
lnc-ACER2-1	RP11-363E7.4	0.55	2.18E-63
lnc-FSIP1-7	RP11-265N7.1	0.57	4.73E-60
lnc-HNMT-2	AC092620.2	0.53	4.80E-59
lnc-KIAA1267-3	MAPT-AS1	0.63	1.52E-54
lnc-C3orf52-1	RP11-757F18.5	0.55	7.24E-54
lnc-ABCB4-1	TP53TG1	0.51	5.38E-53
lnc-C4orf39-2	LINC01207	0.62	3.86E-46
lnc-C2orf27A-1	LINC01087	0.60	2.27E-45
lnc-ATMIN-1	RP11-303E16.2	1.86	1.08E-41
lnc-CALCOCO2-8	HOXB-AS1	0.63	1.50E-41
lnc-SLC39A11-1	LINC00511	2.83	3.14E-41
lnc-SERF1B-1	#N/A	0.43	4.17E-40
lnc-CCDC117-1	CTA-292E10.6	0.62	1.16E-39
lnc-ACOX3-1	RP11-774O3.3	0.40	2.76E-39
lnc-HIST4H4-1	RP11-174G6.5	0.54	5.76E-39
lnc-RP11-809H16.2.1-2	EWSAT1	0.66	1.16E-37
lnc-NKIRAS1-2	THRB-IT1	0.58	2.72E-37
lnc-NFAM1-2	LINC01315	2.05	2.76E-36
lnc-TAF9-2	#N/A	0.50	5.31E-36
lnc-TFEB-1	FOXP4-AS1	2.60	3.99E-33
lnc-INADL-2	#N/A	0.40	2.70E-32
lnc-PPP1R14D-1	RP11-532F12.5	0.56	2.33E-31
lnc-TAOK3-1	ncTr-VSIG10	0.66	2.33E-31
lnc-IL17B-2	CTB-89H12.4	0.54	8.28E-31
lnc-CLGN-2	#N/A	0.55	1.45E-30
lnc-VEZF1-1	RP11-159D12.2	0.66	1.91E-30
lnc-BMS1-3	LINC00839	1.97	2.01E-30
lnc-TC2N-1	RP11-747H7.3	0.49	5.90E-30
lnc-NRL-1	DHRS4-AS1	0.66	1.61E-29
lnc-FBN2-2	LINC01184	0.62	6.45E-28
lnc-CCDC54-2	DUBR	0.64	1.06E-27
lnc-HOXC4-3	RP11-834C11.4	0.62	1.28E-27
lnc-multi-AL445989.1-1	RP11-394A14.2; LINC01198	2.05	2.04E-27
lnc-AC016745.1-2	#N/A	0.63	7.12E-27
lnc-ZNF207-1	#N/A	0.56	1.01E-26
lnc-CDK12-1	RP11-390P24.1	0.58	4.24E-26
lnc-ZFP37-1	ZNF883	1.98	4.99E-26

table S7D. Basal-like specific lncRNAs (continued)			
lncRNA name	HUGO Name	Fold Change	FDR
lnc-C1QTNF8-3	RP11-161M6.2	2.39	5.77E-26
lnc-AL137145.1-2	#N/A	1.61	8.21E-26
lnc-FAM83B-1	RP3-523K23.2	2.08	1.24E-25
lnc-PTTG1-1	MIR3142HG	1.80	2.30E-24
lnc-FAM20C-3	AC093627.10	1.77	4.92E-23
lnc-multi-AP000769.1-2	NEAT1	0.49	6.03E-22
lnc-SLC7A6OS-2	RP11-96D1.11	0.66	6.24E-22
lnc-CCDC148-1	AC005042.4	1.75	7.22E-22
lnc-STOML1-1	LOXL1-AS1	0.60	8.35E-22
lnc-FRG2C-1	LINC00960	1.87	2.86E-21
lnc-C5orf39-2	#N/A	1.57	3.16E-21
lnc-OXR1-1	RP11-395G23.3	1.60	3.22E-21
lnc-PIGM-1	RP11-226L15.5	0.65	3.60E-21
lnc-CAGE1-1	RP3-512B11.3	1.60	2.98E-20
lnc-CRELD1-1	PRRT3-AS1	1.63	4.26E-19
lnc-GTF2F2-2	TPT1-AS1	1.53	1.08E-18
lnc-HEATR6-1	WFDC21P	1.51	6.50E-18
lnc-SNURF-1	PWAR6	0.51	9.13E-18
lnc-multi-SNURF-3	SNHG14	0.57	3.33E-17
lnc-IRX3-4	CRNDE	0.61	6.30E-17
lnc-SCYL1-1	MALAT1	0.52	7.10E-16
lnc-MEF2C-2	LINC00461	1.61	8.45E-16
lnc-LTBP3-2	#N/A	0.56	1.74E-15
lnc-TFAP2A-2	LINC00518	1.63	2.06E-13
lnc-MICB-3	HCP5	1.52	3.28E-13
lnc-NAP1L2-1	XIST	0.57	3.44E-13
lnc-LRRC17-1	NFE4	1.84	7.27E-12
lnc-JAM2-2	MIR155HG	1.65	2.61E-11
lnc-KCNJ9-2	LINC01133	1.56	4.08E-11
lnc-RPIA-1	ANKRD36BP2	1.60	5.76E-11
lnc-multi-POTEG-4	#N/A	1.64	1.17E-09

Legend for table S1 to S3, S5, S8 and S9 (Excel data set files)

table S1. Clinical information of the 823 primary tumors and the 172 normal tissues. Clinical information includes age, size, lymph node status, ER status, HER2 status, grade, PAM-50 associated subtype and relapse information.

table S2. Reannotation of the Affymetrix Human Genome U133 Plus 2.0 array. This annotation table includes information regarding nomenclature, genomic position and categories of lncRNAs.

table S3. Description of the 215 dysregulated lncRNAs in breast cancer. This table provides, for each lncRNA, the percentage of the samples in which expression was analyzed (see Materials and Methods), the fold change between cancer and normal samples, the FDR (false discovery rate) and the dysregulation significance status by classical t-test.

table S5. Significant enrichment scores from the guilt-by-association analysis. This table provides, for each lncRNA (column), significant enrichment scores of its association with particular gene set (row) (see Materials and Methods).

tables S8 and S9. Survival analysis. **Table S8** provides Cox univariate results. **Table S9** provides Cox multivariate results with size, lymph node, ER and HER2 status and grade as covariates. Hazard ratio values are provided on a log₂ scale.

