

Investigating the association between long non-coding RNAs and ductal carcinoma *in situ*

Julia Samson
BSc, MSc

Submitted in fulfilment of the requirements for the degree of Doctor of
Philosophy in Biochemistry and Cell Biology



National University of Ireland, Cork
School of Biochemistry and Cell Biology

11th May 2020

Head of School: Professor Rosemary O'Connor
Supervisors: Dr Kellie Dean and Prof Rosemary O'Connor

TABLE OF CONTENTS

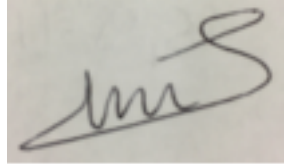
<i>Declaration.....</i>	<i>6</i>
<i>Acknowledgements</i>	<i>7</i>
<i>Abstract</i>	<i>9</i>
<i>List of abbreviations</i>	<i>11</i>
<i>List of figures.....</i>	<i>13</i>
<i>List of tables</i>	<i>18</i>
<i>Publications and presentations.....</i>	<i>20</i>
<i>Licenses for the use of published figures</i>	<i>22</i>
 <i>Chapter 1: Introduction</i>	 <i>24</i>
1.1 Normal breast.....	25
1.1.1 Anatomy of the breast	25
1.1.2 Development of the breast in humans	25
1.2 Invasive breast cancer	27
1.2.1 Definition and origin of IBC.....	28
1.2.2 IBC classifications	30
1.2.3 IDC incidence and statistics in Ireland.....	33
1.3 Ductal carcinoma <i>in situ</i>.....	33
1.3.1 Definition and origin of ductal carcinoma <i>in situ</i>	33
1.3.2 DCIS classifications	35
1.3.3 Detection and treatment of DCIS, an on-going debate.....	37
1.4 DCIS-known markers	39
1.4.1 Genomic alterations specific to DCIS	39
1.4.2 Epigenetic modifications in DCIS.....	41
1.4.3 Gene expression panels for the prediction of DCIS 10-year risk of recurrence	43
1.4.4 Protein markers of DCIS	44
1.4.5 miRNAs markers.....	46
1.5 Non-coding RNA	50
1.5.1 Definition and classification of lncRNAs.....	51
1.5.2 Function of lncRNAs	54
1.5.3 Examples of lncRNAs in the normal context	58
1.6 lncRNAs in disease	59
1.6.1 lncRNAs in neurologic diseases	60
1.6.2 lncRNAs in cancer	61
1.7 Aims of the project	67
 <i>Chapter 2: Material and Methods.....</i>	 <i>70</i>
2.1 Materials	70
2.1.1 Cell lines	70
2.1.2 Bacterial strain	70
2.1.3 Chemicals and consumables	71
2.1.4 Antibodies	73

2.1.5 Oligonucleotides	74
2.1.6 Plasmid	75
2.1.7 siRNAs	76
2.1.8 Antisense oligonucleotides	77
2.2 Methods	77
2.2.1 Tissue culture	77
2.2.2 Transfection	78
2.2.3 Plasmid construction.....	79
2.2.4 Agarose gel electrophoresis	80
2.2.5 RNA	81
2.2.6 Western blotting	84
2.2.7 β -actin immunofluorescence	85
2.2.8 Cell growth monitoring	85
2.2.9 Plating efficiency assay	86
2.2.10 Clonogenic growth assay	86
2.2.11 MTT assay.....	86
2.2.12 Wound healing assay	87
2.2.13 Subcellular fractionation.....	87
2.2.14 RNA affinity purification and mass spectrometry analysis.....	88
2.2.15 RNA immunoprecipitation	88
 Chapter 3: Molecular and cellular characterization of three models of ductal carcinoma in situ (DCIS)	92
3.1 Introduction	92
3.2 Results.....	94
3.2.1 Morphologic characterization of the DCIS cell lines MCF10DCIS.com, ETCC-006 and ETCC-010	94
3.2.2 Molecular subtype of the DCIS cell lines	97
3.2.3 Analysis of the expression of Oncotype DX DCIS test markers in the DCIS cell lines	100
3.2.4 Genome-wide comparison of the ETCC-006 and ETCC-010 DCIS model to MCF-10A	103
3.2.5 Monitoring of the cell proliferation of the DCIS cell lines	111
3.2.6 Assessment of migration and anchorage-independent growth potential of the DCIS cell lines.....	113
3.2.7 ETCC-006 and ETCC-010 have undergone epithelial to mesenchymal transition	116
3.2.8 Signaling profile of the DCIS cell lines	119
3.3 Discussion.....	122
 Chapter 4: Identification of long non-coding RNAs associated with ductal carcinoma in situ (DCIS)	126
4.1 Introduction	126
4.2 Results.....	128
4.2.1 Quality control pre-lncRNA capture	128
4.2.2 Assessment of the efficiency of the lncRNA capture	133
4.2.3 Selection of six lncRNAs candidates for further characterization	142
4.2.4 Validation of the differential expression of the six lncRNAs candidates.....	145
4.2.5 Previous cancer association of the six lncRNAs candidates	148
4.2.6 Bioinformatic characterization of the six lncRNAs candidates.....	149
4.2.7 Recruitment of a DCIS patient cohort	169
4.2.8 lncRNA capture on DCIS patient RNA.....	171
4.2.9 Expression of the six lncRNAs candidates in our DCIS patient cohort	179

4.2.10 Identification of dysregulated lncRNAs in the transcriptomics study of ETCC-006 and ETCC-010 presented in the Chapter 3	181
4.3 Discussion.....	184
 <i>Chapter 5: In vitro characterization of the role of LINC00473, LINC00313, LINC00707, MIR210HG, LOC100134040 and LOC729970 in DCIS</i>	<i>188</i>
5.1 Introduction	188
5.2 Results.....	190
5.2.1 Subcellular localization of the six lncRNAs candidates	190
5.2.2 Effect of the overexpression of the six lncRNAs candidates on breast cells behavior	194
5.2.3 Knockdown of the six lncRNAs candidates.....	203
5.2.4 Effect of LOC729970 knockdown on breast cancer cell behavior using antisense oligonucleotides (ASOs)	206
5.2.5 Expression of the six lncRNA candidates under stress conditions	211
5.2.6 Identification of LOC729970 protein interactors	216
5.2.7 Attempt at confirming LOC729970 interaction with CYR61 and GAPDH.....	224
5.3 Discussion.....	239
 <i>Chapter 6: General Discussion</i>	<i>244</i>
 <i>Bibliography.....</i>	<i>250</i>

Declaration

This thesis submitted is my own work, except where explicit reference is made to the contribution of others, and has not been submitted for another degree, either at University College Cork or elsewhere.

A handwritten signature in dark ink, appearing to be 'Julia Samson', written on a light-colored background.

Signed:

(Julia Samson)

Date: 11th May 2020

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisor, Dr. Kellie Dean, for giving me the opportunity to undertake a PhD in the Biochemistry department at University College Cork under her guidance. Throughout those three years, she has been a constant source of support, guidance and motivation. Being her first PhD student in the laboratory has given me the opportunity to learn independence and organization. In addition to the opportunity of being able to work with Dr. Dean, I am extremely grateful for the chance to conduct my PhD in Ireland. This has allowed me to discover a new country with a completely different culture compared to the French culture. This has broadened my horizons and allowed me to grow as a person.

I would also like to thank the people I have been lucky to work alongside with. First, the supergroup lab formed by Dr. Kellie Dean, Dr. Paul Young and Dr. Justin McCarthy lab for useful joint lab meetings and fun gathering. A special thank you goes to Shona Cronin, Eoghan O’Leary, Leanne O’Sullivan, Averil Kiely, Claire McSweeney, Mehul Patel and Dr. John Williams for the social support and scientific advices in time of doubts. I would like to acknowledge all the people working in the biochemistry department that have helped me along the way, with special thoughts to Pat Allen, Noreen Casey, Jenny Duane, Tricia Fowler, Dr. Gary Loughran and Dr. Orla Cox. I would also add a special thank you to Dr. Sudipto Das (RCSI) without whom, most of this work would not have been possible. Finally, I would also like to thank Dr. Stephanie Davies, Dr. Sinead O’Loughlin, Dr. Janinah Baclaocos, Dr. Anjali Pai and Dr. Ioanna Tzani for the warm welcome when I first came to Ireland and the constant support throughout this experience.

I wish to acknowledge my housemates Dr. Sarah Riis and Dr. Leonie Rieger that both started their PhD at the same time as me, without them this experience would not have been the same.

Finally, I would like to acknowledge the people that supported me despite the distance, my friends Jennifer, Delphine, Clémence, Laëtitia, Aude and Aline and most importantly my family, my Mom, Dad and my sister. All of this work would not have been possible without their support.

ABSTRACT

Breast cancer is the most common cancer in women globally, with incidence rates increasing and survival rates varying widely depending on early detection and access to treatment. In Ireland, the number of diagnoses is increasing (National Cancer Registry, 2016), with 89% of new cases being invasive breast cancer. To reduce the number of individuals with invasive cancer, there is an urgent need for specific and sensitive diagnostic biomarkers for the earliest stages of breast cancer.

Ductal carcinoma *in situ* (DCIS) is a non-obligate precursor to invasive ductal carcinoma. With an increasing number of studies linking long, non-coding RNAs (lncRNAs) to various cancers, we have specifically selected to examine lncRNAs as novel DCIS biomarkers and to characterise their biological roles. To do so, we aimed to perform a transcriptomic study focused on lncRNAs using two patient-derived DCIS cell lines (ETCC-006 and ETCC-010) and one cellular model of DCIS, MCF10DCIS.com. However, since ETCC-006 and ETCC-010 cells have been poorly studied in the literature, we started by conducting a detailed molecular characterization of those cells, comparing them to other breast cancer cell lines. Contrary to previously published results, we found that ETCC-006 and ETCC-010 cells in culture are more similar to triple negative breast cancer cell lines based on hormone receptor status. Next, we performed a lncRNA-capture RNA sequencing (RNAseq) on those characterized cells. From those experiments, we identified 92 common differentially expressed lncRNAs in those cells compared to MCF-10A cells. Using data from The Cancer Genome Atlas (TCGA), we reduced our targets to six lncRNAs whose altered expression is associated with adverse breast cancer patient outcomes. Next, we aimed to functionally characterize those targets by monitoring proliferation and migration

after altering their expression levels. This led us to focus specifically on LOC729970, an uncharacterized lncRNA in the literature. To investigate its molecular function, we identified LOC729970 specific proteins interactors by mass spectrometry. Then, we chose to focus on the glycolytic enzyme GAPDH and the angiogenic factor CYR61 to confirm their interaction with LOC729970.

Ultimately, our goal is to identify and characterise lncRNA biomarkers in DCIS with the hope of establishing a list of lncRNAs involved in the transition toward IDC. This could then lead to the development of diagnostic tools that could change how early breast cancer is detected and treated in patients worldwide.

LIST OF ABBREVIATIONS

ASO	Antisense oligonucleotide
asRNA	Antisense RNA
AD	Alzheimer's disease
ADH	Atypical ductal hyperplasia
BC	Breast cancer
BCSCs	Breast cancer stem cells
BCS	Breast conserving surgery
CSCs	Cancer stem cells
ciRNA	Circular intronic RNA
CPAT	Coding potential assessment tool
ceRNAs	Competing endogenous RNAs
CLIP	Crosslinking and immunoprecipitation
DCIS	Ductal carcinoma <i>in situ</i>
ESCs	Embryonic stem cells
eRNA	Enhancer-associated RNA
EMT	Epithelial to mesenchymal transition
ECM	Extracellular matrix
FBS	Fetal bovine serum
FFPE	Formalin-fixed paraffin-embedded
GO	Gene Ontology
iPSCs	Induced pluripotent stem cells
IBC	Invasive breast cancer
IDC	Invasive ductal carcinoma
ILC	Invasive lobular carcinoma

lincRNAs	Long intergenic non-coding RNAs
lncRNAs	Long non-coding RNAs
miRNAs	MicroRNAs
NSCs	Normal stem cells
Oligo	Oligonucleotide
PD	Parkinson's disease
P/S	Penicillin-streptomycin
piRNAs	Piwi-associated RNAs
PBS	Phosphate buffered saline
PDT	Population doubling time
qRT-PCR	Quantitative Reverse Transcriptase-PCR
RNA-FISH	RNA fluorescent <i>in situ</i> hybridization
RIP	RNA immunoprecipitation
RNAseq	RNA sequencing
siRNAs	Short interfering RNAs
snRNAs	Small nuclear RNAs
snoRNAs	Small nucleolar RNAs
TCGA	The Cancer Genome Atlas
TNBC	Triple negative breast cancer
TNM	Tumor, node, metastasis

LIST OF FIGURES

Figure 1.1: Anatomy of the human female breast.

Figure 1.2: Next generation of hallmarks of cancer.

Figure 1.3: Tumorigenesis process in the duct of the breast.

Figure 1.4: Sub-classification of long non-coding RNAs.

Figure 1.5: Long non-coding RNAs and their mechanism of action in cells.

Figure 2.1: tRSA construct and RNA-affinity purification procedure.

Figure 3.1: Phenotypic analysis of MCF10DCIS.com, ETCC-006 and ETCC-010 by comparison with MCF-10A, MCF-7 and MDA-MB-231 cell lines.

Figure 3.2: Molecular subtype of the DCIS cell lines.

Figure 3.3: Expression level of Oncotype DX DCIS score markers in the DCIS cell lines analyzed by qRT-PCR.

Figure 3.4: Bioanalyzer analysis of the ETCC-006 and ETCC-010 RNA.

Figure 3.5: Transcriptomic comparison of the DCIS cell lines ETCC-006 and ETCC-010 to MCF-10A.

Figure 3.6: Proliferation assessment of the DCIS cell lines.

Figure 3.7: Migration and anchorage-independent growth assessment of the DCIS cell lines.

Figure 3.8: Epithelial to Mesenchymal (EMT) transition in the DCIS cell lines.

Figure 3.9: Protein levels of signaling proteins in the DCIS cell lines analyzed by Western blotting.

Figure 4.1: Long non-coding RNA hybridization-based capture enrichment approach for lncRNA specific RNA sequencing.

Figure 4.2: Assessment of RNA quality of RNA from MCF-10A, MCF10DCIS.com, ETCC-006 and ETCC-010.

Figure 4.3: Verification of successful library preparation with MCF-10A, MCF10DCIS.com, ETCC-006 and ETCC-010 RNA.

Figure 4.4: Differentially expressed genes in ETCC-006 and ETCC-010 compared to MCF-10A.

Figure 4.5: Sorting of differentially expressed genes in DCIS cell lines from target-capture RNAseq.

Figure 4.6: Selection of six lncRNAs candidates for further analysis.

Figure 4.7: Expression of the six lncRNAs candidates in breast cancer cell lines.

Figure 4.8: Differential expression analysis of LINC00707 based on subtype in an invasive breast cancer cohort from TCGA generated in the TANRIC database.

Figure 4.9: Epigenetic landscape of the LINC00707 locus in HeLa cells using the UCSC genome browser annotation hg19.

Figure 4.10: Differential expression analysis of LINC00473 based on subtype in an invasive breast cancer cohort from TCGA generated in the TANRIC database.

Figure 4.11: Epigenetic landscape of the LINC00473 locus in HeLa cells using the UCSC genome browser annotation hg19.

Figure 4.12: Differential expression analysis of LINC00313 based on subtype in an invasive breast cancer cohort from TCGA generated in the TANRIC database.

Figure 4.13: Epigenetic landscape of the LINC00313 locus in HeLa cells using the UCSC genome browser annotation hg19.

Figure 4.14: Differential expression analysis of MIR210HG based on subtype in an invasive breast cancer cohort from TCGA generated in the TANRIC database.

Figure 4.15: Epigenetic landscape of the MIR210HG locus in HeLa cells using the UCSC genome browser annotation hg19.

Figure 4.16: Differential expression analysis of LOC729970 based on subtype in an invasive breast cancer cohort from TCGA generated in the TANRIC database.

Figure 4.17: Epigenetic landscape of the LOC729970 locus in H1 human Embryonic Stem Cells (hESCs) using the UCSC genome browser annotation hg19.

Figure 4.18: Differential expression analysis of LOC100134040 based on subtype in an invasive breast cancer cohort from TCGA generated in the TANRIC database.

Figure 4.19: Epigenetic landscape of the LOC100134040 locus in K562 cells using the UCSC genome browser annotation hg19.

Figure 4.20: Characterization of the DCIS patient cohort.

Figure 4.21: Verification of successful library preparation with normal and tumor RNA from the DCIS patient cohort.

Figure 4.22: Sorting of the differentially expressed genes in the tumor compared to normal.

Figure 4.23: Differential expression analysis of MEG8 based on subtype in an invasive breast cancer cohort from TCGA generated in the TANRIC database.

Figure 4.24: qRT-PCR expression quantification of the six lncRNAs candidates in the DCIS patient cohort.

Figure 4.25: Dysregulated lncRNAs in ETCC-006 and ETCC-010 identified using the dataset generated and described in Chapter 3.

Figure 5.1: Subcellular fractionation of MCF-10A and MDA-MB-231 cells.

Figure 5.2: Effect of the overexpression of LINC00473, LOC729970, MIR210HG, LOC10013404, LINC00707, LINC00313 on MCF-7 and ETCC-006 cells proliferation.

Figure 5.3: Migration of MCF-7 and ETCC-006 cells after overexpression of the lncRNA candidates.

Figure 5.4: Expression of EMT markers after overexpression of the lncRNA candidates.

Figure 5.5: Analysis of LOC729970 as a competing endogenous RNA.

Figure 5.6: Analysis of the knockdown of the lncRNA candidates using siRNAs by qRT-PCR.

Figure 5.7: Analysis of antisense oligonucleotides targeting LOC7229970 in ETCC-006 cells by qRT-PCR.

Figure 5.8: Analysis of cell proliferation and migration of ETCC-006 cells after LOC729970 knockdown with ASOs.

Figure 5.9: Protein level of AKT and p-AKT measured by Western blotting after LOC729970 knockdown using ASOs.

Figure 5.10: Expression of the six lncRNA candidates in stress conditions.

Figure 5.11: Expression of the six lncRNA candidates in hypoxic conditions.

Figure 5.12: Controls to monitor the successful binding of the RNA to the beads and binding of proteins.

Figure 5.13: Attempt at confirming the interaction between LOC729970 and CYR61 or GAPDH using the tRSA-system.

Figure 5.14: RNA immunoprecipitation of myc-tagged version of PAPBP, GAPDH and CYR61 to confirm GAPDH/CYR61 interaction with LOC729970.

Figure 5.15: RNA immunoprecipitation of myc-tagged version of PAPBP, GAPDH and CYR61 after formaldehyde crosslinking to confirm GAPDH/CYR61 interaction with LOC729970.

Figure 5.16: Attempt at confirming the interaction between LOC729970 and CYR61/GAPDH using the tRSA-system with myc-tagged version of GAPDH and CYR61.

Figure 5.17: Bioinformatic analysis to support the data identified in the mass spectrometry experiment.

LIST OF TABLES

Table 1.1: Summary of the DCIS-known-markers presented in the section above.

Table 2.1: Chemical and consumables used for the experiments described in this thesis.

Table 2.2: Antibodies used for Western blotting and indirect immunofluorescence.

Table 2.3: DNA primers used for the quantitative qRT-PCR experiments.

Table 2.4: DNA plasmids used in the experiments described in the thesis.

Table 2.5: Sequences of siRNAs used for the knock-down experiments.

Table 2.6: Sequences of antisense oligonucleotides (ASOs) used for LOC729970 knock-down experiments.

Table 3.1: List of genes with higher expression in ETCC-006 cells.

Table 3.2: List of genes with higher expression in ETCC-010 cells.

Table 3.3: List of genes with lower expression in ETCC-006 cells.

Table 3.4: List of genes with lower expression in ETCC-010 cells.

Table 4.1: RNA Integrity value of RNA samples used for the lncRNA-capture experiment.

Table 4.2: Performance metrics of the RNA sequencing after lncRNA capture with MCF-10A, MCF10DCIS.com, ETCC-006 and ETCC-010.

Table 4.3: List of differentially expressed lncRNAs in MCF10DCIS.com, ETCC-006 and ETCC-010 compared to MCF-10A identified by lncRNA-capture RNA sequencing.

Table 4.4: Performance metrics of the RNA sequencing after lncRNA capture with normal and tumor RNA from the DCIS patient cohort.

Table 5.1: List of proteins binding specifically to LOC729970, identified by mass spectrometry.

Table 5.2: Gene ontology analysis of the proteins binding specifically to LOC729970.

Table 5.3: PANTHER protein class analysis of the proteins binding specifically to LOC729970.

Table 5.4: Reactome pathway analysis of the proteins binding specifically to LOC729970.

Table 5.5: List of CYR61 protein interactors identified in the database IntAct.

Table 5.6: List of GAPDH protein interactors identified in the database IntAct.

PUBLICATIONS AND PRESENTATIONS

Publications

J. Samson, et al., Molecular and cellular characterization of two patient-derived carcinoma in situ (DCIS) cell lines, ETCC-006 and ETCC-010, under revision in *BMC Cancer*.

Samson J, Cronin S, Dean K. BC200 (BCYRN1) – The shortest, long, non-coding RNA associated with cancer. *Non-coding RNA Res.* 2018;3(3):131–43.

Zaheed O, **Samson J**, Dean K. A bioinformatics approach to identify novel long, non-coding RNAs in breast cancer cell lines from an existing RNA-sequencing dataset. *Noncoding RNA Res.* 2020;5(2):48-59

Presentations

Talk titled “Investigating the association between long non-coding RNAs and ductal carcinoma in situ (DCIS)” at the New Horizons In Medical Research Conference 2019 held on 3rd of December 2019 at UCC, Cork.

Poster presentation titled “Identification and characterization of the biological roles of long non-coding RNAs in early stage breast cancer, ductal carcinoma in situ (DCIS)” at RNA 2019 held from 11th to 16th of June 2019 in Krakow, Poland.

Poster presentation titled “Investigating the association between long non-coding RNAs and ductal carcinoma in situ (DCIS)” at the YCRN 2018 held on 21st and 22nd of June 2018 at UCC, Cork.

Poster presentations at the IACR 2017, 2018 and 2019.

LICENSES FOR THE USE OF PUBLISHED FIGURES

Figure 1.2

This Agreement between Ms. Julia Samson ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	4883130903928
License date	Aug 06, 2020
Licensed Content Publisher	Elsevier
Licensed Content Publication	Cell
Licensed Content Title	Hallmarks of Cancer: The Next Generation
Licensed Content Author	Douglas Hanahan, Robert A. Weinberg
Licensed Content Date	Mar 4, 2011
Licensed Content Volume	144
Licensed Content Issue	5
Licensed Content Pages	29
Start Page	646
End Page	674
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Title	Investigating the association between long non-coding RNA and ductal carcinoma in situ
Institution name	University College Cork
Expected presentation date	Sep 2020
Portions	Figure 6: Therapeutic Targeting of the Hallmarks of Cancer

Figure 1.5

Genetics Society of America - License Terms and Conditions

This is a License Agreement between Julia Samson ("You") and Genetics Society of America ("Publisher") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Genetics Society of America, and the CCC terms and conditions.

All payments must be made in full to CCC.

Order Date	06-Aug-2020	Type of Use	Republish in a thesis/dissertation
Order license ID	1053796-1	Publisher	Genetics Society of America
ISSN	1943-2631	Portion	Image/photo/illustration

LICENSED CONTENT

Publication Title	Genetics	Publication Type	e-Journal
Article Title	Long noncoding RNAs: past, present, and future.	Start Page	651
Author/Editor	Genetics Society of America.	End Page	669
Date	01/01/1916	Issue	3
Language	English	Volume	193
Country	United States of America	URL	http://www.genetics.org/
Rights holder	Genetics Society of America		

REQUEST DETAILS

Portion Type	Image/photo/illustration	Distribution	Worldwide
Number of images / photos / illustrations	1	Translation	Original language of publication
Format (select all that apply)	Electronic	Copies for the disabled?	No
Who will republish the content?	Academic institution	Minor editing privileges?	No
Duration of Use	Life of current edition	Incidental promotional use?	No
Lifetime Unit Quantity	Up to 499	Currency	EUR
Rights Requested	Main product		

Chapter 1

Introduction

Breast carcinoma is the most common cancer affecting women globally according to the World Health Organization (1); indeed, it is responsible for over 500,000 cancer deaths every year. Mammography has been historically, and remains presently, the main method for detection of breast cancer (2). While painless and cost-effective, mammography has detection limits, showing low resolution and sensitivity (3). This implies that using only mammography for the screening of breast cancer can result in either false negative where the tumor is not detected or false positive where there is actually no tumor. It is estimated that false negative results represent up to 20% of the screenings (4). Alternatively, false positive results can lead to an overtreatment of the patient that would have health implications in the future (4). The identification of novel biomarkers for early detection has been a focus in cancer research, in particular serum biomarkers (3). Long non-coding RNAs (lncRNAs) are part of the panel of serum biomarkers and will be the focus of this thesis in early breast cancer. In this introduction, we aim to give a brief overview of key features of cancer and particularly breast cancer (BC) and ductal carcinoma *in situ* (DCIS), lncRNAs and to highlight the gap of research between the two. We will start with a short introduction of the normal breast which is essential to understand the changes that occur in breast cancer. Next, we will focus on breast cancer, its origin and its characteristics. Finally, we will conclude with a presentation on lncRNAs, their diversity and their role both in normal and disease context.

1.1 Normal breast

The breast is a specialized structure, specific to mammals, that contain the mammary glands (5). Mammary glands, whose function is to produce and deliver milk to newborns, have been an essential part of the evolutionary success of mammals (6).

1.1.1 Anatomy of the breast in humans

The human breast is mainly composed of adipose and glandular tissue, both connected by a fibrous tissue called Cooper's ligament (5). The glandular tissue is functional and is responsible for the production and secretion of milk. Delivery of milk to the nipple occurs through a complex network of lobes and ducts. Each lobe contains several lobules, themselves divided into small alveoli (Figure 1.1). Small ductules extract the milk from the alveoli and transport it to a main duct to which they are connected, itself connected to the nipple (7). Interestingly, the duct walls are composed of two layers of epithelial cells. The inner layer contains epithelial cells with the ability to differentiate into milk secretory cells, while the outer layer of cells have a phenotype close to smooth muscle cells (8).

1.1.2 Development of the breast in humans

Breast development in humans is cyclic and divided in several phases. It starts at the neonatal stage where a system of small-branched channels is established in the mammary adipose tissue (9). The next growth phase happens around the puberty where the ducts expand and further branch to form secondary ducts. Puberty is also the stage where alveolar buds develop. Successful menstrual cycles allow the duct network to additionally branch to form a structure more and more complex (10). Breast

development is not complete until the pregnancy stage where the functional differentiation of the gland occurs to form an organ capable of producing milk. After the pregnancy, when the lactation is stopped, the breast undergoes through a remodeling which results in the loss of those functional structures (11). The breast morphology then resembles the one observed pre-pregnancy. However, the structure of the mammary gland is different from the pre-pregnancy stage, with an increase in lobe number. All the morphologic changes are tightly regulated by endocrine interactions involving mostly estrogen, progesterone and prolactin (5).

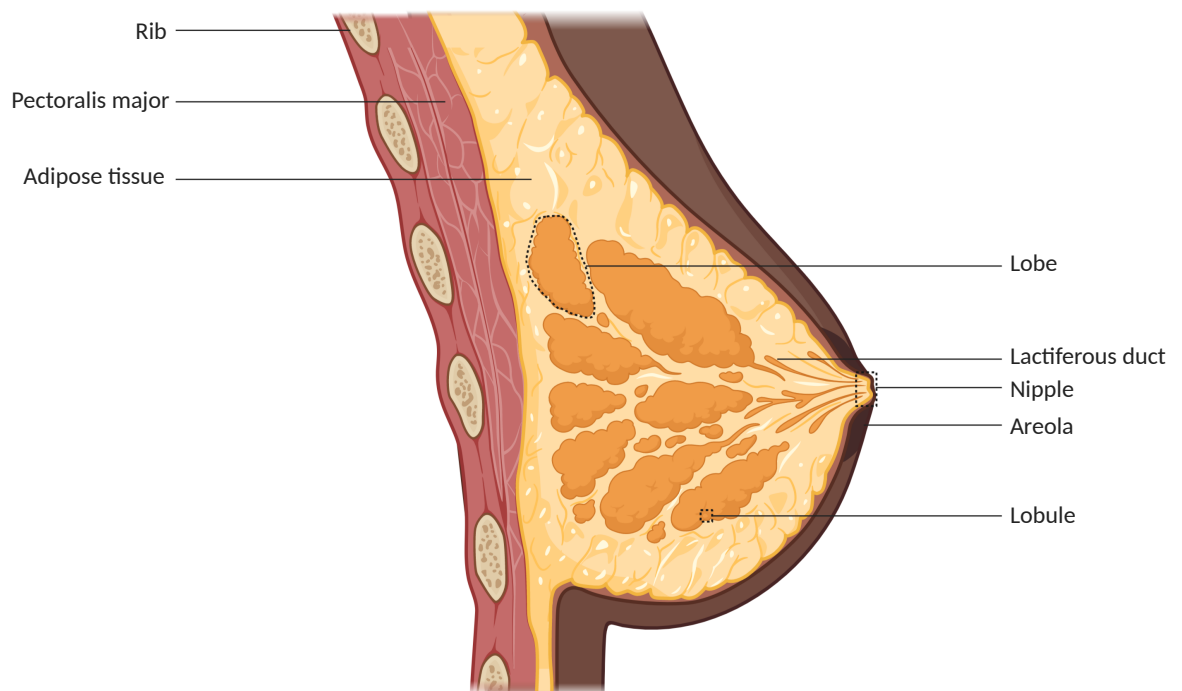


Figure 1.1: Anatomy of the human female breast. The mammary gland is composed of lobes interconnected through ducts. Each lobe is a complex structure subdivided in smaller lobules themselves subdivided into alveoli. During pregnancy, the lobes undergo differentiation to produce a mammary gland capable of producing and secreting milk through the network of ductules and ducts up to the nipple. Figure adapted from brainkart.com – Chapter: Essentials of Anatomy and Physiology – The Reproductive Systems (12) and drawn in BioRENDER.com.

1.2 Invasive breast cancer

Before focusing on invasive breast cancer (IBC) specifically, we need to define what cancer is, its causes and consequences. Cancer is a genetic disease (13) that can also be defined as the process that transform normal cells into malignant cells (14). During their transformation, malignant cells acquire capacities called hallmarks of cancer, recapitulated in many review articles (14–16), that allow them to grow abnormally (Figure 1.2). Indeed, malignant cells are capable of evading apoptosis (17), of avoiding the host immune system (18). Moreover, they show no susceptibility to growth suppressors factors that regulate the proliferation of normal cells. In fact, malignant cells do not require growth factors; indeed, signaling pathways responsible for cell growth are mis-regulated to maximize the cellular proliferation potential (19). By triggering expression of the telomerase, they acquire replicative immortality (20). Moreover, to further promote their growth, metabolic changes are induced to stimulate energy production (21). Angiogenesis is also promoted around the tumor mass for the transport of oxygen and nutrients (22). Interestingly, the host immune system attempts at fighting the cancer can result in a local inflammation, favorable for the cancer growth. Indeed, factors stimulating the angiogenesis and allowing apoptosis evasion are secreted, thus promoting tumor growth (23). Here it is important to highlight that a tumor can be either a benign or malignant mass. This difference relies mostly on the ability of malignant cells to invade other organs by a process called metastasis (24). High genome instability, including both genetic mutations and epigenetic modifications, is observed in cancer. The DNA damage response system is impaired which partially explains how cancer cells acquire those characteristics, making them perfectly adapted for an uninterrupted growth (25).

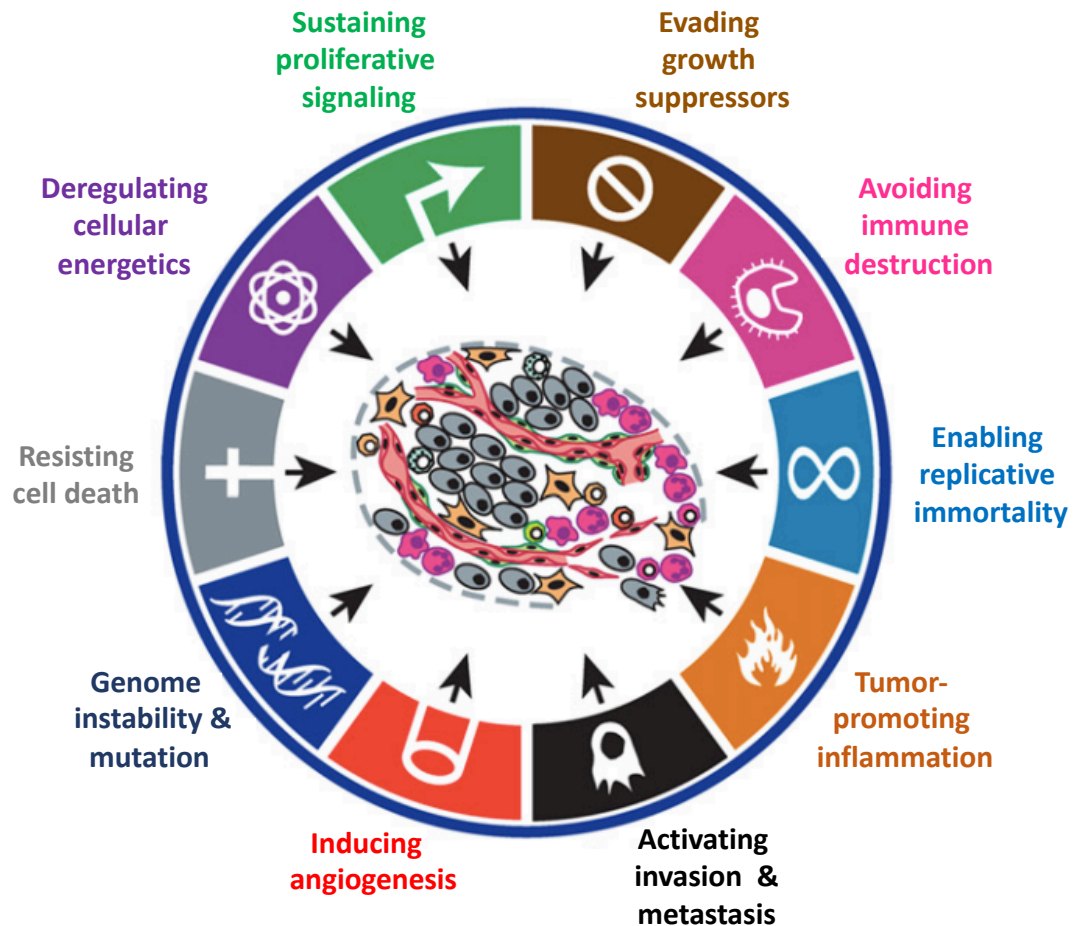


Figure 1.2: Next generation of hallmarks of cancer. This schematic highlights the strategies that cancer cells use to grow in the host organism. It summarizes cancer research conducted until 2011 and lays the foundations of the cancer research that has been conducted since then. Used with permission from Hanahan and Weinberg, Cell (2011) (15).

1.2.1 Definition and origin of IBC

IBC is simply defined as the development of a malignant tumor in the breast. It is associated with several risk factors that include environmental factors (diet, physical activity, obesity and alcohol consumption) and intrinsic factors such as age and genetic predisposition (26).

The genetic predisposition to IBC has been identified since 1990 with the discovery of the pathogenic mutations in the genes *BRCA1* (BRCAst CAncer gene 1) and *BRCA2* (BRCAst CAncer gene 2). Both genes contribute to DNA repair and transcriptional regulation in response to DNA repair (27). Mutations in *BRCA1* and *BRCA2* confer an increased risk of up to 85% to develop breast cancer (28). Other mutations associated with an increased risk of breast cancer have been identified since then. A few examples include *TP53* (Tumor protein 53 or p53), a master regulator of the G1/S checkpoint or *PTEN* (Phosphatase and Tensin homolog), a negative regulator of PI3K (phosphatidylinositol-3-kinase) that controls pathways involved in cell growth including cell proliferation and migration (29).

For cases where no genetic component is identified, cancer researchers have proposed a role for breast cancer stem cells (BCSCs) (30). Indeed, those show properties similar to those observed in normal stem cells (NSCs), including self-renewal, unlimited proliferation, slow replicative rate, high DNA repair capacity and the ability to divide into daughter cells with various differential potential. Cancer stem cells (CSCs), however, show little control of the self-renewal or differentiation program. Because they are resistant to chemotherapy, CSCs represent a challenge to completely eradicate the cancer (30). Interestingly, a hypothesis that the perinatal period of life influences the risk of developing breast cancer has been proposed (31). Indeed, breast stem cell number is decided *in utero* and the larger that number is, the more likely it is, to develop IBC during adulthood. Supporting this theory, a study has shown that breast cancer incidence is lower in women going through breast augmentation surgery (32). Moreover, *in utero* exposure to estrogen and IGF-1 (Insulin-like growth factor-1), a growth hormone, increases breast cancer risk. Indeed, an association between an *in*

utero low estrogen environment and a lower breast cancer incidence of the offspring has been shown in pre-eclampsia cases (33) or when the mother smokes during the pregnancy (34) – both notorious for inducing a reduction in maternal estrogen levels. Furthermore, *in utero* exposure to carcinogens in rat models leads to the development of breast cancer in the adult animal (35), thus supporting the hypothesis that BCSC play an essential role in breast cancer development as early on as the prenatal period. However, that theory does not fully explain what drives breast tumorigenesis. Indeed, a high degree of heterogeneity is observed in malignant tumors, yet according to the BCSC theory tumor growth should occur by clonal expansion. In reality, tumorigenesis is probably the result of the combination of clonal expansion from BCSC and from more differentiated progenitors cells (36–38).

1.2.2 IBC classifications

IBC can be classified differently depending on the criteria analyzed. One type of classification are based on the tumor grade with the histological grading system – the Tumor, Node, Metastasis (TNM) classification or the molecular subtypes (39).

Histological grading can be divided into four grades (grade 1 to 4), with grade 1 breast cancer displaying tumor cells that are most similar to normal cells, moving to grade 4 where tumor cells proliferate rapidly and show mobility (39). In breast cancer specifically, the Nottingham histological system is the most commonly used. It is determined by the evaluation of tubule formation, the degree of nuclear pleomorphism and mitotic count and is divided in three grades, from grade 1 associated with better prognosis to grade 3 associated with the worst prognosis (40). In comparison, the TNM classification takes into consideration the size of the tumor (from 0 no tumor, to 4

where the tumor has grown into the chest wall and skin), whether it has spread to the lymph nodes (from 0 no tumor cells in the lymph nodes, to 3 tumor cells are present in more than ten nodes) and whether the tumor has spread to other organs (39). Taken together, those criteria divide tumors in five stages, stage 0 applying to non-invasive BC and stage 1-4 applying to IBC.

IBC develops mainly in two part of the breast (Figure 1.1), the duct where it is called invasive ductal carcinoma (IDC). It can also develop in the lobes where it will be called invasive lobular carcinoma (ILC), this only represents 10% of IBC cases. ILC differs from IDC in epidemiology, through molecular alterations, the treatment required and patient outcome, ILC showing a worse outcome for stage-matched IDC (41). We will focus on IDC from this point.

IDC is classified clinically based on the expression of three hormone receptors: estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2/ERBB2) (42). Hormone receptor (HR) positive IDC is characterized by expression of ER/PR, and it represents up to 60% of IDC cases. It is treated with hormonal therapy including tamoxifen, a molecule that blocks estrogen binding to the estrogen receptor to ultimately block cell proliferation (43) and aromatase inhibitors, agents blocking estrogen production (44). HR positive IDC is subdivided into two subtypes, luminal A and B. Luminal A tumors tend to be ER and/or PR positive with negative HER2; whereas, luminal B tumors are mostly ER and/or PR positive with positive HER2 (or high expression of the proliferation marker Ki67 and negative HER2). Tumors where HER2 is overexpressed, with or without expression of ER/PR, constitute their own subtype called HER2 positive. They

represent about 20% of all IBC cases (42). Those are treated with drugs targeting HER2. Finally, triple negative breast cancer (TNBC) applies to any tumor that show no expression of ER/PR/HER2. Those represents up to 15% of IBC cases. Since no expression of HR is observed, hormonal therapy is not helpful for those cases. Treatment of TNBC includes a combination of breast surgery removal, radiotherapy and chemotherapy (42). Recently, several studies have highlighted the efficacy of immunotherapy treatment for TNBC treatment. Immunotherapy induces T-cell activation, in charge of recognizing cancer cells and activating the adaptive immune system response, through the targeting of programmed cell death protein 1 (PD-1) (45). The combination of immunotherapy and chemotherapy has shown promising results in TNBC treatment (45).

In addition to being characterized by the expression of different molecular markers, each IDC subtype is associated with different patient outcome. HR positive IDC is associated with the best overall survival after diagnosis with about 85% of patients experiencing more than 5-years survival. HER2 positive subtype is usually more aggressive than HR positive and has been associated with a poorer prognosis until recently. Indeed, advances in treatment comprising a combination of immunotherapy targeting HER2 and antibody-drug conjugates have greatly improved the prognosis of HER2 positive BC (46). TNBC, however is associated with the worst prognosis and a high risk of recurrence compared to other subtypes (42). Interestingly, with tumor heterogeneity, it has been speculated that cells belonging to different subtypes cohabit and cooperate together within the same breast tumor (47).

1.2.3 IDC incidence and statistics in Ireland

In Ireland, it is estimated that for 100,000 people from the general population, 123.5 women and 1 man will develop breast cancer every year. This translates into a risk of 1/10 to develop breast cancer for women before the age of 75 (48). Those statistics include non-invasive breast cancer cases; however, the National Cancer Registry Ireland estimates that almost 30% of invasive cancer occurs in the breast in women (48). Regarding mortality, for 100,000 women 25 will die of breast cancer, translating into a risk of 1/54 of dying from breast cancer before 75. While the mortality rate associated with breast cancer in women is not negligible, breast cancer only comes second in the list of death associated to cancer after lung cancer (48). This highlights the quality of treatments available for breast cancer patients and the importance of early diagnosis, with over 70% of breast cancer cases first diagnosed at stage I and II (48).

1.3 Ductal Carcinoma *In Situ*

1.3.1 Definition and origin of ductal carcinoma *in situ*

Ductal carcinoma *in situ* (DCIS) is a non-obligate precursor to invasive breast cancer. It results from an abnormal proliferation of epithelial cells in the duct of the breast and is associated with a lack of invasive growth in the local stroma (49). The first step toward DCIS tumorigenesis is called atypical ductal hyperplasia, corresponding to the stage where epithelial cells start to abnormally proliferate (47, 48) (Figure 1.3). DCIS stage is reached when the duct is mostly filled with tumor cells. If the tumor cells break out from the duct to invade local tissues, it then turns into IDC. While the reason

remains still mostly unknown, DCIS transitions into IDC in 15-50% of cases (52). Yeong et al. recently reviewed several factors possibly implicated in DCIS progression to IDC with a special attention given toward the tumor microenvironment (53). Hence, a remodeling of the stromal extra cellular matrix, a change in expression of certain genes involved in normal myoepithelial cell function and a role for cytokine signaling between DCIS cells and cancer-associated fibroblasts has been proposed, although each remains to be further investigated.

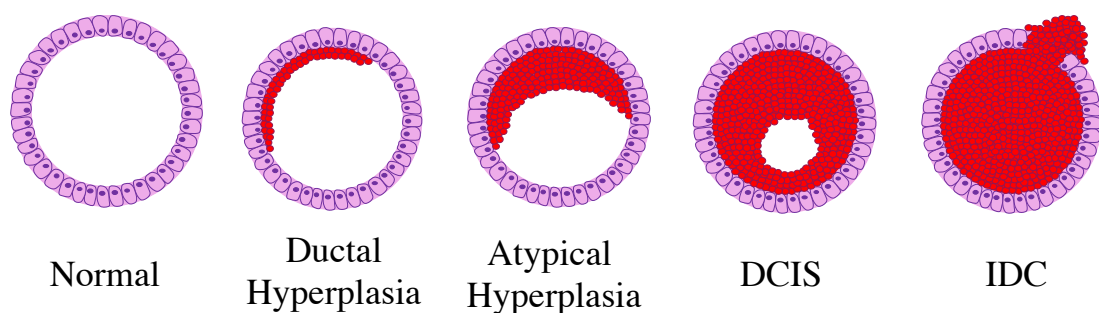


Figure 1.3 Tumorigenesis process in the duct of the breast. Epithelial cells abnormally proliferate and colonize the duct as stages progress until the DCIS stage where the duct is mostly full. Once tumor cells invade neighboring tissue, it evolves into IDC. Adapted from Casasent, Edgerton and Navin (2016) (51).

Since 1982, the use of mammography for screening purposes has translated into an increase in DCIS incidence. DCIS incidence varies with age. While it still remains low in women under the age of 40, partially due to a lack of screening, it increases with age and peaks around 70-79. DCIS-associated risk factors, and more generally what causes its development, remains poorly understood. Some studies seem to indicate that DCIS shares common environmental risk factors with IDC. Hence, alcohol intake and hormone replacement therapy are associated with an increased risk of developing DCIS (49). However, no clear association between obesity and DCIS has been shown (54).

Genetic predisposition to DCIS has been poorly studied. While mutations in *BRCA1* and *BRCA2* are associated with an increased risk of developing DCIS, other genetic associations remained unclear until recently (55). In 2016, Petridis et al. (56) aimed to compare if DCIS and IDC share similar genetic predispositions and whether those association are of similar strength. They found that 67% known IDC-associated loci are also associated with DCIS, with ER positive, IDC-associated loci also associating with ER positive DCIS. Case-only analysis showed no difference in the strength of the association. Moreover, no loci associated specifically to DCIS were found. This large-scale study, comprising 5,067 cases of DCIS, 24,584 cases of IDC and 37,467 controls pooled from 38 studies, highlights DCIS and IDC common genetic predispositions, in addition to sharing common environmental risk factors. Taken together, this suggests that DCIS and IDC are similar; however, the fact remains that DCIS does not transition to IDC in every patient. This implies that genetic predisposition alone is not responsible for the transition from DCIS to IDC and that other factors should be investigated to understand that transition.

1.3.2 DCIS classification

DCIS is characterized by a broad spectrum of cases, which made necessary the need for a specific classification system. Traditionally, DCIS is classified based on the predominant architectural pattern of the lesion, including comedo, solid, cribriform and micropapillary (57). Comedo DCIS is the most recognizable DCIS type due to the high presence of cells with abnormal morphology. At the center of duct, comedo DCIS is associated with necrosis and microcalcification, most likely due to a lack of oxygen and nutrients (58). Moreover, comedo DCIS lesions tend to be larger than other types of DCIS and more often associated with microinvasion.

As for other types, solid DCIS is formed by a compact proliferation of cells that fill almost entirely the duct. It can be associated with some necrosis and microcalcification and can be present in combination with comedo DCIS (57). Cribriform DCIS is identified by the presence of small duct-like structures in the duct. Those sub-ducts are round and rigid and may present some calcifications (57). Micropapillary DCIS is characterized by the presence of papillary structures that extend into the ductal lumen. Microlumens can be formed in micropapillary DCIS, sometimes associated with calcifications, similar to those observed in cribriform DCIS (57).

As suggested, several types of DCIS can be found within the same lesion. This pushed the scientific community to find a more appropriate classification system (59). They settled on nuclear grading, which is more reproducible than the previous classification system and associated with better prognostic values. This grading system is divided into three categories: low, intermediate and high-nuclear grade (60). Low nuclear grade is characterized by the proliferation of cells with small round/oval nuclei and associated with rare mitoses. It is the closest phenotypically to normal cells (57). On the opposite, high-nuclear grade is associated with large cells, their nuclei showing different morphologies. Cells are highly mitotic and necrosis and calcification are often present (57). As expected, intermediate nuclear grade lies in between. Although necrosis can be seen in all stages, its likelihood increases with the grade (61). It extends from a single cell to a whole region like what is observed in comedo DCIS. On the opposite, cell polarization is more common in low nuclear grade DCIS since those look the most like normal cells and its likelihood decreases with the stage (61).

1.3.3 Detection and treatment of DCIS, an on-going debate

As mentioned previously, DCIS is first diagnosed by mammography then confirmed by analysis of a core needle biopsy. Once the diagnosis is done, the lack of knowledge allowing to determine which DCIS cases will develop IDC if untreated forces the medical body to proceed to DCIS treatment.

DCIS treatment involves a combination of breast surgery and post-surgery therapy (62). DCIS surgery treatment ranges from a local excision (also referred to as breast conserving surgery - BCS), where the tumor and surrounding tissue is removed, to a mastectomy, where the whole breast is removed. While mastectomy was used traditionally, the recent success of BCS has now orientated treatment toward this approach, especially in cases where the DCIS lesion is well localized (49). However, in patients associated with additional BC risk-factor, such as carrying mutations in a *BRCA* gene, mastectomy is still preferred. Post-surgery treatment mainly includes radiation therapy and hormonal therapy. Radiation therapy is recommended when BCS is performed to reduce the risk of local recurrence. This is especially necessary in high-nuclear grade DCIS considered as high-risk DCIS (49).

Some studies have investigated if radiation therapy could be bypassed in low-risk DCIS cases. However, the identification of a DCIS group showing no benefit from radiation therapy is still unclear. The benefit of hormonal therapy in DCIS treatment is still being discussed. Indeed, tamoxifen has been investigated in several clinical trials post-BCS with inconsistent results. While a significant reduction of BC risk occurrence was shown in the group with tamoxifen treatment, the follow up study twelve years later showed no difference in patient overall survival between the treated

and control group (40, 50, 51). This is largely explained by the study design, as tamoxifen treatment was not given according to the patient's ER status. Hence, a recent study has now shown that tamoxifen treatment in ER positive DCIS is associated with a decreased risk of BC occurrence and better overall survival, while no effect is evidenced in the ER negative DCIS group (65). However, long term intake of tamoxifen is associated with several side effects including hot flashes, vaginal secretion, thrombosis and endometrial cancer (66). This suggests that tamoxifen overtreatment of DCIS can have dramatic consequences, highlighting the need for clinical studies to strongly categorize patients that would benefit from tamoxifen treatment. Alternatively, treatment with aromatase inhibitors are under study. While those inhibitors seem to benefit ER positive DCIS cases as much as tamoxifen, they are also associated with important side effects including an increased risk of bone fractures (54, 55).

Some studies estimate that 20% to 40% of DCIS cases are HER2 positive ; however, it is not routinely tested in DCIS. In 2013, a trial was initiated to investigate the benefits in HER2 testing in DCIS (69). This could have clinical implications as an anti-HER2 therapy and could then be implemented in DCIS treatment.

As suggested previously, DCIS treatment is surrounded by many controversies centered on the same question – whether it is over-treated (57, 58). Indeed, improvement in screening tools has translated into an increase in the number of women diagnosed with DCIS (72). However, despite assumptions of the opposite, IDC incidence has not decreased. This implies that DCIS is over-diagnosed and over-treated. Thus, it has been proposed that instead of being treated, DCIS patients should

be under active surveillance to monitor whether the DCIS lesion progresses (73). Indeed, in lobular carcinoma *in situ*, a lesion with similar risk of progression to IBC as DCIS, active surveillance is offered to patients (74). However, overtreatment of DCIS may be a necessary step until DCIS progression to IDC is better understood. There has been a research focus of the DCIS to IDC transition through the identification of biomarkers that will allow patient stratification.

1.4 DCIS-known markers

In this section, we will review some of DCIS-known markers. DCIS has been characterized at different levels mostly related to DNA with genomic alterations, epigenetic modifications and dysregulation of gene expression. Some specific RNAs and cancer-associated protein coding genes have also been investigated in the context of DCIS. A few examples will be presented here.

Interestingly, several types of cancer, including BC, have been associated with tumor infiltrating lymphocytes, which are under investigation for their value as a prognostic tool and their role in therapy (75). Similarly, DCIS has been characterized by the presence of a specific population of tumor infiltrating lymphocytes (76). However, since the exact function of these cells and their link to risk of recurrence remains unclear, we will not be presenting those results.

1.4.1 Genomic alterations specific to DCIS

Most studies investigating genomic alterations in DCIS seem to hint that changes occur mostly in high nuclear grade DCIS (77,78). However, low-grade DCIS is also characterized by some genomic alterations, especially with the chromosome 16q loss

(79) and 1q gain (80). In high-grade DCIS, there tends to be a frequent gain of 5p, 8q, 17q and 20q, frequent amplifications of 11q13, 17q12 and 17q22-24 (76, 77) and a loss of 8p, 11q, 13q and 14q (79). Interestingly, 8p loss differs among DCIS cases. Indeed, in high-grade DCIS a whole arm loss is observed in 65% of cases. In low to intermediate DCIS, a partial loss of the 8p arm occurs in combination with a proximal gain in 29% of cases. The whole arm 8p loss observed in high-grade DCIS can happen in low to intermediate DCIS but that only concerns 12% of cases (78). Poorly differentiated DCIS is characterized by a loss of heterozygosity in chromosome 17, in addition to regions 6q25-q27, 8q24, 9p21, 13q14 and 17p13.1 (77, 79).

The work presented here is mostly focused on low and high nuclear grade since intermediate grade seems to share genomic alterations with both low and high-grade and no clear pattern specific to intermediate nuclear grade DCIS has been evidenced (78).

Regarding individual genes present within chromosomal regions altered, *MYC* (8q24) and *HER2* (17q12) have been shown to be amplified in high nuclear grade DCIS (78). Moreover, high nuclear grade DCIS is also specifically characterized by a copy number alteration of *CCNE1* (19q12) and *AURKA* (20q13) and amplification of *CCND1*. Somatic mutations have been evidenced in the gene *PIK3CA* in both mixed DCIS, where a mix between of DCIS and IDC is observed, and pure DCIS but findings suggest that those mutations do not drive the transition into IDC (73, 80, 81).

1.4.2 Epigenetic modifications in DCIS

With epigenetic modifications proven to be drivers in the carcinogenesis process, those have been studied in DCIS to investigate whether they would constitute a better tool to evaluate overdiagnosis and overtreatment. DNA methylation, a reversible DNA modification, is deposited onto CpG islands, GC-rich regions present around gene promoters, to regulate gene expression (87). Indeed, dense methylation induces the silencing of the promoter. Loss of methylation is characteristics of many cancers to allow expression of oncogenes. However, hypermethylation of the promoter of tumor suppressor genes is most commonly observed (87). Two different studies comparing DCIS to normal and IDC (88,89), identified 72 common loci with differential methylation statuses, with mostly homebox and polycomb genes. This needs to be further confirmed in additional studies. However, methylation patterns comparing normal tissue from healthy people, adjacent normal tissue in DCIS patients and DCIS tissue have shown that adjacent normal tissue already shows wide-spread changes in methylation (85, 86). This implies that DNA methylation plays a role in early BC progression.

Histone modifications constitute a major regulator of gene expression, along with DNA methylation. Those modifications mostly include mono- di- tri-methylation and acetylation, but phosphorylation and ubiquitination are also observed. Histone modifications form a complex and very plastic system controlled by protein “writers” that deposit the marks, “readers” that read them and “erasers” that erase them (87).

Each modification has a unique role. For example, mono-methylation of H3K9, H3K27 or H3K79 is mostly associated with active transcription; whereas, di-

methylation is mostly associated with a repressive role (85, 86). However, when a di-methyl mark is deposited onto H3K79, it is associated with activation. Tri-methylation of H3K27 is repressive. Interestingly, the protein responsible for its deposition shows expression already at the DCIS stage. This supports the hypothesis that epigenetic remodeling occurs early in the breast carcinogenesis (87, 88). Similarly, lysine-specific demethylase 1 (*LSD1*), a histone methylation eraser, is expressed at increasing levels from grade 1 DCIS to grade 3 DCIS (96).

Acetylation is mostly associated with active transcription. Compared to histone methylation, histone acetylation is fairly stable between normal and adjacent normal tissue. The most noticeable change in DCIS is a loss of H4 acetylation in normal adjacent, suggesting again an early event in cancer progression (97). However, histone modifications studies in cancer have mostly focused on IBC and very few studies have investigated those early events in the breast carcinoma development (87).

Super enhancers are formed when multiple enhancers are clustered together through chromatin interactions induced by changes in chromosome conformation. A study has shown that in the BC cell line MCF-7, *RACK7*, a tumor suppressor gene, occupies the most active super-enhancers regions to repress expression of the *S100A* (calcium-binding protein) family of oncogenes (98). Analysis of *RACK7* expression in matched DCIS and IDC biopsy of six patients shows a lower expression in IDC. Although, more studies on the subject need to be conducted, mis-regulation of super enhancers might play an important role in tumor development and the downregulation of *RACK7* might participate in the transition from DCIS to IDC (98).

1.4.3 Gene expression panels for the prediction of DCIS 10-year risk of recurrence

Some gene expression panels have been developed to predict the 10-year risk of BC of recurrence. Formerly Genomic Health, a commercial laboratory has established a test specific for DCIS, for patients undergoing BCS without radiation (99). This test called Oncotype DX DCIS test (100) is derived from the Oncotype DX test (101) that is characterized by a twenty-one gene signature list with sixteen cancer-related genes and five reference genes. Since those tests have been getting a lot of attention from the medical body, we will briefly present the test specific to DCIS.

Using a subset of twelve genes from the twenty-one gene signature, the 10-year risk of DCIS recurrence is calculated. Among those genes, five are related to proliferation (*MKI67* – Ki67, *STK15* also called *AURKA*, *BIRC5* – Survivin, *CCNB1* – Cyclin B1 and *MYBL2* - MYB Proto-Oncogene Like 2). The Oncotype DX DCIS test also includes *PR* and *GSTM1* (Glutathione S-Transferase M1) and five references genes (102). Risk of 10-year recurrence is then divided into three categories, low, intermediate and high risk.

The adequacy of the DCIS Oncotype test to predict BC recurrence was analyzed by comparing histopathologic features of DCIS lesions with their respective calculated score (99). Hence, high risk scores are associated with dense chronic inflammation around the lesion, a symptom often found in more advanced cases, high mitotic count and no expression of *PR*. On the opposite, low risk scores are associated with no sign of inflammation, expression of *PR* and low mitotic count, all associated with good prognosis (99). Overall, this study suggests that those histopathologic features

(inflammation, mitotic count and expression of *PR*) correlate with the DCIS oncotype score for low and high-risk cases. Since histopathologic features cannot identify intermediate risk cases, they could not be compared to the Oncotype DX DCIS test.

Interestingly, two studies show a discrepancy between histological features and the Oncotype DX DCIS score with around a fifth of the cases where the score and the histopathologic features do not correlate (103,104). However, an analysis was conducted on two different cohorts (104–106): ECOG-ACRIN E5194 with 561 cases of low to intermediate grade DCIS according to histopathological features and 104 cases with high-grade DCIS and the Ontario DCIS cohort ($n = 3,762$). They show that for 10% to 15% of cases, low risk clinicopathological features had an intermediate to high risk of local recurrence with the Oncotype DX DCIS test. This implies that the DCIS Oncotype score can identify cases with low risk features, but that are associated with high risk of local recurrence, if no additional treatment after BCS is carried out.

1.4.4 Protein markers of DCIS

With DCIS being a non-obligate precursor to IDC, many proteins characteristic of IDC's aggressiveness are assessed at the DCIS stage. Hence, many studies have detected expression of some cancer-associated protein in DCIS and we have chosen to present only a few key examples.

Hence, ER, PR, HER2 and Ki67 are used to determine the aggressiveness of the DCIS tumor (71). DCIS lesions that show no expression of ER, PR and/or high expression of HER2 and/or high Ki67 are associated with aggressiveness and a higher risk of local

recurrence. On the opposite, expression of ER, PR and low Ki67 is associated with low grade DCIS.

In 2014, McNamara et al. (107) investigated the role of the androgen receptor (AR), a steroid hormone nuclear receptor (108) and related metabolic enzymes in DCIS aggressiveness. They showed that low levels of AR and androgen-metabolizing enzymes are associated with an increased aggressive behavior in TNBC and DCIS. However, this needs to be further investigated.

p53 is a tumor suppressor induced in response to DNA damage (109) and responsible for multiple cell cycle checkpoints control (110). In cancer, *TP53* is the most frequently mutated gene (111) and evidence has shown that mutations in the *TP53* gene can be found in some DCIS cases (66, 107). Those are associated with a more aggressive behavior and should therefore be treated accordingly with a combination of radiation therapy and BCS (112).

Levels of p16, a checkpoint regulator of the G1 to S phase transition (113), have been measured by immunohistochemistry in 41 DCIS cases (109, 110). High p16 levels are associated with a higher risk to evolve into IDC.

p21 expression is activated by p53 in response to DNA damage. It inhibits the transition from the G1 to S phase, and in some cases can potentially inhibit the transition from the G2 to M phase (116). Hence, while p21 is expressed in a third of DCIS tumors, only one study showed an association between its expression and local recurrence (117).

Interestingly, expression of another regulator of the G1 to S phase transition, cyclin D1, (118) correlates with a lower risk of local recurrence in DCIS. It is most likely due to some pro-apoptotic function of the protein (112, 114).

Bcl-2 is a negative regulator of apoptosis (120). While it is expressed in half of DCIS cases, Bcl-2 shows no association with local recurrence (117).

COX-2 is a regulator of angiogenesis in normal breast. In tumors, COX-2 function is unknown, although it seems to be linked with oncogenic functions such as increased proliferation and decreased apoptosis (115). Hence, in IDC, COX-2 expression has been shown to be associated with a shorter disease-free survival (115). COX-2 is also overexpressed in DCIS. Therefore, it has been assumed that this may increase the risk of DCIS local recurrence. However, the role of COX-2 as a prognostic marker for DCIS recurrence remains to be investigated with studies only showing a correlation with cell proliferation (121) and studies concluding on an association with DCIS recurrence (115).

1.4.5 miRNA markers

miRNAs are a class of small RNAs with regulatory functions in gene expression. They have mostly been shown to negatively regulate gene expression by binding to their RNA transcript. This induces the recruitment of enzymes involved in mRNA decay and ultimately mRNA degradation (122). miRNAs have been associated with cancer where they can have either oncogenic or tumor suppressor functions (123). While studies have mostly focused on IDC, the role of miRNAs in DCIS and its progression towards IDC is now under investigation.

The first study exploring miRNAs in breast epithelial cells compared the expression of miRNAs in epithelial cells of healthy normal tissue, IDC tissue and cell lines (123). Hannafon et al. (2017) identified miRNAs whose expression is restricted to a specific cell type (123). From this work, miR-145 and miR-205 were expressed specifically in the myoepithelial cell layer of normal tissue, while let-7a, miR-21 and miR-214 were expressed in the luminal epithelial cell layer. Interestingly, let-7a was present in normal tissue, and miR-21 was present in tumor tissue. Unlike other miRNAs, miR-214 was associated with a complex expression pattern (123).

In 2011, Hannafon et al. (124) had previously compared the miRNA profile of healthy normal tissue, matched adjacent normal and DCIS tissue. They identified thirty-five miRNAs differentially expressed between those groups, of those 18 had already been implicated in IDC (125). This suggests that changes in miRNA expression could constitute an early event in BC development.

DCIS seems to be characterized by an increase in miR-21 expression and a decrease in many miRNAs including miR-98 and let-7, with those changes being conserved in IDC (126). A reanalysis of this same study identified 66 miRNAs showing lower expression in DCIS compared to normal whereas only nine showed differential expression in IDC compared to DCIS. This suggests that most changes in miRNA expression occur early in the carcinogenesis process, at a stage as early as atypical ductal hyperplasia (ADH). Indeed, a study comparing ADH, DCIS and IDC shows that the most significant changes in miRNAs expression occur between the transition from normal to ADH (127).

We recapitulate the markers described in this section in Table 1.1.

<i>Marker</i>	Tissue	Type of alteration	Target	Reference
<i>Genomic alteration</i>	Low grade DCIS	Loss	16q	74
			whole 8p arm loss (12% of cases)	78
		Gain	1q	75
			partial loss of 8p with proximal gain (29% of cases)	78
	High-grade DCIS	Loss	8p, 11q, 13q and 14q	74
			whole 8p arm loss (65% of cases)	78
		Gain	5p, 8q, 17q and 20q, 11q13, 17q12 and 17q22-24	76,77
			MYC (8q24) and HER2 (17q12), CCND1	73
			copy number alteration of CCNE1 (19q12) and AURKA (20q13)	73
	Poorly differentiated DCIS	Loss	hetetozygosity in chromosome 17 and in regions 6q25-q27, 8q24, 9p21, 13q14 and 17p13.1	77, 79
<i>Epigenetic modification</i>	Mixed DCIS	Mutation	PIK3CA gene	73, 80, 81
	Comparison DCIS and IDC	Differential methylation	72 loci including mostly homebox and polycomb genes	83, 84
		Super enhancer	RACK7 expression higher in DCIS	93
	Comparison DCIS and normal adjacent	Differential methylation		85, 86
		Histone modification	Loss of H4 acetylation in adjacent normal	92
	DCIS		H3K27	87, 88

Gene expression panel	DCIS grade 1 to 3		LSD1, methylation eraser, level increasing with DCIS grade	91
	DCIS – Oncotype DX DCIS		MKI67, STK15, BIRC5, CCNB1, MYBL2, PR, GSTM1 and five references genes	95, 97
	DCIS		ER, PR, HER2 and Ki67	66
Androgen receptor and related metabolic enzymes			102	
p53			66, 107	
p16			109, 110	
p21			112	
Cyclin D1			112, 114	
Bcl-2			112	
COX-2			110, 116	
miRNA	Myoepithelial cell layer of normal tissue	Differential expression	miR-145 and miR-205	119
	Luminal epithelial cell layer of normal tissue		let-7a	119, 120
	Luminal epithelial cell layer of tumor tissue		miR-21	119
	DCIS		miR-98 and let-7	120

Table 1.1: Summary of the DCIS-known-markers presented in the section above.

A few examples of genomic alterations, epigenetic modifications, protein and miRNA markers of DCIS are recapitulated in this table.

1.5 Non-coding RNA

Advances in high throughput sequencing technologies have challenged the old mRNA-centric dogma by revealing that 62% to 75% of the human genome is transcribed (122, 123). Those studies have shown that transcription is not specific to protein coding genes and that it, actually, initiates in many genomic regions. This transcriptional activity, also called pervasive transcription in reference to its widespread nature (129), is regrouped under the term of non-coding RNA since it does not encode for a protein. While the function of some non-coding RNAs such as tRNAs and rRNAs have been known since the 1950s, the role of most of that transcriptional activity is still unknown (130). Non-coding RNAs have been subclassified based on their size with RNAs shorter than 200 nucleotides called small non-coding RNAs and transcripts longer than 200 nucleotides called long non-coding RNAs. Small non-coding RNAs are mostly regulatory molecules (131). MicroRNAs (miRNAs) and endogenous short interfering RNAs (siRNAs) are both responsible for repression of gene expression (131). Piwi-associated RNAs (piRNAs) are responsible for the repression of transposable elements in the germ line (132). Small nuclear RNAs (snRNAs) are involved in splicing events and small nucleolar RNAs (snoRNAs) act as guides for chemical modifications of other types of RNAs (132).

Long non-coding RNAs constitute a large portion of the transcriptional activity observed in humans. Those are involved in a wide variety of functions in normal and pathologic contexts that we will describe in the next section (132).

1.5.1 Definition and classification of lncRNAs

As mentioned, lncRNAs are defined by a length over 200 nucleotides and generally lacking protein-coding potential. As a class of RNAs, they include a large variety of transcripts implicated in many functions. Since their discovery is relatively recent, their sub-classifications have mostly been based on empirical attributes originally used to detect them (128). Thus, lncRNAs are subdivided in different classes of transcripts including long intergenic non-coding RNAs (lincRNAs), circular intronic RNAs (ciRNAs), antisense RNAs (asRNAs) and enhancer-associated RNAs (eRNAs) (Figure 1.4) (133,134). In the next section, we will focus on the most studied classes of lncRNAs.

RNA sequencing (RNAseq) experiments using total RNA instead of the polyadenylated fraction, led to the discovery of very long intergenic transcripts of up to 1 meganucleotides in size (135). Corresponding to 10% of the human genome (136), those transcripts are involved in numerous biological functions including pluripotency, cancer, apoptosis, cell cycle progression and senescence (126, 129).

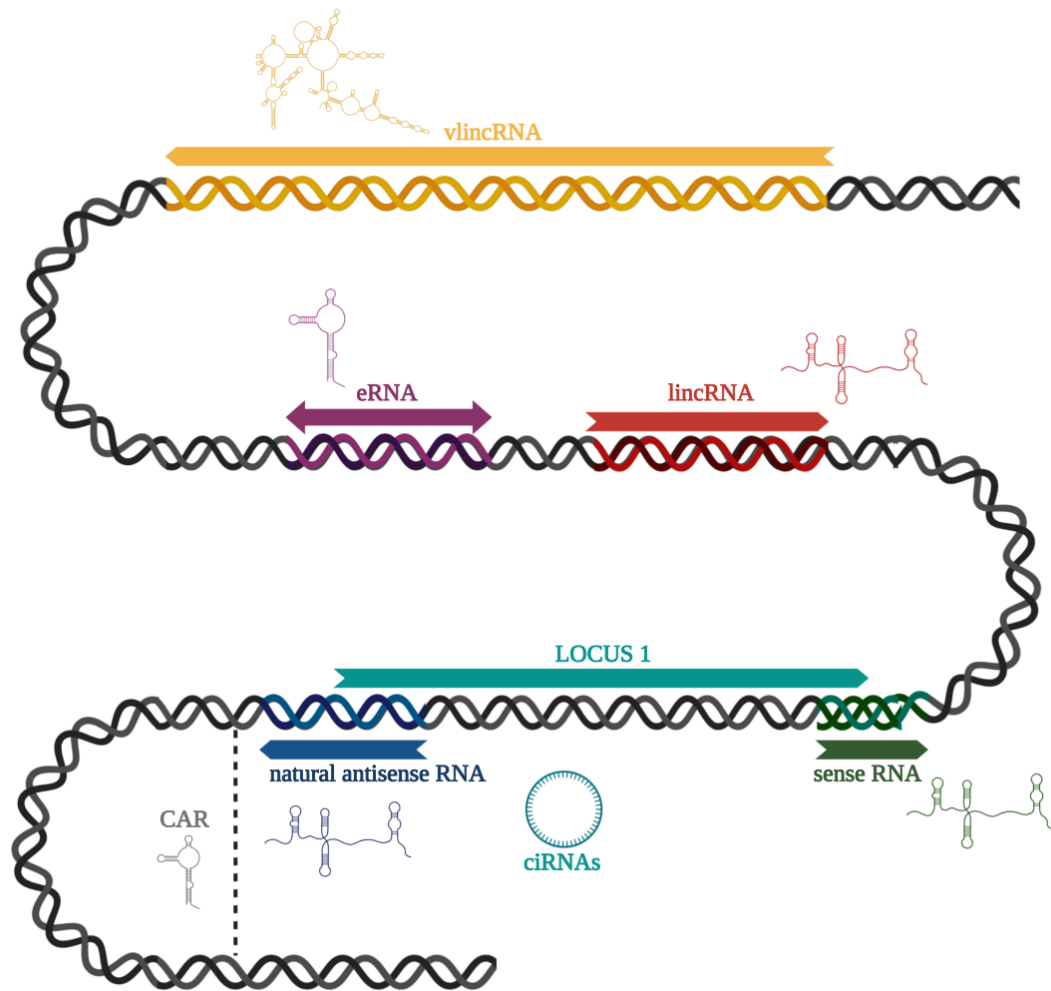


Figure 1.4: Sub-classification of long non-coding RNAs. A hypothetical protein coding gene locus is represented in green (locus 1). Different classes of non-coding RNAs are represented. CAR: chromatin-associated RNA; ciRNA: circular RNA; eRNA: enhancer-associated RNA; lincRNA: long intergenic non-coding RNA; vlinRNA: very long intergenic non-coding RNA. Figure adapted from St Laurent et al. (134).

Some lncRNAs are transcribed using the same sequence as protein-coding genes. In consequence, they share some sequence similarity; however, they do not code for a protein. Those are called sense overlapping. They can be composed of a combination of exons and introns from the protein-coding gene (138). In some cases, the lncRNA

sequence is totally encoded in the intronic sequence of the protein-coding gene (139). A single-molecule sequencing experiment has estimated that those lncRNAs constitutes 40 to 50% of all cellular RNAs by mass, if ribosomal RNA is excluded (140).

In 50 to 70% of protein-coding genes, lncRNAs transcripts are generated from the antisense DNA strand of the one encoding the protein-coding gene (141). Some promoter and enhancer elements have also been shown to produce non-coding transcripts (137). Enhancer producing RNAs have a better chance to be functional in reporter assays compared to those that do not (142). This suggests that the eRNA has a functional role.

lncRNAs have a strong relationship with repeated regions and transposable elements. Indeed, many lncRNA promoters are presents in a repeated region, especially those associated with the pluripotent context (137). Interestingly, lncRNAs in cancer and pluripotency share common regulatory architectures, adding support to the stem cell theory in BC development (143). This strong relationship is also supported by the choice of the RNA polymerase responsible for their transcription. Indeed, while lncRNAs that resemble mRNAs and have a polyadenylated tail are mostly transcribed by the RNA polymerase II, lncRNAs with a strong association and/or ancestry with repeated elements are transcribed by the RNA polymerase III, also responsible for the transcription of Alu, B1 and B2 elements (144).

Competing endogenous RNAs (ceRNAs) are regulatory transcripts competing with a given mRNA for regulatory molecules including miRNAs. Hence, any transcripts with

sequence similarity to an mRNA is potentially a ceRNA, including transcribed pseudogenes (145).

Some lncRNAs that serve as miRNA or piRNA precursors can have dual functions, one as a small regulatory molecule after processing and one as an lncRNA (146). Similarly, small peptides encoded by lncRNAs have been identified in cells (147). Although the extent of this phenomenon remains to be examined, those lncRNAs could also have dual functions, one as a lncRNA and one with an encoded peptide (148).

As lncRNAs classification encompasses many categories, based either on the functionality or the physical attributes of the transcripts, a main limitation of this system remains the likelihood that lncRNAs will be classified in several categories.

1.5.2 Function of lncRNAs

lncRNAs are involved with many biological functions (149) (Figure 1.5). As highlighted previously, they show mainly regulatory function, both at the post-transcriptional level in the cytoplasm and at the transcriptional level in the nucleus. Interestingly, many studies have shown that lncRNAs are predominately nuclear (150–152).

Nuclear lncRNAs are mostly involved in the regulation of gene expression, but through different mechanisms. Importantly, the act of transcription of the lncRNA itself can promote chromatin remodeling. This results in chromatin accessibility of enhancers regions thus regulating gene expression (153). lncRNAs can also act as

scaffolds around enhancers regions to promote the formation of a protein complex, bridging together the enhancer and promoter region (154). They also act at the epigenetic level by recruiting chromatin modifiers. For example, PRC2 (polycomb repressive complex-2) is a histone methyltransferase responsible of repressing gene expression. It has been shown to be recruited to its target loci by several lncRNAs including HOTAIR (HOX Transcript Antisense RNA) (155). Some lncRNAs have been implicated in the regulation of the maintenance of the nuclear architecture (132). Firre is a famous example; its deletion inducing a loss of the normal nuclear architecture (156). Moreover, lncRNAs also modulate gene expression by interacting with transcription factors. One example is the binding of the lncRNA RMST to SOX2 (SRY-box 2) target genes to induce SOX2 recruitment (157). On the other hand, the binding of a lncRNA to a promoter region can also prevent the binding of transcription factors to control gene expression (152, 153).

In the cytoplasm, lncRNAs mostly fulfill their roles by interacting with proteins. For example, some STAU1 (Staufen 1)-binding sites are present in a class of lncRNAs. This allows those transcripts to interact with some mRNA and to form double stranded RNAs. Ultimately, this will induce the recruitment of STAU1 and mRNA degradation through STAU1-mediated mRNA decay (149, 150). Moreover, translation of mRNAs can be regulated by lncRNAs through interactions with the mRNA followed by the recruitment of translational repressors (161,162). Protein stability can also be affected by lncRNAs. For example, lncRNA-p21 binding to von Hippel-Lindau (VHL) protein prevents VHL-mediated ubiquitination of HIF-1 α (hypoxia-induced protein-1 α), thus inducing accumulation of HIF-1 α (161,163). LncRNAs also regulate RNA-binding protein availability, thus preventing them from exerting their original function (161).

We have evoked ceRNAs previously and interestingly, many lncRNAs have been shown to function as ceRNAs to modulate miRNAs activity, especially in cancer (164–166). Interestingly, some cytoplasmic lncRNAs have been shown to modulate signal transduction pathways. For example, the lncRNA lnc-DC binds to STAT3 (signal transducer and activator of transcription 3) and promotes its phosphorylation by preventing the inhibitory function of SHP1 (tyrosine phosphatase Src-homology 2 domain (SH2)-containing PTP-1) (161,167). The lncRNA LINK-A has been shown to modulate the AKT pathway through LINK-A binding with the PIP3 (phosphatidylinositol-3,4,5-trisphosphate). This suggests that the modulation of signaling pathways by lncRNAs is not regulated only through protein interactions (168). In addition, many studies have shown that lncRNAs interact with ribosomes (169). Carlevaro-Fita et al. suggests that cytoplasmic lncRNAs are sent to the ribosome to induce their degradation (170).

As mentioned, lncRNAs have roles both in the nucleus and the cytoplasm. Thus, in addition to having a system that regulates their expression and degradation, their nuclear export and subcellular localization also needs to be tightly regulated. In some cases, this is critical for the functional integrity of organelles, including the mitochondria (161,171).

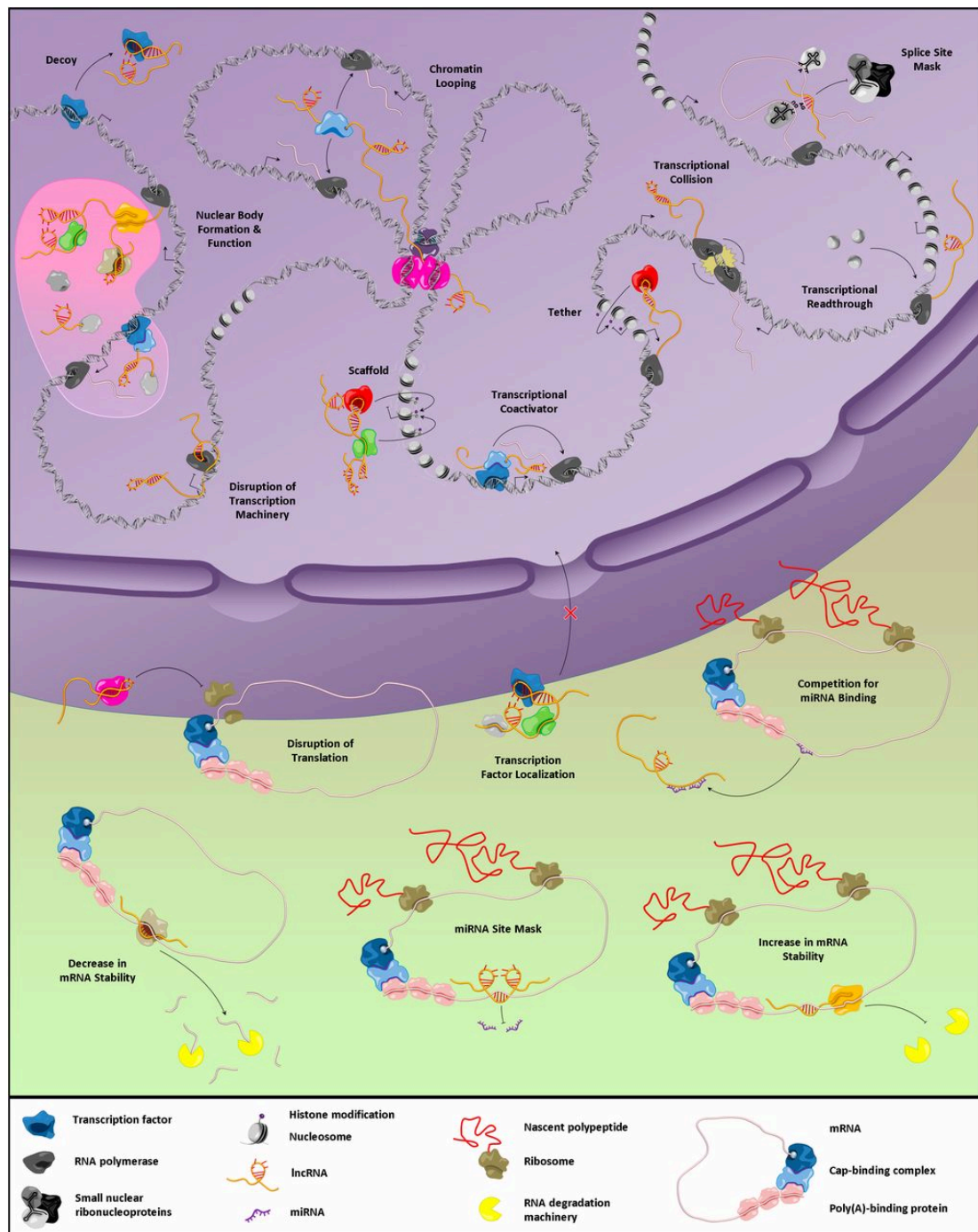


Figure 1.5: Long non-coding RNAs and their mechanism of action in cells.

LncRNAs are regulatory molecules that intervene at different stages to regulate mostly transcription and translation. This figure recapitulates different mechanism in which lncRNAs are involved both in the nucleus and in the cytoplasm. Used with permission from Kung et al. (154).

1.5.3 Examples of lncRNAs in the normal context

Since lncRNAs expression is tightly regulated both in time - some lncRNAs are expressed at specific developmental stages (172), and space - expression of some lncRNAs is limited to a given tissue (173), the section below discusses some examples of lncRNA function in a normal physiological context.

X chromosome inactivation, a phenomenon specific to therian mammals that allows dose compensation of genes present on the X chromosome in females, has been largely studied as an example of the role of lncRNAs in the normal context. Although in mouse, X inactivation is well understood, in humans it remains elusive. However, it is known that it is controlled by lncRNAs, especially XIST (X-inactive specific transcript) and XACT (X-active coating transcript) that are capable of coating the X chromosome, to either induce its inactivation or to prevent it (174).

Cell differentiation is another example of a physiological event also regulated by lncRNAs. In haemopoietic differentiation for example, lncHSC-2, an enhancer-derived lncRNA, controls the activity of TCF3 (transcription factor 3 also called E2A) to regulate haemopoietic stem cells self-renewal and maturation into myeloid-lymphoid progenitors (175). In brain development, several lncRNAs have been associated with neural cell fate determination. Indeed, Dlx1 (distal-less homebox 1) antisense lncRNA is upregulated during GABAergic (γ -aminobutyric acid) neuron differentiation whereas it is downregulated during oligodendrocyte differentiation. One study has suggested that Dlx1 antisense modulates neuronal differentiation by regulating its homeobox gene neighbors (176).

The function of lncRNA in the maintenance of the pluripotent state of embryonic stem cells (ESCs) was first suggested in 2008 when Dinger et al showed that pluripotency-associated transcription factors regulate the expression of some lncRNAs (69, 70). The lncRNA AK028326 is involved in a regulatory feedback loop for Oct4 (octamer-binding transcription factor 4), a pluripotency-associated transcription factor (179). In mice, AK028326 was identified as critical for the maintenance of the pluripotency state. Indeed, knockdown of AK028326 induced a simultaneous downregulation of pluripotent markers and upregulation of mesodermal markers (180). Several examples of lncRNAs with a function in the pluripotent state maintenance of ESCs have since been demonstrated. Interestingly, lncRNAs are also critical in induced pluripotent stem cells (iPSCs) for the reprogramming. For example, lncRNA-p21 interaction with SETDB1 (SET Domain Bifurcated Histone Lysine Methyltransferase 1) and DNMT1 (DNA (cytosine-5)-methyltransferase 1) controls, respectively, the maintenance of H3K9me3 histone modification and CpG methylation of pluripotency-associated gene promoters to prevent reprogramming (181).

1.6 LncRNAs in disease

Since lncRNAs are so integrated into the biology of the cell, it is only natural that they have been implicated in disease development. Although the functions of most lncRNAs in relation to disease have been studied mostly in neurological and cancer contexts, they have also been implicated in endothelial disorders, including diabetes and hypertension-related endothelial dysfunction (182).

1.6.1 LncRNAs in neurologic diseases

As mentioned previously, many examples of lncRNAs associated with neurologic diseases have been described. For example, BACE1-AS (Beta-Secretase 1 Antisense RNA), is transcribed from an intron of the *BACE1* ((Beta-Secretase 1) gene in an antisense manner. It increases BACE1 mRNA stability, ultimately stimulating an increase of A β 1-42, an early event in Alzheimer's disease (AD) (74, 75). Another lncRNA, BC200, which is elevated in AD brain tissue, shows lower expression in dendrites and higher expression in somata. This change in expression affects the microtubule-dependent transport and possibly the axonal and dendritic blockage, another early event of AD pathogenesis (185).

MALAT1, which is highly expressed in neurons, has been linked with Parkinson's disease (PD) where it is suggested to promote alpha-synuclein expression. Indeed, aggregation of alpha-synuclein has been associated with neurotoxic pathways that ultimately leads to neurodegeneration (186). Additionally in PD, the lncRNA HOTAIR promotes LRRK2 (Leucine-rich repeat kinase 2) mRNA levels (187). LRRK2 is a protein kinase involved in the control of neuronal cell morphology, proliferation and motility. HOTAIR stabilizes *LRRK2* mRNA which leads to neuronal degeneration. Interestingly, HOTAIR's role in neuronal degeneration has also been observed through the regulation of caspase-3 activity. Indeed, Wang. et al (2017) (188) have shown that HOTAIR knockdown induces a reduction of apoptosis. Liu et al. (2018) suggests that HOTAIR functions as a ceRNA for miR-143-3p, thus modulating caspase-3 expression (189).

Preliminary studies now suggest an association between dopamine signaling and lncRNAs, with implications to afflictions such as substance dependence, depression and schizophrenia. Hence, in 2015, Bannon et al. identified some differentially expressed lncRNAs between midbrain dopamine neurons of cocaine addicts and healthy-matched controls post-mortem (185, 186). They proposed a role for TRAF3IP2 (TRAF3 Interacting Protein 2) antisense as a regulator of epigenetic modifications. Another study shows that Gomafu, an lncRNA associated with schizophrenia, regulates dopaminergic transmission in mice (192).

1.6.2 LncRNAs in cancer

The association between lncRNAs and cancer has now been established and is well studied, although the function of the large majority of lncRNAs associated with cancer remains elusive. In 1999, Bussemakers et al identified the first lncRNA associated with cancer, Prostate Associated Cancer 3 (PCA3), which is now used as a biomarker for prostate cancer (188, 189). A year later, PCGEM1 (prostate cancer gene expression marker 1) was also linked with prostate cancer (195) then MALAT1 (metastasis associated lung adenocarcinoma transcript 1) was identified as a prognostic marker for lung cancer survival. Since that time, with the variety of genome-wide technologies available (microarray and RNAseq), the number of lncRNAs linked with cancer is continually increasing.

In most studies, to associate a lncRNA with cancer, a comparison of the level of expression of the lncRNA is done between cancer and normal cells. However, differential expression is not enough to characterize a lncRNA as oncogene or tumor suppressor. Thus, experiment analyses used to study the function of protein-coding

genes in cancer are now applied to study the function of lncRNAs. Since those are not always appropriate, we have seen an increase in the development of novel tools more adapted to lncRNAs study like antisense oligonucleotides (ASO) or locked nucleic acids (LNA) GapmeRs (Qiagen) which induce the transcript degradation using the RNase H1 degradation pathway.

LncRNAs study has also revealed novel challenges. Human lncRNAs show very little conservation in other species with 80% orthologous transcripts in chimpanzee but only 38% in mouse (196). This limited level of conservation restricts the use of animal models. Knockout animal models cannot be generated. Alternatively, xenografts of human knockout cells are used to question the function of a lncRNA in tumor formation.

Fundamentally, the role of lncRNAs in cancer are no different than what is observed in a normal context. For example, if a lncRNA promotes expression of an oncogene, this lncRNA will be silenced in a normal context. In cancer, oncogenic-associated transcription factors might induce the expression of this lncRNA, which in turn will induce expression of its oncogene target (197). Hence, most lncRNAs associated with cancer are involved in the regulation of factors implicated in one or more hallmarks of cancer (198). For example, lncRNAs in cancer can be involved in the regulation of signaling pathways such as NF- κ B, AKT and Notch (199), of the epithelial to mesenchymal transition (EMT) (200), or metastasis (201).

Interestingly, lncRNA can be packed in exosomes – lipid vesicles that can contain nucleic acids, proteins and lipids and are utilized by cells to send signals to other cells

(202). Research indicates that exosome-packed lncRNA UCA1 (Urothelial Cancer Associated 1) increases tamoxifen-resistance in breast cancer cells (203), exosome-packed MALAT1 promotes tumor cell migration in non-small-cell lung carcinoma (204) and exosome-packed H19 promotes angiogenesis in the tumor microenvironment (199, 200). Moreover, in addition to promoting tumor progression, those exosome-packed lncRNAs carry value as diagnostic and prognostic biomarkers (199, 201, 202).

Researchers are now considering targeting lncRNAs as a novel form of treatment for cancer (208). In 2016, Leucci et al. carried out intravenous injections of GapmeRs against the lncRNA SAMMSON (Survival Associated Mitochondrial Melanoma Specific Oncogenic Non-Coding RNA) in patient-derived xenograft melanoma models (209). They showed a reduction in cell proliferation and the induction of apoptosis. However, the study lacked a serious number of controls. Indeed, reduction in SAMMSON expression was not shown. Additionally, there was no evidence for the presence of the GapmeRs in the tissue. Further studies will thus have to be carried out, but those preliminary results are promising for a novel form of therapy (209).

1.6.2.1 Examples of lncRNAs associated with breast cancer

lncRNAs involved in breast cancer development are numerous. Among those, several promote cell proliferation either alone like H19 (210) and LSINCT5 (Long Stress-Induced Non-Coding Transcript 5) (211) or in combination with other carcinogenic factors like steroid receptor RNA activator (212). Some lncRNAs stimulate breast carcinogenesis by affecting EMT, like GHET1 (Gastric Carcinoma High Expressed

Transcript 1) (213) or through the promotion of cell drug resistance, such as UCA1 (214), as discussed previously.

Interestingly, the lncRNA CRALA (Chemoresistance-Associated LncRNA) has been found to be elevated in chemoresistant BC cell lines, and its knockdown is capable of resensitizing the cells to chemotherapy *in vitro*. High expression of CRALA is associated with poor prognosis in patients, suggesting that cells are not responsive to chemotherapy. Clinical trials targeting CRALA would have to be performed before concluding on its role in breast cancer therapy (215).

As hinted earlier, many lncRNAs have an important role in the regulation of pluripotency and with stemness being linked with cancer development, lncRNAs involved in the regulation of pluripotency deserve to be studied in the context of cancer. Hence, SOX2, a transcription factor involved in the regulation of stemness in human ESCs and iPSCs (216), has its coding sequence present within the intron of SOX2OT, a multi exon lncRNA. Both have been shown to be upregulated in BC, with expression of SOX2OT *in vitro* inducing an upregulation of SOX2 (217). This could constitute an early event in BC carcinogenesis. Another example is BMI1 (Proto-Oncogene, Polycomb Ring Finger), a master regulator of stemness of stem cells present in organs at the adult stage (218). Indeed, the lncRNA FAL1 (focally amplified lncRNA on chromosome 1) regulates BMI1 protein stability thus regulating stemness. Additionally, FAL1 has also been found to regulate the transcription of several genes involved in regulating proliferation and apoptosis (219).

LncRNAs can also have protective roles in BC development. For example, GAS5 (Growth Arrest Specific 5) shows higher expression in normal cells compared to BC cells *in vitro*. Its presence facilitates apoptosis when cells undergo UV radiation or are treated with cisplatin (220). In support of this, *in vitro* experiments have shown that GAS5 expression is upregulated under stress conditions, such as genotoxic stress or stress induced under serum starvation conditions, thus making the cells more susceptible to apoptosis (221).

1.6.2.2 LncRNA in DCIS

We have previously described some of the research conducted in the identification of miRNAs either specific to DCIS or important in the transition from DCIS to IDC. We have also described some examples of lncRNAs involved in breast cancer progression. Now, we want to address where lncRNAs fit in DCIS research. A research in PubMed associating the terms ‘DCIS’ and ‘lncRNA’ results in only five results. When the research is extended to ‘ductal carcinoma *in situ*’ and ‘long non-coding RNA’ or ‘preinvasive breast cancer’ and ‘RNA’, hundreds of results are available, although, very few actually relate to DCIS and lncRNAs. This highlights a lack of research on the biological function of lncRNA in DCIS.

In 2004, Iacoangeli et al (222) identified BC200 as the first lncRNA associated with DCIS. BC200 has since been linked with numerous types of cancer (223) but this was the first study that showed BC200 association with cancer by comparing *in situ* hybridization staining of normal human breast, IBC and DCIS tissue. BC200 expression was detected in nuclear high-grade nuclear DCIS, which is associated with the worst prognosis in patients. The same team has recently shown that BC200

expression can be detected in blood of IBC patients with high sensitivity and specificity, making of BC200 a candidate for diagnostic purposes (224). Next, Abba et al proposed to characterize molecularly nuclear high-grade DCIS. They performed an RNAseq experiment using thirty samples of high-grade DCIS fresh frozen tissue with matched adjacent normal tissue (77). Among the differentially expressed genes identified, 193 correspond to lncRNAs, with 127 antisense lncRNAs and 66 intergenic lncRNAs. In their RNAseq data, HOTAIR and HOTAIRM1 (HOXA Transcript Antisense RNA, Myeloid-Specific 1) were present. The authors suggest that those may regulate HOX gene expression in *trans*- and *cis*- in early stage of BC development. Unlike work from Iacoangeli et. al., BC200 is not part of the Abba et al. dataset. We suspect that this results from the technique used to prepare the library for RNAseq. Indeed, BC200 is an RNA polymerase III transcript so when polyadenylated enrichment is used during the library preparation, the probability of detecting it is low.

Instead of investigating which lncRNAs are associated with DCIS at a genome-wide level, Zhang et al. (225) decided to analyze the expression level of a limited panel of lncRNAs in BC using chromogenic *in situ* hybridization. This includes four lncRNAs associated with oncogenesis (HOTAIR, H19, KCNQ1OT1 and MALAT1) and two with tumor suppressor functions (MEG3 and ZFAS1). From this work, they detected higher expression of HOTAIR, H19 and KCNQ1OT1 in IBC and DCIS compared to normal, with HOTAIR and H19 showing a slight increase in IDC compared to DCIS. Interestingly, HOTAIR and KCNQ1OT1 were found to be expressed in tumor cells; whereas, H19 was expressed in stromal microenvironment cells.

Recently, RNAseq has been used to identify novel lncRNAs linked with DCIS. This time, instead of using polyadenylated enrichment, Rao et al (226) used a ribosomal RNA removal approach to capture as many lncRNAs as possible. Sixty-nine lncRNAs with differential expression between DCIS and normal and 12 between DCIS and IDC were identified. The authors focused on 21 lncRNAs with MIAT, FAM83H-AS1, ST8SIA6-AS1 and AC109826.1 showing higher expression in DCIS and IDC compared to normal, and EPB41L4A-AS1, WDFY3-AS2, RP11-392O17.1 and RP11-295M3.4 showing lower expression in DCIS and IDC compared to normal.

FAM83H-AS1 (FAM83H antisense RNA 1), WDFY3-AS2 (WDFY3 antisense RNA 2), ST8SIA6-AS1 (ST8SIA6 antisense RNA 1), ADIPOQ-AS1 (ADIPOQ antisense RNA 1) and LINC00861 are the only lncRNAs differentially expressed in both Rao et al. and Abba et al. dataset. However, only FAM83H-AS1, ST8SIA6-AS1 and LINC00861 show a higher expression in DCIS compared to normal for both datasets. Indeed, ADIPOQ-AS1 and WDFY3-AS2 expression is higher in DCIS compared to normal in the Abba et al. dataset, whereas it is lower in DCIS in the Rao et al dataset.

More recently, expression of the lncRNA BHLHE40-AS1 has been shown to increase with disease progression. Through the regulation of IL-6 (interleukin-6)/STAT3 signaling, BHLHE40-AS1 promotes proliferation, motility and invasive capacities of DCIS cells (227).

1.7 Aims of the project

In this introduction chapter, we have highlighted the importance of investigating novel biomarkers for DCIS. Protein-coding genes markers have proven to be disappointing

at best, particularly as biomarkers for early detection and stratification of DCIS patients that would require treatment. We have, in a second part, shown the importance of lncRNAs in the biology of the cell where they are major actors in the tight regulation of each biological process at different stages of development. In disease, lncRNAs also play an important role, in particular in cancer development where dysregulation of the expression of some lncRNAs can have dramatic effects. However, investigation of the lncRNAs associated with cancer and their function has mostly focused on invasive cancer, despite the evidence that regulatory changes occur early in the carcinogenesis process. We have presented some research focusing particularly on lncRNAs in DCIS; however, it remains limited. Indeed, with about 16,000 to 90,000 different lncRNAs identified (228), many of those associated with cancer, we are convinced that only a very small subset of DCIS-linked lncRNAs has been uncovered. This is why we have chosen to focus on the identification of lncRNAs associated with DCIS, with hope of filling this gap.

In the laboratory, we have obtained several cell lines developed as DCIS models. However, limited research has been conducted on those cell lines, so the first aim of this thesis focused on the study of those cells by conducting a molecular and cellular characterization of the DCIS cell lines. Those results will be presented in Chapter 3. This allowed us to establish a baseline of the cells' behavior. This work is a necessary step for the planning and interpretation of the experiments in the next chapters. The second aim of this thesis was to identify DCIS-linked lncRNAs and this work will be described in Chapter 4. We have reviewed, in this introduction, two approaches that have been used to identify lncRNAs associated with DCIS. We have chosen to use a capture-based approach in our study onto our DCIS cell lines and a DCIS patient

cohort. Our hope with using this approach is to uncover a larger number of DCIS-linked lncRNAs without the “mRNA noise pollution”. Finally, the last aim of this thesis was to investigate the function of some DCIS-linked lncRNAs. In Chapter 5, we will examine the function of six lncRNAs candidates whose expression is associated with adverse overall survival in an IDC patient cohort from The Cancer Genome Atlas (TCGA). By altering the level of expression of those lncRNAs in our DCIS cell lines, we planned to determine if they show any oncogenic or tumor suppressor like functions and to identify the mechanism in which they intervene. Ultimately, we are hoping that we will identify lncRNAs with clinical relevance with the ambition that those could be used in DCIS diagnosis and treatment.

Chapter 2

Materials and Methods

2.1 Materials

2.1.1 Cell lines

MCF-10A, MCF-7 and MDA-MB-231 originally from ATCC were a kind gift from Prof Rosemary O'Connor (UCC, Cork, Ireland). MCF10DCIS.com cells were obtained from Wayne State University (Michigan, USA). ETCC-006 and ETCC-010 were purchased from Leibniz-Institut DSMZ (Germany). Cell lines were authenticated using short, tandem repeat (STR) profiling (Eurofins Genomics) and were mycoplasma-free.

2.1.2 Bacterial strain

Strain DH5 α of *Escherichia coli* (*E.coli*) from the Doctor Dean lab was used for all cloning applications (transformation and plasmid preparation). The bacterial cultures were grown in Luria-Bertani (LB) medium or agar with selective antibiotics as required.

2.1.3 Chemicals and consumables

Table 2.1: Chemical and consumables used for the experiments described in this thesis.

Product	Company	Reference number
Tissue Culture		
Dulbecco's Modified Eagle Medium (DMEM)	Sigma-Aldrich	D6429
Roswell Park Memorial Institute (RPMI) medium	Sigma-Aldrich	D8758
DMEM/Nutrient Mixture F-12 Ham (DMEM/F12)	Sigma-Aldrich	D8062
Penicillin-Streptomycin	Sigma-Aldrich	P4333
Trypsin-EDTA	Sigma-Aldrich	T4049
Fetal Bovine Serum	Sigma-Aldrich	F7524
Horse Serum	Gibco	16010159
4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)	Gibco	15630-080
CaCl ₂	Sigma-Aldrich	C3881
Insulin	Sigma-Aldrich	I6634
Epidermal Growth Factor	Peptrotech	100-15
Cholera Toxin	Sigma-Aldrich	C9903
Hydrocortisone	Sigma-Aldrich	H0135
Phosphate Buffered Saline	Sigma-Aldrich	P5493
Oestradiol	Sigma-Aldrich	E2758
(Z)-4-Hydrotamoxifen	Sigma-Aldrich	H7904
Dishes/flasks/pipettes/cell scrapers	Sarstedt	
Dimethylsulfoxide (DMSO) – MTT	Sigma-Aldrich	D8418
Dimethylsulfoxide (DMSO) – sterile	Sigma-Aldrich	D4540
100% Ethanol	Central Stocks	
Trypan Blue solution	Sigma-Aldrich	T8154
Culture-Inserts 2 Well for self-insertion	Ibidi	80209
Lipofectamine RNAiMax Transfection Reagent	ThermoFisher Scientific	13778075
Opti-MEM I Reduced Serum Medium	ThermoFisher Scientific	31985070
siRNAs	GenePharma	
Crystal Violet	Sigma-Aldrich	C0775
Molecular Biology		
Plastic consumables	Sarstedt	
dNTP	ThermoFisher Scientific	R0193
Phusion High-Fidelity DNA Polymerase	New England Biolabs	M0530S
DreamTaq Green PCR Master Mix (2X)	ThermoFisher Scientific	K1081

FIREpol DNA Polymerase	Solis BioDyne	01-01-0000S
HyperLadder 1kb	Bioline	BIO-33053
HyperLadder 100bp	Bioline	BIO-33056
Agarose	Bioline	BIO-41025
SafeView Nucleic Acid Stain	NBS Biologicals	NBS-SV5
Restriction Endonucleases	New England Biolabs	
T4 DNA ligase	ThermoFisher Scientific	EL0014
LB medium	Sigma-Aldrich	51208
LB agar	Sigma-Aldrich	L2897
Carbenicillin disodium salt	Sigma-Aldrich	C1389
Kanamycin sulfate	Sigma-Aldrich	K4000
Zeocin	InvivoGen	Ant-zn-05
DNA oligonucleotides	Integrated DNA Technologies	
Antisense oligonucleotides	Integrated DNA Technologies	
PCR/gel extraction kit	Invitrogen	K220001
DNA extraction (miniprep)	New England Biolabs	T100L
DNA extraction (midiprep)	Macherey-Nagel	740412.50

Quantitative Real-Time PCR

RNeasy FFPE kit	Qiagen	73504
TRIzol Reagent	ThermoFisher Scientific	15596026
Chloroform:Isoamyl alcohol	Sigma-Aldrich	C0549
Isopropanol	Sigma-Aldrich	I9516
Glycogen	Roche	10901393001
Ethyl alcohol, Pure	Sigma-Aldrich	E7023
UltraPure Distilled Water	Invitrogen	10977-035
Turbo DNA-free kit	ThermoFisher Scientific	AM1907
SuperScrip II Reverse Transcriptase	ThermoFisher Scientific	18064014
Random Primers	ThermoFisher Scientific	48190011
SYBR Green JumpStart Taq ReadyMix	Sigma-Aldrich	S4438
MicroAmp Optical Adhesive Film	ThermoFisher Scientific	4311971
MicroAmp Fast Optical 96-Well Reaction Plate with Barcode, 0,1 mL	ThermoFisher Scientific	4346906
Step One Plus Real Time PCR system	Applied Biosciences	

Western Blotting

SDS-Page Running Buffer	Central Stocks	
Methanol	Sigma-Aldrich	32213
Sodium dodecyl sulphate (SDS)	Sigma-Aldrich	L3771
Glycerol	Sigma-Aldrich	G6279
β -mercaptoethanol	Sigma-Aldrich	63689
Bromophenol Blue powder	Sigma-Aldrich	B0126
30% Acrylamide	Sigma-Aldrich	A3574
Ammonium persulfate	Sigma-Aldrich	A3678
Tetramethylethylenediamine (TEMED)	Sigma-Aldrich	T9281
Protein ladder	ThermoFisher Scientific	26616

Immobilon-FL Transfer Membranes	Merck Millipore	IPFL00010
Dried milk	Marvel	
Tween 20	Sigma-Aldrich	P2287
Bovine serum albumin (BSA)	Sigma-Aldrich	A7906
NP-40 alternative	Merck	492016
Protease inhibitor cocktail, EDTA-free	Roche	04 693 159001
Pierce BCA Protein Assay Kit	ThermoFisher Scientific	23225
Odyssey	Li-Cor	

2.1.4 Antibodies

Table 2.2: Antibodies used for Western blotting and indirect immunofluorescence.

Protein targeted	Company	Reference number
β Actin	Sigma-Aldrich	A5441
β Catenin	Santa Cruz	sc7963
β 1 Integrin	Millipore	#1952
Cyclin B1	Cell Signaling Technology	#4135
Cyclin D1	Abcam	ab16663
E-cadherin	BD Biosciences	#610181
Estrogen Receptor α	Cell Signaling Technology	#8644
GAPDH	Santa Cruz	sc47724
HER2	Cell Signaling Technology	#4290
IGF-1R	Cell Signaling Technology	3027
N-cadherin	Abcam	ab19348
Pan AKT	Cell Signaling Technology	#2920
Pan ERK	Cell Signaling Technology	#4696
Pan p-AKT	Cell Signaling Technology	#4060
Pan p-ERK	Cell Signaling Technology	#4377
p-EGFR	Cell Signaling Technology	#3777
Progesterone Receptor A/B	Cell Signaling Technology	#8757
Vimentin	Santa Cruz	sc32322
IRDye 800CW or 680 RD	Li-Cor	

DyLight 488 Donkey anti-mouse	Jackson Immunoresearch	
----------------------------------	------------------------	--

2.1.5 Oligonucleotides

Table 2.3: DNA primers used for the quantitative qRT-PCR experiments.

Gene targeted	Forward Primer (5'-3')	Reverse Primer (5'-3')
βActin	CAGAGCAAGAGAGGCAT CCT	ACGTACATGGCTGGGGTG
BC200	TAATCCCAGCTCTCAGGG AGGCTAA	GGTTGTTGCTTTGAGGGAAGTTA CGC
CCAT1	GCAGGCAGAAAGCCGTA TCT	TCCCAGGTCCTAGTCTGCTT
CCNB1	CCAAATCAGACAGATGG AAAT	GCCAAAGTATGTTGCTCGA
cMYC	CAAACCTCCTCACAGCCC ACT	TGACACTGTCCAACCTGACCC
CTSL2	GCTATGGATGCAGGCCAT TC	ACACCATGATCCAGGTTTTTGC
DSCAM-AS1	ACCACAACAACAACAAC AG	ATGATGAGACCAGAACTTCC
ESR1	CCACCAACCAGTGCACC ATT	GTCTTTCCGTATCCCACCTTTC
GAPDH	GAGTCAACGGATTTGGTC GT	TTCCCGTTCTCAGCCTTG
GSTM1	CTATGATGTCCTTGACCT CCACCGTA	ATGTTACGAAGGATAGTGGGTA GCT
HER2	TCCTGTGTGGACCTGGAT GAC	CCAAAGACCACCCCCAAGA
Ki67	CCACACTGTGTCGTCGTT TG	CCGTGCGCTTATCCATTCA
LINC00313	CAGGCTGACTGAACCCA GAG	CTGTACAGAGGGAAGAAGATTTC A
LINC00313_spliced	GTCCTTCTCCTTACATA GCCA	GCAGGGTCTAAGTGCTTCCA
LINC00473	ATGCGCGTCAGCATACTT TG	CCAAAGCACAACGAGCACCA
LINC00707	AGGCTCCTTCTCCTTACA TAGC	TTCATCGAGGTTTCCCGCTG
LINC00707_spliced	GGGAAAACAGAGGCAGA CTCAAGAGTCTGG	GTGCTCAGTCTGACTTGCACA
LINC02035	CCAGGTGACACTTGATGC CT	CAGCCTGCTGGAGCACTATT
LOC100131564	CCAGCCGTTCCAGTACAA CT	CACCCTGGCCCCAAAAGTTA
LOC100134040	TCAGTCTTGGCAAGTATC TGTACC	TCCTCCAGGAACAAAAGCTTCA

LOC100134040_spliced	AGCTGCGTTCGGACCCGTCTGATCGTA	AACAGCTTCGCTGCCTTGGTCA
LOC105372815	TCTTCAACATGGCGGTCTGAT	GTGGCAGAAGTGGAGTGGAG
LOC541471	AGGACTGGAGTGCCACCTAT	AGGCACGAGATTTTCTGGCA
LOC729970	ACAGGGAACAGAACCTTCGC	CACCCCTCAACGGACATGAA
LOC729970_spliced	CCCCTTGTGCTCTCAGAACT	CTGCACTGTGGTCAAAGGCT
MGC45800	AGCTGTTGTCAAGAGAGGGC	GCTTGCAAAACAGACAGCGA
MIR100HG	TCGAACTTTGGAGTGTGGCA	TCCCTGGTTACTGCCAGAT
MIR210HG	AGTTCCTGTTGCCAAGCTGA	GGATGGTCTGTTGGCTGAA
MMP13	TCCCAGGAATTGGTGATAAAGTAGA	CTGGCATGACGCGAACAATA
MUC5B-AS1	CTCTGTGAGGATCCAGTGGAACG	TGTGCTTTGCTGTGACGACT
MYBL2	AAAACAGTGAGGAGGAAC	CAGGGAGGTCAAATTTAC
PR	TACCCGCCCTATCTCAACTACCTG	GGACACCATAATGACAGCCTGATG
STK15	GGAGAGCTTAAAATTGCAGATTTT	GCTCCAGAGATCCTTCTCAT
Survivin	ACCGCATCTCTACATTCAAG	CAAGTCTGGCTCGTTCTC
TRPC1	ATGTCGTGGTTGTGATTGTGCTT	CTTTGGTGAGGGAATGATGTTGA
UBC	TCGAGAATGTCAAGGCAAAATC	GAGTGGACTCTTTCTGGATGTTGTA
ZNF667-AS1	TGTGACAAGTTCTTCAGGCG	GGATGAATGCCGATTGCAGAC

2.1.6 Plasmid

Table 2.4: DNA plasmids used in the experiments described in the thesis.

Plasmid	Origin
pcDNA3	Dr. Justin McCarthy lab (UCC)
pcDNA3_tRSA_G-quartet_BC200	Dr. Kellie Dean lab (UCC)
pcDNA3_BC200	Digested BC200 from pcDNA3_tRSA_G-quartet_BC200 and cloned into pcDNA3
pUC57_LINC00473	GenScript
pUC57_LINC00707	GenScript
pUC57_MIR210HG	GenScript

pcDNA3_LINC00473	Digested LINC00473 from pUC57_LINC00473 and cloned into pcDNA3
pcDNA3_LINC00707	Digested LINC00707 from pUC57_LINC00707 and cloned into pcDNA3
pcDNA3_MIR210HG	Digested MIR210HG from pUC57_MIR210HG and cloned into pcDNA3
pcDNA3_LOC100134040	LOC100134040 amplified from gBlock Gene Fragment (IDT) and cloned into pcDNA3
pcDNA3_LOC729970	LOC729970 amplified from gBlock Gene Fragment (IDT) and cloned into pcDNA3
pcDNA3_LINC00313	LINC00313 amplified from gBlock Gene Fragment (IDT) and cloned into pcDNA3
gRNA_cloning-vector	Addgene #41824
SP_dCas9_VPR	Addgene #63798
1.1_gRNA_cloning-vector	Annealed 1.1_sgRNA cloned into gRNA_cloning-vector
1.3_gRNA_cloning-vector	Annealed 1.3_sgRNA cloned into gRNA_cloning-vector
2.3_gRNA_cloning-vector	Annealed 2.3_sgRNA cloned into gRNA_cloning-vector
2.9_gRNA_cloning-vector	Annealed 2.9_sgRNA cloned into gRNA_cloning-vector
pcDNA3_tRSA_G-quartet_LOC729970	Dr Kellie Dean Lab (UCC)
pcDNA3_tRSA_G-quartet_LOC729970_AS	Dr Kellie Dean Lab (UCC)

2.1.7 siRNAs

Table 2.5: Sequence of siRNAs used for the knock-down experiments.

Gene targeted	Sequence (5'-3')	Region targeted
Negative Control-FAM	GGCUCAACUACGCAACACUAUAATT	
Negative Control	GGCUCAACUACGCAACACUAUAATT	
βActin	UGAAGAUC AAGAUCAUUGCTT	1058-1077
LINC00473	GCGCCGGGAGAUGCAUCACGAUGAATT	749-774
	CCCUGUCUGCAAAGAUAUCCAGUUUAATT	995-1020
LINC00707	AGUGUGCUCUAUGCUGGCUUGAAUAATT	1036-1061
	ACAGGUGGAUAAGACUAACACUGAATT	1298-1323
MIR210HG	CAGGUGUGGUCAUAUCUUCTT	161-180

	CCCACUUGGCCUAUGCAUUTT	1674-1693
LOC100134040	GGACCCGUCUGAUCGUAAATT	17-36
	GCUUUAAAUUUGGGUAGCATT	468-487
LOC729970	GCAGCUGGUAGACUAGAUATT	47-66
	CCAUUUCAGUUCCUGUUAUTT	667-686
LINC00313	GGCUUCCUGGAUUGCAUAAATT	35-54
	CCCACUAUCUGCUAGGAAATT	386-405
MEG8	GCGAGGAAGAUCAUGGUUGTT	1627-1646
	GGAUCAAGAUUGUUGUGUUTT	3183-3202

The sequence in red represents a dTdT overhang to increase siRNA efficiency.

2.1.8 Antisense oligonucleotides

Table 2.6: Sequence of antisense oligonucleotides (ASOs) used for the LOC729970 knock-down experiments.

Antisense oligonucleotide	Sequence (5'-3')	Region Targeted
Negative Control	TTATAGTGTTGCGTAGTTGAGCC	
ASO1	TATCTAGTCTACCAGCTGC	47-65
ASO2	AAATCTGTCCTAGGTGTGC	219-237
ASO3	CTTCCTTGTTTATGTGACAACAGGGG	536-561
ASO4	ATAACAGGAACTGAAATGG	667-685

2.2 Methods

2.2.1 Tissue Culture

MCF-10A were maintained in DMEM/F12 supplemented with 5% horse serum, 10 µg/ml insulin, 20 ng/ml EGF, 100 ng/ml cholera toxin and 0.5 µg/ml hydrocortisone. MCF-7 and MDA-MB-231 were cultured in DMEM supplemented with 10% FBS and 1% P/S. MCF10DCIS.com were cultured in DMEM/F12 supplemented with 5% horse serum, 1.05 mM calcium chloride and 10 mM HEPES. ETCC-006 and ETCC-010

were cultured in RPMI supplemented with 10% FBS and 1% P/S. All cell lines were grown at 37°C in 5% CO₂ humidified atmosphere.

Cells were routinely passaged every three days. Briefly, after a wash in PBS, cells were incubated in trypsin-EDTA until detached, resuspended in media to inactivate the trypsin and plated at densities varying depending on the cell line. To make cellular stocks, cells were washed with PBS, trypsinised, spun at 800g for 3 minutes at room temperature then resuspend in freezing media. For MCF-7, MDA-MB-231, ETCC-006 and ETCC-010 freezing media consisted in 50% FBS, 40% media (DMEM or RPMI) and 10% of sterile DMSO. MCF-10A were frozen in a mixture of 45% media, 45% horse serum and 10% sterile DMSO. MCF10DCIS.com were frozen in a mixture of 70% DMEM/F12, 20% calf serum and 10% sterile DMSO. Cells resuspended in freezing media were stored at -80°C overnight in an insulated box before being transferred into liquid nitrogen for long-term storage. Cells were thawed at 37°C, immediately after resuspended in media and cultured overnight at 37°C in 5% CO₂ humidified atmosphere. Media was changed the next day.

2.2.2 Transfection

2.2.2.1 siRNA transient transfection

siRNA transfection was performed using Lipofectamine RNAiMAX reagent. Briefly, siRNA stock was diluted to obtain a 10 µM working solution. For each 24-well plate, 3 µl of the siRNA working solution was added to 96 µl of OptiMEM I Reduced medium; then 1.5 µl of lipofectamine RNAiMAX was added to the mix and resuspended gently. The mix was plated into the wells and left to incubate for 20

minutes. In the meantime, cells were trypsinised and resuspended in media that did not contain P/S. 65,000 cells were added to the mix per 24 well-plate. The next day, media was changed. Cells were harvested after 24 hours, 48 hours or 72 hours.

2.2.2.2 Plasmid DNA transient transfection

Lipofectamine 3000 Reagent (Invitrogen) was used to transfect DNA into MCF-10A, MCF-7, MDA-MB-231 or ETCC-006 cell lines as per manufacturer instructions. The media was changed six to eight hours after transfection.

2.2.3 Plasmid construction

2.2.3.1 LINC00473, LINC00707, MIR210HG, LINC00313, LOC729970, LOC100134040 and BC200 overexpression constructs

A gBlock gene fragment corresponding to the cDNA sequence of LOC729970 (NR_033998.1), LOC100134040 (NR_148939.1) and LINC00313 (NR_026863.1) was purchased from Integrated DNA Technologies. The cDNA sequence of LINC00473 (NR_026860.1), LINC00707 (NR_038291.1) and MIR210HG (NR_038262.1) cloned in pUC57 was purchased from GenScript. BC200 was digested from pcDNA3_tRSA_G-quartet_BC200 and inserted into the *EcoRI* and *XhoI* restriction sites of pcDNA3. LINC00473 and LINC00707 were digested from pUC57_ LINC00473 and pUC57_ LINC00707 and inserted into the *BamHI* and *NotI* restriction sites of pcDNA3. MIR210HG was digested from pUC57_ MIR210HG and inserted into the *EcoRI* and *XhoI* restriction sites of pcDNA3. LINC00313 was amplified from the gBlock gene fragment and inserted into the restriction sites *BamHI* and *XhoI* of

pcDNA3. LOC729970 was amplified from the gBlock gene fragment and inserted into the restriction sites *KpnI* and *NotI* of pcDNA3. LOC100134040 was amplified from the gBlock gene fragment and inserted into the restriction sites *BamHI* and *NotI* of pcDNA3. All constructs generated are summarized in the Table 2.5.

2.2.3.2 Ligation

50 ng of digested vector DNA treated with Antarctic phosphatase (New England Biolabs) was used in a ligation reaction with digested insert DNA at a molar ratio varying between 1:3 to 1:5 using T4 DNA ligase (New England Biolabs).

2.2.3.3 Transformation

1 µl to 3 µl of ligation reaction was transformed into *Escherichia coli* DH5α cells by heat shock for 45 seconds at 42°C. Cells were then grown in super optimal broth (SOB) media for an hour, centrifuged for 3 minutes at 3000 rpm and plated onto a LB agar plate with the appropriate antibiotic.

2.2.3.4 Sequencing

Purified plasmid DNAs were confirmed by Sanger sequencing (GATC or Eurofins).

2.2.4 Agarose gel electrophoresis

Agarose was dissolved in Tris-Acetate EDTA (TAE) buffer (40 mM Tris-Acetate and 1 mM EDTA) by microwave heating. The mixture was left to cool to a temperature close to room temperature and SafeView Nucleic Acid Stain (NBS Biologicals) was added (2.5 µl for 50 ml agarose gel). After mixing, gel was casted and left to set.

Samples were prepared in 6X DNA loading dye and ran in TAE buffer at 100V at room temperature for 30 min.

2.2.5 RNA

2.2.5.1 TRIzol RNA isolation

Samples were lysed using TRIzol (Thermo Fisher Scientific). Briefly, 200 µl of chloroform was added per 1 ml of TRIzol reagent, samples were homogenized and then left at room temperature for 3 min. The aqueous phase was separated by centrifugation and RNA was precipitated using isopropanol. After two washes using 75% ethanol, the RNA pellet was airdried, resuspended in water and incubated at 58°C for 10 min to fully resuspend the RNA.

2.2.5.2 RNeasy FFPE RNA isolation kit (Qiagen)

Tissue sections were incubated in an oven for one hour at 60°C, then washed three times in xylene, twice in 100% ethanol, once in 90% ethanol, once in 80% ethanol and twice in water. Each wash was left for 3 minutes. Tissue sample were scrapped off the slides, resuspended in buffer PKD (provided in the kit) with proteinase K and left overnight to incubate at 56°C. The remainder of the protocol was performed as per manufacturer's instructions.

2.2.5.3 cDNA synthesis

1 µg of RNA was treated with DNase to eliminate contaminating DNA using TURBO DNase (Invitrogen) then used in a cDNA synthesis reaction using Superscript II

(Thermo Fisher Scientific) as per manufacturer instructions. Reactions lacking reverse transcriptase enzyme were also performed in the same condition as controls. The synthesized cDNA was diluted 1:5 and used for quantitative reverse transcriptase PCR (qRT-PCR).

2.2.5.4 Quantitative Reverse Transcriptase PCR

The diluted cDNA was used in qRT-PCR reactions. Briefly, 1 μ l of cDNA (2.5 ng/ μ l) was combined with SYBR Green JumpStart Taq ReadyMix (Sigma-Aldrich) in 20 μ l reactions and run using the following conditions:

- 95°C 10 min
- 94°C 30 sec, 57°C 45 sec, 72°C 1 min repeated 39 times
- 94°C 30 sec, 57°C 45 sec, 72°C 15 min
- Melting curve stage

After analysis of the melting curve, data was analyzed using the $\Delta\Delta C_T$ method.

2.2.5.5 RNAseq at EMBL, library preparation and sequencing

Total RNA was extracted using TRIzol reagent according to the manufacturer's instructions. The RNA integrity was measured using the RNA 6000 nano kit (Agilent). Libraries were prepared using the NEB Ultra II Directional RNA Library Prep Kit for Illumina (NEB #E7760) and NEBNext Poly (A) mRNA Magnetic Isolation Module (NEB #E7490). The bioinformatic analysis was performed using the Galaxy platform (230). FASTQ reads were trimmed and filtered for quality and adapter using Trimmomatic (231) and aligned to the Human reference genome (hg19) using RNA Star (232). featureCounts was used to quantify gene expression (233) and differential

expression analysis was performed with DESeq2 (234) respectively between ETCC-006 and ETCC-010 against MCF-10A (study: ERP000799; experiment: ERX016686, biosample: SAMEA1484687). Genes with p-value < 0.05 were selected for reactome pathway-based analysis (235).

2.2.5.6 Library preparations and capture

RNA extracted from ETCC-006, ETCC-010, MCF10DCIS.com and MCF-10A from three different replicates was used to prepare libraries using the KAPA Stranded RNA-Seq Library Preparation kit (Roche) and SeqCap RNA Enrichment System (Roche) following the protocol described by Feng and Zhang (2016) (236).

A cohort of DCIS patients was recruited in collaboration with Dr Conleth Murphy, Bon Secours Hospital, Cork, Ireland. FFPE (formalin-fixed paraffin-embedded) blocks of DCIS tissue and normal-adjacent tissue were obtained from seven patients. Only patients showing pure DCIS lesions were recruited, patients showing a mix of DCIS and IDC were thus discarded from the study. All participants to the study signed an informed consent form. The study was approved by the research ethics committee of the cork teaching hospitals (CREC).

The lncRNA capture approach was utilized using RNA extracted from DCIS lesion and adjacent-normal tissue from DCIS patients to prepare RNA sequencing libraries. Only RNA from lesions showing over 30% of tumor cells ($n = 5$) was used for the library preparation.

2.2.5.7 Sequencing and bioinformatics

RNA sequencing libraries prepared using RNA extracted from cell lines were sent for Illumina HiSeq 4000 75 bp pair end sequencing to Edinburgh Genomics. RNA sequencing libraries prepared using RNA extracted from tumor and normal patient tissue were sent for Illumina HiSeq 4000 100 bp pair end sequencing to Genome Technology Center, NYU. FastQ reads were trimmed and filtered based on quality then aligned to the Human reference genome (hg19) using TopHat (237). Genes and transcript levels were quantified using cufflinks (238) and differential expression analysis was conducted using cuffdiff (238). Volcano plots and heatmaps were generated using CummeRbund (238).

2.2.6 Western blotting

Samples were lysed using NP-40 lysis buffer (150 mM NaCl, 50 mM Tris HCl pH=8, 1% NP-40 and 1X cOmplete EDTA-free Protease Inhibitor Cocktail (Sigma-Aldrich)) for 30 min on ice, then centrifuged for 30 min at 12,000g at 4°C. Protein levels were quantified using Pierce BCA Protein Assay Kit (Thermo Fisher Scientific) and equal amounts of protein samples (40-50 µg) were resuspended in SDS loading buffer and run in SDS-PAGE gel at 100V for 2 hours. Proteins were transferred onto a PDVF (polyvinylidene difluoride) membrane (Merck-Millipore) for 1 hour at 100V then blocked with 5% milk or BSA for 1 hour at room temperature and incubated with the primary antibody overnight at 4°C. The next day, the membrane was washed with TBST (tris buffered saline tween) for 5 min 3 times, incubated with secondary antibody for 1 hour at room temperature, washed with TBST for 5 min 3 times and visualized using Odyssey Imaging Systems (LI-COR Biosciences).

2.2.7 β -actin immunofluorescence

Cells were seeded in a 6-well plate onto sterile coverslips and allowed to grow for 48 hours. Cells were then fixed with 4% PFA (paraformaldehyde) in PHEM buffer (containing PIPES, HEPES, EGTA and MgCl_2), quenched with 50 mM ammonium chloride and permeabilized with 0.1% Triton X in PHEM buffer for 5 min. Non-specific antibody binding was blocked by incubating the coverslips with 5% goat serum/PHEM for 30 min. Cells were stained with anti- β actin (1:400) overnight at 4°C in an humidified chamber and washed with PHEM. The coverslips were then incubated with anti-mouse 488-fluorophore (1:1000) conjugated secondary antibody and DAPI (4',6-diamidino-2-phenylindole) at room temperature for 1 h then washed with PHEM. The coverslips were mounted onto Fluoroshield histology mounting medium (Sigma Aldrich). Images were acquired using a Leica DM6000 upright fluorescence microscope.

2.2.8 Cell growth monitoring

MCF-10A, MCF-7, MDA-MB-231, MCF10DCIS.com, ETCC-006 and ETCC-010 cells were seeded at a density of 50,000 cells per 12-well plate in triplicate. Media was replaced every third day. Cells were trypsinized, resuspended into medium and counted every 24 h over a period of 6 days. Trypan blue was used to monitor the viability of the cells. Population doubling time was calculated (16) using values at day 2 and day 5 with the following formula:

$$\text{Population Doubling Time} = \frac{\text{duration} \times \log(2)}{\log(\text{final cell number}) - \log(\text{initial cell number})}.$$

2.2.9 Plating efficiency assay

MCF-10A, MCF-7, MDA-MB-231, MCF10DCIS.com, ETCC-006 and ETCC-010 cells were seeded at a density of 2,000 cells per 6-well plate in triplicate. Media was replaced every third day. Cells were washed with PBS (phosphate buffered saline) and fixed with 96% ethanol for 10 min, then stained with 0.05% crystal violet in 20% ethanol for 30 min. Staining was assessed using an Odyssey Scanner and quantified using ImageJ software.

2.2.10 Clonogenic growth assay

MCF-10A, MCF-7, MDA-MB-231, MCF10DCIS.com, ETCC-006 and ETCC-010 cells were seeded at a density of 300,000 cells in an agar suspension (0.4% agar) over an agar underlay (0.8% agar) for assessment of anchorage-independent growth. Media was changed every 5 days. After 3 weeks of growth, the agar layers were stained with 1% crystal violet at 4°C overnight and destained through water washes every 24 hours until single colonies appeared. Staining was assessed using an Odyssey Scanner and counted by hand.

2.2.11 MTT assay

Cell viability of cells transfected with either vectors to overexpress, or siRNAs to knockdown, expression of LINC00313, LINC00473, LINC00707, MIR210HG, LOC729970 and LOC100134040 was assessed by MTT assay. Briefly, cells were cultured in 96-well plate and thiazolyl blue tetrazolium bromide reagent (Sigma-Aldrich) was added at a final concentration of 0.5 mg/ml then left to incubate for 90

min at 37°C and 5% CO₂. The media was removed and 100 µl of DMSO was added to each well to dissolve the purple formazan crystals. Absorbance was measured at 570 nm using an Infinite 2000 spectrophotometric plate reader (Tecan).

2.2.12 Wound healing assay

MCF-10A, MCF-7, MDA-MB-231, MCF10DCIS.com, ETCC-006 and ETCC-010 cells were seeded as to form a monolayer in culture-inserts 2 well (Ibidi, #80209) the day before performing the assay. Removal of the insert created a wound of 500 µm width. Wound closure was monitored 5 hours, 9 hours and 12 hours after removal of the insert. Images were acquired using the EVOS FL Auto Imaging System (Thermofisher Scientific). Wound width was quantified using the ImageJ software.

2.2.13 Subcellular fractionation

To determine the localization of lncRNAs into nuclear and cytoplasmic compartments, approximately 10 million cells were trypsinized and washed with cold PBS. Half of the sample was kept as the total fraction and half was used for subcellular fractionation. Both samples were resuspended in buffer A (10 mM Tris-HCl pH = 7.5, 1.5 mM MgCl₂, 140 mM NaCl, 0.05% Nonidet P40, 1X cOmplete EDTA-free Protease Inhibitor Cocktail (Sigma-Aldrich) and RNasin at 100U/ml (Promega)) and incubated 10 minutes on ice before SDS loading dye or TRIzol was added to the total fraction. The other fraction was added on top of a layer of buffer A plus sucrose (10 mM Tris-HCl pH = 7.5, 1.5 mM MgCl₂, 140 mM NaCl, 0.05% Nonidet P40, 50% sucrose) and centrifuged for 10 minutes at 12,000 g. The upper phase (cytoplasmic fraction) was collected and resuspended in SDS loading dye or TRIzol, the total nuclear fraction was

resuspended in buffer B (10 mM Tris-HCl pH = 7.5, 100 mM NaCl, 1 mM EGTA, 300 mM sucrose, 0.5 mM sodium metavanadate, 50 mM sodium fluoride, 1 mM PMSF, 0.5% Triton X-100, 1X cOmplete EDTA-free Protease Inhibitor Cocktail (Sigma-Aldrich) and RNasin at 100U/ml (Promega)) and left 10 minutes to incubate on ice. Next, it was centrifuged at 2,000 g for 5 minutes at 4°C, the supernatant (nuclear soluble fraction) and pellet (chromatin fraction) were collected and resuspended into SDS loading dye or TRIzol for either protein or RNA analysis, respectively.

2.2.14 RNA affinity purification and mass spectrometry analysis

LOC729970 protein interactors were isolated using the tethered RNA affinity capture system described by Iioka et al. (239) following the protocol published by Iioka et al (240). Briefly, LOC729970 RNA tagged with a tRNA scaffold and a streptavidin aptamer (Figure 2.1.A) was synthesized in an *in vitro* transcription reaction and bound to streptavidin beads. In parallel, 80% confluent cells (15 cm dishes) were harvested, protein lysates are prepared and blocked with white egg avidin (10 µg/mg protein) and yeast RNA (0.5 mg/mg protein) then combined with the tethered RNA-bound streptavidin beads for the affinity purification as shown in Figure 2.1.B. An antisense version of the RNA was used as a negative control in parallel to allow for the identification of specific protein interactors. Samples were resuspended in SDS loading dye and sent for mass spectrometry analysis at the proteomics facility of the center of excellence for mass spectrometry in King's College London, UK.

2.2.15 RNA immunoprecipitation

The cDNA sequence of GAPDH (glyceraldehyde 3-phosphate dehydrogenase), CYR61 (cysteine-rich angiogenic protein 61) and PABP (poly(A)-binding protein)

was cloned into a mammalian vector to allow for tagging of the proteins with MYC at the C-terminus. Briefly, six million ETCC-006 cells were plated the day before transfection on a 15 cm dish. 30 µg of plasmid DNA was transfected using Lipofectamine 3000 as per manufacturer's instructions, and the media was changed 6 hours after transfection. 48 hours after transfection, the cells were trypsinized and washed twice with cold PBS. They were resuspended in equal volume of polysome lysis buffer (100 mM KCl, 5 mM MgCl₂, 10 mM HEPES-NaOH pH 7, 0.5 % Nonidet P-40 (NP-40), 1 mM dithiothreitol (DTT), 200 units/ml RNase OUT and 1X cOmplete EDTA-free Protease Inhibitor Cocktail (Sigma-Aldrich)) and incubated for 5 min on ice. Protein lysates were stored at - 80°C to promote cell lysis. Myc-Trap magnetic beads (Chromotek) were washed for 5 min three times at 4°C in NT-2 buffer (25 mM Tris-HCl pH 7.4, 75 mM NaCl, 1 mM MgCl₂, 0.05 % NP-40, 20 mM EDTA pH 8.0, 1 mM DTT, 200 units/ml RNase OUT and 1X cOmplete EDTA-free Protease Inhibitor Cocktail (Sigma-Aldrich)). After the washes, beads were resuspended in 900 µl NT-2 and kept on ice. The protein lysates were centrifuged for 10 min at 4°C at 20,000 g. 1% of the total sample was kept as input fraction and resuspended in SDS loading dye or TRIzol and the rest was added to the beads. The mix lysate/bead was incubated on a rotating wheel at 4°C for 2 hours. After the immunoprecipitation, beads were washed three times with 1 ml NT-2 and resuspended in SDS loading dye or TRIzol for protein or RNA analysis, respectively.

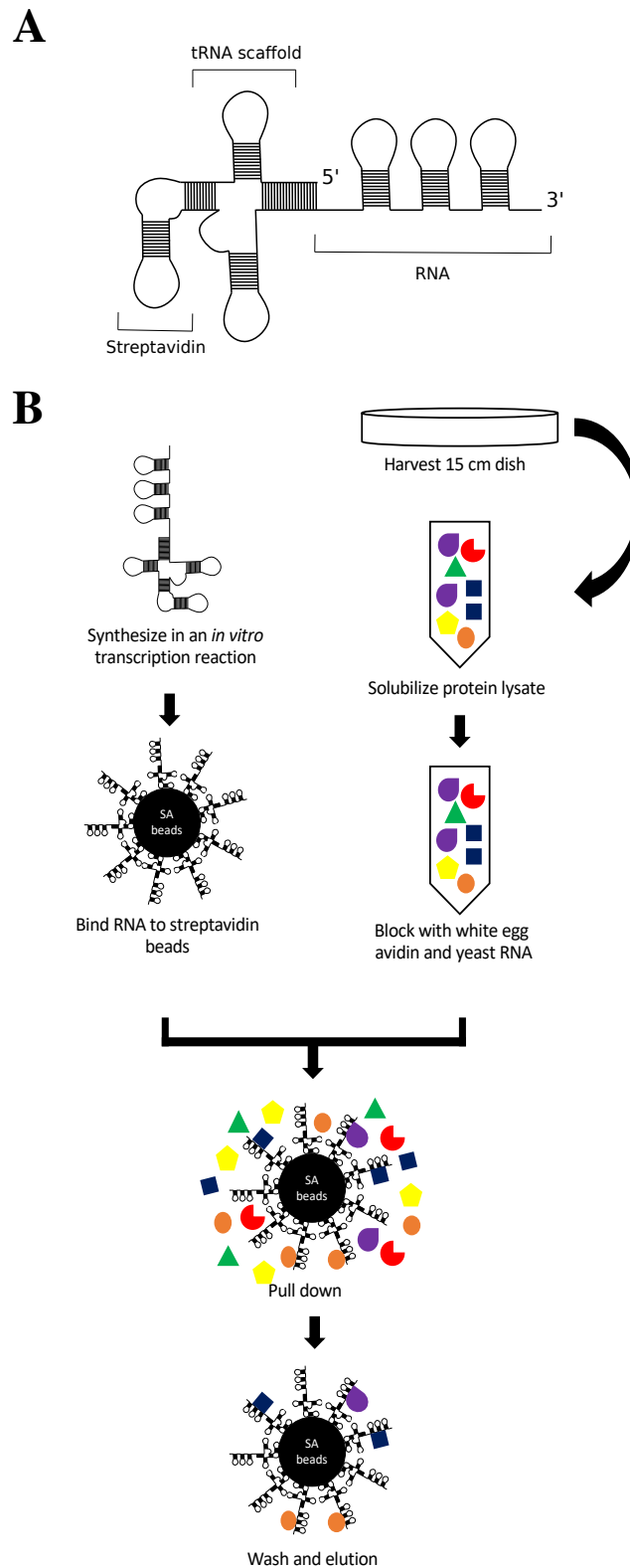


Figure 2.1: tRSA construct and RNA-affinity purification procedure. (A) Schematic of the tRSA construct. The target RNA is tagged by tRNA scaffold and a streptavidin aptamer. (B) Schematic of RNA affinity purification procedure. The

previous construct is synthesized in an *in vitro* transcription reaction and bound to streptavin beads. Protein lysates are prepared and blocked with white egg avidin and yeast RNA. The RNA bound to the beads is incubated with the blocked protein lysate. After several washes, only the proteins bound to the RNA are recovered.

Chapter 3

Molecular and cellular characterization of three models of ductal carcinoma *in situ* (DCIS)

3.1 Introduction

Medical research relies widely on the use of cellular *in vitro* models. In cancer research, a wide variety of cell lines are available. They allow researchers to study the details of cancer initiation and progression and are enormously beneficial to develop and test therapeutics drugs. In breast cancer research, researchers have been using a limited number of cell lines that are well characterized in the literature. Indeed, MCF-7, T47D and MDA-MB-231 are utilized in more than two thirds of the studies using cell lines (241). However, a cell line has to be chosen accordingly to the study that will be conducted (242). As our focus is on the study of early breast cancer, in particular DCIS, the use of cell lines that are models of invasive breast cancer is not relevant in our case.

Some cellular models of early stage breast cancer have been developed. MCF10DCIS.com is the DCIS cell line most used by researchers according to the literature. It was derived from a xenograft lesion following two trocar passages of the premalignant cell line MCF-10AT (243). SUM225CWN is another available model for DCIS; it is derived from a chest wall recurrence in a patient diagnosed and treated

for DCIS 6 years prior (244). While useful, those cell lines are not directly derived from a primary DCIS tumor and perhaps do not replicate the *in vivo* tumor as closely as other cell lines.

In 2014, Yong et al. (245) developed five DCIS sub-cell lines derived from a primary tumor by overexpressing the human telomerase reverse transcriptase (hTERT). hTERT expression allows primary cells to become immortalized without causing cancer-associated changes or altering phenotypic properties (246). Since those cells are directly derived from the primary tumor, we expect them to be models more accurate for DCIS study. For this study, we acquired two of those five sub-cell lines, ETCC-006 and ETCC-010. Those two sub-cell lines were chosen based on the original publication where the authors showed that each cell line had different abilities in proliferation and tumor growth when injected into mice, with ETCC-010 showing the most aggressive phenotype *in vivo* and ETCC-006 *in vitro*. Those cell lines have not been used in the literature other than in the original publication (247).

Yong et al. showed by immunoblotting and immunofluorescence that those cell lines express the estrogen receptor (245). Based on those observations, we aimed to stimulate those cells with estradiol and to analyze the changes in gene expression of estrogen responsive genes. We saw no difference in expression of estrogen responsive genes after estradiol stimulation (data not shown). We considered whether those results might be explained by a lack of expression of the estrogen receptor in those cells. In consequence, we decided to conduct a molecular and cellular characterization of those cells to, simultaneously, confront the results published and have a baseline of those cells cancer phenotype for our subsequent studies. In detail, we determined the

molecular subtype of those cells and we analyzed the expression level of several signaling proteins implicated in cancer progression. We also described their cellular behavior in terms of proliferation rate and migration ability. This work was conducted by comparing ETCC-006 and ETCC-010 to MCF10DCIS.com, the conventional model of DCIS, and to the widely characterized breast cancer cell lines, MCF-7, a hormone receptor positive breast cancer model, MDA-MB-231, a triple negative breast cancer model and MCF-10A, a normal immortalized breast epithelial cell line.

3.2 Results

3.2.1 Morphologic characterization of the DCIS cell lines MCF10DCIS.com, ETCC-006 and ETCC-010

For this work, MCF10DCIS.com, ETCC-006 and ETCC-010 DCIS cell lines were used. Before comparing the molecular details of those cell lines, we examined the similarity and differences in morphology. Despite being models of DCIS, they showed evident morphologic differences. We studied their morphologic differences while growing onto a culture plate (Figure 3.1 A) or onto coverslips after β -actin immunofluorescence and DAPI staining (Figure 3.1 B). Indeed, β -actin is a highly conserved and expressed protein implicated in the control of cell growth and migration (248). It is important for the maintenance of the cellular architecture and it is a good marker to analyze cell morphology. Differences in size, shape and organization can be noticed.

Similar to MCF-7, MCF10DCIS.com cell growth displayed high cellular contact between cells, the cells forming a cell hive-like structure with each other. This suggests

that those cells might be highly dependent on cellular contacts for growth to be possible. Although cell size was not quantified, observation under the microscope suggests that MCF10DCIS.com cells are of similar size as MCF-7 cells.

On the opposite, ETCC-006 cells seem to show the largest cell size of all cells analyzed in this study. Many cytoplasmic projections were observed, possibly indicative of side branching (249,250). Moreover, ETCC-006 cells were phenotypically striking because of their apparent flatness when visualized under the microscope; thus, observation under bright field microscopy proved to be more difficult than for other cell lines (Figure 3.1 A).

ETCC-010 cells were derived from the same DCIS tumor as ETCC-006, yet they show little morphologic similarities to each other. Instead, ETCC-010 cells displayed a higher degree of resemblance to triple negative, invasive MDA-MB-231. The cells are elongated with the nucleus occupying most of the available space within the cell. The difference in morphology between ETCC-006 and ETCC-010 is interesting since they both have the same genetic background. This is possibly explained by a difference in origin of the ETCC-006 and ETCC-010 cells in the DCIS tumor and highlights the high degree of tumor heterogeneity within DCIS lesions (251).

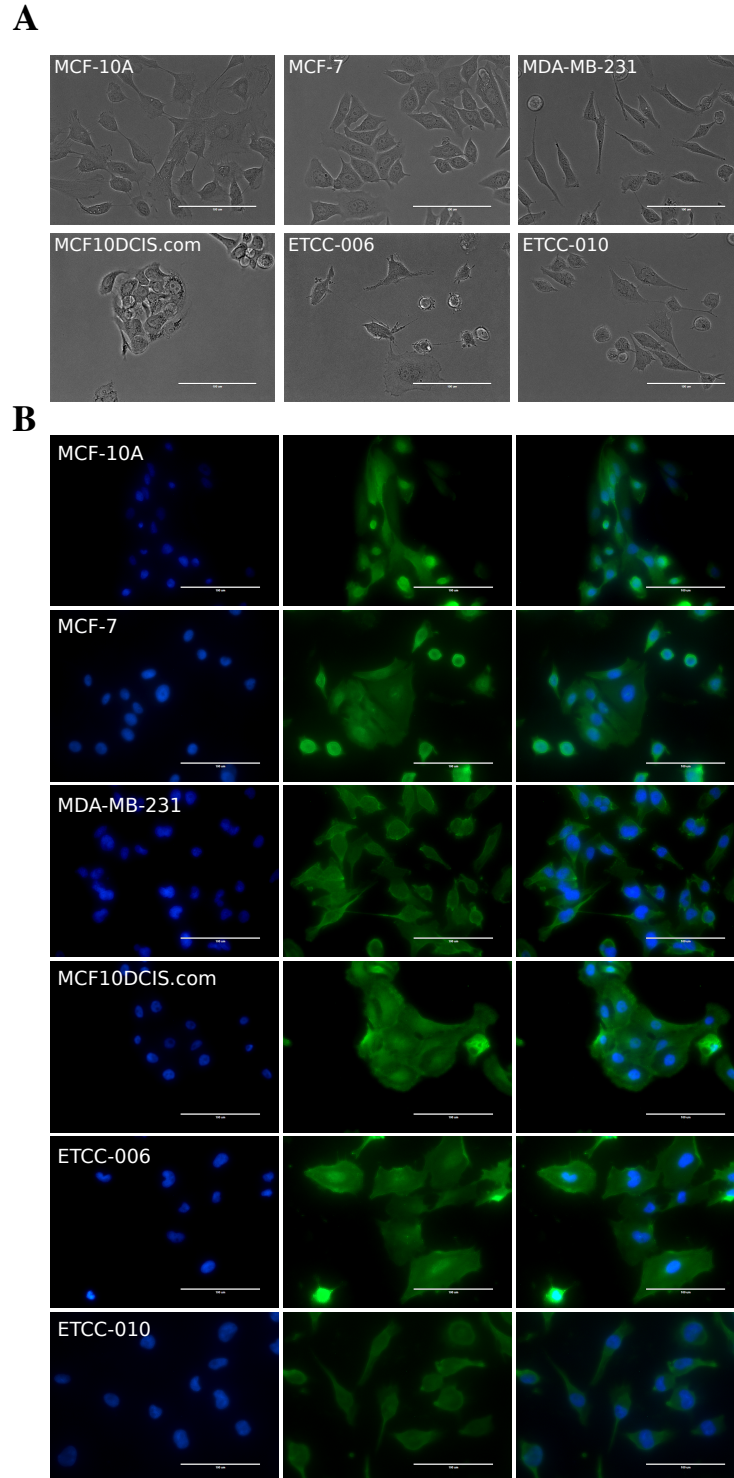


Figure 3.1: Phenotypic analysis of MCF10DCIS.com, ETCC-006 and ETCC-010 by comparison with MCF-10A, MCF-7 and MDA-MB-231 cell lines by (A) bright field microscopy and (B) immunofluorescence using an antibody against β actin coupled with an anti-mouse 488 secondary antibody. DAPI was used to stain cell

nuclei in the immunofluorescence analysis. Images were captured on the EVOS FL Auto Imaging System at 40X magnification. Scale bar represents 100 μ m.

3.2.2 Molecular subtype of the DCIS cell lines

In breast cancer care, testing of the ER, PR and HER2 receptor status is routinely done to assess the prognosis and the treatment that will be given to the patient (241). In breast cancer research, the ER, PR and HER2 receptor status of cell lines is verified to determine which breast cancer subtype those cell lines are a model of. Before analyzing the ER, PR and HER2 receptor status of the DCIS cell lines, we confirmed the published molecular subtype of the control cell lines by immunoblotting analysis. MCF-10A and MDA-MB-231 are known to have no expression of ER, PR and HER2. MCF-7 is a hormone receptor breast cancer cell line so it expresses ER, PR and has normal HER2 level (252). Those results were confirmed at both protein and RNA levels (Figure 3.2).

MCF10DCIS.com has the same genetic background as MCF-10AT, which was derived from a lesion of MCF-10A cells that have undergone an oncogenic transformation with an activated *HRAS* gene (253, 254). Therefore, it is not surprising that similar to MCF-10A, this cell line shows little expression of ER at the protein and RNA level. Low expression of PR was detected at the protein level (Figures 3.2 A and 3.2 B) but not at the RNA level (Figure 3.2 C). Since levels of ER and PR in MCF10DCIS.com are low, those receptors are likely not functionally relevant. High levels of HER2 were detected in MCF10DCIS.com, explained by a HER2 duplication in that cell line (255).

While ETCC-006 showed no expression of ER, PR and HER2, ETCC-010 showed low levels of ER and no expression of PR and HER2 at protein and RNA level. These results contradict those published by Yong et al. (245), where they show expression of ER and PR both by Western blotting and immunofluorescence staining. Our results are supported by an additional experiment where both ETCC-006 and ETCC-010 cells were deprived from estradiol in the media for 24 hours and then stimulated with the addition of estradiol to induce expression of estrogen responsive genes (data not shown). Expression in characterized estrogen responsive genes did not change through the estradiol starvation or reintroduction of estradiol in the media, confirming the lack of ER expression in ETCC-006 and that the ER in ETCC-010 cells is not functionally active.

Taken together, this suggests that ETCC-006 and ETCC-010 are functionally similar, based on hormone receptor status, to triple negative breast cancer cell lines.

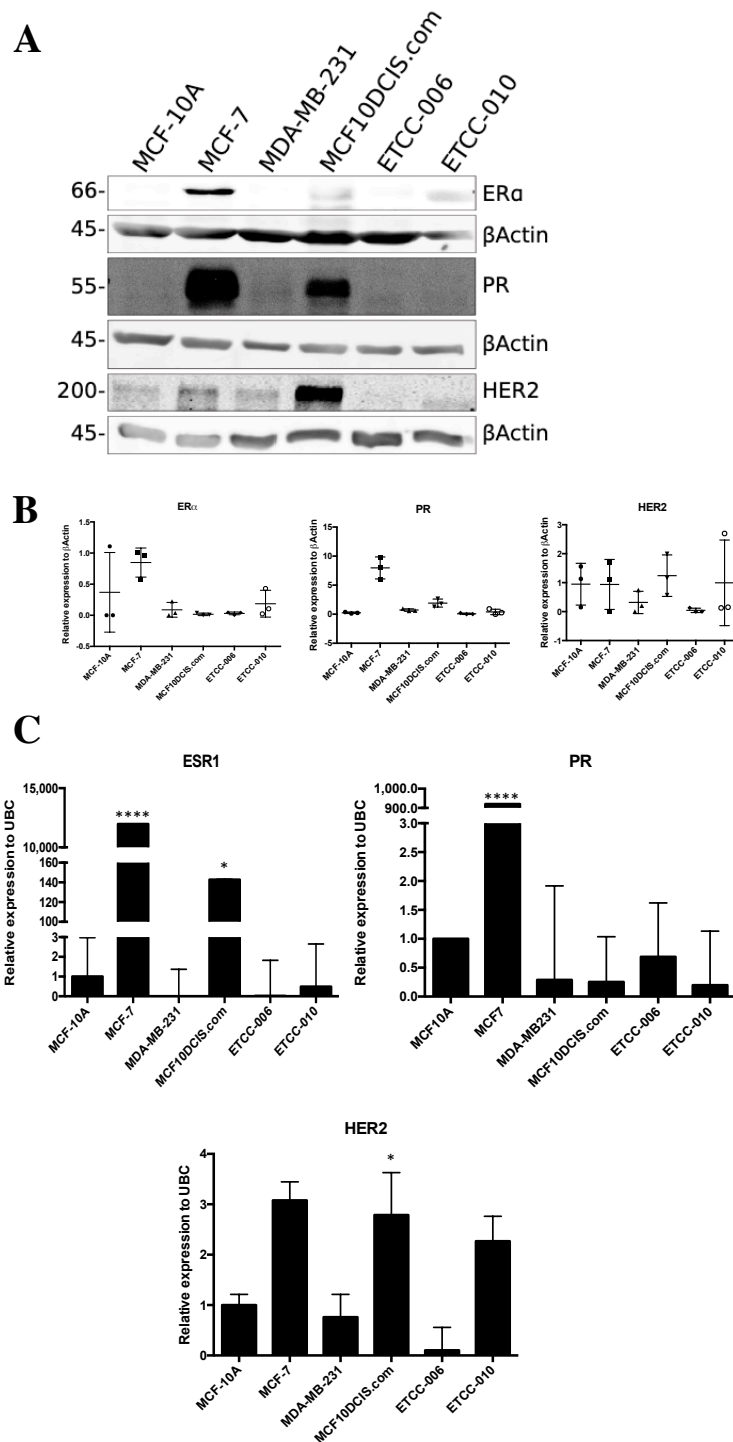


Figure 3.2: Molecular subtype of the DCIS cell lines. Estrogen receptor alpha (ERalpha), progesterone receptor (PR) and receptor tyrosine kinase ErbB-2 (HER2) levels were measured at the protein level (A) and (B) and RNA level (C) by immunoblotting and qRT-PCR. For the immunoblot, 42 µg of total protein was loaded per lane onto the gel. Protein level was quantified and normalized against βactin using

ImageJ. This experiment was repeated in triplicate, this is a representative immunoblot.

3.2.3 Analysis of the expression of Oncotype DX DCIS test markers in the DCIS cell lines

Since we have found that ETCC-006 and ETCC-010 display different characteristics to what was published by Yong et al. (2014) (245), we decided to characterize the cell lines using previously suggested DCIS markers. However, defining DCIS using a small set of known markers has been difficult (256), mainly because medical research on DCIS is limited in regards to that of IDC.

As we mentioned in the introduction, the Oncotype DX DCIS Score is a 12-gene panel that predicts the risk of DCIS recurrence. It includes five gene markers of proliferation (*MKI67* – Ki67, *STK15* also called *AURKA*, *BIRC5* – Survivin, *CCNB1* – Cyclin B1 and *MYBL2*), *PR* (progesterone receptor) and *GSTM1* (Glutathione S-Transferase M1) and five reference genes. The calculation of the DCIS score is as follows, first an average of the five proliferative markers is calculated and included in the following: DCIS score $\mu = 0.31 \times$ proliferation group score $- 0.08 \times PR - 0.09 \times GSTM1$ (102). Then, the DCIS score is calculated as follows: DCIS score = $(66.7 \times \text{DCIS score } \mu) + 10$. When the score is below zero, it is equal to zero and when it is greater than a hundred, it equals to a hundred. Risk of recurrence is then divided in three categories: low risk (DCIS score < 39), intermediate risk ($39 \leq \text{DCIS score} \leq 54$) and high risk (DCIS score ≥ 55).

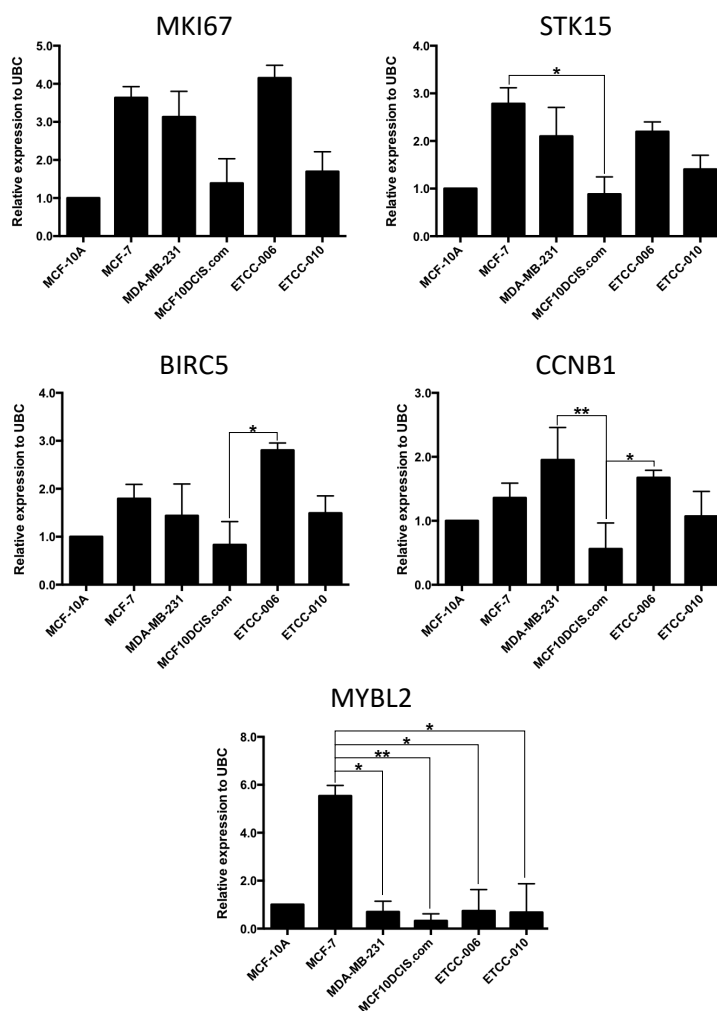
However, calculating a score that estimates a risk of recurrence would have no meaning for cell lines. Therefore, we decided to assess the expression levels of the markers used to calculate that score in the DCIS cell lines, ETCC-006 and ETCC-010 in addition to the other cell lines in our panel. From that, we could possibly hypothesize a correlation between expression of the proliferation markers and a more aggressive proliferative behavior of the cell line. Moreover, since the Oncotype DX DCIS Score was initially developed for ER positive IDC, results have to be interpreted with caution.

Figure 3.3 shows the results of a qRT-PCR to analyze expression of *MKI67*, *STK15*, *BIRC5*, *CCNB1*, *MYBL2*, *PR* and *GSTM1*. MCF-7 shows high expression of the proliferation markers (Figure 3.3 A) along with high expression of *PR* and *GSTM1* (Figure 3.3 B). MDA-MB-231 shows high expression of four proliferation markers, low expression of *PR* and a lower *GSTM1* expression than in MCF-7 cells. The expression level of this limited panel of markers correlates with the literature that shows that MDA-MB-231 is a model of TNBC associated with the worst prognosis whereas MCF-7 is a model for HR positive associated with a better prognosis (42).

Regarding MCF10DCIS.com, we observed low expression of the proliferation markers and low *PR* and *GSTM1* expression. Although no conclusion can be drawn from such a limited panel of genes, this might suggest that MCF10DCIS.com is less aggressive in term of proliferation than the two previously described lines. On the opposite, ETCC-006 shows high expression of four proliferative markers and low expression of *PR* and *GSTM1*. Following the same methodology as previously, this could suggest that ETCC-006 cells are more aggressive in term of proliferation than

MDA-MB-231. ETCC-010 shows a similar trend with a slight increase in *MKI67*, *STK15* and *BIRC5*. However, it shows lower expression of *PR* and higher expression of *GSTM1*, again possibly indicative that ETCC-010 is less aggressive in regard to proliferation than ETCC-006.

A



B

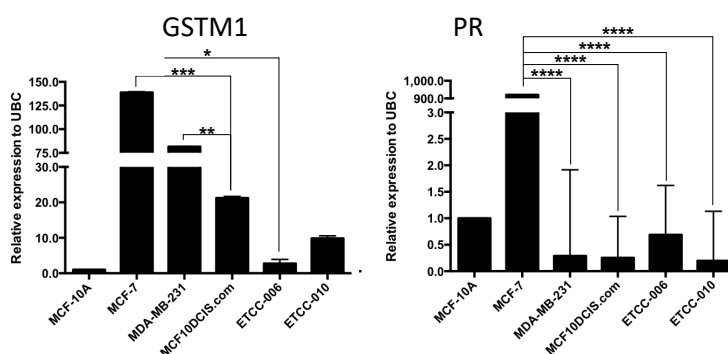


Figure 3.3: Expression level of Oncotype DX DCIS score markers in the DCIS cell lines analyzed by qRT-PCR. (A) Relative expression of the proliferation markers is associated with an increased risk of DCIS recurrence whereas (B) GSTM1 and PR have a protective role against the DCIS risk of recurrence. qRT-PCR was performed using cDNA synthesized from total RNA isolated from MCF-10A (normal-like), MCF-7 (triple positive breast cancer), MDA-MB-231 (TNBC) and MCF10DCIS.com, ETCC-006 and ETCC-010 (DCIS). All data was normalized to GAPDH and statistical analysis was done using one-way ANOVA (GraphPad v.8.3.0). **** p-value < 0.0001; *** p-value < 0.001; ** p-value < 0.01; * p-value < 0.05.

3.2.4 Genome-wide comparison of the ETCC-006 and ETCC-010 DCIS model to MCF-10A

In parallel to analyzing the molecular subtype of our cells, we performed RNAseq of ETCC-006 and ETCC-010 cell lines. RNAseq allows us to identify the transcripts present in the ETCC-006 and ETCC-010 cells at a genome-wide level. By comparing the level of expression of those transcripts in ETCC-006 and ETCC-010 to MCF-10A, we aimed to identify if genes in some cancer related pathways show altered expression. Before performing RNAseq, we analyzed the quality of the total RNA extracted from our cell lines. Figure 3.4 shows the bioanalyzer traces of the RNA used to prepare libraries for RNAseq. The two bands (Figure 3.4 A) and peaks (Figure 3.4 B) correspond to the 28S and 18S ribosomal RNA. Those values are integrated to calculate the RNA Integrity (RIN) value (Figure 3.4 C) that is indicative of the level of degradation of the RNA. RIN values range from 0 (degraded) to 10 (intact). We obtained RIN values of 10 for RNA extracted from ETCC-006 and ETCC-010. This was indicative of RNA of good quality that can be used for library preparation.

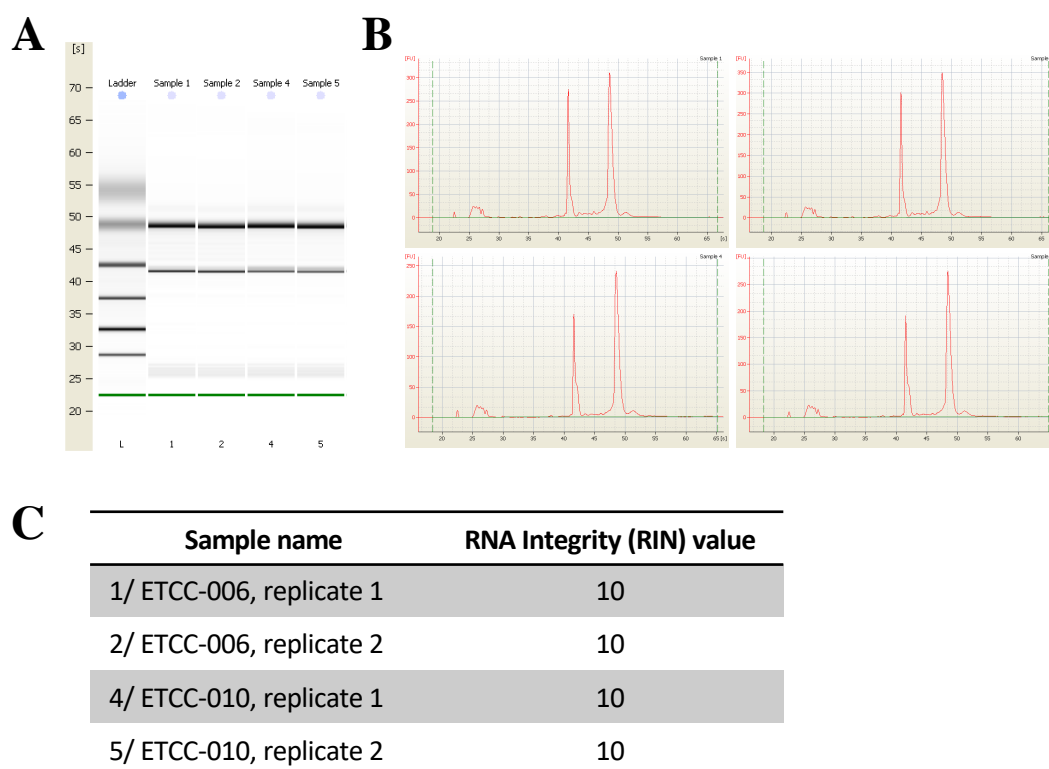


Figure 3.4: Bioanalyzer analysis of the ETCC-006 and ETCC-010 RNA. (A) Gel image and (B) electropherogram for each sample (Agilent). Those two bands and peaks correspond to the 28S and 18S ribosomal RNA. (C) RIN values for each sample. The RIN values for each sample is 10, which suggest that it can be used for library preparation.

After sequencing, reads were trimmed and filtered for quality with Trimmomatic (231) and aligned to the human reference genome (hg19) with RNA Star (232). Gene expression was quantified with featureCounts (233) and differential expression was performed using DESeq2 (234). We compared the transcriptome of the ETCC lines to that of MCF-10A, using an available dataset (EBI - study: ERP000799; experiment: ERX016686, biosample: SAMEA1484687). After the differential expression, we excluded genes with a p-value over 0.05. Figure 3.5 A shows a Venn diagram of the number of genes with higher expression in ETCC-006 and ETCC-010 in green and

lower expression in ETCC-006 and ETCC-010 in red. The most significant genes (p-value over 0.05 and a log₂(fold change) between -2 and 2) are recapitulated in the Tables 3.1, 3.2, 3.3 and 3.4. 235 and 246 genes were respectively up-expressed in ETCC-006 and ETCC-010, with 99 up-expressed genes in common. 194 and 240 genes were respectively down-expressed in ETCC-006 and ETCC-010 with 91 down-expressed genes in common (Figure 3.5 A). Next, reactome pathway-based analyses were performed to identify pathways and functions specific to ETCC cell lines (Figure 3.5 B and 3.5 D) (235). This bioinformatic tool allows the identification of enriched pathways in a dataset. The most enriched pathways common to ETCC-006 and ETCC-010 related to cell proliferation: mitotic G1-G1/S phases, G1/S transition, unwinding of DNA (S phase), etc. Integrin-cell surface interaction gene products were enriched in both cell lines compared to MCF-10A. Since those regulate cell motility, it possibly indicates that ETCC-006 and ETCC-010 have migratory abilities. The reactome bioinformatic tool also allowed us to generate a String analysis of the genes implicated in the pathways we just mentioned (Figure 3.5.C and 3.5 E). Proliferation in both cell lines seems to rely mostly on commonly upregulated genes (*CCND1*, *MCM* family, *MYBL2*, etc.). *CCND1* is an important regulator of the G1/S transition (257), the *MCM* proteins constitute a family of protein involved in the regulation of DNA replication (258), and *MYBL2* is a transcription factor involved in the regulation of the G1/S transition (259). Interestingly, an enrichment of “signaling by VEGF” was observed in ETCC-006 and VEGF signaling has been shown to be upregulated in DCIS (260). Overall, most enriched pathways in ETCC-006 and ETCC-010 are related to DNA replication and proliferation.

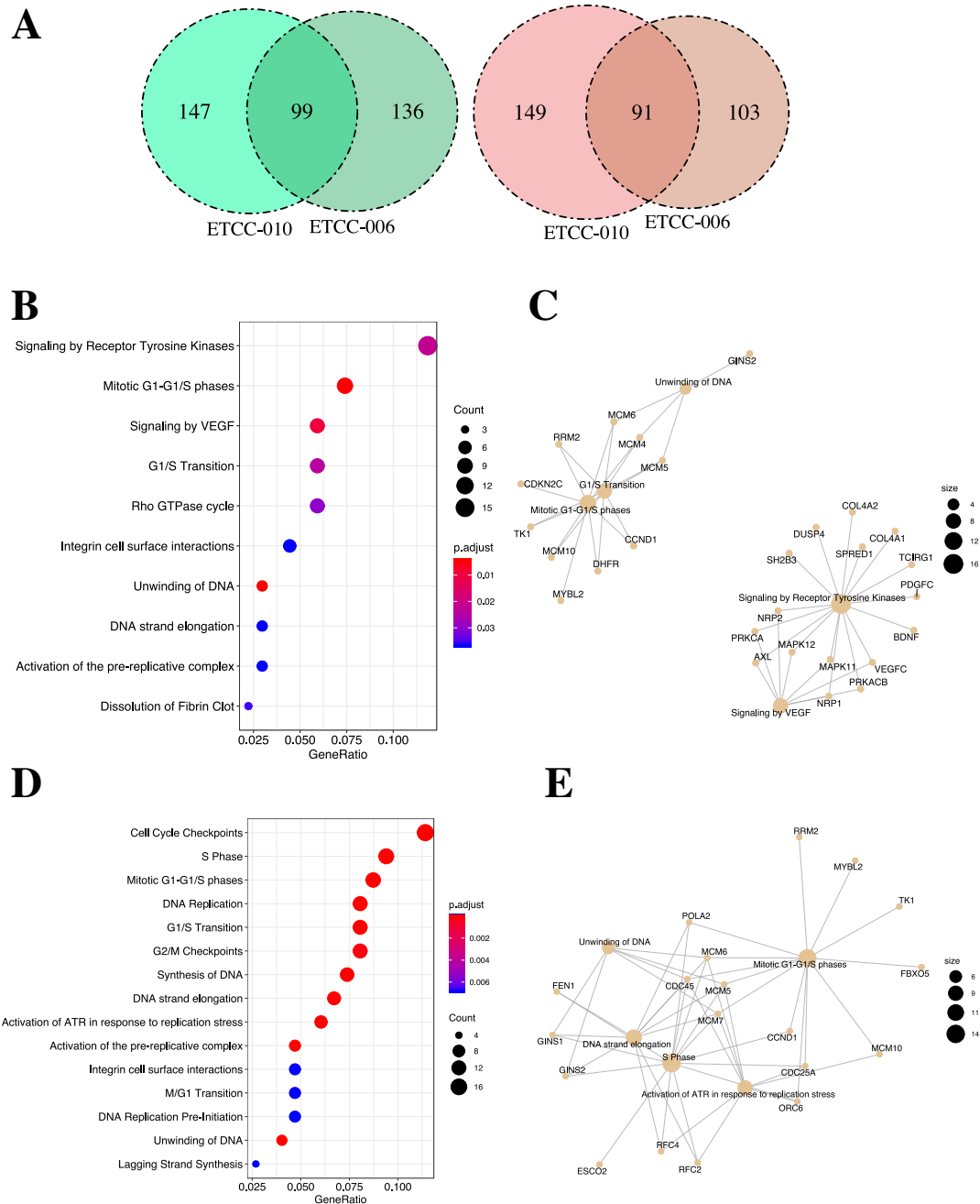


Figure 3.5: Transcriptomic comparison of the DCIS cell lines ETCC-006 and ETCC-010 to MCF-10A. (A) Venn diagram of the number of genes with higher expression in green and lower expression in red in ETCC-006 and ETCC-010 compared to MCF-10A (p -value < 0.05). (B) and (D) Reactome and (C) and (E) String pathway analysis of the differentially expressed genes in ETCC-006 and ETCC-0010, (p -value < 0.05) to identify enriched pathways in those cells and the genes associated with those pathways.

Table 3.1: List of genes with higher expression in ETCC-006 cells.

Ensembl_gene_ID	Gene_name	log2(Fold_Change)	Fold_Change	P-value	Q-value
ENSG00000166147	FBN1	2,00537175	4,014921414	0,032692976	0,602882305
ENSG00000134013	LOXL2	2,006146867	4,017079092	0,03310005	0,602882305
ENSG00000253276	CCDC71L	2,009497319	4,026419022	0,030633694	0,602882305
ENSG00000187210	GCNT1	2,012598435	4,035083241	0,029187788	0,602882305
ENSG00000186866	POFUT2	2,01676227	4,04674592	0,028678916	0,602882305
ENSG00000167552	TUBA1A	2,017546423	4,04894606	0,028245522	0,602882305
ENSG00000140859	KIFC3	2,019589062	4,054682816	0,028740868	0,602882305
ENSG00000188130	MAPK12	2,021745127	4,060746949	0,030288866	0,602882305
ENSG00000084731	KIF3C	2,022011711	4,061497372	0,03015363	0,602882305
ENSG00000218336	ODZ3	2,026081009	4,072969495	0,029368067	0,602882305
ENSG00000147852	VLDLR	2,026814777	4,075041571	0,03061371	0,602882305
ENSG00000169851	PCDH7	2,027478515	4,076916801	0,029011038	0,602882305
ENSG00000168874	ATOH8	2,028606309	4,080105082	0,029042771	0,602882305
ENSG00000102401	ARMCX3	2,031926492	4,089505749	0,028945403	0,602882305
ENSG00000069424	KCNAB2	2,034759699	4,097544731	0,02817666	0,602882305
ENSG00000187498	COL4A1	2,035041349	4,098344752	0,02880624	0,602882305
ENSG00000166068	SPRED1	2,035691421	4,100191865	0,028463789	0,602882305
ENSG00000134871	COL4A2	2,036482812	4,102441643	0,027385248	0,602882305
ENSG00000196923	PDLIM7	2,038674766	4,108679412	0,027188263	0,602882305
ENSG00000165801	ARHGEF40	2,039692184	4,11157796	0,027835476	0,602882305
ENSG00000174938	SEZ6L2	2,040209254	4,113051837	0,028654408	0,602882305
ENSG00000182534	MXRA7	2,042794136	4,120427812	0,026906935	0,602882305
ENSG00000154127	UBASH3B	2,046772852	4,131806958	0,028120406	0,602882305
ENSG00000139278	GLIPR1	2,048898637	4,137899589	0,027484945	0,602882305
ENSG00000186480	INSIG1	2,052693324	4,148797736	0,028775947	0,602882305
ENSG00000162520	SYNC	2,056535431	4,159861315	0,028682792	0,602882305
ENSG00000115232	ITGA4	2,061168429	4,173241557	0,028085463	0,602882305
ENSG00000099822	HCN2	2,063374763	4,179628636	0,028059552	0,602882305
ENSG00000121900	TMEM54	2,066401748	4,188407313	0,027666793	0,602882305
ENSG00000118257	NRP2	2,070477392	4,200256382	0,025519771	0,602882305
ENSG00000171067	C11orf24	2,074443476	4,211819109	0,025329752	0,602882305
ENSG00000082781	ITGB5	2,074965826	4,213344339	0,027528586	0,602882305
ENSG00000149256	ODZ4	2,079611491	4,226933721	0,025982485	0,602882305
ENSG00000120875	DUSP4	2,08056251	4,229721019	0,025359227	0,602882305
ENSG00000135318	NT5E	2,084937415	4,242566915	0,024861902	0,602882305
ENSG00000122870	BICC1	2,087564266	4,250298797	0,025993719	0,602882305
ENSG00000169902	TPST1	2,088392696	4,252740119	0,026513383	0,602882305
ENSG00000168010	ATG16L2	2,090337741	4,258477542	0,02634444	0,602882305
ENSG00000140682	TGFB11	2,093114821	4,266682685	0,025070812	0,602882305
ENSG00000154229	PRKCA	2,093792942	4,268688659	0,025134817	0,602882305
ENSG00000127920	GNG11	2,09759249	4,279945692	0,025815911	0,602882305
ENSG00000163577	EIF5A2	2,098257523	4,281919053	0,02572627	0,602882305
ENSG00000165434	PGM2L1	2,099257279	4,284887359	0,025790411	0,602882305
ENSG00000111252	SH2B3	2,099823358	4,286568976	0,02520783	0,602882305
ENSG00000162073	PAQR4	2,100729266	4,289261475	0,025705031	0,602882305
ENSG00000110031	LPXN	2,107428441	4,309225037	0,024643927	0,602882305
ENSG00000177706	FAM20C	2,108707842	4,313048208	0,024925926	0,602882305
ENSG00000172379	ARNT2	2,11174182	4,322128065	0,024984557	0,602882305
ENSG00000149948	HMG2A	2,117171096	4,338424119	0,024108414	0,602882305
ENSG00000142875	PRKACB	2,117610781	4,339746527	0,023872563	0,602882305
ENSG00000012171	SEMA3B	2,119196129	4,344518005	0,024447928	0,602882305
ENSG00000173068	BNC2	2,121355793	4,351026467	0,024331361	0,602882305
ENSG00000099250	NRP1	2,122932792	4,355785141	0,02422998	0,602882305
ENSG00000102038	SMARCA1	2,125101431	4,362337621	0,023915001	0,602882305
ENSG00000117525	F3	2,126041528	4,365181157	0,02370369	0,602882305
ENSG00000183287	CCBE1	2,126574368	4,366793676	0,023634797	0,602882305
ENSG00000198910	L1CAM	2,126792023	4,367452531	0,023690863	0,602882305
ENSG00000175602	CCDC85B	2,130056785	4,377347096	0,023618617	0,602882305
ENSG00000176697	BDNF	2,130914315	4,379949739	0,023494994	0,602882305
ENSG00000148516	ZEB1	2,131738726	4,382453325	0,023456462	0,602882305
ENSG00000105137	SYDE1	2,132945412	4,386120391	0,023568861	0,602882305
ENSG00000166073	GPR176	2,133175669	4,38682048	0,02350937	0,602882305
ENSG00000171617	ENC1	2,136966426	4,398362237	0,023308196	0,602882305
ENSG00000168785	TSPAN5	2,138158378	4,401997658	0,023217441	0,602882305
ENSG00000078098	FAP	2,138238404	4,402241844	0,022968995	0,602882305
ENSG00000107984	DKK1	2,138579761	4,403283585	0,022782067	0,602882305
ENSG00000131711	MAP1B	2,140389514	4,408810639	0,023002107	0,602882305
ENSG00000251493	FOXDI	2,140757541	4,409935458	0,022428731	0,602882305
ENSG00000080823	MOK	2,14513454	4,423335096	0,022636967	0,602882305
ENSG00000133816	MICAL2	2,146324096	4,426983806	0,022010291	0,602882305
ENSG00000132561	MATN2	2,146901478	4,428755888	0,022649037	0,602882305
ENSG00000139289	PHLDA1	2,146987995	4,429021483	0,022095994	0,602882305
ENSG00000150630	VEGFC	2,151822526	4,443888213	0,022354465	0,602882305
ENSG00000214530	STARD10	2,153108308	4,447850534	0,021934448	0,602882305
ENSG00000152377	SPOCK1	2,153998813	4,450599905	0,021870838	0,602882305
ENSG00000076706	MCAM	2,1616674	4,474316772	0,02172499	0,602882305
ENSG00000164741	DLC1	2,161699123	4,474415158	0,021732194	0,602882305
ENSG00000134986	NREP	2,163055601	4,478624155	0,021575179	0,602882305
ENSG00000188549	C15orf52	2,167320498	4,491883461	0,021289819	0,602882305
ENSG00000166396	SERPINB7	2,173643219	4,511612683	0,020961749	0,602882305

Table 3.2: List of genes with higher expression in ETCC-010 cells.

Ensembl_gene_ID	Gene_name	log2(Fold_Change)	Fold_Change	P-value	Q-value
ENSG00000186340	THBS2	2,00469372	4,01303495	0,03150945	0,57140492
ENSG00000070081	NUCB2	2,00820782	4,02282178	0,02659296	0,57140492
ENSG00000213190	MLLT11	2,01046726	4,02912695	0,02993689	0,57140492
ENSG00000092470	WDR76	2,01060145	4,02950173	0,02840451	0,57140492
ENSG00000168785	TSPAN5	2,01075963	4,02994355	0,03015257	0,57140492
ENSG00000050344	NFE2L3	2,01448919	4,04037496	0,02909438	0,57140492
ENSG00000166147	FBN1	2,01471083	4,04099575	0,02600722	0,57140492
ENSG00000087086	FTL	2,01474196	4,04108294	0,03176084	0,57140492
ENSG00000100558	PLEK2	2,01565076	4,04362934	0,03006859	0,57140492
ENSG00000173334	TRIB1	2,01673621	4,04667281	0,02957293	0,57140492
ENSG00000101236	RNF24	2,02380513	4,06654938	0,02753213	0,57140492
ENSG00000178773	CPNE7	2,02416544	4,0675651	0,02836701	0,57140492
ENSG00000005189	LOC81691	2,03038507	4,08513872	0,02815101	0,57140492
ENSG00000132561	MATN2	2,03055085	4,08560818	0,02848678	0,57140492
ENSG00000138356	AOX1	2,03440134	4,09652705	0,02514018	0,57140492
ENSG00000101400	SNTA1	2,03521053	4,09882538	0,02826021	0,57140492
ENSG00000129226	CD68	2,0373767	4,10498429	0,02512443	0,57140492
ENSG00000109814	UGDH	2,03931757	4,11051046	0,02488166	0,57140492
ENSG00000101057	MYBL2	2,03973847	4,11170986	0,0248432	0,57140492
ENSG00000176697	BDNF	2,04395894	4,12375589	0,0274949	0,57140492
ENSG00000182871	COL18A1	2,04811004	4,13563836	0,02774289	0,57140492
ENSG00000175602	CCDC85B	2,05017562	4,14156382	0,02439289	0,57140492
ENSG00000182054	IDH2	2,05474728	4,15470857	0,0242124	0,57140492
ENSG00000102401	ARMCX3	2,05481588	4,15490612	0,02626156	0,57140492
ENSG00000157193	LRP8	2,05499635	4,15542589	0,02449899	0,57140492
ENSG00000169902	TPST1	2,0570914	4,1614647	0,02790849	0,57140492
ENSG00000183287	CCBE1	2,0637386	4,18068284	0,02607397	0,57140492
ENSG00000163577	EIF5A2	2,07074284	4,20102927	0,02667545	0,57140492
ENSG00000184254	ALDH1A3	2,0719329	4,20449607	0,02771302	0,57140492
ENSG00000134013	LOXL2	2,07596336	4,2162586	0,0240474	0,57140492
ENSG00000187240	DYNC2H1	2,07929711	4,22601273	0,02541928	0,57140492
ENSG00000135931	ARMC9	2,08008983	4,22833544	0,02393452	0,57140492
ENSG00000135541	AH11	2,0803603	4,22912822	0,02302476	0,57140492
ENSG00000135318	NTSE	2,08040195	4,22925032	0,02329376	0,57140492
ENSG00000105011	ASF1B	2,08045314	4,22940038	0,02404728	0,57140492
ENSG00000162073	PAQR4	2,08185499	4,23351202	0,02629446	0,57140492
ENSG00000188130	MAPK12	2,08305031	4,23702109	0,02541861	0,57140492
ENSG00000176485	PLA2G16	2,08512762	4,24312629	0,02585199	0,57140492
ENSG00000166073	GPR176	2,08766312	4,25059003	0,02526509	0,57140492
ENSG00000134138	MEIS2	2,09166204	4,26238834	0,02300469	0,57140492
ENSG00000131016	AKAP12	2,09638611	4,27636831	0,02431271	0,57140492
ENSG00000137692	DCUN1D5	2,09720035	4,27878251	0,02540125	0,57140492
ENSG00000145335	SNCA	2,0974576	4,27954553	0,02225117	0,57140492
ENSG00000152465	NMT2	2,10344673	4,29734834	0,02199495	0,57140492
ENSG00000198832	SELM	2,10560429	4,30377988	0,02288946	0,57140492
ENSG00000140859	KIFC3	2,10741956	4,30919851	0,02236106	0,57140492
ENSG00000172731	LRRC20	2,10827588	4,31175703	0,02451587	0,57140492
ENSG00000079819	EPB41L2	2,11711232	4,33824736	0,02160503	0,57140492
ENSG00000138166	DUSP5	2,11930376	4,34484212	0,02374424	0,57140492
ENSG00000171848	RRM2	2,12116846	4,35046153	0,02309256	0,57140492
ENSG00000188549	C15orf52	2,12676159	4,3673604	0,0220717	0,57140492
ENSG00000164741	DLC1	2,12834483	4,37215586	0,02244496	0,57140492
ENSG00000065328	MCM10	2,13385099	4,38887441	0,02240237	0,57140492
ENSG00000130956	HABP4	2,13569377	4,39448398	0,02297145	0,57140492
ENSG00000166780	C16orf45	2,14284929	4,41633402	0,02245198	0,57140492
ENSG00000011028	MRC2	2,1458657	4,42557742	0,02066116	0,57140492
ENSG00000164292	RHOBTB3	2,1463887	4,42718206	0,02030158	0,57140492
ENSG00000131153	GINS2	2,14914848	4,43565908	0,0215529	0,57140492
ENSG00000178531	CTXN1	2,15326764	4,44834178	0,0219255	0,57140492
ENSG00000181649	PHLDA2	2,15435017	4,45168087	0,02065523	0,57140492
ENSG00000025156	HSF2	2,15649294	4,45829766	0,01978918	0,57140492
ENSG00000105137	SYDE1	2,15788207	4,4625925	0,02151174	0,57140492
ENSG00000147852	VLDLR	2,15928284	4,46692751	0,02160048	0,57140492
ENSG00000123485	HJURP	2,16393162	4,48134445	0,01949647	0,57140492
ENSG00000144366	GULP1	2,16545971	4,48609356	0,02094779	0,57140492
ENSG00000153904	DDAH1	2,16915622	4,49760267	0,02075234	0,57140492
ENSG00000022267	FHL1	2,17059683	4,50209602	0,01965239	0,57140492
ENSG00000184371	CSF1	2,17134891	4,50444361	0,01979885	0,57140492
ENSG00000174938	SEZ6L2	2,17661155	4,52090485	0,02007934	0,57140492
ENSG00000150630	VEGFC	2,17718366	4,52269799	0,02023496	0,57140492
ENSG00000120875	DUSP4	2,17797799	4,52518883	0,01920357	0,57140492
ENSG00000129514	FOXA1	2,1812084	4,53533273	0,02035573	0,57140492
ENSG00000187498	COL4A1	2,18481526	4,54668566	0,01955353	0,57140492
ENSG00000100739	BDKRB1	2,18813016	4,55714465	0,0198854	0,57140492
ENSG00000152377	SPOCK1	2,19508546	4,57916789	0,01832655	0,57140492
ENSG00000146350	C6orf170	2,19519769	4,57952412	0,01910033	0,57140492
ENSG00000143515	ATP8B2	2,20107967	4,5982333	0,01933562	0,57140492
ENSG00000131711	MAP1B	2,20263936	4,60320712	0,01903347	0,57140492
ENSG00000159167	STC1	2,20376751	4,60680812	0,01805036	0,57140492
ENSG00000099250	NRP1	2,20694306	4,61695944	0,01798732	0,57140492
ENSG00000128342	LIF	2,21368407	4,63858273	0,01850091	0,57140492
ENSG00000134871	COL4A2	2,21697929	4,64918969	0,01754789	0,57140492
ENSG00000111252	SH2B3	2,22310844	4,66898335	0,01801359	0,57140492
ENSG00000079257	LXN	2,2331087	4,70145954	0,017502	0,57140492
ENSG00000080823	MOK	2,23835837	4,71859833	0,0169924	0,57140492

Table 3.3: List of genes with lower expression in ETCC-006 cells.

Ensembl_gene_ID	Gene_name	log2(Fold_Change)	Fold_Change	P-value	Q-value
ENSG00000142178	SIK1	-2,172414786	0,221839045	0,0210484	0,602882305
ENSG00000183943	PRKX	-2,162687663	0,22333981	0,021520798	0,602882305
ENSG00000164070	HSPA4L	-2,162616571	0,223350816	0,021559321	0,602882305
ENSG00000143850	PLEKHA6	-2,160758751	0,223638619	0,021510307	0,602882305
ENSG00000168497	SDPR	-2,159664442	0,223808317	0,021486477	0,602882305
ENSG00000168672	FAM84B	-2,158833763	0,22393722	0,021917583	0,602882305
ENSG00000141441	FAM59A	-2,158723892	0,223954275	0,021780102	0,602882305
ENSG00000125848	FLRT3	-2,157869398	0,22408696	0,021520268	0,602882305
ENSG00000120899	PTK2B	-2,155106687	0,22451649	0,022024211	0,602882305
ENSG00000197375	SLC22A5	-2,153594056	0,224752013	0,02195771	0,602882305
ENSG00000135378	PRRG4	-2,148661302	0,225521784	0,022545592	0,602882305
ENSG00000140022	STON2	-2,148260888	0,225584385	0,021923201	0,602882305
ENSG00000163624	CDS1	-2,146127262	0,225918252	0,022721881	0,602882305
ENSG00000076944	STXBP2	-2,143120482	0,226389589	0,022907562	0,602882305
ENSG00000003436	TFPI	-2,141626249	0,226624187	0,022692812	0,602882305
ENSG00000102890	ELMO3	-2,141198084	0,226691455	0,023036762	0,602882305
ENSG00000110042	DTX4	-2,140507847	0,226799938	0,023081136	0,602882305
ENSG00000186716	BCR	-2,135819834	0,22753812	0,023233552	0,602882305
ENSG00000162512	SDC3	-2,131902035	0,228156864	0,023625049	0,602882305
ENSG00000165804	ZNF219	-2,129688618	0,228507177	0,02351414	0,602882305
ENSG00000143320	CRABP2	-2,124919176	0,229263855	0,023979111	0,602882305
ENSG00000073350	LLGL2	-2,123151046	0,229545007	0,023895094	0,602882305
ENSG00000173898	SPTBN2	-2,12052662	0,229962955	0,023541058	0,602882305
ENSG00000152229	PSTPIP2	-2,116832559	0,230552536	0,02452484	0,602882305
ENSG00000080546	SESN1	-2,116579558	0,230592971	0,023667771	0,602882305
ENSG00000143126	CELSR2	-2,114593468	0,230910636	0,023423845	0,602882305
ENSG00000102554	KLF5	-2,114250927	0,230965468	0,024420739	0,602882305
ENSG00000119771	KLHL29	-2,1131123	0,231147826	0,024886128	0,602882305
ENSG00000129595	EPB41L4A	-2,11293947	0,231175519	0,024446505	0,602882305
ENSG00000025708	TYMP	-2,112679849	0,231217124	0,024896179	0,602882305
ENSG00000117266	CDK18	-2,110763913	0,23152439	0,025047769	0,602882305
ENSG00000130940	CASZ1	-2,109647196	0,231703671	0,025093975	0,602882305
ENSG00000126822	PLEKHG3	-2,107218471	0,232094064	0,025194226	0,602882305
ENSG00000102854	MSLN	-2,105054928	0,232442386	0,024791224	0,602882305
ENSG00000153094	BCL2L11	-2,103870277	0,232633331	0,024602153	0,602882305
ENSG00000152056	AP1S3	-2,098799968	0,233452352	0,025650272	0,602882305
ENSG00000184363	PKP3	-2,098396418	0,233517663	0,02417241	0,602882305
ENSG00000150907	FOXO1	-2,098252034	0,233541034	0,024181244	0,602882305
ENSG00000196850	PPTC7	-2,09495006	0,234076164	0,024659736	0,602882305
ENSG00000115339	GALNT3	-2,088860859	0,235066221	0,025866557	0,602882305
ENSG00000057294	PKP2	-2,087957887	0,235213393	0,026031104	0,602882305
ENSG00000095203	EPB41L4B	-2,087839501	0,235232695	0,026263	0,602882305
ENSG00000138162	TACC2	-2,08521805	0,235660513	0,024858473	0,602882305
ENSG00000157557	ETS2	-2,083789607	0,235893961	0,026189736	0,602882305
ENSG00000196743	GM2A	-2,082498544	0,236105157	0,027046702	0,602882305
ENSG00000112297	AIM1	-2,079808133	0,236545868	0,026276367	0,602882305
ENSG00000023445	BIRC3	-2,076805094	0,237038763	0,025262441	0,602882305
ENSG00000143674	KIAA1804	-2,075456067	0,237260515	0,026812978	0,602882305
ENSG00000148400	NOTCH1	-2,072608437	0,237729289	0,027584464	0,602882305
ENSG00000148671	C10orf116	-2,070331044	0,238104857	0,02601608	0,602882305
ENSG00000116016	EPAS1	-2,069334142	0,238269444	0,026738088	0,602882305
ENSG000000021762	OSBP15	-2,060161206	0,239789234	0,026047688	0,602882305
ENSG00000115008	IL1A	-2,057258105	0,240272242	0,028330732	0,602882305
ENSG00000184205	TSPYL2	-2,055557046	0,240555711	0,026764588	0,602882305
ENSG00000164251	F2RL1	-2,049537144	0,241561569	0,027600893	0,602882305
ENSG00000111679	PTPN6	-2,049105852	0,241633795	0,029159258	0,602882305
ENSG00000151690	MFSD6	-2,042370054	0,242764596	0,027460915	0,602882305
ENSG00000130707	ASS1	-2,040281116	0,24311636	0,030200165	0,602882305
ENSG00000116741	RGS2	-2,039469787	0,24325312	0,027142442	0,602882305
ENSG00000186395	KRT10	-2,038304722	0,243449641	0,030416539	0,602882305
ENSG00000197565	COL4A6	-2,032076322	0,244502933	0,027788459	0,602882305
ENSG00000197747	S100A10	-2,0294011	0,244956741	0,030903442	0,602882305
ENSG00000133056	PIK3C2B	-2,028018775	0,24519156	0,02978733	0,602882305
ENSG00000054654	SYNE2	-2,026791284	0,245400266	0,031428476	0,602882305
ENSG00000168209	DDIT4	-2,026389381	0,245468638	0,031441439	0,602882305
ENSG00000099204	ABLM1	-2,023883506	0,245895373	0,030884537	0,602882305
ENSG00000168118	RAB4A	-2,022647514	0,246106128	0,028220292	0,602882305
ENSG00000108679	LGALS3BP	-2,019906523	0,246574152	0,028378807	0,602882305
ENSG00000143772	ITPKB	-2,017300295	0,247019991	0,031459112	0,602882305
ENSG00000170421	KRT8	-2,015761619	0,247283585	0,029665467	0,602882305
ENSG00000148175	STOM	-2,012704899	0,247808074	0,032334391	0,602882305
ENSG00000162599	NFIA	-2,010215826	0,248235985	0,029969165	0,602882305
ENSG00000170385	SLC30A1	-2,005202245	0,249100143	0,029726465	0,602882305
ENSG00000157514	TSC2D3	-2,001827202	0,24968357	0,032737518	0,602882305
ENSG00000170412	GPRC5C	-2,001239525	0,249785299	0,0315974	0,602882305

Table 3.4: List of genes with lower expression in ETCC-010 cells.

Ensembl_gene_ID	Gene_name	log2(Fold_Change)	Fold_Change	P-value	Q-value
ENSG00000115129	TP53I3	-2,2297759	0,21319184	0,01716659	0,57140492
ENSG00000021762	OSBP15	-2,2168707	0,21510744	0,01754044	0,57140492
ENSG00000102890	ELMO3	-2,2162669	0,21519749	0,01845101	0,57140492
ENSG00000206538	VGLL3	-2,2134331	0,21562059	0,0177735	0,57140492
ENSG00000135378	PRRG4	-2,2074185	0,2165214	0,01857679	0,57140492
ENSG00000164938	TP53INP1	-2,2048481	0,21690751	0,0188975	0,57140492
ENSG00000129595	EPB41L4A	-2,2015821	0,21739911	0,01934556	0,57140492
ENSG00000196743	GM2A	-2,2010976	0,21747212	0,01927803	0,57140492
ENSG00000115339	GALNT3	-2,1865904	0,21966998	0,02020171	0,57140492
ENSG00000058085	LAMC2	-2,1863177	0,2197115	0,01928744	0,57140492
ENSG00000067141	NEO1	-2,1790192	0,22082582	0,02063976	0,57140492
ENSG00000137393	RNF144B	-2,1779408	0,22099095	0,01995197	0,57140492
ENSG00000120549	KIAA1217	-2,1765495	0,22120417	0,01916116	0,57140492
ENSG00000182541	LIMK2	-2,1715359	0,22197423	0,01934531	0,57140492
ENSG00000102854	MSLN	-2,1682924	0,22247384	0,0204938	0,57140492
ENSG00000186395	KRT10	-2,1662546	0,22278831	0,02127307	0,57140492
ENSG00000025708	TYMP	-2,162642	0,22334688	0,02157751	0,57140492
ENSG00000143061	IGSF3	-2,1617748	0,22348118	0,02166635	0,57140492
ENSG00000132357	CARD6	-2,1601579	0,22373177	0,02171538	0,57140492
ENSG00000019549	SNAI2	-2,1559704	0,22438211	0,02163053	0,57140492
ENSG00000147813	NAPRT1	-2,1537326	0,22473043	0,02156027	0,57140492
ENSG00000134762	DSC3	-2,1512188	0,22512234	0,02229557	0,57140492
ENSG00000082438	COBL1	-2,1500635	0,2253027	0,02179109	0,57140492
ENSG00000054654	SYNE2	-2,1462029	0,22590641	0,02232942	0,57140492
ENSG00000136848	DAB2IP	-2,1461375	0,22591665	0,02025751	0,57140492
ENSG00000163629	PTPN13	-2,1424266	0,2264985	0,02034183	0,57140492
ENSG00000115109	EPB41L5	-2,140839	0,22674788	0,02296832	0,57140492
ENSG00000071051	NCK2	-2,1406575	0,22677641	0,02297271	0,57140492
ENSG00000158769	F11R	-2,1390271	0,22703283	0,02154798	0,57140492
ENSG000000257315	ZBED6	-2,1382603	0,22715353	0,0210139	0,57140492
ENSG00000169926	KLF13	-2,1302351	0,22842064	0,0208523	0,57140492
ENSG00000113083	LOX	-2,1274656	0,22885955	0,02119151	0,57140492
ENSG00000114796	KLHL24	-2,1268972	0,22894973	0,02205605	0,57140492
ENSG00000104361	NIPAL2	-2,1227336	0,22961143	0,02395196	0,57140492
ENSG00000109458	GAB1	-2,1217635	0,22976588	0,02403511	0,57140492
ENSG00000189060	H1FO	-2,1189467	0,23021493	0,02332118	0,57140492
ENSG00000187193	MT1X	-2,115648	0,23074192	0,02285913	0,57140492
ENSG00000173801	JUP	-2,1144059	0,23094065	0,02452848	0,57140492
ENSG00000151726	ACSL1	-2,1125911	0,23123134	0,02429489	0,57140492
ENSG00000182580	EPHB3	-2,109221	0,23177213	0,02447731	0,57140492
ENSG00000148400	NOTCH1	-2,0965267	0,2338205	0,02366244	0,57140492
ENSG00000120899	PTK2B	-2,0929652	0,23439842	0,02300843	0,57140492
ENSG00000146278	PNRC1	-2,0928498	0,23441718	0,02246347	0,57140492
ENSG00000142178	SIK1	-2,0866357	0,23542906	0,02280332	0,57140492
ENSG00000085276	MECOM	-2,0853422	0,23564024	0,0264329	0,57140492
ENSG00000178184	PARD6G	-2,0841209	0,2358398	0,02628026	0,57140492
ENSG00000178695	KCTD12	-2,0840992	0,23584335	0,02497979	0,57140492
ENSG00000156869	FRS1	-2,0834083	0,23595632	0,02667201	0,57140492
ENSG00000104332	SFRP1	-2,0775598	0,23691479	0,02420454	0,57140492
ENSG00000161249	DMKN	-2,0766551	0,2370634	0,02560208	0,57140492
ENSG00000164904	ALDH7A1	-2,0758695	0,23719253	0,02333318	0,57140492
ENSG00000169855	ROBO1	-2,0748693	0,23735703	0,02493553	0,57140492
ENSG00000166347	CYB5A	-2,0733502	0,23760709	0,02717259	0,57140492
ENSG00000170421	KRT8	-2,0727408	0,23770748	0,02398145	0,57140492
ENSG00000173530	TNFRSF10D	-2,0702992	0,23811011	0,02370585	0,57140492
ENSG00000100003	SEC14L2	-2,065848	0,2388459	0,02362883	0,57140492
ENSG00000160712	IL6R	-2,0638303	0,23918017	0,02422123	0,57140492
ENSG00000152104	PTPN14	-2,0619448	0,23949296	0,0273139	0,57140492
ENSG00000176046	NUPR1	-2,0579691	0,24015387	0,02436449	0,57140492
ENSG00000143162	CREG1	-2,0566234	0,24037797	0,02776522	0,57140492
ENSG00000102554	KLF5	-2,0544444	0,24074131	0,02490771	0,57140492
ENSG00000150593	PDCD4	-2,0528367	0,24100974	0,02530866	0,57140492
ENSG00000096060	FKBP5	-2,051527	0,24122862	0,02884563	0,57140492
ENSG00000162757	C1orf74	-2,0509875	0,24131885	0,02699209	0,57140492
ENSG00000115041	KCNIP3	-2,0483629	0,24175826	0,02643521	0,57140492
ENSG00000265972	TXNIP	-2,0458999	0,24217135	0,02908556	0,57140492
ENSG00000120915	EPHX2	-2,0455501	0,24223008	0,02939693	0,57140492
ENSG00000197632	SERP1NB2	-2,0440941	0,24247467	0,0288551	0,57140492
ENSG00000196754	S100A2	-2,0401551	0,2431376	0,02984464	0,57140492
ENSG00000158246	FAM46B	-2,0355459	0,24391562	0,02519051	0,57140492
ENSG00000196323	ZBTB44	-2,0325035	0,24443055	0,02631464	0,57140492
ENSG00000134824	FADS2	-2,0324643	0,24443719	0,0266689	0,57140492
ENSG00000111679	PTPN6	-2,0318093	0,24454819	0,02889711	0,57140492
ENSG00000112378	PERP	-2,0306151	0,24475069	0,03101489	0,57140492
ENSG00000011198	ABHD5	-2,0293541	0,24496473	0,02529951	0,57140492
ENSG00000164050	PLXNB1	-2,0292994	0,24497402	0,02554009	0,57140492
ENSG00000100100	PIK3IP1	-2,0251206	0,24568461	0,03049944	0,57140492
ENSG00000139679	LPAR6	-2,0153421	0,2473555	0,03153241	0,57140492
ENSG00000104419	NDRG1	-2,0106943	0,24815367	0,0269255	0,57140492
ENSG00000168300	PCMTD1	-2,0091842	0,24841355	0,02733719	0,57140492
ENSG00000116741	RGS2	-2,0052053	0,24909961	0,02662019	0,57140492

3.2.5 Monitoring of the cell proliferation of the DCIS cell lines

To confirm the observation from the transcriptomic study that suggest that proliferation is promoted in the ETCC lines, cell growth was monitored to determine population doubling time (PDT) using trypan blue staining. MCF-10A, MCF-7 and MDA-MB-231 showed different rates of proliferation with PDT of 20 hours, 24.4 hours and 36.2 hours respectively. The rate of proliferation of MCF10DCIS.com was comparable to that of MCF-10A with 20.5 hours PDT. ETCC-006 and ETCC-010 showed intermediates proliferative rates to that of MCF-7 and MDA-MB-231 and slower to that of MCF-10A, with 30.7 hours and 29.7 hours respectively. Interestingly, while most cell lines seemed to undergo a lag phase after plating, the ETCC-010 growth profile displayed high proliferative abilities even at low cell number (Figure 3.6 A).

Next, we used a plating efficiency assay to examine cell survival *in vitro* in the ETCC lines. Plating efficiency measures the cells' ability to go through cell division with no limit in time (261, 262). Following growth and regular refreshment of media, cells were fixed after 10 days and colonies were stained with crystal violet. We saw similar signal intensities for both MCF10DCIS.com and MCF-7 cells (Figure 3.6 B). In comparison, MDA-MB-231 cells and the ETCC lines showed reduced crystal violet staining compared to MCF-7 cells. Interestingly, while MCF-7 and MDA-MB-231 showed distinguishable colonies, all DCIS cell lines appeared more like a monolayer, with an observable reduction in colony formation. This might be explained by an inability to form colonies due to either the importance of cellular contacts or the ability to migrate on the plate.

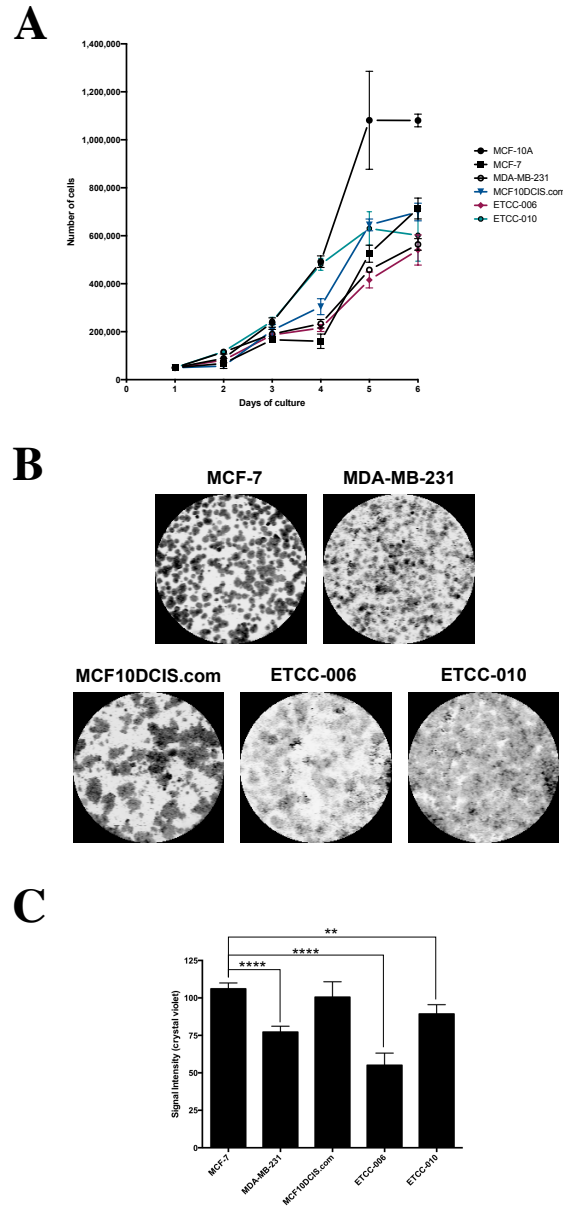


Figure 3.6: Proliferation assessment of the DCIS cell lines. Cell proliferation was monitored by trypan blue staining over six days (A) or plating efficiency assay with crystal violet staining (B) and its quantification using ImageJ (C). Statistical analysis was done using one-way ANOVA (GraphPad v.8.3.0). **** p-value < 0.0001; *** p-value < 0.001; ** p-value < 0.01; * p-value < 0.05.

3.2.6 Assessment of migration and anchorage-independent growth potential of the DCIS cell lines

Since the ‘integrin cell surface interactions’ pathway was enriched in ETCC-006 and ETCC-010, we monitored the migratory abilities of the DCIS cell lines using a wound healing assay. Wound width and wound area were examined after removal of the culture insert at time zero hour (T_0), five hours after insert removal (T_5), and then at 9 and 12 hours post-insert removal. Using bright-field microscopy, we saw wound closure within 12 hours in MCF-10A, MCF-7 and ETCC-006 cells (Figure 3.7 A and 3.7 B). In ETCC-010, after 12 hours the wound was close to closure. MCF10DCIS.com and MDA-MB-231 cells showed slower wound closure.

Anchorage-independent growth, a property of most malignant cells to grow independently of a solid surface, was evaluated by soft agar colony formation assay (263). As expected, MCF-7 and MDA-MB-231 showed the highest ability to form colonies in soft agar. The three DCIS cell lines MCF10DCIS.com, ETCC-006 and ETCC-010 showed similar abilities to form colonies, but with number intermediate to MCF-10A (lowest) and MCF-7 or MDA-MB-231 (highest) (Figure 3.7 C and 3.7 D).

Interestingly, results from the wound healing assay and the anchorage-independent growth assay showed discrepant results. The first assay suggests that wound closure occurred at a higher rate in the ETCC lines compared to the MDA-MB-231 cells whereas with the anchorage independent growth assay the ETCC lines showed lower crystal violet staining intensity than other cells. This might be explained partially by the fact that in the wound healing assay, we do not separate proliferation from migration. To do that we could have removed the serum from the media and measured

wound closure. However, to correctly measure the invasion abilities of those cells, a Transwell system could be used with the addition of an extra layer of extracellular matrix.

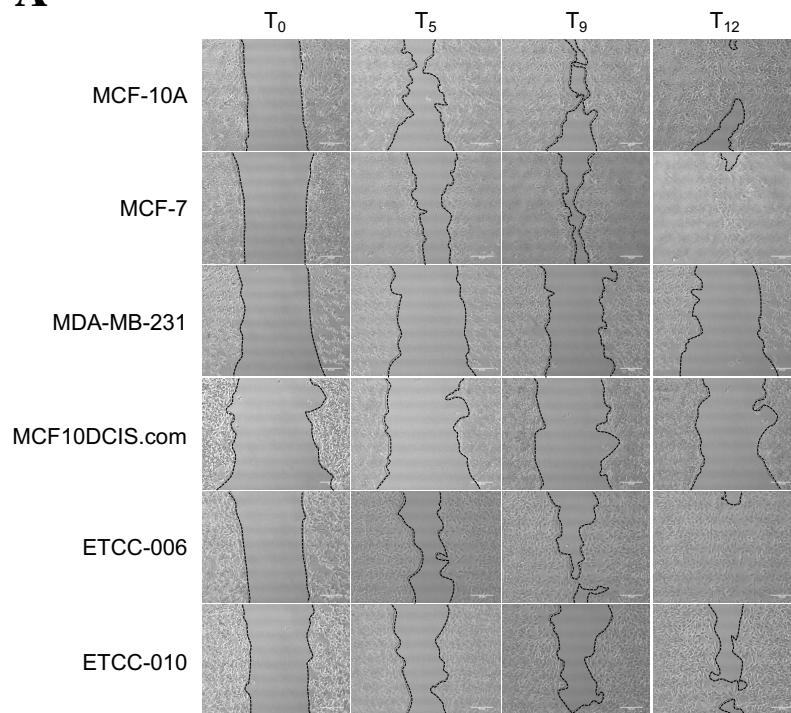
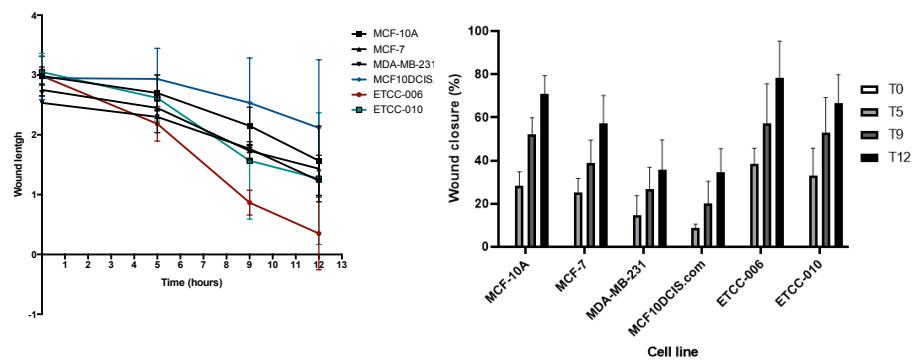
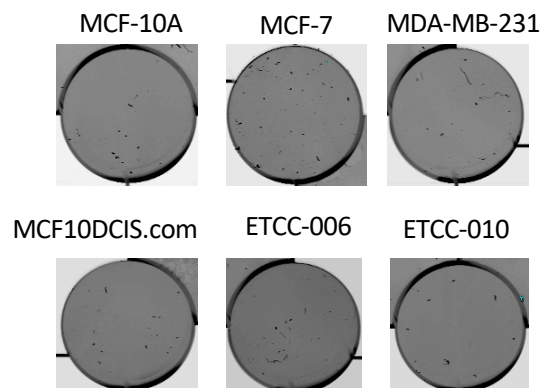
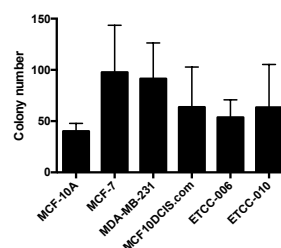
A**B****C****D**

Figure 3.7: Migration and anchorage-independent growth assessment of the DCIS cell lines. (A) Migratory capacity of the DCIS cell lines was examined by wound healing assay using bright field microscopy at 20X. (B) The wound width and wound area were quantified in ImageJ in an arbitrary unit. Anchorage-independent growth was measured by soft agar assay; wells were imaged using the Odyssey Li-Cor system (C) and scored manually (D). The experiments were repeated in triplicate.

3.2.7 ETCC-006 and ETCC-010 have undergone epithelial to mesenchymal transition

Taken together, our data has shown that the DCIS cell lines ETCC-006 and ETCC-010 express migration markers (Figure 3.5) and have migratory abilities (Figure 3.7). We next sought to consider whether those cell lines have undergone epithelial to mesenchymal transition (EMT), especially considering the morphologic similarities between ETCC-010 and MDA-MB-231 cells (Figure 3.1).

EMT is a common event in cancer progression where epithelial cells lose their epithelial phenotype, characterized by tight and adherent cells, to gain a mesenchymal phenotype (Figure 3.8 A). Moreover, during EMT, cells lose expression of epithelial markers such as E-cadherin to gain expression of mesenchymal markers (vimentin, N-cadherin, etc.). Along with cell motility (264), EMT constitutes an early event into the formation of metastasis (265).

To examine the level of expression of key EMT markers, whole cell lysates were prepared. Vimentin, E-cadherin, N-cadherin, β -catenin and β 1-integrin expression was analyzed by Western blotting. Interestingly, several bands were detected when using

the anti-vimentin antibody. We suspect that those bands correspond to different fragments of vimentin. Indeed, vimentin has been shown to be digested by caspase 3 and to produce a similar pattern in sera of patients with idiopathic pulmonary fibrosis and non-specific interstitial pneumonia (266). High expression of vimentin was observed in ETCC-006 and ETCC-010 at similar levels to those observed in MDA-MB-231, while low expression was detected in MCF10DCIS.com to level comparable to those in MCF-10A (Figure 3.8 B and 3.8 C).

Moreover, while MCF10DCIS.com cells express E-cadherin, we could not detect expression of E-cadherin in ETCC-006 and ETCC-010. Instead, expression of N-cadherin was observed in ETCC-006 and ETCC-010. β -catenin is a transcription factor induced by the WNT pathway that stimulates transcription of genes involved in the EMT (267). Its expression is associated with poor prognosis in breast cancer (268). In our analysis, β -catenin protein levels were elevated in MCF-10A, MCF10DCIS.com and ETCC-006 while levels are more balanced in MCF-7, MDA-MB-231 and ETCC-010. β 1-integrin is a transmembrane protein whose function is to bind extracellular matrix (ECM) (269), regulating cells' ability to invade (270). β 1-integrin was detected at high levels in MCF-10A, MDA-MB-231 and ETCC-006 while ETCC-010, MCF10DCIS.com and MCF-7 show lower protein levels.

Overall, the co-expression of β -catenin, vimentin, β 1 integrin and N-cadherin suggests that ETCC-006 and ETCC-010 have undergone EMT. While MCF10DCIS.com expresses some markers of EMT, overall it represents an epithelial phenotype.

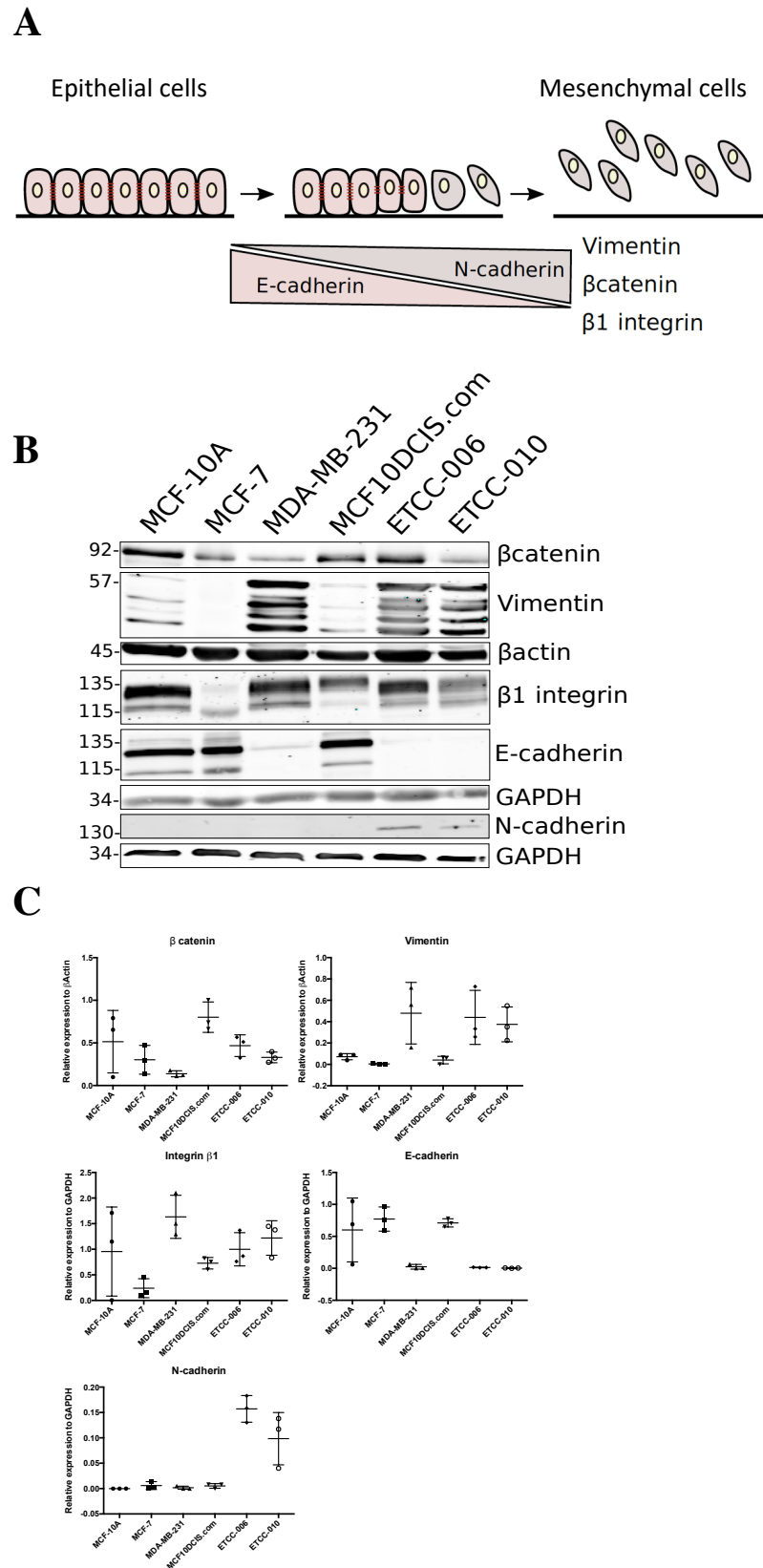


Figure 3.8: Epithelial to Mesenchymal (EMT) transition in the DCIS cell lines.

(A) Schematic of the markers associated with EMT, (B) and (C) protein levels of EMT

markers in the DCIS cell lines analyzed by immunoblot. 42 µg of total protein was loaded per lane onto the gel. Protein level was quantified and normalized against β actin or GAPDH using ImageJ. This experiment was repeated in triplicate, this is a representative immunoblot.

3.2.8 Signaling profile of the DCIS cell lines

Since VEGF signaling seemed enhanced in ETCC-006 cells (Figure 3.5), and ETCC-006 and ETCC-010 have undergone EMT (Figure 3.8), we decided to investigate levels of major signaling proteins in cancer. We measured Epidermal Growth Factor Receptor (EGFR) and Insulin-like Growth Factor 1 Receptor (IGF-1R), both critical in the process leading to invasion in breast cancer (271,272). To do this, cell lysates were prepared using cells grown in normal conditions. Total levels of IGF-1R, EGFR and the active phosphorylated-EGFR were examined by Western blotting (Figure 3.9 A and 3.9 B). Low protein levels of EGFR were detected in MCF-7, with the other cell lines showing similar levels of EGFR. In MCF-10A cells, phosphorylation of EGFR was detected; whereas, p-EGFR was not detected in other cell lines under normal conditions. IGF-1R showed high levels in MCF-7 and low levels in MDA-MB-231 cells. The DCIS cell lines showed similar levels of IGF-1R, with intermediates levels to those observed in MDA-MB-231 and MCF-7 cells.

Common downstream targets of EGFR and IGF-1R include the Ras/Raf/MEK/ERK and Ras/PI3K/PTEN/AKT pathways, known to stimulate proliferation, migration and EMT when activated after being phosphorylated (267). While total levels of ERK were similar across DCIS cell lines (Figure 3.9 C and 3.9 D, normalization to GAPDH), phosphorylated ERK (p-ERK) levels seem slightly higher in MDA-MB-231,

MCF10DCIS.com and ETCC-006. Similarly, total AKT levels were similar across cell lines; however, phosphorylated AKT (p-AKT) showed high disparity (normalization to β actin). Low p-AKT was observed in MCF-7 and MDA-MB-231 cells. Higher levels of p-AKT were observed in MCF-10A, ETCC-006 and ETCC-010, while MCF10DCIS.com showed consistently high levels of p-AKT. We aimed to investigate whether AKT is constitutively activated in MCF10DCIS.com cells. To do this, we compared the phosphorylation levels of AKT under normal and serum starvation conditions. Indeed, p-AKT is lost after serum starvation conditions. We confirmed the constitutional activation of AKT in MCF10DCIS.com cells, indeed p-AKT levels remained unaffected under serum starvation conditions (Figure 3.9 E, 3.9 F and 3.9 G).

Since high levels of cyclin D1 are associated with DCIS (273), we examined the levels of cyclin B1 and cyclin D1, two proliferation markers (Figure 3.9 H and 3.9 I). While the cyclin B1 level seemed slightly lower in the DCIS cell lines compared to the control cell lines, cyclin D1 showed high expression in MCF-7 and ETCC-006, confirming our RNAseq results (Figure 3.4).

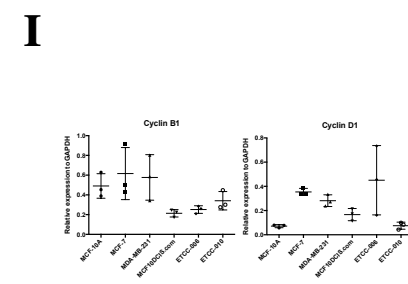
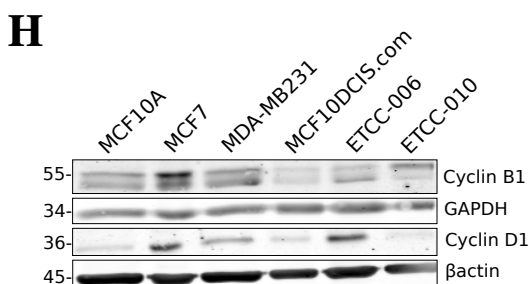
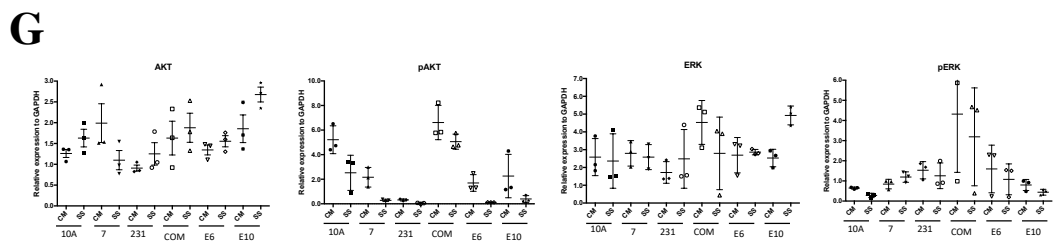
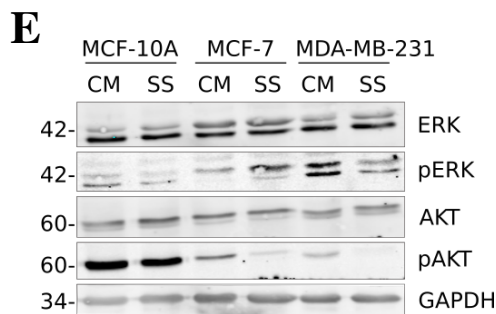
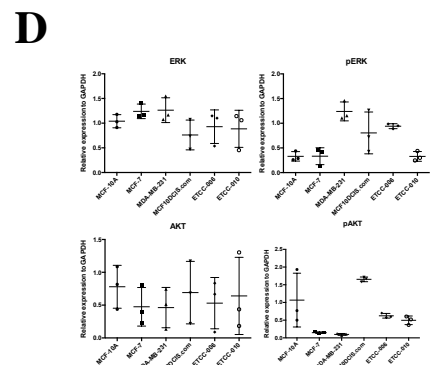
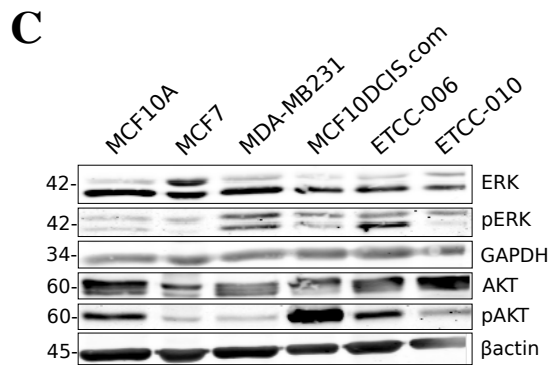
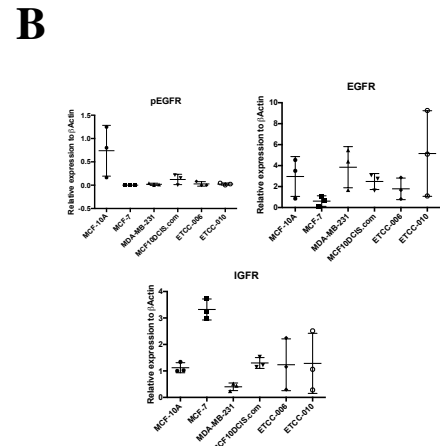
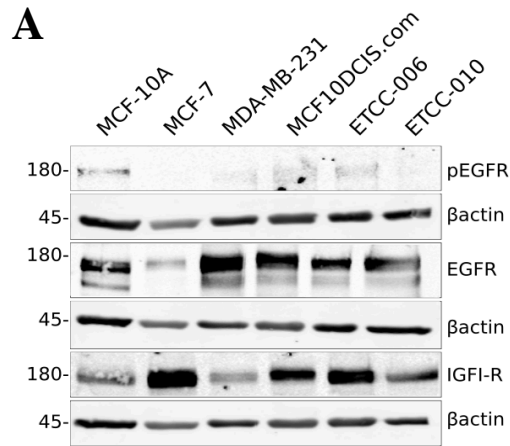


Figure 3.9: Protein levels of signaling proteins in the DCIS cell lines analyzed by Western blotting. (A) and (B) IGF-1R, EGFR and p-EGFR protein level. (C) ERK, p-ERK, AKT and p-AKT protein levels under normal conditions. (D), (E) and (F) Comparison of ERK, p-ERK, AKT and p-AKT protein levels under normal conditions and serum starvation conditions. (H) and (I) Cyclin B1 and D1 protein levels. Protein level was quantified and normalized against β actin or GAPDH using ImageJ. This experiment was repeated in triplicate, this is a representative immunoblot.

3.3 Discussion

Choosing the correct model, animal or cellular, is critical in research. To study DCIS in an *in vitro* cell model, we choose to use three DCIS cell lines: MCF10DCIS.com, a well characterized DCIS cell line model, and ETCC-006 and ETCC-010, two patient-derived cell lines. Since those last two cell lines were largely uncharacterized, and we had previously obtained preliminary results contradicting the published results (245), we conducted a broad phenotypal and molecular characterization of those lines.

Despite some differences highlighted in the original publication, ETCC-006 and ETCC-010 share the same genetic background. In consequence, we expected to observe more similarities between those two lines than with the rest of the cell lines included in the study. Analysis of the hormone receptor status suggested that ETCC-006 and ETCC-010 are similar to triple negative breast cancer cell lines, contrary to what was described in the original publication. At the time of our study, we started with cells on passages 15 for ETCC-006 and 19 for ETCC-010, so it can be argued that expression of the estrogen and progesterone receptor was lost over time. If that is the case, those cells are highly sensitive to transformation, which makes them prone

to variation between hands and labs. ETCC-006 and ETCC-010 cells are expressing markers of EMT (vimentin, N-cadherin and $\beta 1$ integrin), suggesting that they have undergone EMT. Moreover, since they showed similar proliferation abilities than those observed in our cell line panel including MCF-10A cells, no specific pattern in term of proliferation rate could be seen. This is surprising in light of the transcriptomic study comparing the ETCC lines to MCF-10A cells suggesting a positive enrichment of genes associated with proliferation. Migration, examined by wound closure, occurred at a similar rate in the ETCC lines and in MCF-7 and MCF-10A cells.

Regarding the differences between ETCC-006 and ETCC-010, it was shown previously that ETCC-006 is more aggressive than ETCC-010 *in vitro* where it forms twice as many colonies in a soft agar assay. This is not the case *in vivo* where tumor volume induced by injection of cells in mice is reduced with ETCC-006 compared to ETCC-010 cells (245). We show that proliferation rates are similar between the two *in vitro*. However, at the transcriptomic level, ETCC-006 shows more enrichment of proliferative markers (those mostly showing positive effects) than ETCC-010, with Cyclin D1 more highly expressed in our RNAseq data and at protein levels (Figure 3.9 H and I).

We should note here a few limitations of the RNAseq study. First, both ETCC lines were compared to an existing dataset from MCF-10A cells that do not share the same genetic background. Moreover, since that dataset was not generated in parallel to the ones for the ETCC lines, variation was introduced by having different people conducting the experiment, by conducting that experiment at a different time and by using different library preparation kits. However, since those datasets were generated

during a RNAseq course, those limitations could not be avoided. This work still constitutes the first transcriptomic study of the ETCC DCIS cell lines.

MCF10DCIS.com has been shown to be similar to a hormone receptor-negative cell line (274). We could not detect expression of ER at the RNA level. By Western blotting, we saw low levels of ER in MCF10DCIS.com cells, although those levels could be too low to be functionally relevant. We show constitutive activation of AKT that might be explained by a gain of function mutation, H1047R, in the catalytic domain of PI3K (274). This mutation is one of the most oncogenic mutations and occurs frequently in other types of cancer (275). Compared to the ETCC lines, MCF10DCIS.com appears to have conserved an epithelial phenotype with cells showing tight contact, no expression of EMT markers and no migratory abilities. Taken together, all this data suggests that in the DCIS spectrum, the MCF10DCIS.com cell line is less aggressive than the ETCC lines. However, where do those cell lines fit in terms of the clinical representation of DCIS?

As previously described, DCIS is an early stage of breast cancer. However, a large spectrum of cases are diagnosed as DCIS. From low grade disease where cells are mostly normal and grow slowly, to high-grade DCIS where cells look more like cancer cells and grow aggressively. Moreover, high-grade DCIS is associated with a higher risk of developing IBC. MCF10DCIS.com cells are defined as a model of high grade comedo DCIS (243). In comparison to MCF10DCIS.com, our ETCC lines seem to show more aggressive characteristics, especially since they have undergone EMT. Based on this small panel of cell lines alone, this would suggest that the ETCC lines are also models of high-grade DCIS. However, additional DCIS cell lines representing

different grades of DCIS would have to be included in the study to fully conclude on the ETCC lines as models of high-grade DCIS. If those lines are indeed model of high-grade DCIS, they are a great tool to study the mechanisms leading to the transition between DCIS and IDC. In addition to ETCC-006 and ETCC-010, four more cell lines were derived from that same DCIS lesion, ETCC-001, ETCC-007, ETCC-008 and ETCC-011. It might be interesting to include those in a future study and to determine if they would behave in a similar manner as ETCC-006 and ETCC-010. Furthermore, it has to be noted that, while interesting, that study would be focused on different cell lines that share the same genetic background. Thus, for larger scale studies, more cell lines as models of DCIS, would need to be included. However, the number of DCIS cell lines is still limited in research as DCIS study is mostly focused at the histological level. This highlights an ongoing need for more DCIS cell lines to be established and characterized.

Regarding the use of MCF10DCIS.com in DCIS research, as mentioned it has been characterized as a model of human comedo DCIS (243). Therefore, it is an interesting model that should be used in parallel to the ETCC lines. Indeed, if the ETCC lines are in fact more aggressive than MCF10DCIS.com cells, the use of the cell line MCF10DCIS.com would allow the confirmation of results observed in the ETCC lines and could suggest that those findings also apply to earlier stages of DCIS. However, since one cell line is not enough to draw conclusions, this again emphasizes the need for more DCIS cell lines to be developed, at the earliest stages of DCIS.

Chapter 4

Identification of long non-coding RNAs associated with ductal carcinoma *in situ* (DCIS)

4.1 Introduction

The association of long non-coding RNAs with pathological contexts, such as cancer has been established for years with the breakthrough of high throughput sequencing technologies (128). Although the function of lncRNAs has been largely investigated in IDC, as evidenced in the introduction, research specific to DCIS is still at its onset. This does not mean that non-coding RNAs have not been studied in DCIS. Indeed, we have presented some of the research focused on miRNAs; however, investigation of the function of lncRNAs in DCIS is still recent and very much limited.

Non-coding RNAs can be present in exosomes. Indeed, many miRNAs have now been detected in exosomes, and by extension in blood, and are proposed as biomarkers for early detection of cancer (276). LncRNAs are also associated with exosomes. While the number of lncRNAs that have been identified in exosomes remains limited compared to the one of miRNAs, researchers have been proposing to use lncRNAs as an alternative for diagnostic and prognostic purposes. Hence, BC200, an lncRNA associated with DCIS, can be detected in blood sample of patients with IDC (224). To our knowledge, no lncRNA has been detected in DCIS patient blood samples as of yet,

but considering the overall lack of research on lncRNAs, this is not really surprising. Because we anticipate that the number of non-coding RNAs present in exosomes represents a very small portion of those dysregulated in the tumor, we need to expand our knowledge of the lncRNA profile of expression in DCIS compared to normal.

In the introduction, we have presented some studies that offer a list of dysregulated lncRNAs in DCIS (72). Those studies mostly use RNAseq with polyA enrichment to study the lncRNAs transcriptome in DCIS. Using this method, all the lncRNAs that do not possess a polyA tail are eliminated. One study uses the ribosomal RNA removal approach to perform the library preparation (226); however, we know that lncRNAs are expressed at lower levels than mRNAs. Therefore, this could translate into a saturation of dysregulated mRNAs in their results when the interest is in lncRNAs only. To avoid those issues, we chose to perform RNAseq using a capture-based approach that specifically targets lncRNAs (Figure 4.1). Briefly, total RNA is extracted and used to make a double stranded cDNA library. Following that, the library is hybridized to probes corresponding to 17 Mb of long non-coding RNAs based on the human genome hg19 annotation. This represents 32,808 lncRNAs and transcripts with unknown potential. After hybridization, the complex formed by the probe and the cDNA library is pulled out and washed. After PCR amplification, the RNA sequencing libraries are then sent for high throughput sequencing.

As a starting point, we used the lncRNA target-capture on the three DCIS cell lines characterized in Chapter 3. This allowed us to, simultaneously, test this approach on RNA of good quality and to identify a panel of lncRNAs associated with DCIS. Next, we performed the lncRNA capture on RNA extracted from a DCIS lesion and normal

adjacent tissue from a small DCIS patient cohort ($n = 5$). Differentially expressed lncRNAs in DCIS were identified. Then, patient overall survival from TCGA data was analyzed depending on the level of expression of differentially expressed lncRNAs. Based on those results, we focused on six lncRNAs candidates that we characterized bioinformatically by studying their expression level in breast cancer and regulation using the University of California Santa Cruz (UCSC) genome browser. All of those results will be presented in this chapter.

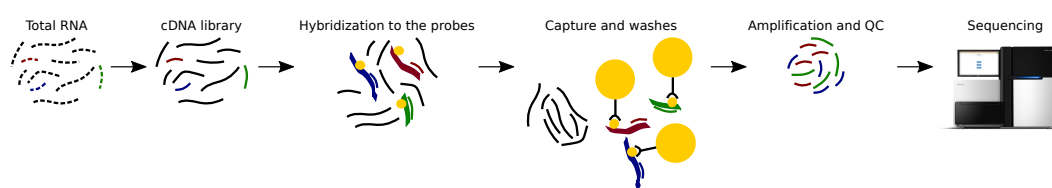


Figure 4.1: Long non-coding RNA hybridization-based capture enrichment approach for lncRNA specific RNA sequencing. Total RNA is extracted from cells and used to make double stranded cDNA libraries. Probes corresponding to lncRNAs are hybridized to the library to capture specifically lncRNAs.

4.2 Results

4.2.1 Quality control pre-lncRNA capture

Before the RNA sequencing library preparation, total RNA was extracted from MCF-10A, MCF10DCIS.com, ETCC-006 and ETCC-010 cells grown in 10 cm dishes in triplicate. We used MCF-10A cells as a control and the expression level of lncRNAs in the DCIS cell lines was compared to the levels present in MCF-10A cells. While MCF-10AT cells could have been used here instead of MCF-10A, we aimed to keep this consistent with the work conducted in the previous Chapter. Moreover, since MCF-10AT cells are derived from MCF-10A, MCF-10A cells constituted an

acceptable control. Next, we confirmed that the quality of the RNA extracted was sufficient to be used for an RNA sequencing experiment by using a bioanalyzer. RNA from three different replicates for each cell line was processed. Figure 4.2 A and 4.2 B shows the bioanalyzer traces. The RNA was of high quality – they were two bands in the gel for each sample corresponding to the ribosomal RNA 28S and 18S, and in 4.2 B there are two distinct peaks corresponding to the 28S and 18S fraction. Moreover, we mentioned previously that the RIN measures the level of degradation of the RNA with an inverse relationship between the number and the level of degradation of the RNA. RIN ranges from 0 to 10. RIN values are shown in Table 4.1.

Table 4.1: RNA Integrity value of RNA samples used for the lncRNA-capture experiment.

Sample name	RNA Integrity (RIN) value
MCF-10A n = 1	9.4
MCF-10A n = 2	8.7
MCF-10A n = 3	7.7
MCF10DCIS.com n = 1	10
MCF10DCIS.com n = 2	10
MCF10DCIS.com n = 3	10
ETCC-006 n = 1	9.4
ETCC-006 n = 2	9.7
ETCC-006 n = 3	9.8
ETCC-010 n = 1	9.8
ETCC-010 n = 2	9.3
ETCC-010 n = 3	10

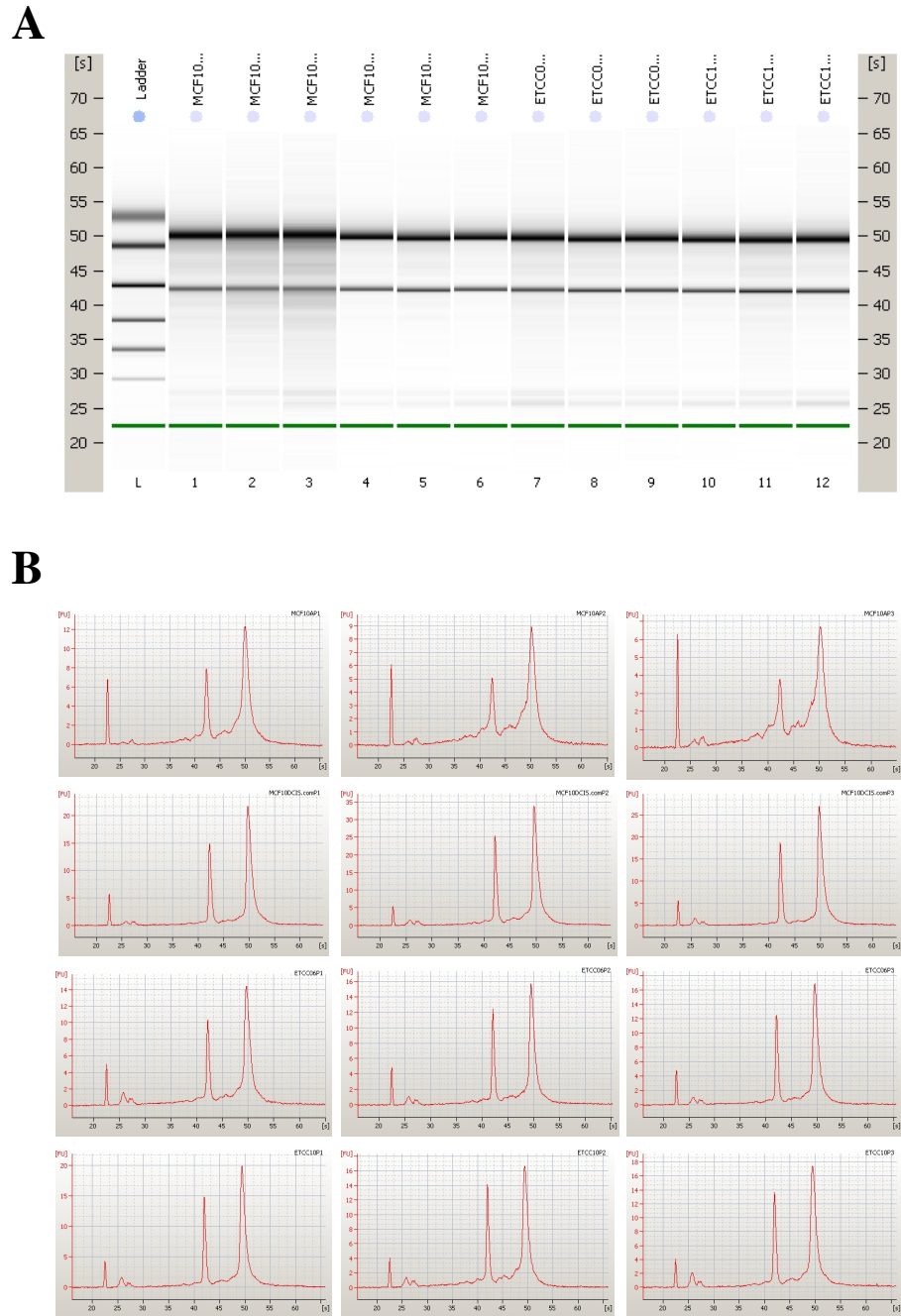


Figure 4.2: Assessment of RNA quality of RNA from MCF-10A, MCF10DCIS.com, ETCC-006 and ETCC-010. (A) Gel image from the bioanalyzer RNA analysis (Agilent). (B) Electropherogram from the bioanalyzer RNA analysis. Those two bands and peaks correspond to the 28S and 18S ribosomal RNA.

Next, the RNA was used to make double stranded cDNA libraries. Success of the library preparation was verified using a bioanalyzer (Figure 4.3 A and 4.3 B). It was

important to check that the library preparation was successful as it was used in the next step to perform the lncRNA capture. The gel shows a light smear between 60 to 100 bp corresponding to the fragmented library for each sample but “ETCC-6 n=3”. Similarly, the electropherogram corresponding to “ETCC-6 n=3” shows no signal. This suggests that the bioanalyzer analysis failed for that sample, most likely due to external environmental disturbance. Regarding the other samples, the electropherogram is similar for all but “MCF10A n=3”. The electropherogram for that sample shows a flat curve and the gel shows a very light smear suggesting that the library preparation in that sample failed. That sample was thus discarded. The rest of the samples were used to perform the lncRNA capture and then sent for high throughput sequencing (Edinburgh Genomics, UK). After the sequencing, the pre-processing and quality control verification of the reads, the alignment of the reads to the reference genome and the differential expression analysis were conducted by Dr Dominic Rose and Dr Xavier Miro from Roche Diagnostics (Mannheim, Germany).

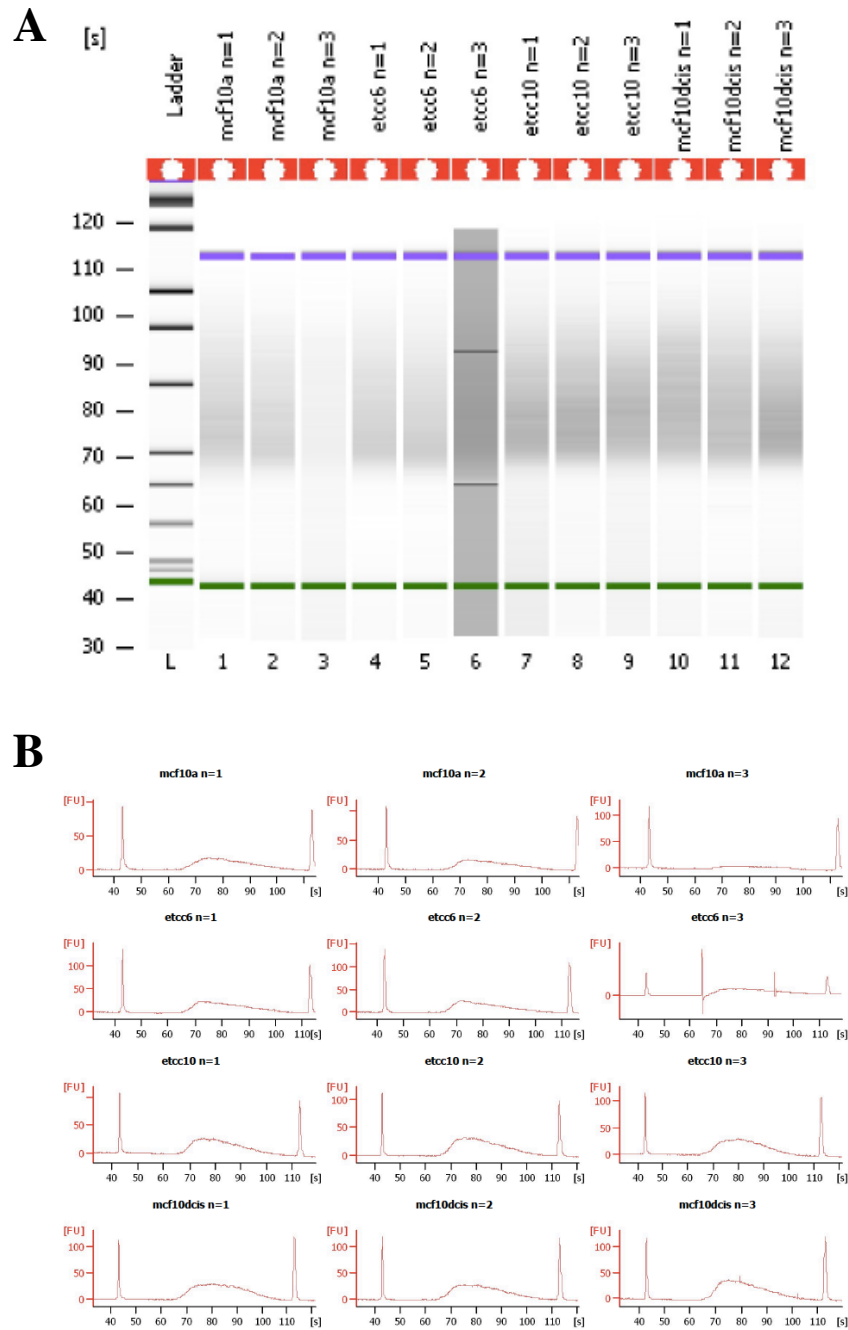


Figure 4.3: Verification of successful library preparation with MCF-10A, MCF10DCIS.com, ETCC-006 and ETCC-010 RNA. (A) Gel image from the bioanalyzer DNA analysis (Agilent). (B) Electropherogram from the bioanalyzer DNA analysis.

4.2.2 Assessment of the efficiency of the lncRNA capture

Table 4.2 shows the performance metrics from the sequencing. Since the probes used to perform the lncRNA capture were designed using the hg19 annotation, that annotation was also used to align the reads after the quality control filtering. For a global view of gene expression, Illumina recommends having about 30 to 60 million reads per sample. For all of our samples, the number of reads was included between 55 million and 89 million reads. Moreover, for human samples, a mean coverage of 30X is recommended, which was achieved for all of our samples. Around 90% of reads were successfully aligned across the samples. The success of the lncRNA capture was assessed by calculating the percentage of bases that aligned to the capture target. This corresponds to 37% to 49% of bases depending on the sample, which corresponds to a 77 to 88-fold enrichment to the target.

Sample	MCF10A_1	MCF10A_2	MCF10M_1	MCF10M_2	MCF10M_3	ETCC0_1	ETCC0_2	ETCC0_3	ETCC6_1	ETCC6_2	ETCC6_3
Reference	hg19	hg19	hg19	hg19	hg19	hg19	hg19	hg19	hg19	hg19	hg19
Primary_target_size [bp]	18 950 280	18 950 280	18 950 280	18 950 280	18 950 280	18 950 280	18 950 280	18 950 280	18 950 280	18 950 280	18 950 280
Capture_target_size [bp]	17 427 025	17 427 025	17 427 025	17 427 025	17 427 025	17 427 025	17 427 025	17 427 025	17 427 025	17 427 025	17 427 025
Total_input_reads	62 279 600	65 866 514	63 593 036	78 631 868	96 019 742	65 203 340	68 192 666	66 288 932	61 102 890	59 565 024	59 565 024
Total_reads_after_filtering	57 728 040	61 428 890	58 034 850	72 853 254	89 142 942	60 552 492	63 552 492	71 736 108	61 655 474	56 962 136	55 360 870
Input_reads_after_filtering [%]	92.69	93.28	91.69	92.66	92.84	92.87	93.20	92.38	93.01	93.22	92.94
Mean_length_read1 [bp]	74.89	74.85	74.89	74.89	74.85	74.89	74.89	74.89	74.89	74.89	74.89
Mean_length_read2 [bp]	74.73	74.75	74.70	74.73	74.74	74.72	74.73	74.72	74.73	74.75	74.72
---TOPIA1---											
Overall_read_mapping_rate [%]	91.1	90.1	89.5	89.9	90.1	91.1	90.6	89.8	90.5	89.9	90.4
Aligned_pairs	25 467 392	26 733 482	24 945 632	31 567 426	38 744 921	26 604 688	27 750 150	30 906 635	26 873 807	24 642 139	24 664 244
Pairs_with_multiple_alignments	3 818 758	3 938 314	3 855 204	4 655 831	5 605 009	2 944 575	3 143 527	3 505 631	3 262 584	2 840 055	3 023 295
Pairs_with_multiple_alignments [%]	15.0	14.7	15.5	14.7	14.5	11.1	11.3	11.3	12.1	11.5	12.6
No_of_junctions	29 961	59 969	28 740	30 761	32 302	32 122	34 909	48 689	54 121	44 800	42 460
Reads_aligned_to_ETCC [%]	6.17	5.62	6.42	7.51	7.39	5.99	5.84	6.47	6.84	6.59	6.29
---Pecand_MarkDuplicates---											
Mapped_non-duplicate_reads	9 797 923	9 786 327	9 258 687	10 766 225	11 723 214	9 715 332	9 924 754	10 094 528	9 840 228	9 029 939	9 179 921
Read_pair_duplicates	15 669 969	16 947 155	15 686 965	20 801 201	27 021 077	16 889 156	17 825 396	20 812 097	17 033 579	15 602 200	14 886 323
Duplicate_rate [%]	62.64	64.52	64.23	67.04	70.72	64.65	65.43	68.54	64.59	64.57	63.18
Optical_duplicates [%]	0.88	0.78	0.78	0.80	0.71	0.82	0.93	0.81	0.89	0.75	0.90
---Pecand_CollectInsertSizeMetrics---											
Mean_insert_size [bp]	278.02	267.15	325.70	301.15	291.50	277.58	278.82	278.13	269.70	247.82	288.48
Insert_size_std_dev [bp]	224.99	215.27	278.03	256.75	240.35	209.88	214.61	210.50	210.93	182.40	226.89
Median_insert_size [bp]	239	230	277	254	248	245	245	246	236	221	253
---bedtools_intersect---											
Reads_on_primary_target_withoutERC	18 254 452	18 308 257	18 490 782	21 473 539	25 223 615	16 917 822	17 443 660	18 406 478	17 187 016	15 731 596	16 478 483
Reads_on_primary_target_withERC	19 192 380	19 701 930	19 471 568	22 684 010	26 443 327	17 824 450	18 660 026	19 388 103	18 168 829	16 638 882	17 398 886
Reads_on_primary_target [%]	75.35	71.71	78.05	71.85	68.25	66.99	66.16	62.73	67.60	67.52	72.29
---GATK_DepthOfCoverage---											
Mean_coverage_primary_target	37.07	35.64	34.33	39.56	43.67	40.46	40.66	40.55	37.68	36.15	36.46
Median_coverage_primary_target	1	1	1	1	1	1	1	1	1	1	1
%bases_above_1x_primary_target	41.6	41.6	42.0	43.1	44.6	43.4	43.4	45.7	44.5	44.5	44.8
%bases_above_10x_primary_target	20.6	20.4	19.5	20.7	22.1	21.5	21.6	21.4	22.0	21.1	21.2
%bases_above_20x_primary_target	15.2	15.0	13.8	15.1	16.2	15.6	15.7	15.4	15.9	15.2	15.3
%bases_above_30x_primary_target	12.2	12.0	10.8	12.5	13.0	12.4	12.5	12.2	12.7	12.1	12.1
%bases_above_50x_primary_target	8.8	8.6	7.6	8.6	9.4	8.9	9.0	8.8	9.2	8.7	8.7
%bases_above_100x_primary_target	5.1	4.9	4.3	5.0	5.6	5.1	5.2	5.0	5.6	5.2	5.2
---Pecand_CollectMetrics---											
Bases_in_primary_target [%]	38.53	36.82	37.93	37.35	37.76	44.63	43.73	42.76	39.96	41.80	42.06
Bases_in_capture_target [%]	45.12	43.40	44.11	43.84	44.37	49.91	49.09	48.00	45.92	48.10	47.65
Bases_near_capture_target [%]	4.63	4.46	4.79	4.53	4.79	4.55	4.47	4.23	4.23	4.08	4.36
Bases_in_padded_target [%]	47.87	47.87	48.90	48.38	49.17	54.47	53.57	52.58	50.16	52.19	52.02
Fold_enrichment	80.16	77.11	78.36	77.89	78.84	88.67	87.21	85.27	81.58	85.46	84.65
---Pecand_CollectNASeqMetrics---											
PF_BASIS [bp]	1 470 298 742	1 469 324 421	1 389 019 564	1 615 589 236	1 758 612 116	1 458 268 473	1 489 726 050	1 515 494 908	1 477 593 497	1 357 635 284	1 378 226 863
PF_ALIGNED_BASIS [bp]	1 470 213 829	1 469 241 067	1 388 934 959	1 615 487 936	1 758 548 009	1 458 174 906	1 489 630 155	1 515 396 017	1 477 499 627	1 357 549 213	1 378 140 939
RIBOSOMAL_BASIS [bp]	15 261 774	16 112 010	16 562 399	18 024 785	18 669 637	14 766 602	15 387 318	17 002 355	15 631 118	15 862 856	15 044 932
CODING_BASIS [bp]	451 991 199	463 161 846	420 646 918	508 083 130	532 692 184	393 364 693	411 685 598	423 968 205	434 280 256	384 738 115	392 587 442
UTR_BASIS [bp]	407 248 382	399 976 509	404 598 553	469 096 836	504 443 900	477 450 844	485 670 071	483 491 016	467 041 060	442 830 960	450 265 441
INTRONIC_BASIS [bp]	129 620 295	140 397 109	119 392 001	133 231 199	153 977 810	112 977 835	116 435 886	120 094 666	140 502 359	124 385 138	123 696 916
INTERGENIC_BASIS [bp]	466 193 534	449 595 049	427 576 683	487 053 748	549 122 354	459 456 309	460 452 630	470 841 388	420 246 148	389 733 522	396 547 525
PCT_RIBOSOMAL_BASIS [%]	1.04	1.10	1.18	1.12	1.06	1.02	1.03	1.12	1.04	1.17	1.09
PCT_CODING_BASIS [%]	30.74	31.52	30.29	31.45	30.29	26.98	27.64	27.98	29.39	28.34	28.49
PCT_UTR_BASIS [%]	27.69	27.22	29.16	29.04	28.69	32.74	32.60	31.91	31.61	32.62	32.67
PCT_INTRONIC_BASIS [%]	8.82	9.56	8.60	8.25	8.73	7.75	7.82	7.92	9.51	9.16	8.98
PCT_INTERGENIC_BASIS [%]	31.71	30.60	30.78	30.15	31.23	31.51	30.91	31.07	28.44	28.71	28.77
PCT_MRNA_BASIS [%]	58.44	58.75	59.44	60.49	58.98	59.72	60.24	59.88	61.00	60.96	61.16
PCT_USABLE_BASIS [%]	58.43	58.74	59.44	60.48	58.97	59.72	60.24	59.88	61.00	60.96	61.15
---ERC_controls---											
Mean_FPKM_targeted_ERCCs	1496.17	1357.23	1644.64	1886.57	1503.70	1533.08	1495.36	1663.68	1666.92	1596.66	1558.44
Mean_FPKM_non_targeted_ERCCs	2.76	2.76	3.29	7.07	6.36	3.98	4.88	4.35	6.19	3.39	4.44
Mean_ERCC_fold_change	453.00	491.00	329.00	266.00	236.00	385.00	306.00	380.00	269.00	470.00	351.00

Table 4.2: Performance metrics of the RNA sequencing after lncRNA capture with MCF-10A, MCF10DCIS.com, ETCC-006 and ETCC-010. The number of reads obtained for each sample before and after filtering, number of reads aligned to the reference genome, number of reads aligned to the target, type of reads are

recapitulated in this table. This table was provided to us by Dr. Xavier Miro and Dr. Dominic Rose (Roche Diagnostics).

Differential expression analysis was performed, comparing ETCC-006, ETCC-010 and MCF10DCIS.com respectively to MCF-10A. Figure 4.4 A shows a hierarchical clustering heatmap based on the differentially expressed genes between ETCC-006 and MCF-10A. From this, all three ETCC-006 replicates clustered together, and the two MCF-10A replicates clustered together. Both cell lines show different overall profiles of gene expression. Figure 4.4 B shows a similar heatmap showing a comparison between ETCC-010 and MCF-10A. Similar to the previous analysis, all ETCC-010 replicates cluster together, and all MCF-10A do as well. Again, this heatmap highlights that ETCC-010 and MCF-10A have different profiles of gene expression. This is confirmed by the volcano plots of significantly differentially expressed genes in Figure 4.4 C and D, showing numerous differentially expressed genes between ETCC-006 and ETCC-010 compared to MCF-10A. When we tried to conduct the same analysis with the RNAseq data obtained with MCF10DCIS.com RNA, there was a technical difficulty, such that the bioinformatic analysis could not be performed.

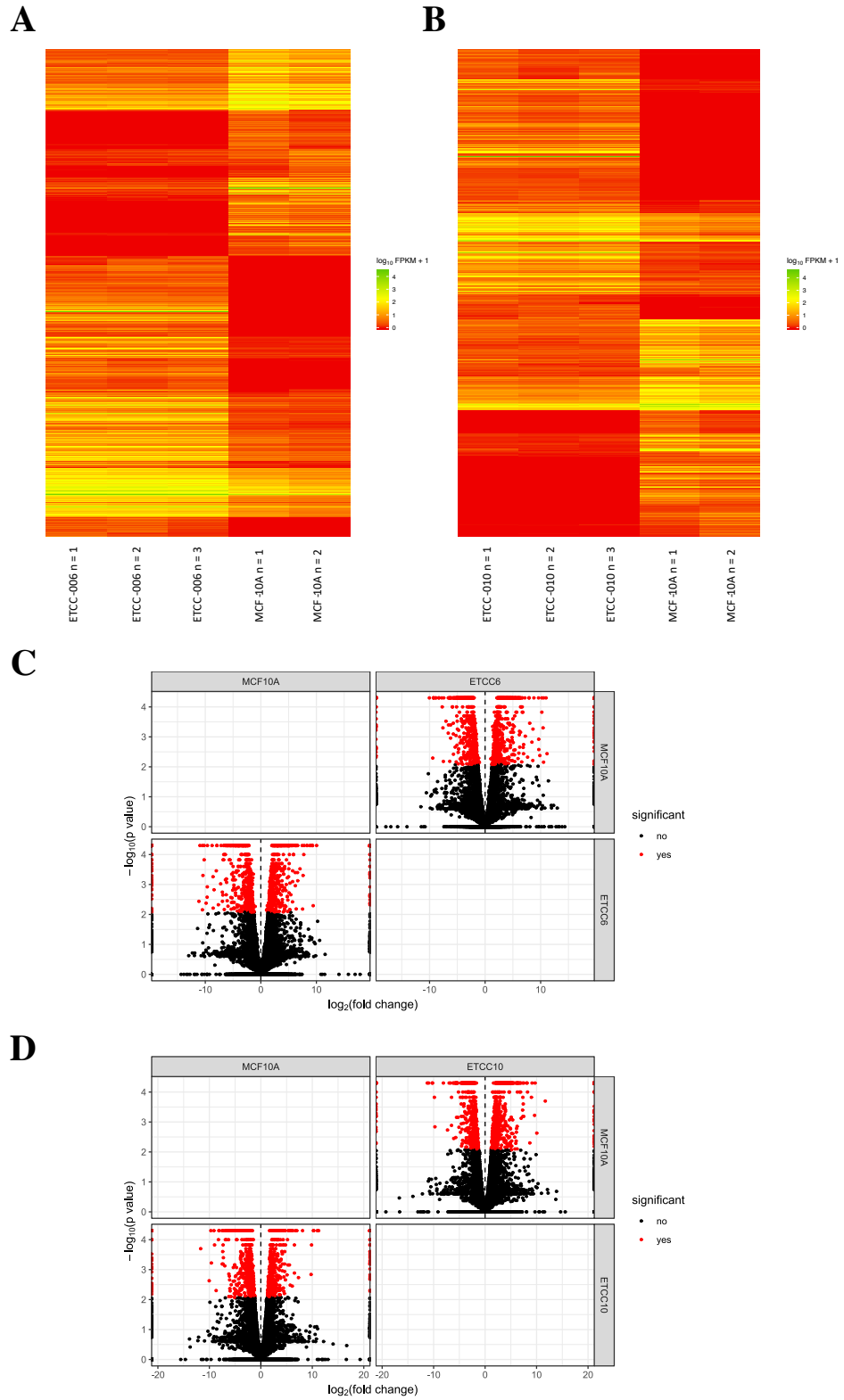


Figure 4.4: Differentially expressed genes in ETCC-006 and ETCC-010 compared to MCF-10A. (A) Heatmap of differentially expressed genes in ETCC-006 compared to MCF-10A. (B) Heatmap of differentially expressed genes in ETCC-010 compared to MCF-10A. (C) Volcano plot of differentially expressed genes in ETCC-006 compared to MCF-10A. (D) Volcano plot of differentially expressed genes in ETCC-010 compared to MCF-10A.

compared to MCF-10A. The different replicates of MCF-10A, ETCC-006 and ETCC-010 cluster together. (C) Volcano plot of differentially expressed genes in ETCC-006 compared to MCF-10A. (D) Volcano plot of differentially expressed genes in ETCC-010 compared to MCF-10A. In red are represented the genes with significant differential expression (p-value adjusted < 0.05).

Next, we aimed to compare the list of differentially expressed genes in each cell line (Figure 4.5 A). Overall, MCF10DCIS.com and ETCC-010 cells have over 1,300 differentially expressed genes, with over 700 genes with higher expression and over 500 with lower expression. ETCC-006 shows over 1,500 differentially expressed genes, half with higher expression and half with lower expression. We grouped the list of differentially expressed genes common to all cell lines, this reduced the number of differentially expressed genes to 112 with 77 with higher expression and 35 with lower expression. That list was sorted by gene type using Biomart in the Ensembl database to keep only the lncRNAs. Following this, we generated a list of 92 differentially expressed lncRNAs, 58 with higher and 34 with lower expression (Table 4.3). By identifying the commonly dysregulated lncRNAs in all DCIS cell lines, we hoped to increase our chances at identifying dysregulated lncRNAs associated with DCIS. Next we wanted to confirm dysregulation of those lncRNAs from RNA extracted from patient samples.

Figure 4.5 B is a Venn diagram of the number of differentially expressed genes common to all three cell lines; in red are the genes with lower expression and in green are the genes with higher expression. As expected from their same genetic background, the overlap between ETCC-006 and ETCC-010 was greater than with

MCF10DCIS.com. Figure 4.5 C shows the proportion of each gene type in the differentially expressed genes, for all cell lines. From that, about 50% of genes are lncRNAs, highlighting again that the lncRNA capture was successful.

Figure 4.5 D represents a Venn diagram of the number of differentially expressed lncRNAs in common to all cell lines. Again, there is more overlap between ETCC-006 and ETCC-010 than with MCF10DCIS.com.

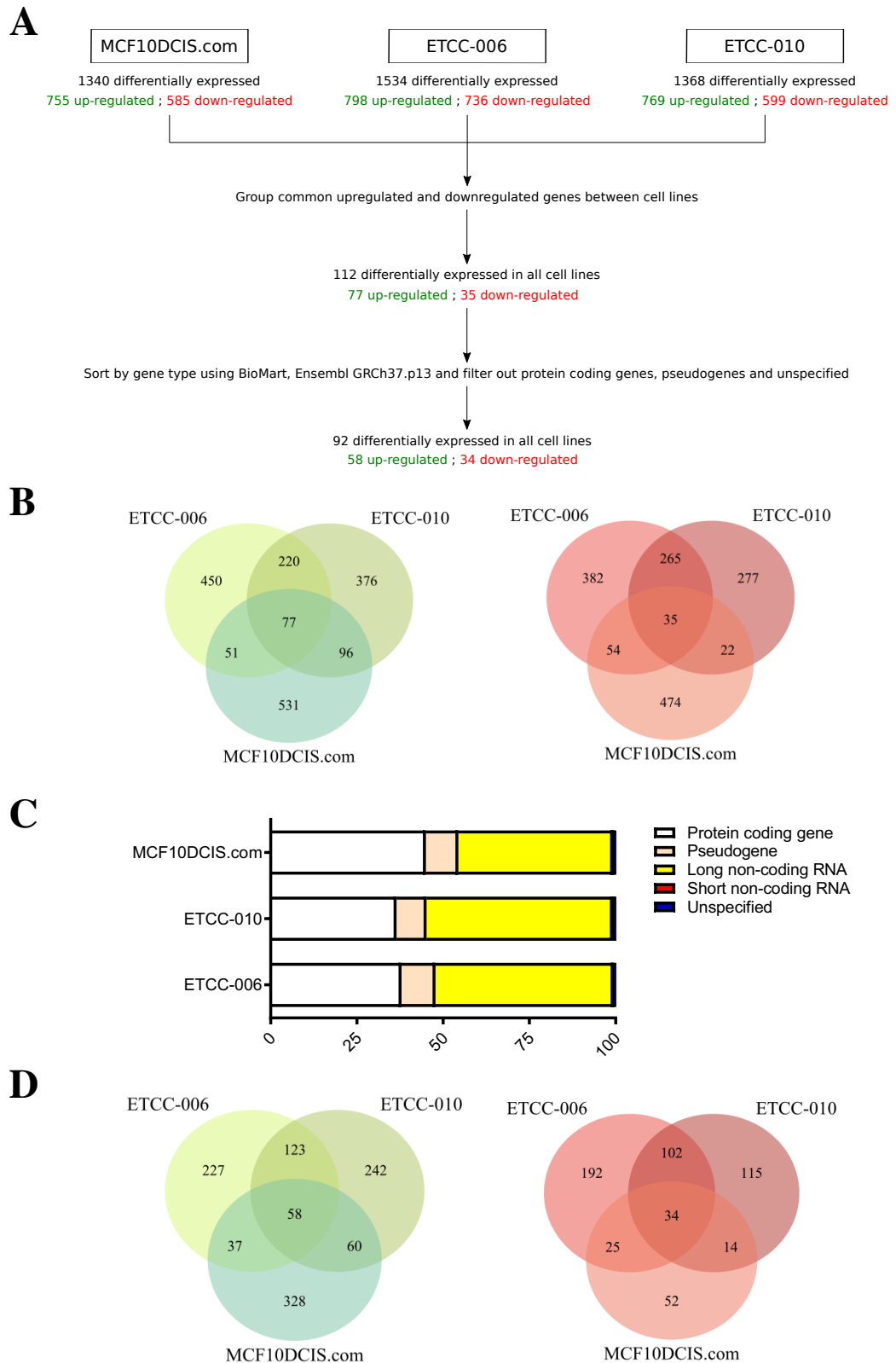


Figure 4.5: Sorting of differentially expressed genes in DCIS cell lines from target-capture with RNAseq. (A) Pipeline for the identification of common differentially expressed lncRNAs in the DCIS cell lines. (B) Venn diagram of the

number of differentially expressed genes common among the DCIS cell lines. In green is represented the genes with higher expression and in red the genes with lower expression. (C) Gene type of the differentially expressed genes in each DCIS cell lines. (D) Venn diagram of the number of differentially expressed lncRNAs common among the DCIS cell lines. In green is represented the lncRNAs with higher expression and in red the lncRNAs with lower expression.

Table 4.3: List of differentially expressed lncRNAs in MCF10DCIS.com, ETCC-006 and ETCC-010 compared to MCF-10A identified by lncRNA-capture RNA sequencing.

Gene name	Gene Ensembl ID	Gene type	Locus	Change in expression in DCIS	Strand	No of transcripts	References
AC008269.2	ENSG00000229321	lincRNA	chr2:207731518-207733013	Down	+	1	
AC074391.1	ENSG00000204929	lincRNA	chr2:65663844-66311773	Up	+	22	
AC093642.3	ENSG00000237940	lincRNA	chr2:242823513-243026289	Up	+	4	
AC104653.1	ENSG00000228857	lincRNA	chr2:114586966-114600933	Up	-	4	
AC105053.4	ENSG00000232504	lincRNA	chr2:86116402-86118103	Up	+	12	
AC107079.1	ENSG00000224844	lincRNA	chr2:238031552-238032740	Down	-	1	
AC144831.1	ENSG00000261888	lincRNA	chr17:81061204-81063860	Up	+	1	
AL137145.1	ENSG00000215244	lincRNA	chr10:6319649-6377938	Up	+	1	
AP000654.4	ENSG00000269895	lincRNA	chr11:86502100-86666433	Up	+	1	
BAIAP2-AS1	ENSG00000226137	lincRNA	chr17:79002932-79008501	Up	-	9	
BCYRN1	ENSG00000236824	lincRNA	chr2:47419543-47572213	Up	+	1	(222-224)
CCAT1	ENSG00000247844	lincRNA	chr8:128220110-128231333	Down	-	2	(277-281)
CTB-32H22.1	ENSG00000253968	lincRNA	chr5:172841037-172878168	Up	+	1	
CTD-2066L21.3	ENSG00000250697	lincRNA	chr5:32925744-33298016	Down	-	3	
CTD-2341M24.1	ENSG00000258733	lincRNA	chr14:86401053-86596117	Up	+	20	
CTD-2369P2.2	ENSG00000175898	lincRNA	chr19:10244020-10341962	Up	-	2	
CTD-3162L10.1	ENSG00000267053	lincRNA	chr19:36795481-36803570	Down	+	6	
GATA6-AS1	ENSG00000266010	lincRNA	chr18:19746858-19748929	Up	-	25	(282,283)
KB-1184D12.1	ENSG00000253585	lincRNA	chr8:95109991-95116550	Up	-	1	
KB-1460A1.1	ENSG00000253395	lincRNA	chr8:102134372-102138386	Up	-	1	
LA16c-380H5.5	ENSG00000272079	lincRNA	chr16:3056120-3057386	Up	+	2	
LINC00163	ENSG00000234880	lincRNA	chr21:46409778-46414001	Up	-	2	(284,285)

LINC00173	ENSG00000196668	processed_transcript	chr12:116971226-116974323	Up	+	9	(286–288)
LINC00313	ENSG00000185186	lincRNA	chr21:44881973-44899414	Down	-	12	(289–291)
LINC00315	ENSG00000184274	lincRNA	chr21:46720159-46725172	Up	-	1	
LINC00319	ENSG00000188660	lincRNA	chr21:44866480-44873773	Down	+	7	(292–295)
LINC00472	ENSG00000232327	lincRNA	chr6:71931159-72130472	Up	-	21	(296–299)
LINC00473	ENSG00000223414	lincRNA	chr6:165740775-166401536	Up	-	5	(300–303)
LINC00665	ENSG00000232677	lincRNA	chr19:36803968-36822620	Down	-	20	(304–306)
LINC00707	ENSG00000238266	lincRNA	chr10:6821559-6884868	Down	+	6	(307–309)
LINC00920	ENSG00000246898	lincRNA	chr16:66442426-66446038	Up	+	6	
LINC00973	ENSG00000240476	lincRNA	chr3:98621267-98737406	Up	+	8	(310,311)
LMCD1-AS1	ENSG00000227110	antisense	chr3:7994491-8653610	Up	-	16	(312–315)
MIR210HG	ENSG00000247095	lincRNA	chr11:565659-568457	Up	-	11	(316–319)
NR2F1-AS1	ENSG00000237187	processed_transcript	chr5:92745064-92930321	Up	-	43	(320,321)
RNU6ATAC35P	ENSG00000221571	lincRNA	chr1:220999117-220999235	Down	+	2	
RP1-140K8.5	ENSG00000260604	lincRNA	chr6:3905143-3912213	Down	-	7	
RP1-142L7.9	ENSG00000270661	lincRNA	chr6:112429962-112672498	Up	-	1	
RP1-170O19.21	ENSG00000270182	lincRNA	chr7:27197962-27198595	Up	+	2	
RP1-261G23.5	ENSG00000236961	lincRNA	chr6:43690522-43705571	Down	-	1	
RP1-272L16.1	ENSG00000224260	lincRNA	chr1:209701799-209741018	Down	-	8	
RP1-27K12.2	ENSG00000231683	lincRNA	chr6:53362138-53481969	Down	-	7	
RP1-37N7.3	ENSG00000220161	lincRNA	chr17:18314472-18317694	Up	+	2	
RP11-11N9.4	ENSG00000247134	lincRNA	chr8:32853695-32902373	Up	+	2	
RP11-120C12.3	ENSG00000234756	lincRNA	chr10:64049279-64063792	Up	+	1	
RP11-125B21.2	ENSG00000236404	antisense	chr9:2422701-2654480	Up	-	19	
RP11-1299A16.3	ENSG00000260641	antisense	chr4:99580054-99585701	Up	+	13	
RP11-145A3.1	ENSG00000227496	antisense	chr1:225888305-225939945	Up	+	5	
RP11-211G23.2	ENSG00000260877	lincRNA	chr11:69186230-69187279	Up	+	2	
RP11-229C3.2	ENSG00000250889	lincRNA	chr5:74343543-74348668	Down	-	2	
RP11-242F11.2	ENSG00000234675	lincRNA	chr6:148338557-148342110	Down	-	1	
RP11-245M24.1	ENSG00000229563	lincRNA	chrX:45364632-45489447	Up	+	10	
RP11-248E9.5	ENSG00000257912	lincRNA	chr12:88189509-88195805	Down	+	1	
RP11-249C24.10	ENSG00000260823	lincRNA	chr16:56643412-56645287	Down	-	1	
RP11-255B23.4	ENSG00000254166	lincRNA	chr8:128197879-128215467	Down	-	84	(322,323)
RP11-326C3.15	ENSG00000270972	lincRNA	chr11:333191-333688	Up	+	1	
RP11-333I13.1	ENSG00000260977	lincRNA	chr2:47004921-47007384	Up	+	1	
RP11-353N14.2	ENSG00000262772	lincRNA	chr17:77797250-77801503	Up	+	1	(324)
RP11-356J5.12	ENSG00000250303	processed_transcript	chr11:112141471-112233257	Up	+	16	
RP11-359M6.1	ENSG00000257474	lincRNA	chr12:79933982-79944315	Down	+	3	
RP11-399D6.2	ENSG00000234840	lincRNA	chr9:22646198-22824212	Up	+	6	
RP11-400K9.4	ENSG00000237807	lincRNA	chr8:54427729-54436491	Up	-	6	
RP11-410A.2	ENSG00000266278	lincRNA	chr18:68047375-68051827	Up	-	1	
RP11-434C1.1	ENSG00000247157	lincRNA	chr12:11694328-11743303	Up	+	3	
RP11-434E6.2	ENSG00000261803	lincRNA	chr16:54399918-54404611	Up	-	2	

RP11-453F18_B.1	ENSG00000213468	lincRNA	chrX:130836677-130964671	Up	-	14	(325–328)
RP11-459O1.2	ENSG00000234426	lincRNA	chr6:89008668-89130752	Down	-	1	
RP11-507B12.1	ENSG00000259675	lincRNA	chr15:61931547-62007370	Down	-	3	
RP11-527H14.3	ENSG00000264301	lincRNA	chr18:14946265-14973769	Up	-	6	
RP11-5407.1	ENSG00000230699	lincRNA	chr1:840213-851356	Up	+	1	
RP11-556E13.1	ENSG00000228651	antisense	chr10:54316461-54515169	Down	-	2	
RP11-575F12.1	ENSG00000249345	lincRNA	chr12:127399765-127544946	Down	-	11	
RP11-63E9.1	ENSG00000259828	lincRNA	chr6:141768147-141772143	Down	+	1	
RP11-66B24.5	ENSG00000259579	lincRNA	chr15:101256453-101368702	Up	+	7	
RP11-71J2.1	ENSG00000226296	lincRNA	chr10:55050831-55070825	Down	-	1	
RP11-7F17.5	ENSG00000246548	lincRNA	chr14:77507406-77552789	Down	+	3	
RP11-809C9.2	ENSG00000258102	lincRNA	chr12:116985910-116992186	Up	+	3	
RP11-834C11.4	ENSG00000250742	lincRNA	chr12:54519881-54526627	Down	+	2	(329)
RP11-95H11.1	ENSG00000260171	lincRNA	chr2:207685913-207687018	Down	-	1	
RP13-539J13.1	ENSG00000272342	lincRNA	chr2:739587-740164	Up	-	1	
RP5-1028K7.2	ENSG00000266088	lincRNA	chr17:38673277-38683254	Down	+	2	(330)
RP5-1063M23.1	ENSG00000250770	lincRNA	chr12:3405367-3409390	Up	-	6	
RP5-1120P11.3	ENSG00000231881	lincRNA	chr6:43963459-44045689	Up	+	1	
RP5-1198O20.4	ENSG00000230615	lincRNA	chr1:44509599-44570647	Up	+	30	
RP5-902P8.10	ENSG00000230415	lincRNA	chr1:1210602-1215800	Up	+	2	
RP5-968J1.1	ENSG00000230839	lincRNA	chr20:1784661-1798252	Up	-	13	
RP6-24A23.7	ENSG00000261409	sense_overlapping	chrX:107963178-107979651	Down	-	1	
SLC2A1-AS1	ENSG00000227533	lincRNA	chr1:43424719-43449029	Up	+	12	(331)
TINCR	ENSG00000223573	lincRNA	chr19:5558177-5568045	Down	-	7	(166,332–334)
TPRG1-AS1	ENSG00000234076	lincRNA	chr3:188659503-189043093	Down	-	1	
TPTEP1	ENSG00000100181	lincRNA	chr22:17082776-17185367	Down	+	29	(335,336)
U73166.2	ENSG00000230454	lincRNA	chr3:50297734-50300790	Up	+	3	

4.2.3 Selection of six lncRNA candidates for further characterization

Overall, we aimed to identify lncRNAs specifically linked to DCIS that have a clinical significance, either for DCIS detection or for the identification of patients more likely to develop IDC. This would imply a functional role for those lncRNAs in DCIS progression to IDC. From targeted-capture and RNAseq analysis, we have compiled a list of 92 lncRNAs commonly dysregulated in all DCIS cell lines compared to MCF-10A. 225 are commonly dysregulated in the ETCC lines, 62 commonly dysregulated in ETCC-006 and MCF10DCIS.com and 74 commonly dysregulated in ETCC-010

and MCF10DCIS.com. Because change in expression does not mean that those lncRNAs are functionally relevant, we sought to find a way to discriminate the lncRNAs most likely to be important for the transition toward IDC. To do that, we used the list of differentially expressed lncRNAs common to all three DCIS cell lines, or common to at least two cell lines. Using that list, we compared patient overall survival in an IBC cohort from The Cancer Genome Atlas (TCGA) with patients divided into two groups, a low expression group versus high expression group for a given lncRNA. By using overall survival in IBC, we aimed to focus our research efforts on lncRNAs possibly important in IBC, with dysregulated expression in our DCIS cell lines. This would hopefully increase our chances at identifying lncRNAs important in both DCIS and IDC and that could be studied in our DCIS cell lines. We identified five lncRNAs whose altered expression correlates with different patient overall survival. Low expression of LINC00473, LINC00313, LOC100134040 and LINC00707 is associated with better overall survival (Figure 4.6 A). This could mean that those lncRNAs are important for breast cancer progression. High expression of LOC729970 is associated with better overall survival, this could mean that this lncRNA has a protective role against breast cancer progression. We also identified an interesting lncRNA MIR210HG. Low expression of MIR210HG is associated with better metastasis-free survival, which suggests that it could be involved in the regulation of cell motility and invasiveness. If an lncRNA is important in cancer progression, we could expect its expression to be higher in a breast cancer sample compared to normal breast. On the opposite, if an lncRNA has a protective role against tumor progression, we could expect its expression to be lower in a breast cancer sample compared to normal breast. For that reason, we analyzed the expression levels of those six lncRNAs in the TCGA IDC cohort (Figure 4.6 B). No difference in expression

between the tumor and normal group was observed for LINC00473, LINC00313, LINC00707 and MIR210HG. LOC729970 showed lower expression in tumor compared to normal, adding support to the hypothesis that LOC729970 correlates with a lower probability of breast cancer progression. In comparison, in our DCIS cell lines, we showed higher expression of LOC729970 in the ETCC lines compared to MCF-10A.

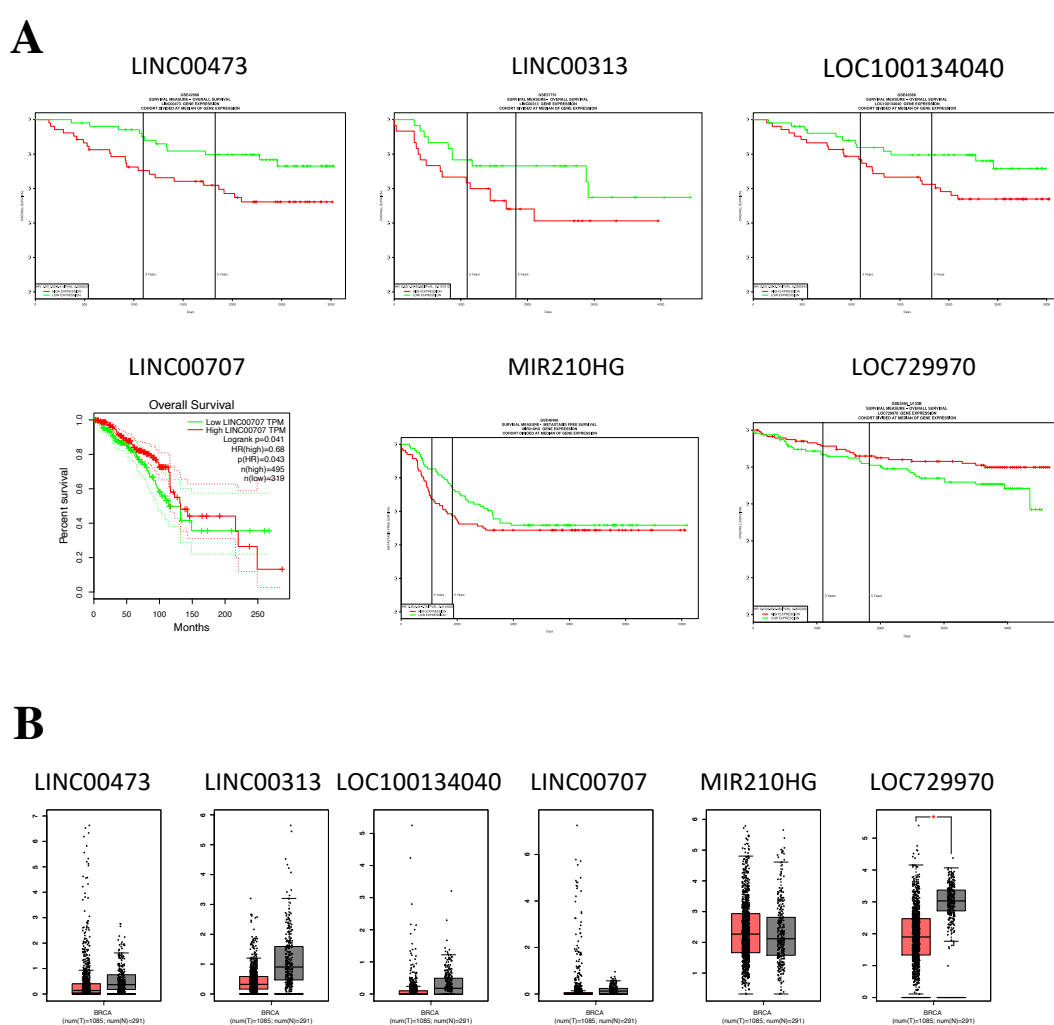


Figure 4.6: Selection of six lncRNA candidates for further analysis. (A) Kaplan–Meier survival plot of six lncRNAs identified in the RNA sequencing experiment whose expression associate with different overall survival in an invasive breast cancer

cohort from The Cancer Genome Atlas (TCGA). (B) Expression level of each lncRNA in an invasive breast cancer cohort from TCGA in tumor and normal tissue. Those plots were generated in the GEPIA2 database.

4.2.4 Validation of the differential expression of the six lncRNA candidates

Following our analysis of lncRNA candidates in TCGA, we aimed to confirm the differential expression identified through the RNAseq experiment and presented in Figure 4.7, panel A by performing qRT-PCR (Figure 4.7 B). In the RNAseq data, LINC00473 and MIR210HG were found to have higher expression in all cell lines. LINC00707 and LINC00313 were found to have lower expression in all cell lines. LOC100134040 expression was found to be higher in MCF10DCIS.com and lower in ETCC-006 and ETCC-010, while LOC729970 expression was found to be higher in ETCC-006 and ETCC-010. We experimentally confirmed those observations from RNAseq data by performing qRT-PCR (Figure 4.7 B). We included MCF-7 and MDA-MB-231 as examples of a hormone receptor positive and triple negative breast cancer cell line. LINC00473 showed high expression in MCF10DCIS.com and ETCC-006. No difference was observed in the other cell lines. LINC00313 low expression was not different in all cell lines compared to MCF-10A as shown in the RNAseq experiments. However, it was shown to be lower in MCF10DCIS.com, ETCC-006 and MCF-7 compared to MDA-MB-231 by qRT-PCR analysis. LOC100134040 expression was higher in MCF10DCIS.com and was not different in ETCC-006 and ETCC-010 compared to MCF-10A, confirming the RNAseq results. Its expression was also higher in MDA-MB-231 cells. LINC00707 low expression was confirmed in all DCIS cell lines and was also shown to be low in MCF-7 and MDA-MB-231.

MIR210HG expression was higher in the three DCIS cell lines compared to MCF-10A. We showed an even higher MIR210HG expression in MCF-7 and MDA-MB-231 than the one observed in the DCIS cell lines. The LOC729970 pattern of expression was confirmed in ETCC-006, while its expression was low in MCF-7 but high in MDA-MB-231. Overall, we confirmed the RNAseq results for LOC100134040, LINC00707, MIR210HG and LOC729970.

A

Gene	Sample	Fold Change (RNAseq)	p_adj	Fold Change (qRT-PCR)
LINC00473	MCFCOM	9,46	8,30E-04	13,04
	ETCC-006	4,89	0.0041	58,95
	ETCC-010	3,51	0.031	2,48
LINC00313	MCFCOM	0,11	8,30E-04	0,52
	ETCC-006	0,030	6,40E-04	0,46
	ETCC-010	0,11	7,30E-04	1,601
LOC100134040	MCFCOM	4,05	0.0075	47,89
	ETCC-006	0,14	0.0077	0,66
	ETCC-010	0,045	0.015	0,87
LINC00707	MCFCOM	0,027	8,30E-04	0,044
	ETCC-006	0,029	6,40E-04	0,105
	ETCC-010	0,39	0.033	1,17
MIR210HG	MCFCOM	8,61	8,30E-04	11,8
	ETCC-006	4,52	0.0092	6,58
	ETCC-010	4,94	0.0049	9,72
LOC729970	MCFCOM	Not differentially expressed		2,95
	ETCC-006	16,32	6,40E-04	29,16
	ETCC-010	18,74	7,30E-04	18,46

B

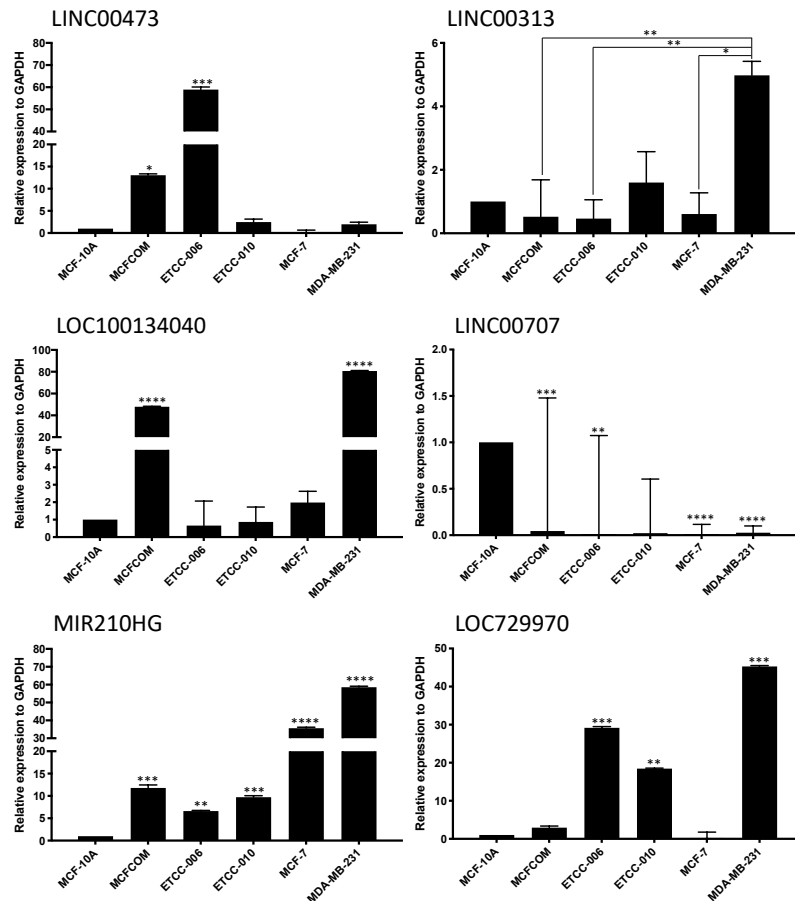


Figure 4.7: Expression of the six lncRNA candidates in breast cancer cell lines.

(A) Table listing the differential expression of the six lncRNA candidates in the

RNAseq experiment and qRT-PCR. (B) Relative expression of the six lncRNA candidates in breast cancer cell lines by qRT-PCR normalized to GAPDH. Statistical analysis was done using one-way ANOVA (GraphPad v.8.3.0). **** p-value < 0.0001; *** p-value < 0.001; ** p-value < 0.01; * p-value < 0.05.

4.2.5 Previous cancer association of the six lncRNA candidates

To gain an insight into if any of the lncRNAs candidates had been previously associated with cancer, we first investigated the lncRNAs in the literature, with a specific attention to any literature related to breast cancer.

LINC00473 is the most studied lncRNA in our candidate list. It shows higher expression in Wilms tumor and is associated with unfavorable histology of the tumor (302). LINC00473 knockdown with siRNAs inhibits cell viability through induction of Bcl-2 dependent apoptosis and G1/S arrest via CDK2 and cyclin D1 (302). It has also been linked with the cAMP signaling pathway, both in liver cancer and non-small cell lung cancer (35, 36). In the latest work, LINC00473 was shown to interact with NONO, a protein of the cAMP signaling pathway, that induces CREB-mediated transcription. In addition to those studies, LINC00473 was also linked with salivary gland cancer (338), osteosarcoma (339) and cervical cancer (303). LINC00473 is now linked with breast cancer through mediation of cyclin D1 (340), and it also acts as a sponge for miR-497 (341) and miR-198 (342). However, there are no publications demonstrating an association with DCIS.

MIR210HG has been shown to be induced under hypoxia conditions (343). It is linked with gliomas, where it can be detected in serum from patients (344). It is upregulated

in osteosarcoma where it promotes cell proliferation, invasion and EMT through the sponging of miR-503 (345). In breast cancer, it has recently been shown to promote cell invasion and metastasis *in vitro* and *in vivo* by inducing the upregulation of MUC1-C when it sponges miR-1226-3p (346). MIR210HG association with DCIS remains uncertain.

Few studies have focused on LINC00313, LOC100134040, LINC00707 and LOC729970 in cancer research. LINC00313 is associated with lung cancer (347) but it has not been associated with breast cancer. LINC00707 is linked with melanoma (348). It has also been shown to be upregulated in lung adenocarcinoma where it promotes cell proliferation through CDC42 regulation (308). It regulates metadherin in breast cancer by acting as a sponge for miR-876 (309). LOC100134040 and LOC729970 are fully uncharacterized in cancer. LOC729970 has been proposed to have a role in drug resistance in pancreatic cancer (349) but no study has linked it to breast cancer yet.

4.2.6 Bioinformatic characterization of the six lncRNA candidates

Since little information is available for most of our lncRNA candidates, especially in the context of breast cancer, we aimed to characterize them bioinformatically to gather information about these potential clinical targets. Using the TANRIC database (https://ibl.mdanderson.org/tanric/_design/basic/main.html), we analyzed the level of expression of each lncRNA candidate in an IDC cohort divided into clinically relevant groups. In Figure 4.8, we show LINC00707 level of expression in different breast cancer subtypes based on the PAM50 classification (luminal A, luminal B, basal and

HER2) (Figure 4.8 A) , therapy-strategy classification (HR positive, TNBC and mix typed) (Figure 4.8 B) or simply ER (Figure 4.8 C) and PR status (Figure 4.8 D). LINC00707 shows a lower expression in luminal breast cancer types and a higher expression in basal type. Supporting this, LINC00707 shows higher expression in TNBC than the levels observed in any other types and higher expression in ER and PR negative patients. This supports the hypothesis that LINC00707 could be important for breast cancer progression toward later stages.

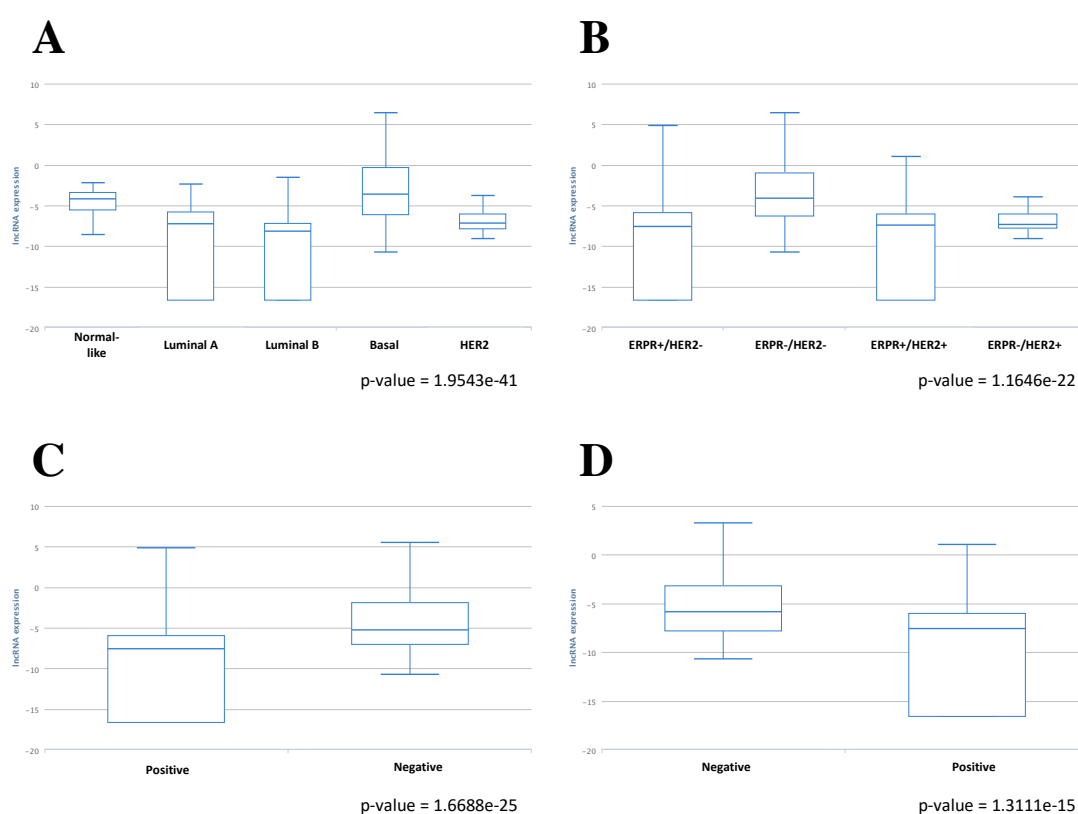


Figure 4.8: Differential expression analysis of LINC00707 based on subtype in an invasive breast cancer cohort from TCGA generated in the TANRIC database.

(A) Expression of LINC00707 in each group of PAM50 classification. (B) Expression of LINC00707 based on the therapy strategy classification. (C) Expression of LINC00707 based on the estrogen receptor status classification. (D) Expression of LINC00707 based on the progesterone receptor status classification.

Next, we aimed to study LINC00707 regulation. To do so, we focused on the promoter region to identify any CpG methylation region, specific chromatin marks and binding of transcription factors, using data available on the UCSC Genome Browser (350). Unfortunately, since no breast cancer cell line dataset was available to study chromatin marks or transcription factor binding to the promoter region, we chose to study those parameters in HeLa cells that show expression of LINC00707 according to RNAseq data from the UCSC genome browser (data not shown). No CpG methylation was observed in HeLa cells around the predicted promoter region suggesting that the promoter is active. Indeed, the chromatin marks associated with active transcription H3K9ac, H3K4me3, H3K27ac and H3K36me3 are present on the promoter of LINC00707 (Figure 4.9 A). Several transcription factors have been identified to bind the promoter region in HeLa cells (Figure 4.9 B) including: BRCA1, MAFK (MAF bZIP Transcription Factor K), CEBPB (CCAAT Enhancer Binding Protein Beta), FOS (Fos Proto-Oncogene), EP300 (E1A Binding Protein P300), NFYA (Nuclear Transcription Factor Y Subunit Alpha), NFYB (Nuclear Transcription Factor Y Subunit Beta), RCOR1 (REST Corepressor 1), TFAP2C (Transcription Factor AP-2 Gamma), JUND (JunD Proto-Oncogene), JUN (Jun Proto-Oncogene), STAT3 (Signal Transducer And Activator Of Transcription 3), RFX5 (Regulatory Factor X5) and TFAP2A (Transcription Factor AP-2 Alpha). Interestingly, most of those transcription factors have been linked with breast cancer, in particular BRCA1 that remains one of the most studied breast cancer susceptibility gene (351). POL2RA binding to the promoter region suggests that LINC00707 is an RNA polymerase II transcript.

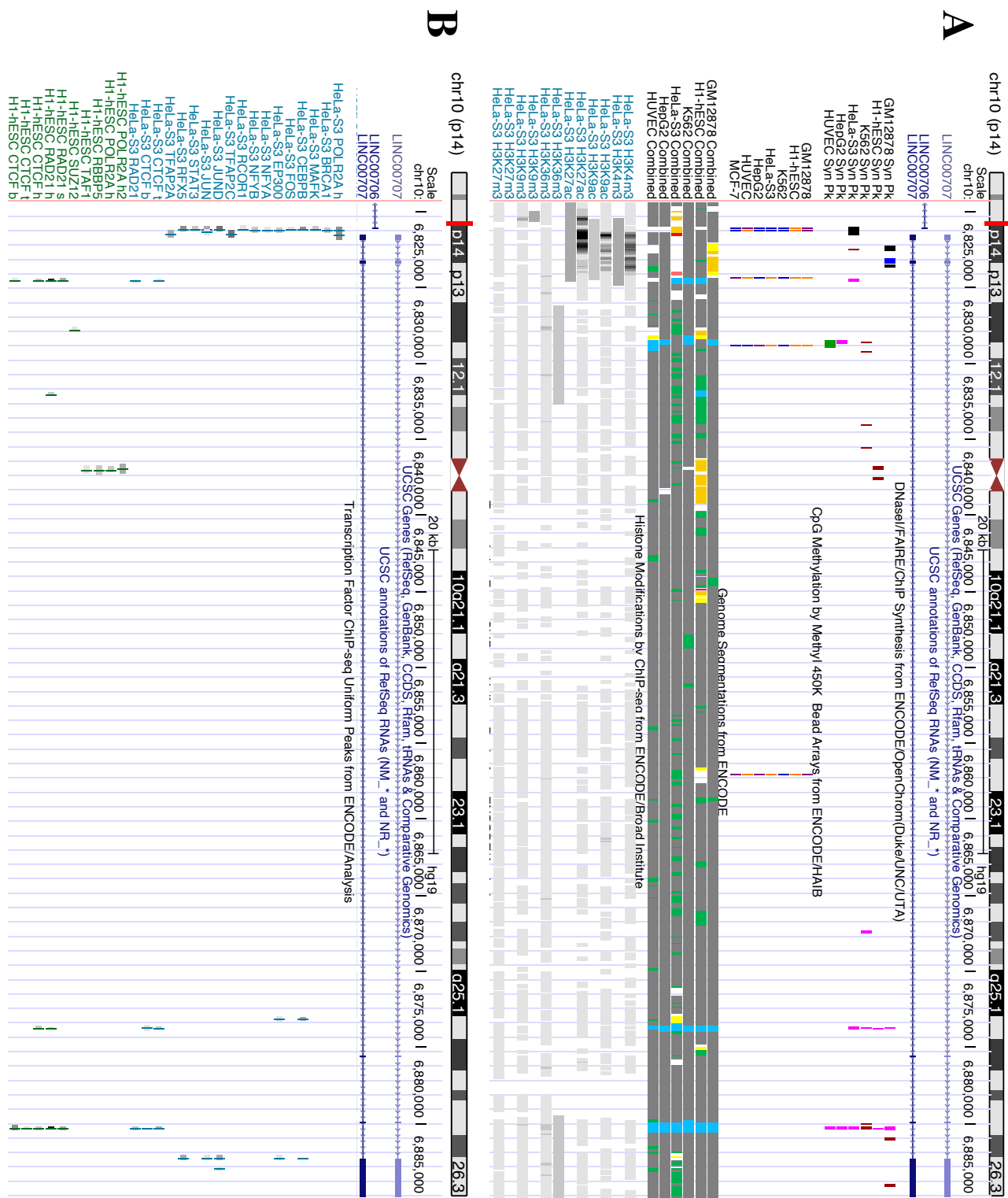


Figure 4.9: Epigenetic landscape of the LINC00707 locus in HeLa cells using the UCSC genome browser annotation hg19. (A) Chromosome compaction, CpG island methylation (orange: methylated; purple: partially methylated; blue: unmethylated; black: NA), Genome segmentation (bright red: predicted promoter region including

TSS; light red: predicted promoter flanking region; orange: predicted enhancer; yellow: predicted weak enhancer or open chromatin cis regulatory element; blue: CTCF enriched element; dark green: predicted transcribed region; gray: predicted repressed or low activity region) and histone modifications present on the locus. (B) Transcription factors binding sites presents on the promoter region of the locus determined using CHIP-seq data.

Similar to LINC00707, LINC00473 shows lower levels in luminal type BC compared to the levels observed in basal types (Figure 4.10 A). This is supported by the therapy strategy analysis where we show higher expression in TNBC (Figure 4.10 B) with higher expression in ER and PR negative patients (Figure 4.10 C and 4.10 D). On the regulatory perspective, we also used HeLa cells as the context to study the regulation of LINC00473 since it is expressed in those cells according to RNAseq data from the UCSC genome browser (data not shown). The CpG island around the promoter region of that lncRNA showed no methylation across the datasets available on the UCSC Genome Browser, suggesting that LINC00473 might be broadly expressed in contexts not limited to breast cancer. Regarding the histone modifications, we identified high levels of H3K9ac, confirming the presence of an active promoter and H3K4me3, H3K27ac and H3K36me3, suggesting active transcription (Figure 4.11 A).

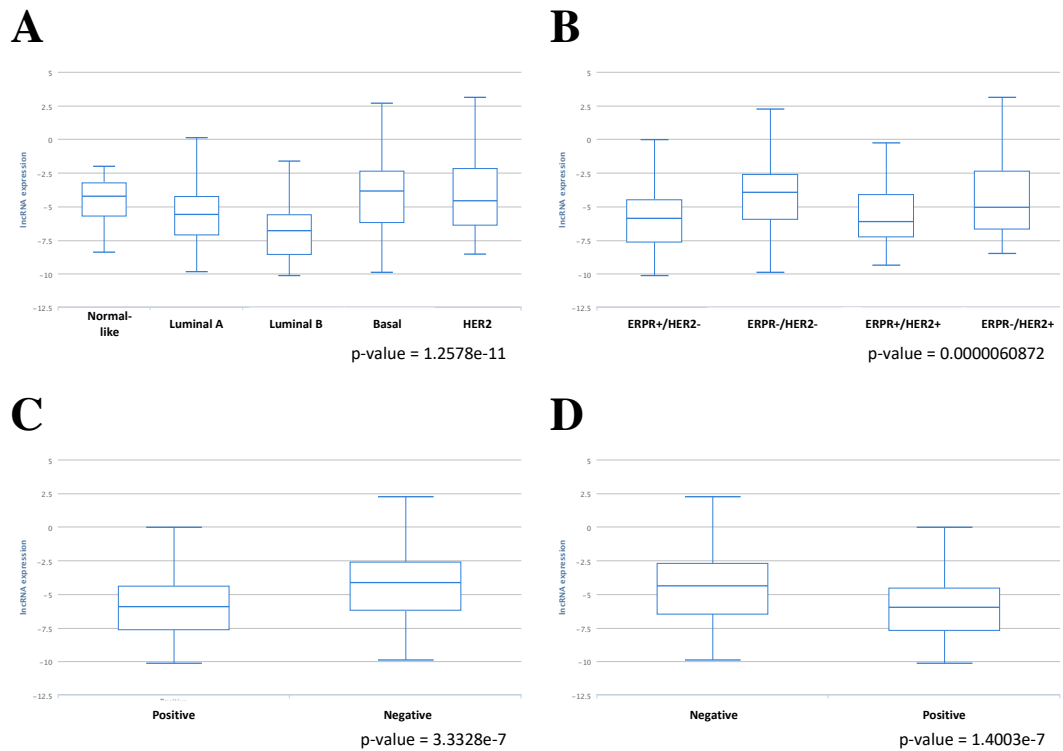


Figure 4.10: Differential expression analysis of LINC00473 based on subtype in an invasive breast cancer cohort from TCGA generated in the TANRIC database. (A) Expression of LONC00473 in each group of PAM50 classification. (B) Expression of LINC00473 based on the therapy strategy classification. (C) Expression of LINC00473 based on the estrogen receptor status classification. (D) Expression of LINC00473 based on the progesterone receptor status classification.

Binding of POLR2A throughout LINC00473 gene suggests that it is an RNA polymerase II transcript (Figure 4.11 B). This is also supported by GTF2F1 binding to the promoter region, an initiation factor for RNA polymerase II transcription (352).

LINC00473 is regulated by several transcription factors, most of them are associated with breast cancer: BRCA1, TFAP2A, TFAP2C, MAX (MYC Associated Factor X), MYC (MYC Proto-Oncogene), MXI (MAX Interactor 1), MAFK, CEBPB, EP300,

RCOR1, JUND, CHD2 (Chromodomain Helicase DNA Binding Protein 2), JUN, RAD21 (RAD21 Cohesin Complex Component), ELK4 (ETS Transcription Factor ELK4), STAT1 (Signal Transducer And Activator Of Transcription 1), GABPA (GA Binding Protein Transcription Factor Subunit Alpha), SMC3 (Structural Maintenance Of Chromosomes 3), TAF1 (TATA-Box Binding Protein Associated Factor 1), TBP (TATA-Box Binding Protein), USF2 (Upstream Transcription Factor 2), TCF7L2 (Transcription Factor 7 Like 2), SMARCA4 (SWI/SNF Related, Matrix Associated, Actin Dependent Regulator Of Chromatin, Subfamily A, Member 4).

CTCF enriched element; dark green: predicted transcribed region; gray: predicted repressed or low activity region) and histone modifications present on the locus. (B) Transcription factors binding sites presents on the promoter region of the locus determined using CHIP-seq data.

LINC00313 expression is lower in HER2-positive type BC; however, no statistically significant difference was seen between the ER/PR positive and the ER/PR negative group (Figure 4.12). Similar to LINC00473, LINC00313 promoter region is unmethylated in most cell lines studied suggesting that it might be expressed in several cell types including in MCF-7, the broadly studied BC cell line. A promoter region was identified around the first exon in red, just after an enhancer region in yellow (Figure 4.13 A). In HeLa cell line, we identified some H3K4 trimethylation and high levels of H3K9ac and H3K27ac, which correlate with an active promoter and active transcription. LINC00313 is suggested to be an RNA polymerase II transcript as evidenced by POLR2A binding to the promoter region. Many transcription factors found to bind to LINC00707 and LINC00473 promoter also show binding to LINC00313 promoter (Figure 4.13 B). Hence, binding of E2F1 (E2F Transcription Factor 1), E2F6 (E2F Transcription Factor 6), TFAP2A, TFAP2C, REST (RE1 Silencing Transcription Factor), MAX, MYC, CEBPB, STAT1, STAT3, EP300, RCOR1, PRDM1 (PR/SET Domain 1), RAD21 (RAD21 Cohesin Complex Component), MAZ (MYC Associated Zinc Finger Protein), JUND, CHD2, SMC3, TAF1, TBP, USF2, TCF7L2 has been reported.

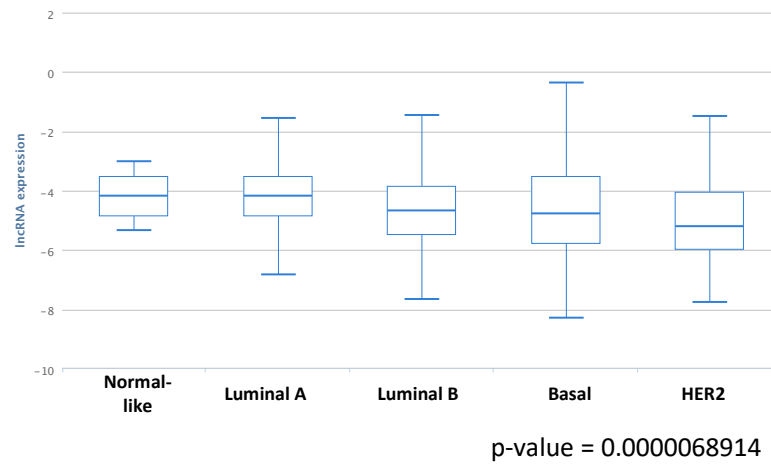
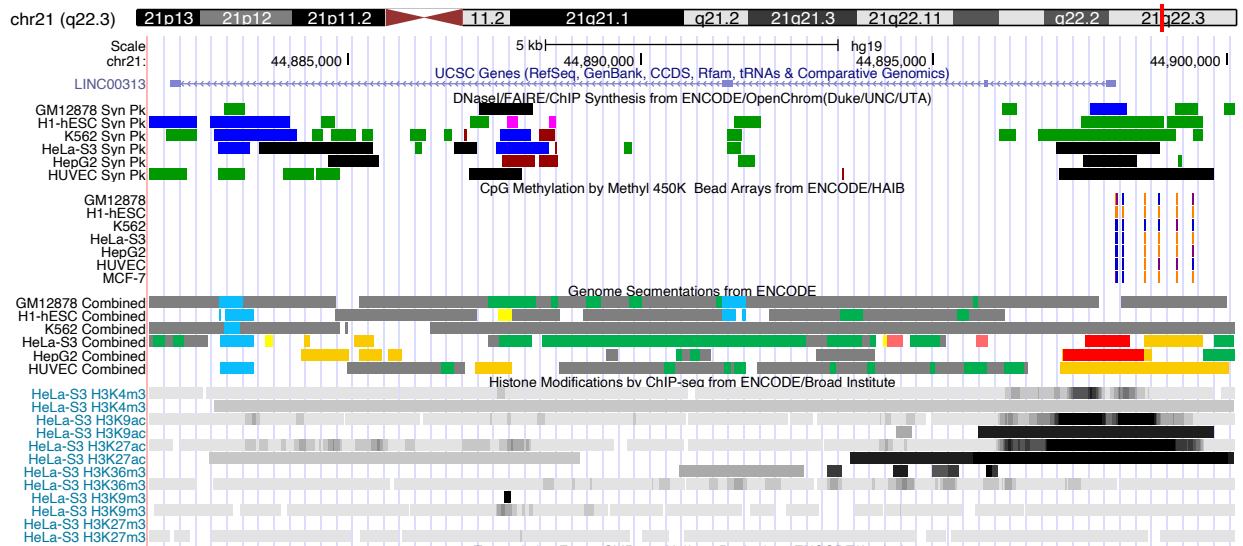


Figure 4.12: Differential expression analysis of LINC00313 based on subtype in an invasive breast cancer cohort from TCGA generated in the TANRIC database.

Expression of LINC00313 in each group of PAM50 classification.

A



B

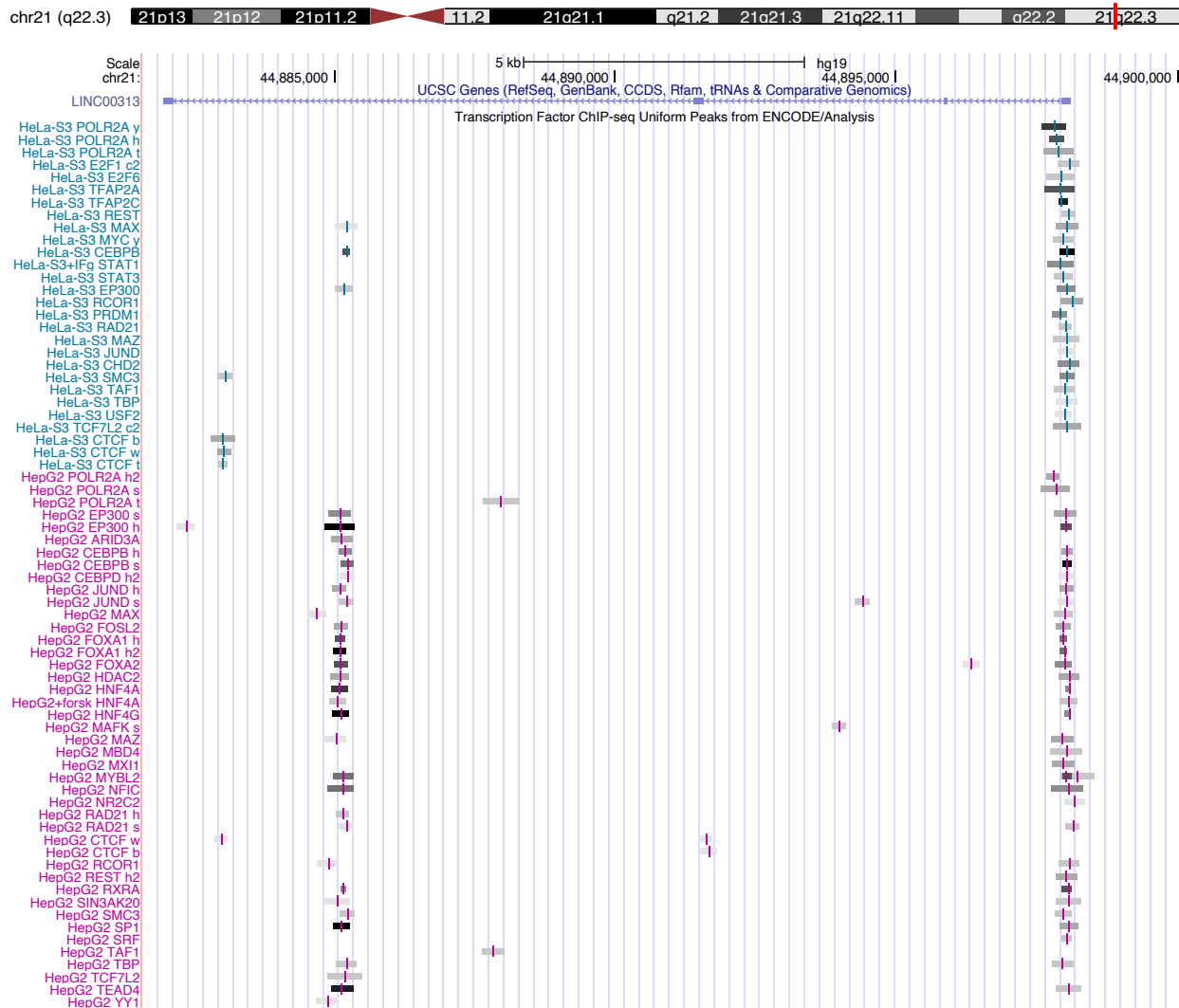


Figure 4.13: Epigenetic landscape of the LINC00313 locus in HeLa cells using the UCSC genome browser annotation hg19. (A) Chromosome compaction, CpG island methylation (orange: methylated; purple: partially methylated; blue: unmethylated; black: NA), Genome segmentation (bright red: predicted promoter region including TSS; light red: predicted promoter flanking region; orange: predicted enhancer; yellow: predicted weak enhancer or open chromatin cis regulatory element; blue: CTCF enriched element; dark green: predicted transcribed region; gray: predicted repressed or low activity region) and histone modifications present on the locus. (B) Transcription factors binding sites presents on the promoter region of the locus determined using CHIP-seq data.

The overlaps of common transcription factors to LINC00707, LINC00473 and LINC00313 (TFAP2A, TFAP2C, CEBPB, EP300, RCOR1, JUND) is not surprising if we consider that those transcription factors are linked with cancer, especially breast cancer.

MIR210HG expression is low in normal breast. We show an increase in expression in the BC types associated with the worst prognosis; indeed, MIR210HG expression is higher in luminal types cancer and even higher in basal type breast cancer (Figure 4.14 A). This is supported by our observation that MIR210HG expression is higher in TNBC (Figure 4.14 B), with ER and PR negative tumors showing higher MIR210HG expression (Figure 4.14 C and 4.14 D). An unmethylated promoter region is evidenced across all cell lines analyzed (Figure 4.15 A). Interestingly, around that promoter region there is a presence of several chromatin marks H3K4me3, H3K9ac and H3K27ac in HeLa cells, a sign of active transcription in the region. MIR210HG is also

suspected to be an RNA polymerase II transcript (Figure 4.15 B); however, unlike LINC00473, LINC00707 and LINC00313, a limited number of transcription factors bind to the promoter region: E2F1, CHD2, MAX, MAZ, NFYA, NFYB and TFAP2C.

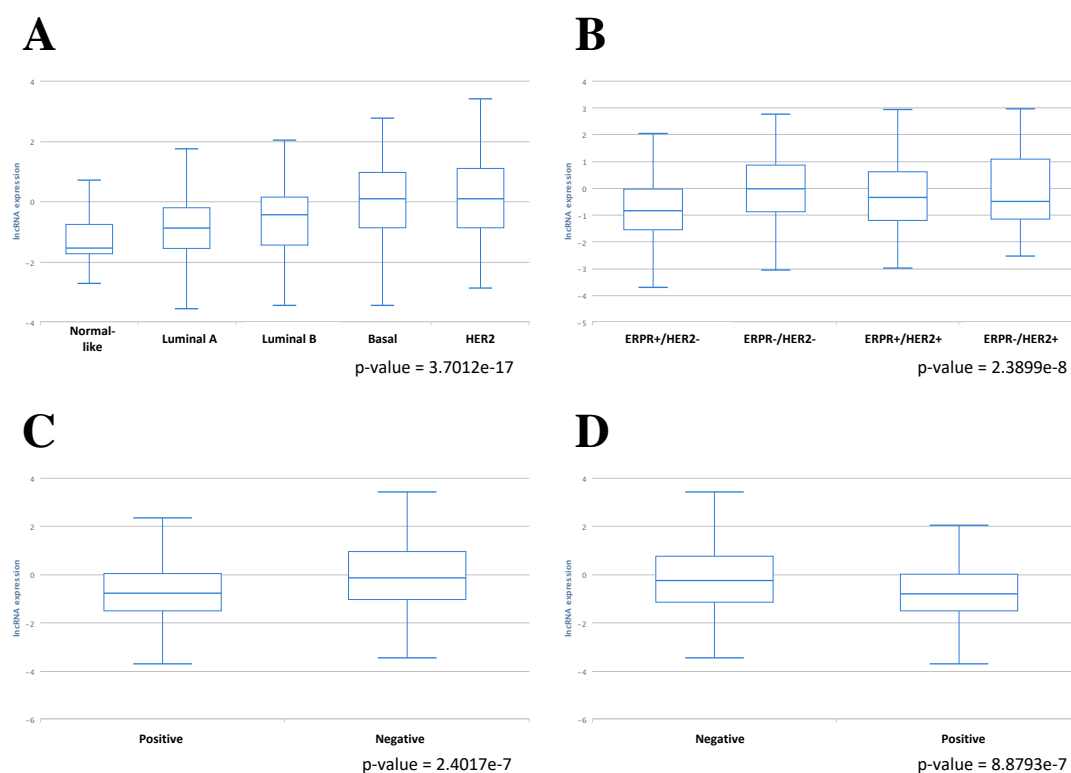


Figure 4.14: Differential expression analysis of MIR210HG based on subtype in an invasive breast cancer cohort from TCGA generated in the TANRIC database. (A) Expression of MIR210HG in each group of PAM50 classification. (B) Expression of MIR210HG based on the therapy strategy classification. (C) Expression of MIR210HG based on the estrogen receptor status classification. (D) Expression of MIR210HG based on the progesterone receptor status classification.

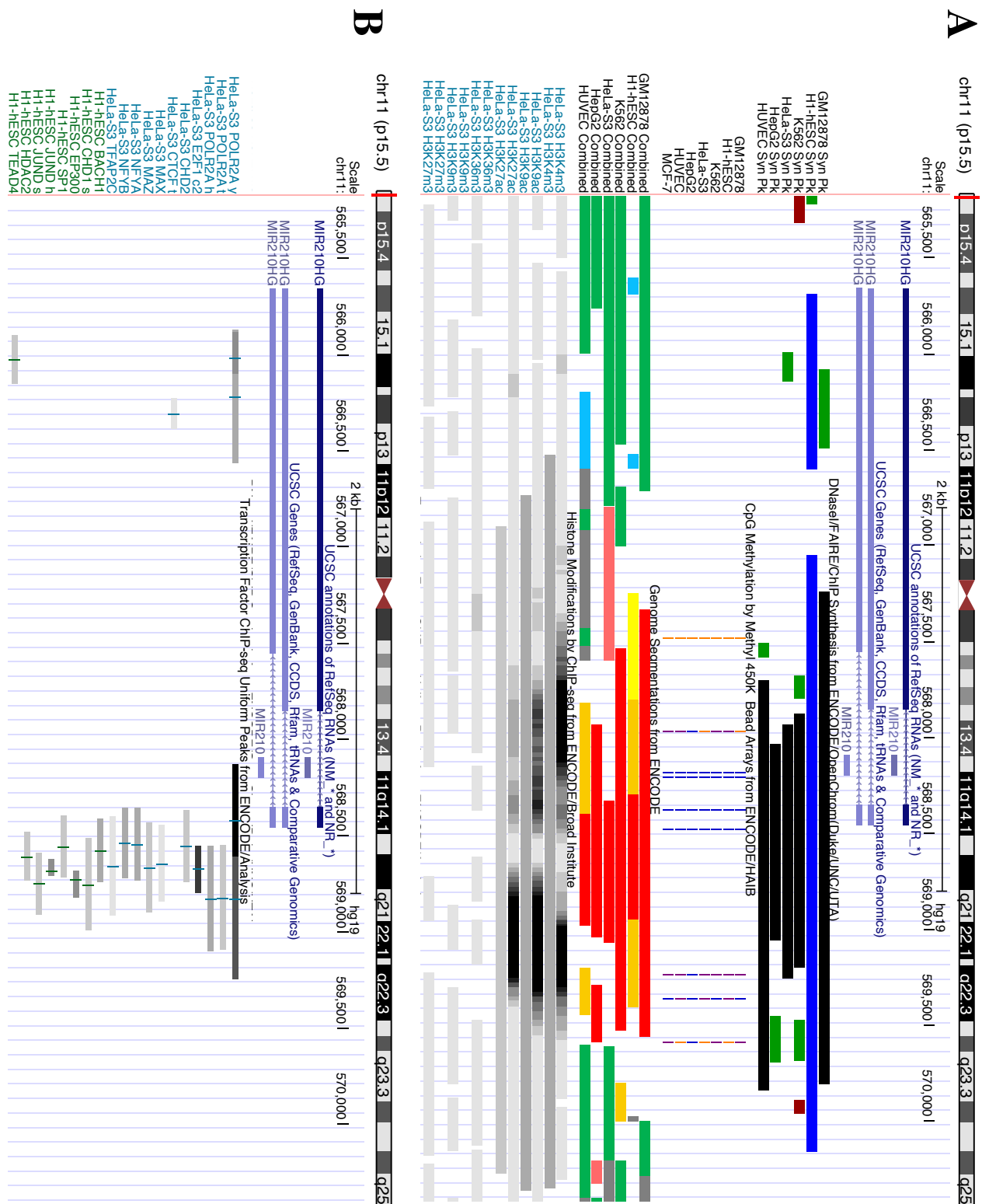


Figure 4.15: Epigenetic landscape of the MIR210HG locus in HeLa cells using the UCSC genome browser annotation hg19. (A) Chromosome compaction, CpG island methylation (orange: methylated; purple: partially methylated; blue: unmethylated;

black: NA), Genome segmentation (bright red: predicted promoter region including TSS; light red: predicted promoter flanking region; orange: predicted enhancer; yellow: predicted weak enhancer or open chromatin cis regulatory element; blue: CTCF enriched element; dark green: predicted transcribed region; gray: predicted repressed or low activity region) and histone modifications present on the locus. (B) Transcription factors binding sites presents on the promoter region of the locus determined using CHIP-seq data.

LOC729970 shows higher expression in normal and luminal type breast cancer, compared to basal or HER2-positive breast cancer, associated with the worst prognosis (Figure 4.16 A). Supporting this, we show that TNBC has the lowest expression of LOC729970 (Figure 4.16 B) with higher expression in ER and PR positive tumors (Figure 4.16 C and 4.16 D). LOC729970 promoter region was found to be unmethylated in several cell lines but not in MCF-7 where the promoter is methylated. This would explain why no expression of LOC729970 was detected in MCF-7 by qRT-PCR (Figure 4.7 B). We chose H1 cells, human embryonic stem cells (hESCs), to study the chromatin marks present on LOC729970 since the lncRNA is expressed in those cells. In red, we identified the promoter region along with some evidence of active transcription in green. In yellow, we can also identify a predicted weak enhancer or open chromatin cis-regulatory sequence. The promoter region is characterized by the presence of H3K27ac and H3K9ac in H1, that correlate with active promoters (Figure 4.17 A). In H1-hESCs a limited number of transcription factors show binding to the promoter region: CHD1 (Chromodomain Helicase DNA Binding Protein 1), MAX, USF1 (Upstream Transcription Factor 1), YY1 (YY1 Transcription Factor) and TAF7 (TATA-Box Binding Protein Associated Factor 7) (Figure 4.17 B).

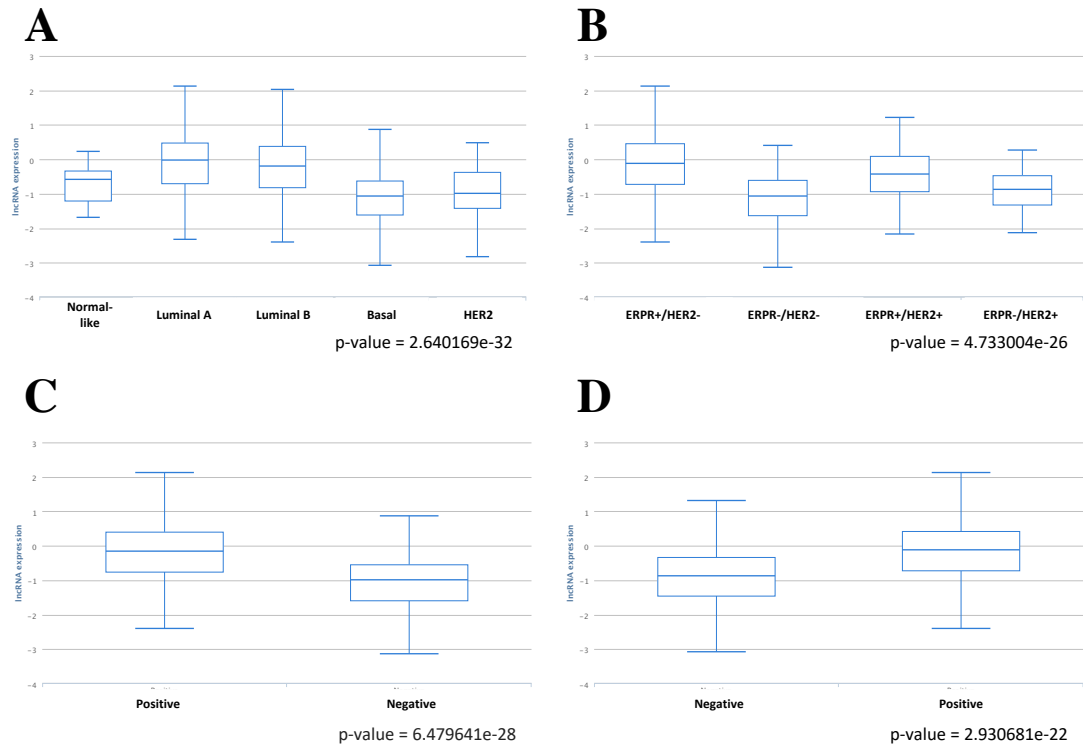


Figure 4.16: Differential expression analysis of LOC729970 based on subtype in an invasive breast cancer cohort from TCGA generated in the TANRIC database.

(A) Expression of LOC729970 in each group of PAM50 classification. (B) Expression of LOC729970 based on the therapy strategy classification. (C) Expression of LOC729970 based on the estrogen receptor status classification. (D) Expression of LOC729970 based on the progesterone receptor status classification.

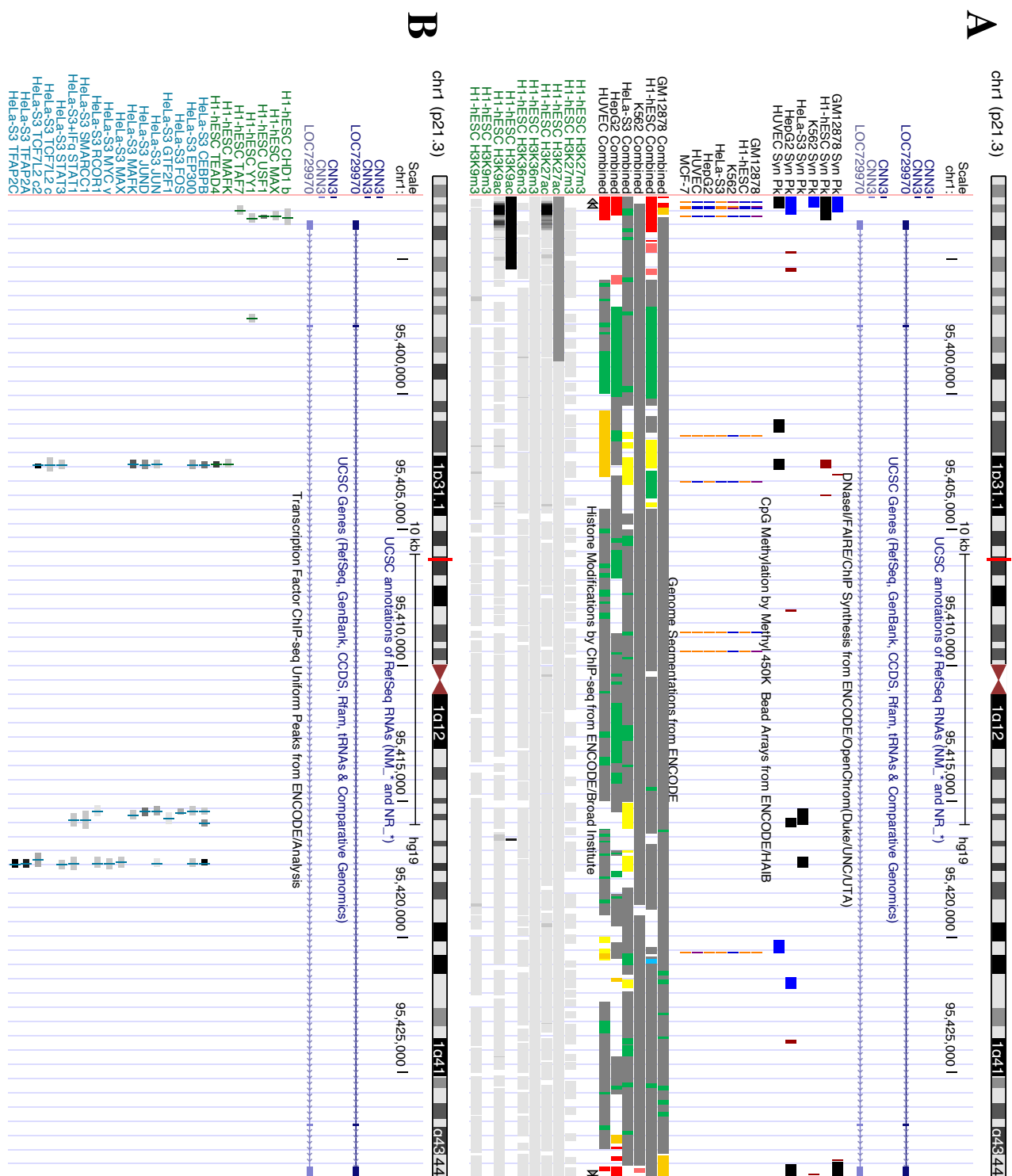


Figure 4.17: Epigenetic landscape of the LOC729970 locus in H1 human Embryonic Stem Cells (hESCs) using the UCSC genome browser annotation hg19. (A) Chromosome compaction, CpG island methylation (orange: methylated; purple: partially methylated; blue: unmethylated; black: NA), Genome segmentation

(bright red: predicted promoter region including TSS; light red: predicted promoter flanking region; orange: predicted enhancer; yellow: predicted weak enhancer or open chromatin cis regulatory element; blue: CTCF enriched element; dark green: predicted transcribed region; gray: predicted repressed or low activity region) and histone modifications present on the locus. (B) Transcription factors binding sites presents on the promoter region of the locus determined using CHIP-seq data.

LOC100134040 expression seems similar between normal breast, basal and HER-positive BC (Figure 4.18 A). A slight increase in expression is observed in luminal type BC. Figure 3.16 C, 3.18 B and 3.18 D suggest that LOC100134040 expression is slightly decreased in TNBC which does not support our previous observations. LOC100134040 showed CpG methylation around the promoter region in most of the cell lines in our study, only K562 shows partial methylation (Figure 4.19 A). No histone modification or transcription factor binding (Figure 4.19 A and 4.19 B) was identified on LOC100134040 promoter but that can be explained by the partial methylation found on that region which silences LOC100134040 expression.

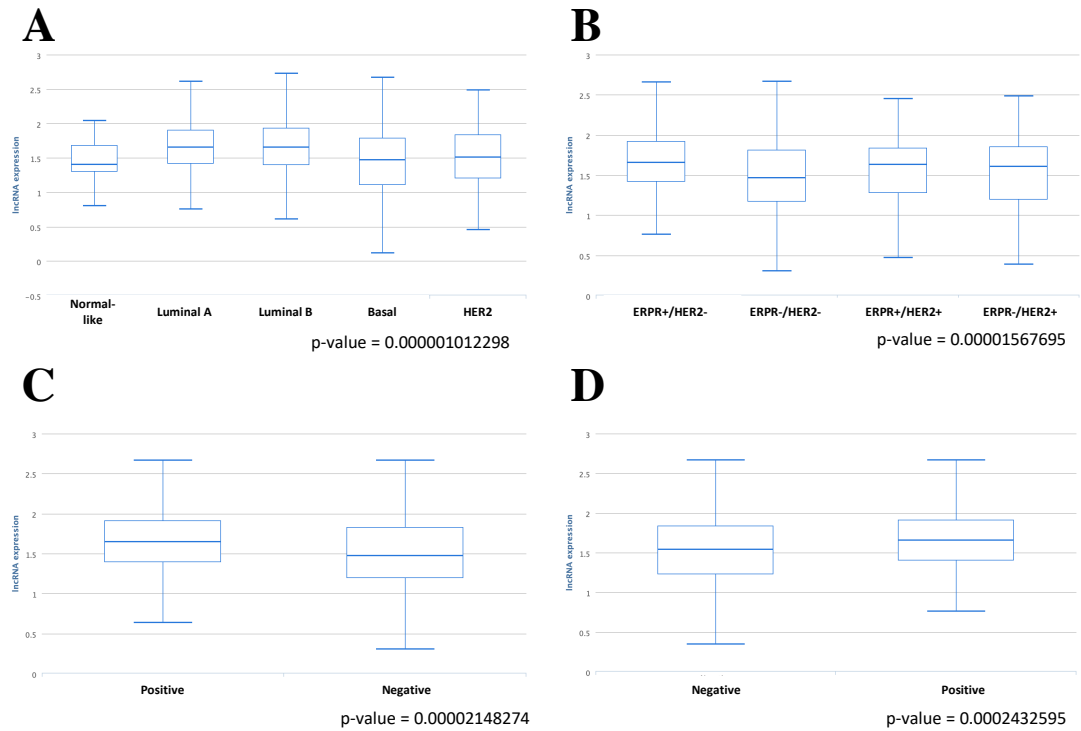
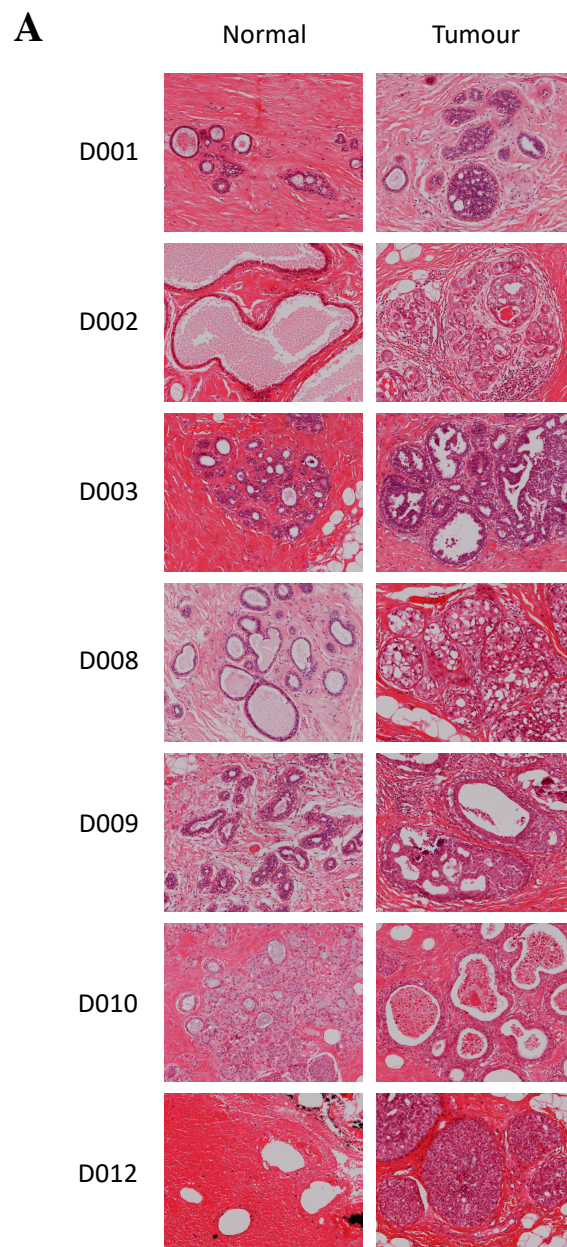


Figure 4.18: Differential expression analysis of LOC100134040 based on subtype in an invasive breast cancer cohort from TCGA generated in the TANRIC database. (A) Expression of LOC100134040 in each group of PAM50 classification. (B) Expression of LOC100134040 based on the therapy strategy classification. (C) Expression of LOC100134040 based on the estrogen receptor status classification. (D) Expression of LOC100134040 based on the progesterone receptor status classification.

blue: CTCF enriched element; dark green: predicted transcribed region; gray: predicted repressed or low activity region) and histone modifications present on the locus. (B) Transcription factors binding sites presents on the promoter region of the locus determined using CHIP-seq data.

4.2.7 Recruitment of a DCIS patient cohort

Next, we aimed to compare our RNAseq results in DCIS cell lines to the lncRNAs expressed in DCIS tumors. Patients with DCIS were enrolled in a study where they were asked for permission following surgery for the tissue to be used. In collaboration with Dr Conleth Murphy (Bon Secours Hospital), the tissues were prepared in Formalin Fixed Paraffin Embedded (FFPE) blocks and sliced. Those slices of tissue were used to extract RNA from the DCIS biopsy and adjacent normal tissue to perform the lncRNA capture, successfully conducted on RNA extracted from DCIS cell lines. Figure 4.20 A shows the haematoxylin and eosin staining of tissue from each patient, with the nucleus stained in purple and the cytoplasm in pink. For each patient sample, we can see in normal tissue with a transversal cut of the duct of the breast, there are no cells inside the duct; whereas in the DCIS areas, we observed that the duct have been completely filled with cells, in particular for the patients D001 and D012. Several clinicopathological parameters in relation to the tumor were collected, including the size of the tumor in mm, the grade of DCIS, the status of ER, PR and HER2, whether some necrosis or microinvasion was observed and the pattern of tumor growth. That information is recapitulated in the table in Figure 4.20 B. Five out of seven tumors were ER and PR positive (D001, D003, D009, D010, D012). Necrosis was present in five out of seven tumors (D002, D008, D009, D010, D012). Microinvasion was only observed in the tumor biopsy of the patient D002.



B

	D0001	D002	D003	D008	D009	D010	D012
Size of DCIS (mm)	15	10	2.1	64	4	20	34
Grade of DCIS	2	3	1	3	2	2	3
ER Status	+	-	+	-	+	+	+
PR Status	+	-	+	-	+	+	+
HER2 Status	NA	NA	NA	NA	NA	NA	NA
Necrosis present	-	+	-	+	+	+	+
Microinvasion	-	+	-	-	-	-	-
Pattern of tumor growth	Micropapillary and cribiform	cribiform	cribiform	solid	solid	solid	solid

Figure 4.20: Characterization of the DCIS patient cohort. (A) Haematoxylin and eosin staining of the biopsy of the DCIS lesion and the normal adjacent tissue of each recruited patient. (B) Clinicopathological features of DCIS samples, including size of the tumor, grade of DCIS, ER/PR/HER2 status, presence of microinvasion or necrosis and pattern of tumor growth for each tumor sample.

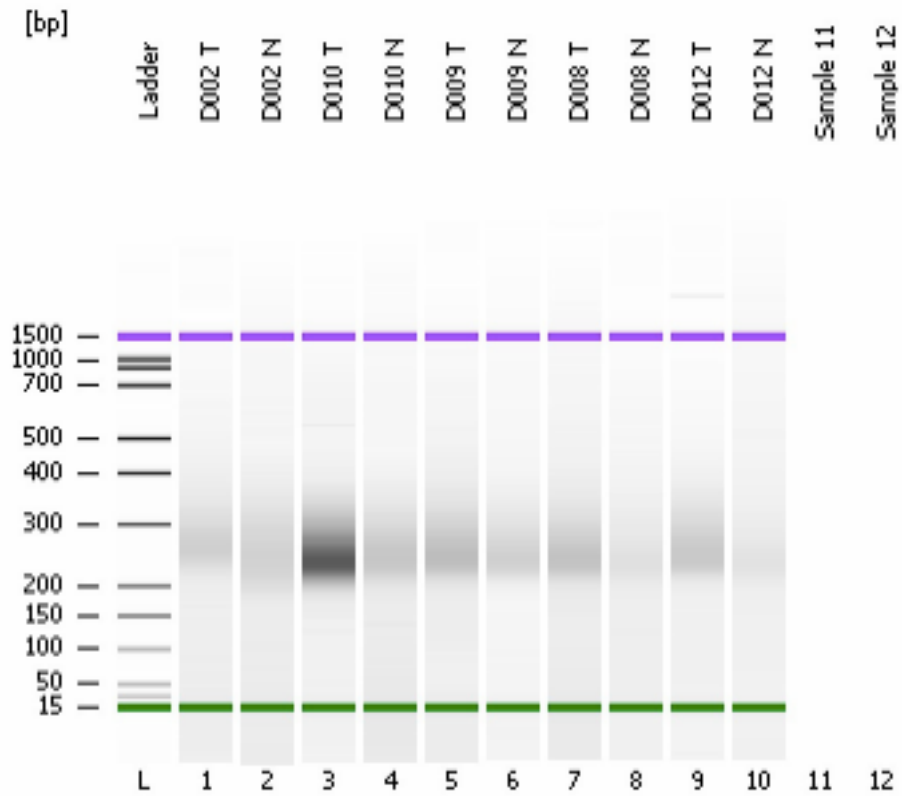
4.2.8 lncRNA capture on DCIS patient RNA

Since limited material in the DCIS lesions was available for RNA extraction, we decided to limit the amount of RNA dedicated to control and only measured RNA concentration with the Nanodrop spectrophotometry. In consequence, we did not analyze RNA quality with the bioanalyzer.

Slides of tissues were marked by Dr. Claudia Aura (Conway Institute, University College Dublin). After deparaffinization, the DCIS lesion and adjacent normal tissue were scrapped from the slide and separated in two different microtubes. RNA was extracted from tissue from all seven patients; however, RNA from only five patients was used to prepare RNA sequencing library. Indeed, the percentage of tumor cells in the samples collected from patients D001 and D003 was under 30% which would have resulted in not enough RNA from tumor cells for RNAseq library preparation.

We verified if the library preparation had been successful (Figure 4.21). We observed a light smear between 200 bp and 350 bp corresponding to the fragmented library. This is confirmed by the electropherogram that resembles those of a successfully prepared library attesting of the success of the library preparation.

A



B

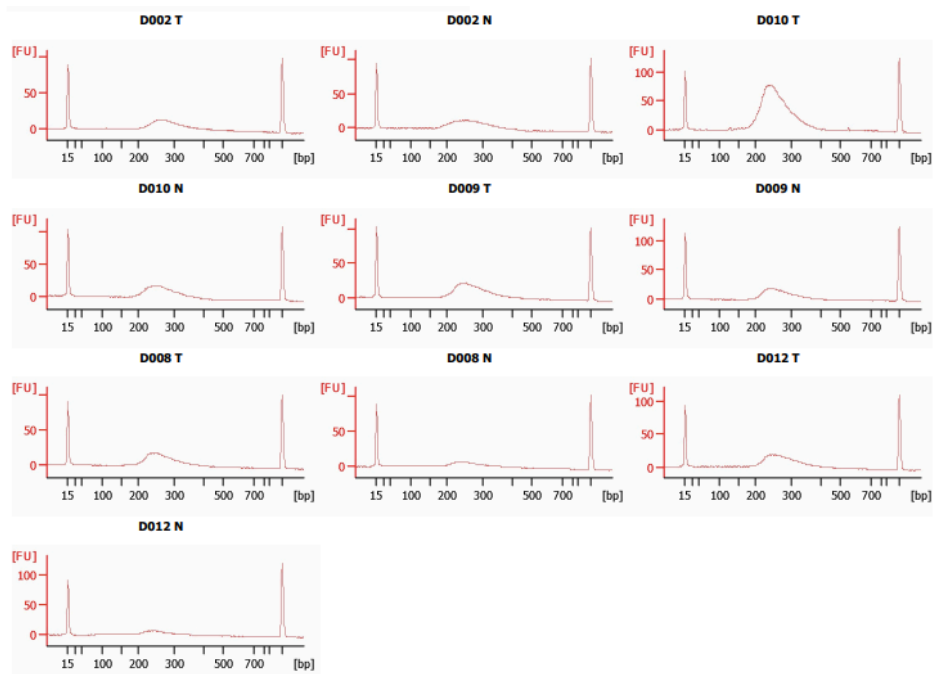


Figure 4.21: Verification of successful library preparation with normal and tumor RNA from the DCIS patient cohort. (A) Gel image from the bioanalyzer DNA analysis. (B) Electropherogram from the bioanalyzer DNA analysis.

Table 4.4 shows the performance metrics from the lncRNA capture using RNA from DCIS patients. All samples showed a mapping rate around 80% apart from D010_T that shows a mapping rate of only 27.9%. Regarding the efficiency of the capture, the percentage of bases in target ranges from 36% to 82% which corresponds to a 64 to 147-fold change enrichment across samples. This is a higher rate than what was obtained with the cell lines which suggests that the lncRNA capture was successful. However, the overall percentage of input reads after quality control was between 36% and 45% which corresponds to half of what was observed in the previous round of sequencing for the cell lines. Moreover, the overall rate of reads that aligned to the human genome ranged from 8% to 42%. This might be explained by the high rate of reads that aligned to ERCC controls (External RNA Controls Consortium), a rate in average two to three times higher than what was observed in our previous study with the cell lines.

Reference	Sample	D002_N	D002_T	D008_N	D008_T	D009_N	D009_T	D010_N	D010_T	D012_N	D012_T
Primary_target_size [bp]	hg19	18 950 280	18 950 280	18 950 280	18 950 280	18 950 280	18 950 280	18 950 280	18 950 280	18 950 280	18 950 280
Capture_target_size [bp]	17 427 025	17 427 025	17 427 025	17 427 025	17 427 025	17 427 025	17 427 025	17 427 025	17 427 025	17 427 025	17 427 025
Total_input_reads	86 297 086	108 970 096	78 520 654	74 438 728	61 721 686	64 915 002	73 044 402	72 508 116	81 376 296	83 451 770	83 451 770
Input_reads_after_filtering	39 337 316	55 294 856	26 094 856	29 064 560	25 006 038	28 778 722	31 533 066	26 976 926	24 906 604	37 996 260	37 996 260
Mean_length_read [bp]	45.58	50.74	33.23	39.04	40.51	44.33	44.37	36.23	30.61	45.53	45.53
Mean_length_read2 [bp]	146.29	146.29	139.44	143.70	145.73	145.73	144.18	144.14	136.56	154.23	154.23
Mean_length_read2 [bp]	138.42	139.84	134.12	138.18	139.38	140.02	138.10	136.65	130.70	139.00	139.00
---TOPHAT---											
Overall_read_mapping_rate [%]	85.3	82.6	89.1	84.3	80.0	84.3	79.4	27.9	88.6	83.8	83.8
Aligned_pairs	16 065 015	21 790 870	11 218 081	11 721 500	9 552 895	11 593 658	11 291 832	3 291 644	10 634 001	15 218 636	15 218 636
Pairs_with_multiple_alignments	874 623	1 113 029	139 358	489 260	501 830	566 311	666 748	99 768	97 762	813 306	813 306
Pairs_with_multiple_alignments [%]											
No_of_junctions	26548	21337	6125	16113	21744	15658	22858	52177	3171	20282	20282
Reads_aligned_to_ERCC [%]	23.93	33.07	82.70	40.45	27.70	28.48	23.72	12.68	87.14	17.77	17.77
---Picard MarkDuplicates---											
Mapped_non_duplicate_reads	3 273 886	2 667 228	629 734	2 167 558	2 513 952	2 615 758	2 876 549	1 333 339	382 577	3 336 144	3 336 144
Read_pair_duplicates	12 789 129	19 123 642	10 588 347	9 553 942	7 038 943	8 977 900	9 055 283	1 958 305	10 251 424	11 882 492	11 882 492
Duplicate_rate [%]	80.11	88.07	94.52	81.96	74.36	78.05	76.56	62.40	96.51	78.69	78.69
Optical_duplicates [%]	2.83	3.01	3.07	2.62	2.84	2.63	2.85	2.46	2.55	2.74	2.74
---Picard CollectInsertSizeMetrics---											
Mean_insert_size [bp]	201.95	201.17	195.53	192.64	196.95	198.98	195.49	190.32	189.67	203.78	203.78
Insert_size_std_dev [bp]	65.25	62.46	58.92	55.57	58.09	57.65	58.68	51.00	53.80	66.81	66.81
Median_insert_size [bp]	196	193	187	186	191	192	190	184	183	197	197
---bedtools intersect---											
Reads_on_primary_target_withoutERCC	4 620 749	3 563 190	557 474	2 751 677	3 303 760	3 585 792	4 001 354	904 210	311 808	4 848 213	4 848 213
Reads_on_primary_target_withERCC	5 406 679	4 506 097	1 368 846	3 660 750	4 023 131	4 421 341	4 776 020	1 183 774	857 778	5 508 815	5 508 815
Reads_on_primary_target [%]	33.65	22.25	12.20	31.50	42.11	38.13	39.60	35.96	8.06	36.19	36.19
---GATK DepthOfCoverage---											
Mean_coverage_primary_target	29.60	23.41	7.81	20.97	24.00	25.95	26.65	7.89	4.71	30.97	30.97
Median_coverage_primary_target	2	1	1	1	2	1	1	1	1	1	1
%bases_above_1x_primary_target	53.4	46.0	13.0	40.8	52.9	45.3	48.8	29.9	4.9	49.7	49.7
%bases_above_1x_primary_target	22.9	18.0	1.9	14.1	20.9	18.4	20.7	6.1	0.6	21.7	21.7
%bases_above_20x_primary_target	14.8	10.8	1.0	8.4	13.2	11.8	13.4	2.7	0.4	14.5	14.5
%bases_above_30x_primary_target	10.9	7.5	0.7	5.9	9.5	8.5	9.9	1.7	0.3	10.9	10.9
%bases_above_50x_primary_target	7.0	4.4	0.5	3.5	5.7	5.3	6.2	1.0	0.3	7.0	7.0
%bases_above_100x_primary_target	3.3	2.0	0.4	1.7	2.6	2.5	2.9	0.6	0.2	3.4	3.4
---Picard CollectHsMetrics---											
Bases_in_primary_target [%]	37.08	35.18	54.48	39.00	38.30	40.00	37.03	24.66	55.11	37.87	37.87
Bases_in_capture_target [%]	57.01	54.32	82.76	61.51	59.95	62.74	58.47	36.54	83.89	58.29	58.29
Bases_near_capture_target [%]	4.74	4.73	2.77	4.03	4.46	4.51	4.71	3.17	3.20	4.73	4.73
Bases_in_padded_target [%]	61.75	59.06	85.54	65.55	64.42	66.77	63.19	39.72	87.10	63.03	63.03
Fold_enrichment	101.27	96.50	147.03	109.27	106.50	110.58	103.88	64.93	149.03	103.55	103.55
---Picard CollectRnaSeqMetrics---											
PF_BASIS [bp]	943 509 194	779 152 044	174 534 383	623 070 440	729 993 096	761 125 574	827 723 081	386 778 574	102 917 175	963 482 165	963 482 165
PF_ALIGNED_BASIS [bp]	943 441 832	779 093 099	174 321 780	623 025 489	729 943 191	761 072 687	827 666 788	386 751 382	102 909 605	963 413 361	963 413 361
RIBOSOMAL_BASIS [bp]	11 436 209	14 767 980	6 832 278	9 114 187	10 807 665	9 979 390	12 438 411	18 414 229	7 504 969	13 632 851	13 632 851
CODING_BASIS [bp]	149 613 684	132 125 826	9 915 105	88 714 189	111 293 137	108 080 879	128 343 384	49 031 926	3 428 077	159 281 284	159 281 284
INTRONIC_BASIS [bp]	263 873 383	193 879 215	17 253 396	158 333 786	208 779 051	220 040 099	234 271 482	75 375 926	5 164 504	300 151 267	300 151 267
INTERGENIC_BASIS [bp]	120 159 164	87 856 539	6 242 917	60 794 783	93 758 582	80 043 356	107 505 143	111 742 476	2 198 952	103 854 876	103 854 876
PCT_RIBOSOMAL_BASIS [%]	398 360 027	350 464 507	134 078 312	306 069 237	305 305 119	342 927 389	345 108 999	132 187 554	84 613 348	386 493 770	386 493 770
PCT_CODING_BASIS [%]	1.21	1.90	3.92	1.46	1.48	1.31	1.50	4.76	7.29	1.42	1.42
PCT_UTR_BASIS [%]	15.66	16.96	5.69	14.24	15.25	14.20	15.51	12.68	3.33	16.53	16.53
PCT_INTRONIC_BASIS [%]	27.97	27.97	28.60	25.41	28.60	28.91	28.30	19.49	5.02	31.16	31.16
PCT_INTERGENIC_BASIS [%]	12.74	11.28	3.58	9.76	12.84	10.52	12.99	28.89	2.14	10.78	10.78
PCT_MRNA_BASIS [%]	42.22	44.98	76.91	49.13	41.83	45.06	41.70	34.18	82.22	40.12	40.12
PCT_USABLE_BASIS [%]	43.82	41.84	15.58	39.65	43.85	43.11	43.81	32.17	8.35	47.68	47.68
---ERCC controls---											
Mean_FPKM_targeted_ERCCs	3447.78	2664.60	7752.06	4489.40	4743.26	4091.83	4182.01	1860.98	7403.39	3430.67	3430.67
Mean_FPKM_non_targeted_ERCCs	16.45	31.92	39.89	46.40	14.30	23.57	8.75	209.83	26.11	5.19	5.19
Mean_ERCC_fold_change	209.00	83.00	194.00	96.00	331.00	173.00	477.00	8.00	283.00	661.00	661.00

Table 4.4: Performance metrics of the RNA sequencing after lncRNA capture with normal and tumor RNA from the DCIS patient cohort. The number of reads obtained for each sample before and after filtering, number of reads aligned to the reference genome, number of reads aligned to the target and type of reads are listed in this table. This table was provided to us by Dr. Xavier Miro and Dr. Dominic Rose (Roche).

Nonetheless, differential expression was performed. Two different analysis were performed, first a pair-wise analysis where for each patient, the tumor sample is compared to the normal sample. The second analysis compared all tumor samples as one group to all normal samples. Regarding the pair-wise analysis, for D002 we identified 127 differentially expressed genes, 83 with higher expression and 44 with lower expression; for D008, 60 differentially expressed genes, 12 with higher expression and 48 with lower expression; for D009, 52 differentially expressed genes, 27 with higher expression and 25 with lower expression; for D010, 73 differentially expressed genes, 27 with higher expression and 46 with lower expression and for D012, 200 differentially expressed genes, 18 with higher expression and 179 with lower expression (Figure 4.22 A). We compared the list of differentially expressed genes in each patient, but none overlapped. The second analysis resulted in 119 differentially expressed genes in the tumor group compared to the normal. After filtering off the protein coding genes and pseudogenes, we obtained five differentially expressed lncRNAs, three with higher expression and two with lower expression. Compared to the previous study with the cell lines, the proportion of differentially expressed lncRNAs is minor. Since the performance metrics of the RNAseq suggest that the lncRNA capture was successful, the lack of lncRNAs in differentially expressed genes might be explained by an insufficient coverage. Indeed, we have mentioned previously that, in average, coverage for the human genome should be around 30X for each sample which we reached with each sample of the first RNAseq experiment (Table 4.2). For the second, however, mean coverage was around 25X, three samples showing a coverage lower than 10X. We also mentioned in the introduction that lncRNAs tend to be expressed at lower levels than mRNAs, so an insufficient coverage could prevent us from analyzing expression of lncRNAs.

Analysis of overall survival of IBC cohort from TCGA was conducted based on the level of expression of a given lncRNA. This allowed us to identify MEG8, an lncRNA with higher expression in tumor sample in comparison to normal (Figure 4.22 B). Low expression of MEG8 is associated with better disease-free survival in the TCGA cohort. This could mean that MEG8 is important in breast cancer progression. However, its expression is lower in invasive breast cancer compared to normal tissue, which does not support a role in breast cancer progression.

expression comparison of the tumor and normal group. In green are represented genes with higher expression and in red genes with lower expression. (B) Identification of MEG8, an upregulated lncRNA in the tumor group whose low expression is associated with better overall survival in an invasive breast cancer cohort.

The low expression of MEG8 in invasive breast cancer is confirmed by data extracted from the TANRIC database that shows lower expression of MEG8 in BC, with basal BC showing the lowest expression (Figure 4.23 A). This is supported by expression analysis based on ER and PR status where ER/PR negative patients show slightly lower expression than ER/PR positive cases (Figure 4.23 B and 4.23 C).

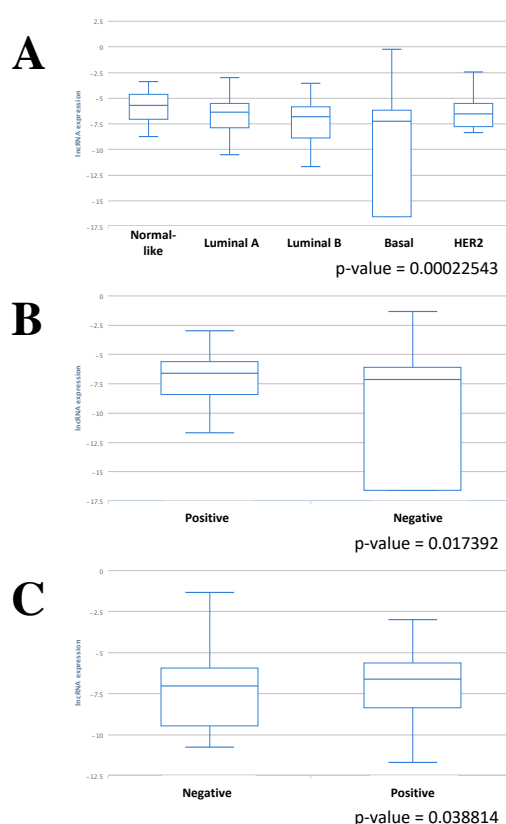


Figure 4.23: Differential expression analysis of MEG8 based on subtype in an invasive breast cancer cohort from TCGA generated in the TANRIC database. (A) Expression of MEG8 in each group of PAM50 classification. (B) Expression of

MEG8 based on the estrogen receptor status classification. (C) Expression of MEG8 based on the progesterone receptor status classification.

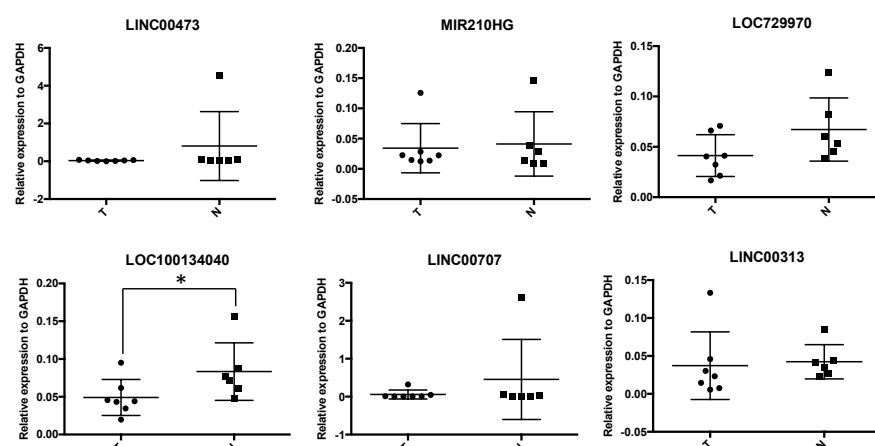
Next, we tried to quantify *MEG8* expression in our breast cancer cell lines by qRT-PCR. However, after performing qRT-PCR the analysis of the melting temperature suggested that the DNA oligonucleotides bound unspecifically to DNA by detecting different templates. On the Ensembl genome browser, the *MEG8* gene is associated with 53 different splice variants, which probably explains the lack of specificity of the DNA oligonucleotides. Since the study of *MEG8* is highly complex, we decided to focus, for the time being, on the six lncRNA candidates we identified with our RNAseq study in the DCIS cell lines.

4.2.9 Expression of the six lncRNA candidates in our DCIS patient cohort

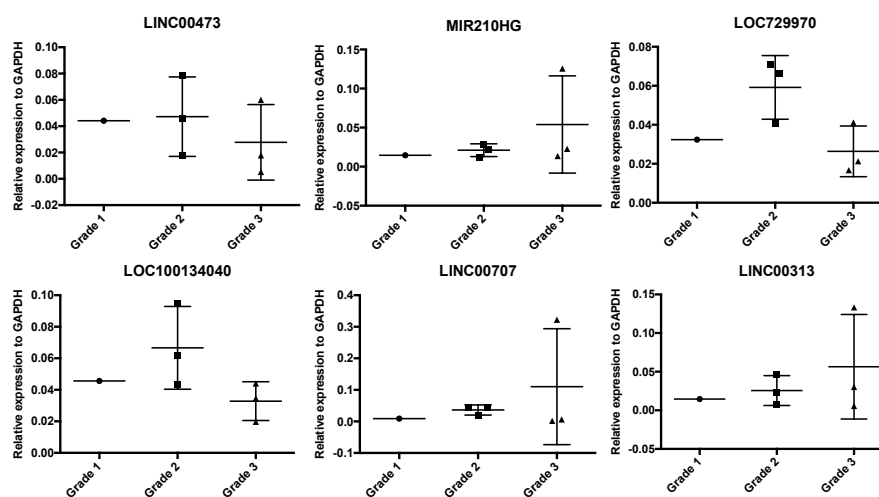
Since the six lncRNA candidates identified in the RNAseq experiment on the DCIS cell lines were not confirmed in the RNAseq experiment on the DCIS patient cohort, we wanted to quantify their expression by qRT-PCR using the DCIS patient samples. For this, we did not exclude any samples, as qRT-PCR using specific primers is a technique more sensitive for low transcript numbers than the RNAseq experiment conducted previously and would therefore allow us to highlight a smaller difference in expression. Comparing expression of each lncRNA candidate in tumor and matched normal tissue, we saw no difference for LINC00473, MIR210HG, LINC00707 and LINC00313. For LOC100134040, we show lower expression in the tumor (Figure 4.24 A). LOC729970 shows a trend toward a lower expression in the tumor group as well, close to significance. This supports what we observed in invasive breast cancer with data from TCGA, where LOC729970 expression is lower in the tumor compared to

the normal tissue (Figure 4.6 B). Expression was also compared based on the DCIS grade (Figure 4.24 B) or the ER/PR status (Figure 4.24 C) but no statistical significance was evidenced.

A



B



C

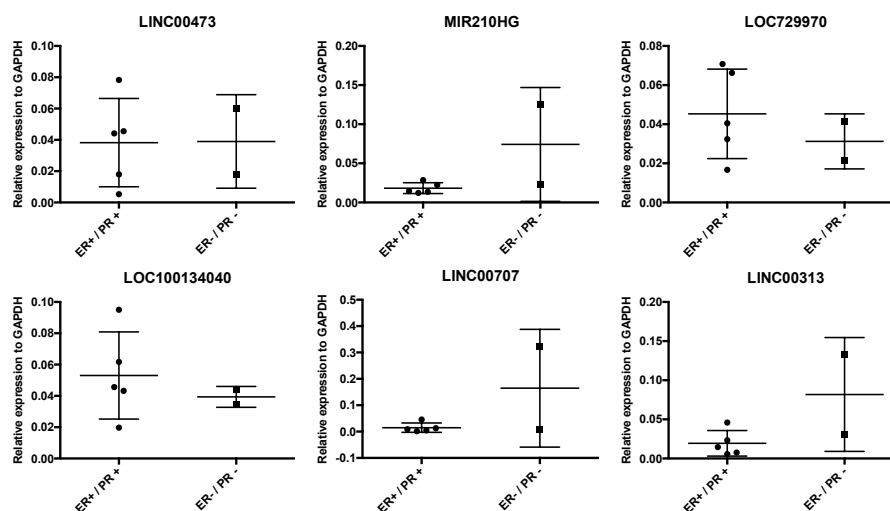


Figure 4.24: qRT-PCR expression quantification of the six lncRNA candidates in the DCIS patient cohort. (A) Comparison of the level of expression of each candidate between the tumor and normal group. (B) Comparison of the expression level of each candidate by DCIS grade. (C) Comparison of the level of expression of each candidate by ER/PR status. The qRT-PCR data was normalized to GAPDH and the $2^{-\Delta CT}$ was plotted. Statistical analysis was done using t-test (GraphPad v.8.3.0). **** p-value < 0.0001; *** p-value < 0.001; ** p-value < 0.01; * p-value < 0.05.

4.2.10 Identification of dysregulated lncRNAs in the transcriptomics study of ETCC-006 and ETCC-010 presented in the Chapter 3

Along with the results described in Chapter 2, we analyzed that RNAseq dataset to determine if dysregulated lncRNAs could be identified in ETCC-006 or ETCC-010 compared to MCF-10A. Five lncRNAs shown in Figure 4.24 A were identified: MIR100HG, CCDC18-AS1, MGC45800, LOC511471 and LINC02035. In that dataset, we showed higher expression of MIR100HG, CCDC18-AS1, MGC45800 in ETCC-006 cells, LOC511471 showed higher expression in ETCC-010 and LINC02035 showed lower expression in ETCC-006 and ETCC-010. We next attempted to confirm those results by qRT-PCR (Figure 4.25 B). MIR100HG showed similar expression in MCF-10A, MCF10DCIS.com and ETCC-006, and lower expression in ETCC-010. CCDC18-AS1 showed lower expression in MCF10DCIS.com and higher expression in ETCC-006, confirming the RNAseq results, no difference was shown in ETCC-010. MGC45800 expression was higher in ETCC-006 cells compared to MCF10DCIS.com cells (p-value = 0.0089). LOC511471 showed lower expression in MCF10DCIS.com and ETCC-010 which contradicts the

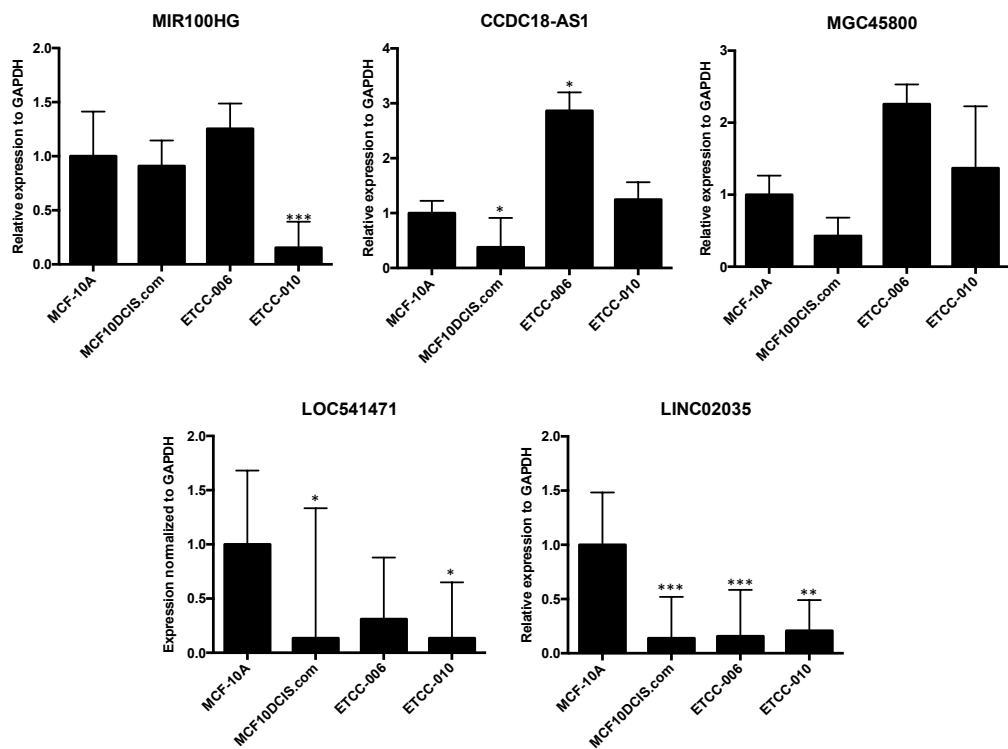
RNAseq results. LINC02035 showed lower expression in all DCIS cell lines, thus confirming the RNAseq results. Overall, the RNAseq results were not confirmed by qRT-PCR. We analyzed the expression of those lncRNAs in the IBC cohort of TCGA and saw lower expression of MIR100HG in the tumor group; however, no difference in patient survival was shown (Figure 4.25 C). The other lncRNAs identified here showed no difference in expression between the tumor and the normal group. Regarding overall survival, only CCDC18-AS1 seemed to show a statistical difference, overall survival being improved when CCDC18-AS1 expression is high. This could suggest that CCDC18-AS1 has a protective role against breast cancer progression. This is supported by the trend of CCDC18-AS1 expression between the tumor and normal group, indeed expression tends to be lower in the tumor group.

The lncRNAs MIR100HG, CCDC18-AS1, MGC45800, LOC511471 and LINC02035 were compared to the RNAseq dataset obtained by the lncRNA capture, but none had been found to be differentially expressed in that subsequent study, so they were discarded.

A

Cell line	Gene_name	Fold_Change (RNAseq)	P-value	P-adjusted	Fold_Change (qRT-PCR)
ETCC-006	MIR100HG	3,4	0,044	0,6	4,1
	CCDC18-AS1	3,7	0,04	0,6	3,9
	MGC45800	3,84	0,037	0,6	3,8
ETCC-010	LOC541471	3,63	0,042	0,57	0,13
ETCC-006 and ETCC- 010	LINC02035	0,28	0,044	0,6	0,44
		0,25	0,03	0,57	0,53

B



C

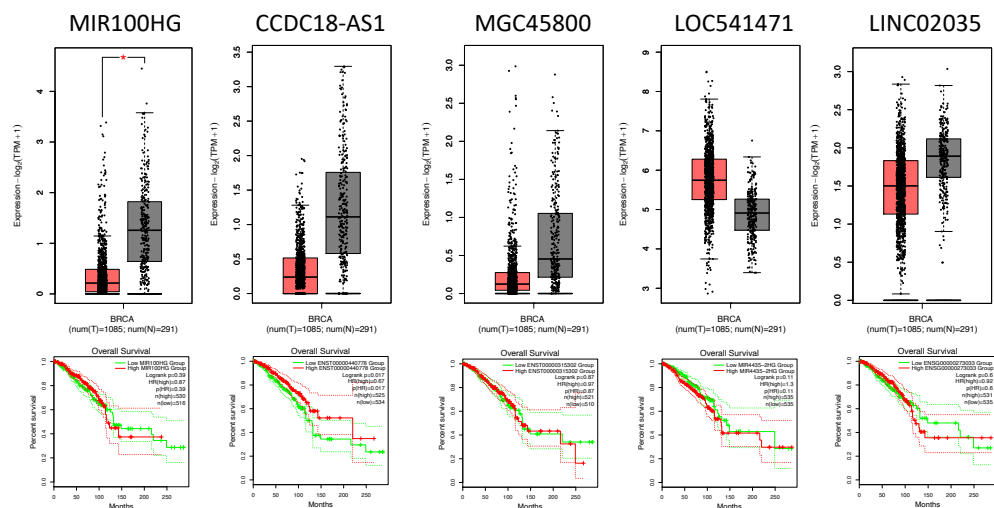


Figure 4.25: Dysregulated lncRNAs in ETCC-006 and ETCC-010 identified using the dataset generated and described in Chapter 3. (A) Table listing lncRNAs dysregulated in ETCC-006 and ETCC-010 compared to MCF-10A by RNA-seq and qRT-PCR. (B) qRT-PCR confirmation of the RNAseq results. (C) Expression of the identified lncRNAs in the IBC cohort from TCGA and overall survival associated plot within the same cohort.

4.3 Discussion

Since lncRNAs could potentially provide clues about DCIS initiation and progression and, in the remote future, be used clinically for the detection and/or treatment of DCIS, we aimed to identify lncRNAs associated with DCIS. We performed two RNAseq experiments, the first using three DCIS cell lines, MCF10DCIS.com, ETCC-006 and ETCC-010. This allowed us to identify 92 differentially expressed lncRNAs with 58 lncRNAs showing higher expression in the DCIS cell lines and 34 showing lower expression in the DCIS cell lines compared to MCF-10A. Although we chose to focus on the lncRNAs that are clinically relevant by their association with overall patient survival, we aim to interrogate our list of 92 lncRNAs in a larger DCIS patient cohort in the future. Unfortunately, those lncRNAs were not confirmed in the RNAseq experiment conducted in our DCIS patient cohort. This is likely explained by the difference in RNA quality between the first and second RNAseq experiment. Indeed, RNA extracted from FFPE blocks is severely degraded in comparison to RNA extracted from frozen tissue. However, obtaining fresh frozen tissue of DCIS tumors is extremely difficult. As mentioned previously, full diagnosis DCIS is done with the analysis of a tumor biopsy and DCIS lesions are often small. In consequence, after the characterization of the lesion by the pathologists, little material is left for research

purposes. However, this does not mean that the lncRNAs we identified do not play a role in DCIS *in vivo*. Indeed, analysis by qRT-PCR of the six lncRNA candidates that we focused on has shown that LOC100134040 is expressed at lower levels in the DCIS cohort. A similar trend is observed for LOC729970 although statistical significance was not reached. We hypothesized that this is due to our small sample number. Genetic diversity is more important when we consider a patient cohort as opposed to different cell lines, particularly considering that ETCC-006 and -010 cells have the same genetic background. With a power calculation based on the difference in LOC729970 expression observed between the tumor and normal group, we estimate that we would need between 126 and 168 samples in each group to see a statistical difference. This corresponds to a sample size twenty-four times bigger than ours. However, recruiting DCIS patients for our study has been challenging. Since patients have to give consent for the study before the surgery, we have been unable to get access to FFPE blocks archived by hospitals, especially in light of General Data Protection Regulation (GDPR) introduced in 2018. As a consequence, we have been working in tight collaboration with Dr. Conleth Murphy at the Bon Secours Hospital and his team that have very successfully recruited DCIS patients. One limitation lies with the hospital itself, since Bon Secours is a private hospital, the turnover in patient number is lower than in a public hospital. A second limitation is with our strict parameters to select patient samples. Indeed, it is not rare for a tumor to show DCIS mixed with IDC; however, we chose to specifically focus on pure DCIS tumors to study DCIS at its earlier stages. Nonetheless, this has forced us to discard DCIS patient samples showing an invasive component from our study. Recruitment of patients is still on-going; in fact, several patients with DCIS have already given consent and their samples will be included in our study shortly.

Until then, we have mentioned in the introduction several transcriptomic studies specific to DCIS. Abba et al. identified 127 antisense lncRNAs and 66 lincRNAs using the polyA fraction to conduct RNAseq (77). In another study, Rao et al. identified 69 lncRNAs with differential expression between DCIS tumor and normal breast using RNA depleted from ribosomal RNA to conduct RNAseq (226). We have highlighted that only five lncRNAs overlapped between those two studies, FAM83H-AS1, WDFY3-AS2, ST8SIA6-AS1, ADIPOQ-AS1 and LINC00861. We compared the overlap between those datasets and our RNAseq study with the DCIS cell lines since our following RNAseq study had shown little overlap across patients. Interestingly, the overlap between our dataset and either Abba's or Rao's dataset was greater than the overlap between Abba and Rao datasets. This might be explained by the method used to perform the RNAseq study. If successful, the lncRNA capture approach is the best method to encompass information about a larger number of lncRNAs. Hence, 42 lncRNAs are common to our dataset and Abba's (TRHDE-AS1, HOTAIR, GATA3-AS1, PCED1B-AS1, NR2F1-AS1, ZBED3-AS1, IQCH-AS1, SLC2A1-AS1, FAM83H-AS1, CEBPA-AS1, PRKG1-AS1, BAIAP2-AS1, FGF14-AS2, LOXL1-AS1, ZNF571-AS1, PSMG3-AS1, PSMG3-AS1, MLLT4-AS1, LINC00277, LINC00578, LINC00426, LINC00885, LINC00478, LINC00173, LINC00939, LINC00271, LINC00899, LINC00514, LINC00842, LINC00202-1, LINC00920, LINC00886, LINC00968, LINC00961, LINC00689, LINC00847, LINC00263, LINC00174, LINC00087, LINC00936, LINC00957, LINC00622 and LINC00152) ; while 10 are common between our study and Rao's (AC005062.2, RP11-834C11.4, RP11-166D19.1, FAM83H-AS1, MIAT, MIR181A1HG, CTB-92J24.3, RP11-392O17.1, AC093642.3 and RP11-624L4.1). Of the 5 common between Abba's and Rao's, FAM83H-AS1, more highly expressed in DCIS, shows lower expression in the

ETCC lines compared to MCF-10A. But overall, the overlap shown between the three studies remains limited and more work to characterize lncRNAs in DCIS will need to be done.

Alternatively, we have considered a novel approach to study lncRNAs. Indeed, in a clinical setting, *in situ* analysis is preferred to a PCR-based test, in consequence, we have been considering using a novel approach in RNA quantification analysis called RNAscope (353). This RNA *in situ* approach allows direct detection of the RNA in FFPE tissue for an accurate quantification. Moreover, it also allows to distinguish the lncRNAs expressed directly in tumor cells as opposed to the surrounding stroma. The RNAscope system is already being used to study lncRNAs in tissue (354,355). Since most of the lncRNAs we identified are uncharacterized, it would require some optimization to find the appropriate probes that allow specific quantification of the signal. This associates with additional time and cost implication. However, this would be an interesting approach to consider for the future. If it could be applied to the 58 lncRNAs with higher expression in the three DCIS cell lines, it might lead to the discovery of novel lncRNAs that can be used clinically for DCIS diagnosis. Moreover, in a study with tissue from both DCIS and IDC patients, those lncRNAs might show differences that could allow us to determine how lncRNAs contribute in the transition towards IDC.

Chapter 5

***In vitro* characterization of the role of LINC00473, LINC00313, LINC00707, MIR210HG, LOC100134040 and LOC729970 in DCIS**

5.1 Introduction

Our study in Chapter 4 allowed us to identify six lncRNA candidates associated with DCIS. We characterized those lncRNAs *in silico* by investigating their regulation through the study of transcription factors binding to the promoter region and epigenetic changes. Those modifications are possibly implicated in their regulation and that could ultimately explain the change in expression observed in DCIS in our study. Epigenetic modifications (CpG island methylation and histone modifications) seem to be important actors in the regulation of most of those lncRNAs candidates. However, we now aim to functionally characterized those candidates in the DCIS context to investigate if they are functionally relevant in DCIS and DCIS progression to IDC.

As described previously, most lncRNAs implicated in cancer, in particular breast cancer, are involved in the regulation of cell proliferation or cell motility. They also

modulate cell response to stress stimuli or resistance to chemotherapy. Hence, in this chapter, we aimed to characterize those lncRNAs further to determine and understand their role in DCIS and DCIS progression to IDC. Using TCGA patient data, we proposed previously that LINC00313, LINC00473, LINC00707, MIR210HG and LOC100134040 could be important for breast cancer progression whereas LOC729970 could have a protective role against breast cancer progression. Interestingly, by investigating the level of expression of those lncRNAs candidates in our cohort of DCIS patient, we found no difference between tumor and normal tissue for all but LOC100134040 and LOC729970, with lower expression in the tumor compared to adjacent normal tissue. Therefore, we aimed to monitor cell behavior including cell proliferation, cell migration and expression of EMT markers after altering the level of expression of the six lncRNAs candidates. Since those processes can contribute to DCIS progression to IDC, we thought that this would allow us to determine if our lncRNAs are involved in DCIS progression to IDC. Moreover, this would give us an indication as to their role in breast cancer progression.

In a second part of this chapter, we focus our research on LOC729970 and the study of its protein interactors. Since lncRNAs function in cells mostly operate through interactions with proteins (356,357), the identification of proteins interactors is one of the first steps in the study of the mechanism of an lncRNA. Here, we choose to concentrate on LOC729970 since we observe lower expression in DCIS (from our DCIS patient cohort) and IDC (TCGA data) compared to normal and that low expression of LOC729970 is associated with better overall survival in IDC (TCGA data). All this could suggest that it has a protective role against breast cancer progression. Using a combination of an RNA-affinity purification and mass

spectrometry, we identified a list of possible protein interactors. Next, we performed a gene ontology analysis to identify specific pathways in which LOC729970 might be implicated. Finally, we aimed to confirm LOC729970 interaction with two specific protein candidates.

5.2 Results

5.2.1 Subcellular localization of the six lncRNA candidates

Some techniques such as RNA fluorescent *in situ* hybridization (RNA-FISH) are used to directly study the localization of a given RNA in cells at the single cell resolution. However, they mostly rely on the identification of a panel of probes specific to the RNA of interest. Therefore, identifying probes for several transcripts can be a lengthy and non-cost-effective process. Hence, we chose to investigate the subcellular localization of our candidates by performing subcellular fractionation in combination with qRT-PCR, a technique largely used for lncRNAs study. The cytoplasmic and nuclear fractions were separated, the nuclear fraction subdivided into a nuclear soluble and chromatin-bound fraction, the last one allowing us to identify lncRNAs interacting with DNA. This was performed using MCF-10A cells for LINC00707 and MDA-MB-231 cells for LINC00473, LINC00313, MIR210HG, LOC100134040 and LOC729970, corresponding to the cell lines showing the highest level of expression of each lncRNA (Figure 4.7 B).

In Figure 5.1 A, we show expression of tubulin α , a cytoplasmic marker, and lamin B1 and histone H3, two nuclear markers, to assess the success of the subcellular

fractionation. While low expression of tubulin α is detected in the nuclear soluble fraction, it is mostly restricted to the cytoplasmic fraction in both MCF-10A and MDA-MB-231. Regarding lamin B1, two isoforms seemed to be detected, with the most expressed isoform restricted to the nuclear insoluble fraction as expected. Similarly, histone H3 expression was detected in the chromatin bound fraction. Taken together, this suggests that while a minor cytoplasmic contamination was observed in the nuclear soluble fraction, the subcellular fractionation was successful.

GAPDH expression was measured as a control of a cytoplasmic mRNA and MALAT1 as a control of nuclear lncRNA (358). We confirm that the expression of *GAPDH* is mostly restricted to the cytoplasm and MALAT1 mostly to the nucleus, as expected (Figure 5.1 B).

Next, we analyzed the relative expression of our lncRNAs candidates in each subcellular fraction. We show that at least 30% of the LINC00707 and LINC00313 total transcripts were present in the cytoplasm and 40% of LINC00473 and MIR210HG total transcripts were present in the cytoplasm. Hence, LINC00707, LINC00313, LINC00473 and MIR210HG transcripts seemed to be mostly localized in the nucleus. The LOC729970 transcript seemed to be present in higher proportion in the cytoplasm, with over 50% of the transcript in the cytoplasmic fraction. LOC100134040 transcript seemed to be mostly localized in the nucleus, with only 10% of the transcript in the cytoplasmic fraction. This suggests that those two lncRNAs might exert their function respectively in the cytoplasm and nucleus.

Since in the previous experiment we examined the expression of our lncRNAs candidates in the different subcellular fractions using DNA oligonucleotides on different exons, both native and spliced transcripts could be detected. To examine specifically the expression of the spliced transcripts in those fractions (Figure 5.1 C) we used DNA oligonucleotides targeting an exon-exon junction. We confirmed that a higher proportion of LOC100134040 is present in the nucleus. In contrast, LOC729970 showed this time a slightly higher proportion in the nucleus. However, taking into account the small cytoplasmic contamination present in the nuclear fraction, we suspect that this is just a byproduct of the contamination. Overall, this could mean that LOC729970 has a cytoplasmic function, although the possibility that it is exported to the cytoplasm to prevent a nuclear function cannot be disregarded.

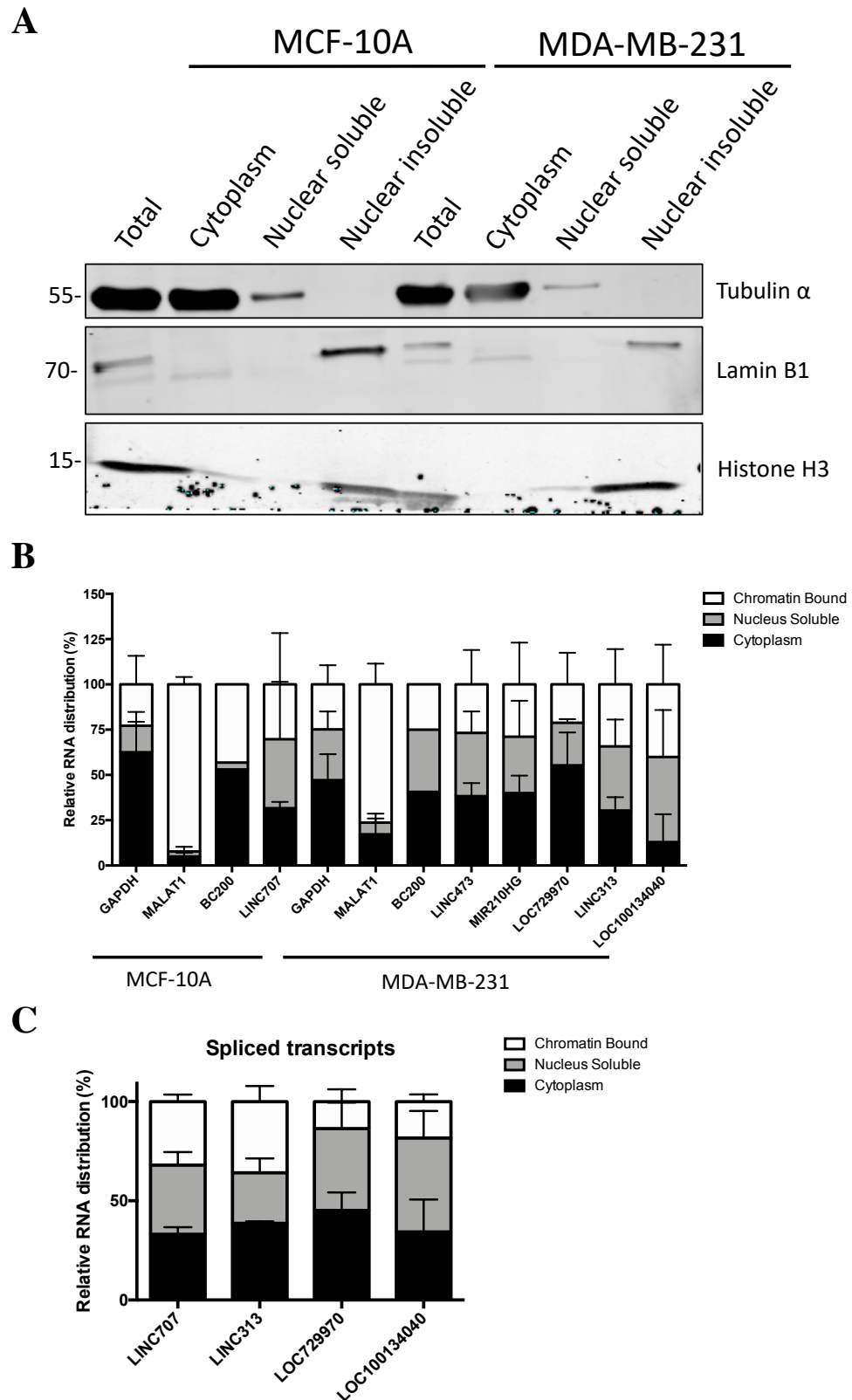


Figure 5.1: Subcellular fractionation of MCF-10A and MDA-MB-231 cells. (A) Assessment of the success of the fractionation by Western Blotting, tubulin alpha is a cytoplasmic marker and lamin B1 and Histone H3 are nuclear markers. (B) qRT-PCR

of the total transcript of the lncRNAs candidates in the cytoplasmic, nuclear soluble and nuclear insoluble fraction. (C) qRT-PCR of the spliced transcript of the lncRNA candidates in the cytoplasmic, nuclear soluble and nuclear insoluble fraction. The qRT-PCR data was plotted on GraphPad v.8.3.0.

5.2.2 Effect of the overexpression of the six lncRNA candidates on breast cancer cells behavior

To functionally characterize our six lncRNA candidates, we chose to alter their expression in breast cancer cells and to analyze how cells respond to that change in terms of proliferation, migration and expression of EMT markers.

The cDNA sequences of LINC00473, LOC729970, MIR210HG, LOC100134040, LINC00707 and LINC00313 were cloned into the pcDNA3 vector under the CMV promoter to induce an overexpression of each respective transcript. Overexpression of LINC00473 and LOC729970 was confirmed in MCF-7 cells, 48 hours after transfection of the construct, by respectively 10 and 45-fold change (Figure 5.2). Since MCF-7 cells show the lowest expression level of LINC00473 and LOC729970, we expected them to show the most dramatic effect in response to the change in expression of those lncRNAs. Similarly, we overexpressed MIR210HG, LOC100134040, LINC00707 and LINC00313 in ETCC-006 cells since those cells show low expression of those lncRNAs. While no overexpression of MIR210HG was detected, a slight increase in LINC00707 expression was observed in ETCC-006. Meanwhile, LINC00313 expression was increased by 30-fold change and LOC100134040 by 60-fold change.

Next, we analyzed the proliferation rate of MCF-7 cells after overexpression of LINC00473 or LOC729970 and of ETCC-006 cells after overexpression of MIR210HG, LOC100134040, LINC00707 or LINC00313. To do this, we performed an MTT assay, a colorimetric assay that assesses cell metabolic activity. Hence, OD reading (at 590 nm) will increase with cell number. These readings were repeated in technical triplicates for each condition across a period of 10 days. Before performing the 10-day experiment, measurements had been taken over a period of three days but since no difference was evidenced (data not shown), longer times points were tested.

Monitoring the effect on cell proliferation for each of our lncRNAs in an overexpression system, we found that LINC00473 overexpression did not affect cell proliferation rate across three biological replicates compared to the empty vector control. LOC729970 overexpression was associated with a slight increase in proliferation at day 7, but it was not statistically significant. Since overall MIR210HG was not successfully overexpressed, no change in proliferation was seen. LOC100134040 overexpression seems to slightly stimulate cells proliferation in the first two biological replicates, but this could not be confirmed in the last replicate. Lastly, LINC00707 and LINC00313 overexpression did not induce a change in cell proliferation. Overall these results suggest that none of our lncRNA candidates induce a consistent and significant change in cell proliferation in our cells.

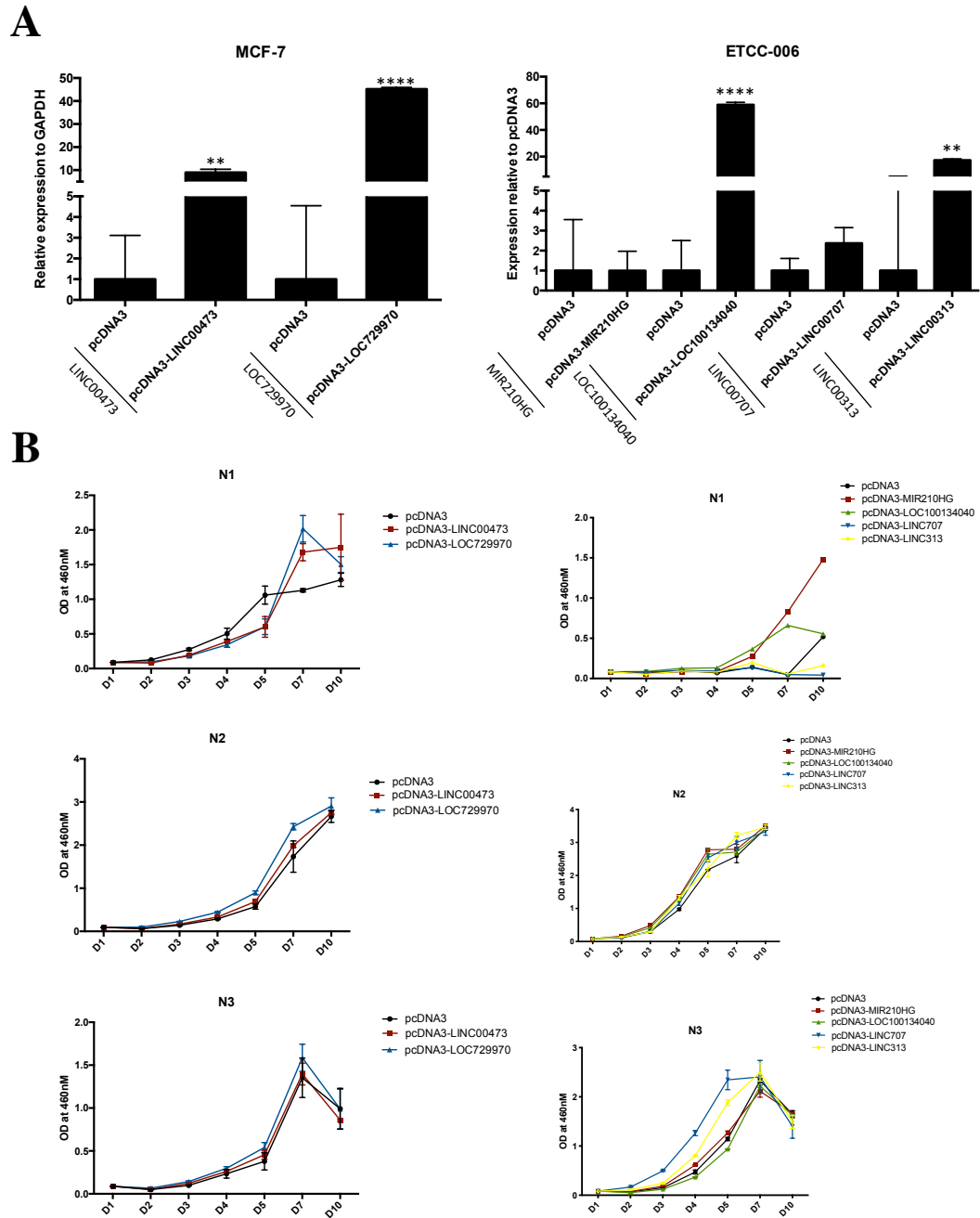


Figure 5.2: Effect of the overexpression of LINC00473, LOC729970, MIR210HG, LOC10013404, LINC00707, LINC00313 on MCF-7 and ETCC-006 cells proliferation. (A) Assessment of the success of the overexpression of lncRNA candidates by qRT-PCR. (B) Analysis of the proliferation of MCF-7 and ETCC-006 cells by MTT assay after overexpression of the lncRNA candidates.

Since no effect on cell proliferation could be seen after overexpression of our lncRNA candidates, we aimed to investigate if migration is affected when we alter the expression of LINC00473, LOC729970, MIR210HG LOC100134040, LINC00707 or LINC00313. Indeed, if migration is altered after overexpression of one of our lncRNA candidates, it could be an indicator that this lncRNA is involved in DCIS progression to IDC. In Figure 5.3, we measured the migration rate of our cells after overexpression of our lncRNA candidates by wound healing assay. While the first biological replicate of LINC00473 and LOC729970 overexpression suggests that those lncRNAs induce a delay in migration with wound closure being slower compared to the control, this could not be replicated. Indeed, the second replicate revealed that cells that overexpress LINC00473 or LOC729970 show similar proliferative rates than those observed in the control. Overexpression of LOC100134040, LINC00707 and LINC00313 did not seem to induce a change in cell motility compared to the empty vector control.

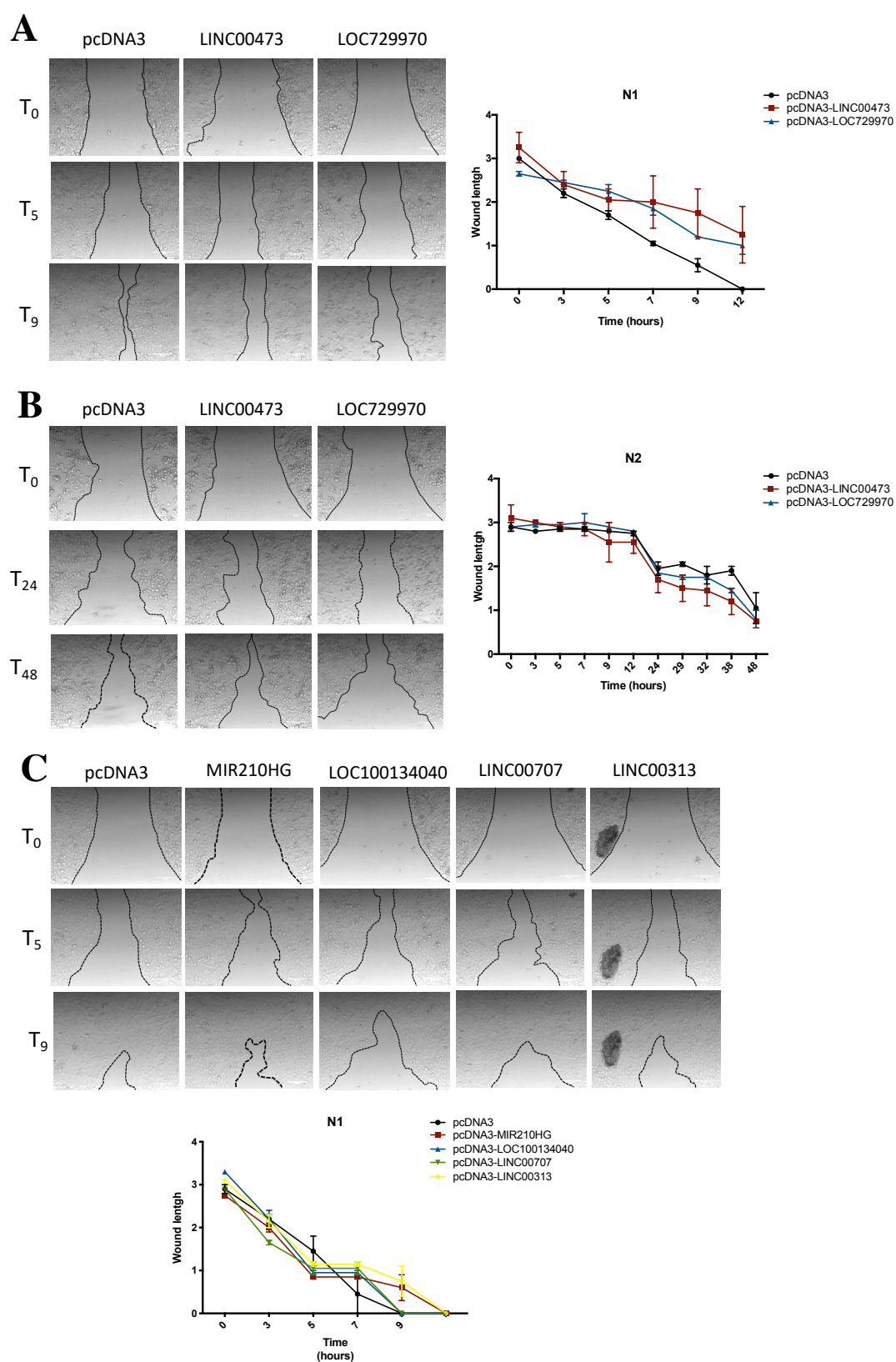
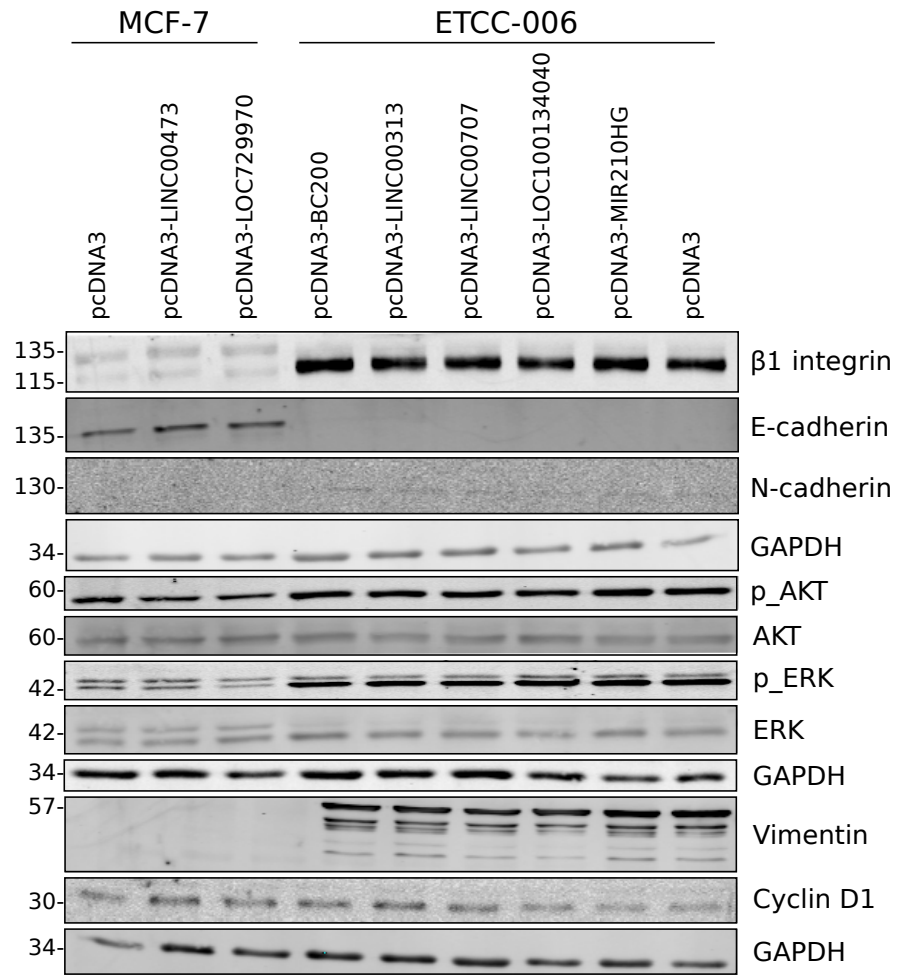


Figure 5.3: Migration of MCF-7 and ETCC-006 cells after overexpression of the lncRNA candidates. (A) and (B) Wound healing assay of MCF-7 cells after

LINC00473 and LOC729970 overexpression. (B) Wound healing assay of ETCC-006 cells after overexpression of MIR210HG, LOC100134040, LINC00707, LINC00313. Images were taken using bright field microscopy. Scale bar represents 50 μ m.

Since we analyzed the expression of EMT markers in our DCIS cell lines characterization in Chapter 3, we chose to examine EMT markers after overexpression of our lncRNA candidates (Figure 5.4). We also overexpressed lncRNA BC200 since it is known to affect EMT in colon cancer (359). No consistent change was observed for β 1 integrin, E-cadherin, N-cadherin, ERK and its active phosphorylated form, Vimentin and Cyclin D1. We noticed, however that active phosphorylated AKT might be slightly reduced after overexpression of LOC729970.

A



B

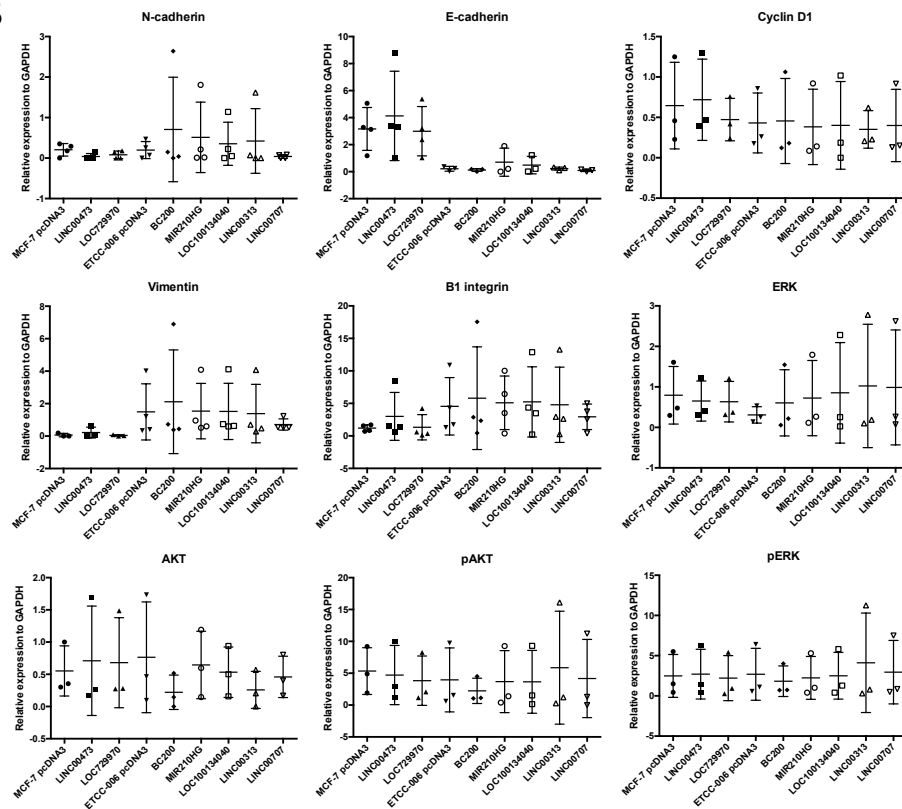


Figure 5.4: Expression of EMT markers after overexpression of the lncRNA candidates. (A) Western blotting analysis of EMT markers after overexpression of the lncRNA candidates in MCF-7 and ETCC-006 cells. (B) Densitometry quantification of EMT markers after overexpression of the lncRNA candidates was done using ImageJ.

We next investigated how LOC729970 might be responsible for the small change in phosphorylated AKT levels. Since we could not identify a direct link between AKT and LOC729970, we examined if LOC729970 might be acting as a competing endogenous RNA (ceRNA) for a miRNA either directly targeting AKT mRNA or a regulator of AKT. We identified miR-129-1-3p and miR-129-2-3p that have been shown to bind to LOC729970 in an Argonaute-2 photoactivatable-ribonucleoside-enhanced crosslinking (AGO2 PAR-CLIP) experiment (Figure 5.5 A and 5.5 B), PAR-CLIP is a biochemical technique that allows the identification of motifs that miRNA bind to with their target RNAs. Next, we investigated the molecular targets of miR-129-1-3p and miR-129-2-3p in the database StarBase (360). While some targets are implicated in the regulation of cell proliferation and cell motility, we could not identify an association with AKT. In parallel, we investigated miR-129-5p and identified it as a potential miRNA targeting PTEN, a negative regulator of AKT (Figure 5.5 C). To support an association between miR-129, LOC729970 and PTEN, we looked at their co-expression in an IDC patient cohort (Figure 5.5 D). No correlation was evidenced between miR-129-1-3p/miR-129-2-3p/miR-129-5p and LOC729970, despite a weak positive correlation evidenced between LOC729970 and PTEN. To further test the link between PTEN and LOC729970, we examined PTEN expression after LOC729970 overexpression, but no consistent change was evidenced (Figure 5.5 E).

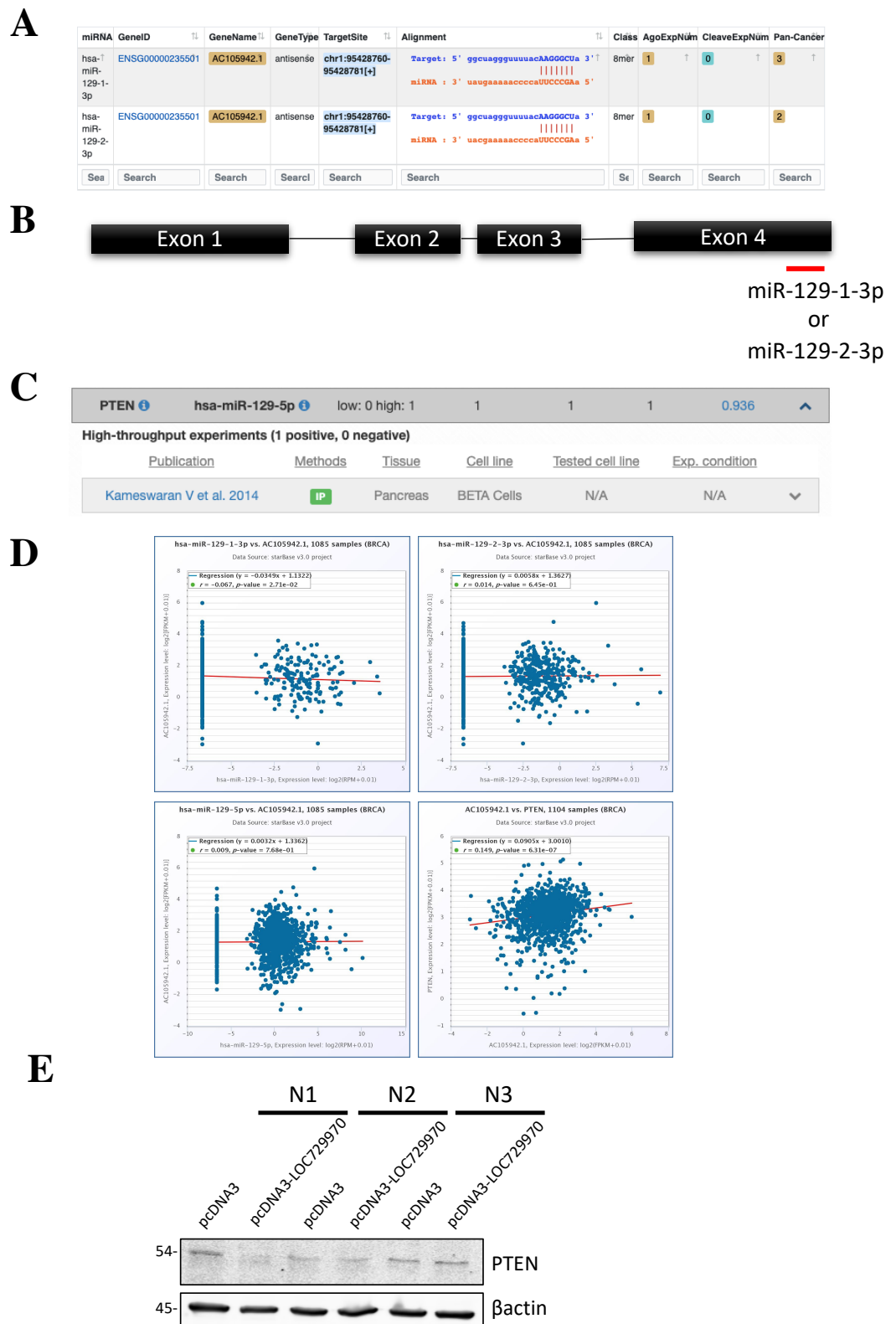


Figure 5.5: Analysis of LOC729970 as a competing endogenous RNA. (A) Bioinformatic determination of miRNAs that bind to LOC729970 in the database StarBase. (B) Determination of the binding site of miR-129-1-3p and miR-129-2-3p

to LOC729970. (C) Investigation of miR-129 mRNA targets in the database TarBase v.8. (D) Co-expression of LOC729970 and miR-129-1-3p, miR-129-2-3p, miR-129-5p and PTEN in an invasive breast cancer cohort of TCGA. (E) Protein level of PTEN analyzed by Western blotting after LOC729970 overexpression in MCF-7 cells.

5.2.3 Knockdown of the six lncRNA candidates using siRNAs

In parallel of the previous overexpression experiments, we also aimed to knockdown the expression of the six lncRNA candidates and monitor cell behavior. Two siRNAs targeting different regions of the transcript of each lncRNA were designed, synthesized and tested at 50 nM. A first round of experiments was performed with treatment of the siRNAs at 24h, 48h and 72h in MCF-7 cells (data not shown). However, those siRNAs were highly inefficient, as evidenced 48h post transfection (Figure 5.6 A). Next, we focused on the effect of siRNAs for a given lncRNA in the cell line that showed highest expression of that lncRNA (Chapter 4, Figure 4.7 B). Hence, siRNAs targeting LINC00707 were tested in MCF-10A cells and siRNAs targeting LINC00473 and LOC729970 were tested in ETCC-006 cells. Disappointingly, the siRNA knockdowns were inefficient in those cell lines, as they had been in MCF-7 cells (Figure 5.6 B). Because the siRNA treatment at 50 nM did not reduce the level of the lncRNAs, different concentrations of siRNAs were tested (Figure 5.6 C). Indeed, while 50nM would be considered a standard concentration for knockdown using siRNAs, knockdown of lncRNAs are known to be particularly difficult in comparison to mRNA knockdowns, especially for nuclear lncRNAs (361). We thus tested several concentrations of siRNAs targeting LOC729970 in ETCC-006

cells, but none showed a consistent knockdown. Similar results were observed with MIR210HG and LOC100134040 (Figure 5.6 C and 5.6 D).

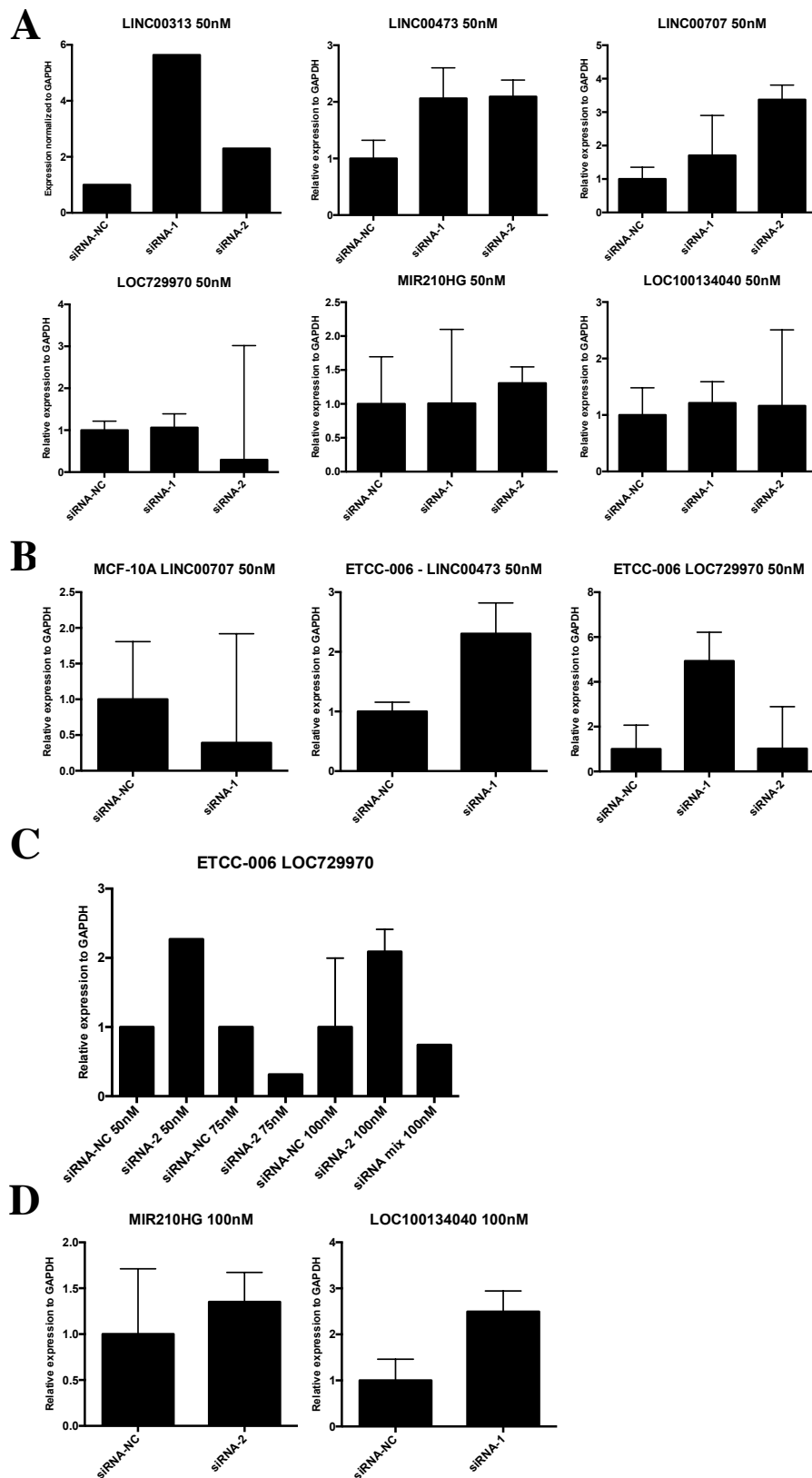


Figure 5.6: Analysis of the knockdown of the lncRNA candidates using siRNAs by qRT-PCR. (A) Test of the siRNAs targeting the lncRNA candidates at 50 nM in

MCF-7 cells. (B) Test of the siRNAs targeting LINC00707 in MCF-10A cells, LINC00473 and LOC729970 in ETCC-006 cells at 50 nM. (C) Test of the siRNA-2 targeting LOC729970 in ETCC-006 cells at 50 nM, 75 nM and 100 nM. (D) Test of the most efficient siRNAs targeting MIR210HG and LOC100134040 in MCF-7 cells at 100 nM. The qRT-PCR data was normalized to GAPDH. Statistical analysis was done using one-way ANOVA (GraphPad v.8.3.0). **** p-value < 0.0001; *** p-value < 0.001; ** p-value < 0.01; * p-value < 0.05.

Since an efficient knockdown of the lncRNA candidates could not be achieved using siRNAs, we decided to use a different approach. However, we could not carry out those detailed analyses for all six different lncRNAs. Since our results suggested that LOC729970 could possibly have a protective role against breast cancer progression, we decided to focus on it for further studies.

5.2.4 Effect of LOC729970 knockdown on breast cancer cell behavior using antisense oligonucleotides (ASOs)

Due to the lack of successful knockdown of our lncRNA candidates using an siRNA-directed approach, we examined antisense oligonucleotides (ASO) (361) as an alternative approach to knockdown LOC729970 expression. ASOs are single-stranded DNA oligonucleotides that target a small region of 18 to 25 nucleotides of an RNA by base complementarity, similar to siRNAs. However, the degradation of the RNA target is then induced through recruitment of RNase H1, which recognizes and digests RNA-DNA hybridized regions (362). Interestingly, ASOs have been shown to be highly efficient against nuclear transcripts that cannot be targeted by the RISC degradation

machinery. Moreover, ASOs are capable of inducing the degradation of nascent transcripts directly in the nucleus, but they can also be efficient against cytoplasmic transcripts (362).

Four ASOs targeting different regions of LOC729970 were designed, the first two corresponding to the same sequence targeted by the siRNAs. ASOs were first tested at a concentration of 50 nM but no knockdown was observed (data not shown). Higher concentrations were tested and ASO1 and ASO2 seemed able to induce a LOC729970 reduction at a concentration between 150 nM and 250 nM (Figure 5.7 A). Next, we tested an intermediate concentration of 200 nM (Figure 5.7 B) and confirmed an efficient knockdown with ASO1 and ASO2. All further experiments were conducted using those conditions.

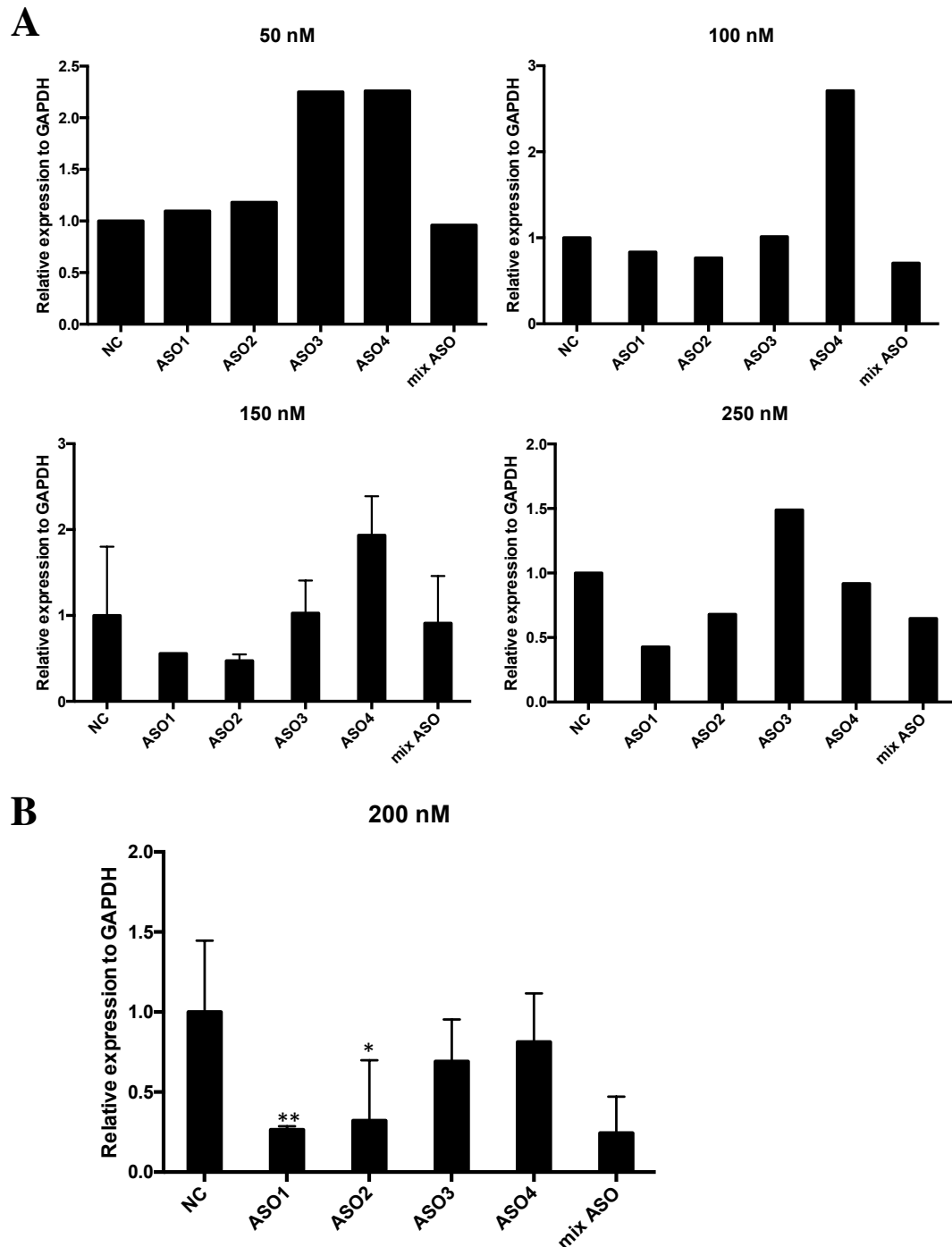


Figure 5.7: Analysis of antisense oligonucleotides targeting LOC7229970 in ETCC-006 cells by qRT-PCR. (A) Test of ASOs targeting LOC729970 in ETCC-006 cells at 50 nM, 100 nM, 150 nM and 250 nM. (B) Knockdown of LOC729970 in ETCC-006 cells using ASOs at 200 nM. The qRT-PCR data was normalized to

GAPDH. Statistical analysis was done using one-way ANOVA (GraphPad v.8.3.0).

**** p-value < 0.0001; *** p-value < 0.001; ** p-value < 0.01; * p-value < 0.05.

Previously, we showed no global effect on cell proliferation or migration after overexpression of LOC729970. However, we do not know how much LOC729970 transcript is enough to induce a change in cell proliferation or migration, if LOC729970 has a role in regulating those processes. Hence, not seeing an effect on cell proliferation or migration after overexpressing LOC729970 does not necessarily mean that LOC729970 has no function in their regulation. To determine if LOC729970 has any role in cell proliferation and motility, we assessed those parameters after LOC729970 knockdown by MTT assay and wound healing assay. Similar to what was observed after overexpression, no consistent change in proliferation (Figure 5.8 A) and no change in the migration rate (Figure 5.8 B) was observed after LOC729970 knockdown using ASOs.

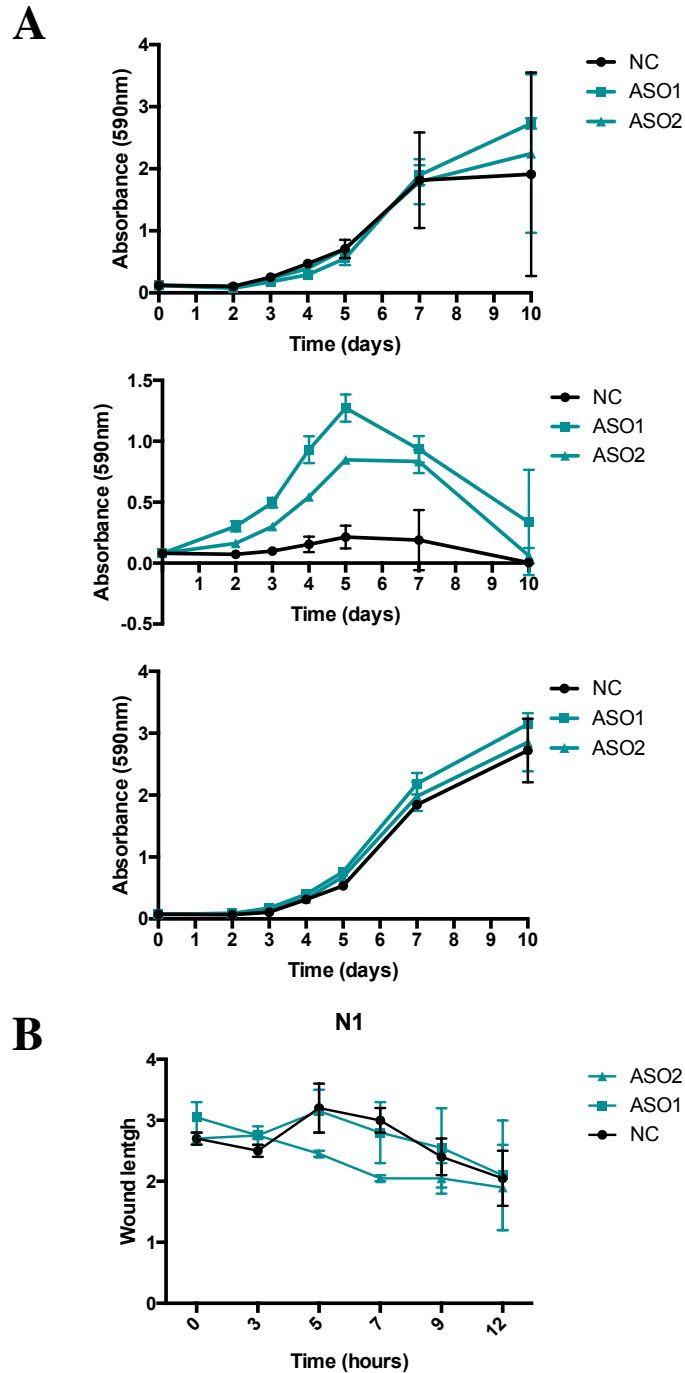


Figure 5.8: Analysis of cell proliferation and migration of ETCC-006 cells after LOC729970 knockdown with ASOs. (A) Proliferation of ETCC-006 cells after LOC729970 knockdown. (B) Migration of ETCC-006 cells after LOC729970 knockdown. The wound length is measured in an arbitrary unit using ImageJ.

Since we could not make any firm conclusions on the association between LOC729970 and phosphorylated AKT, we analyzed AKT phosphorylation status after LOC729970 knockdown (Figure 5.9). We observed a slight increase of p-AKT after LOC729970 knockdown with ASO2, supporting our previous observations. However, since that change is only seen with ASO2, which displayed the weakest knockdown of LOC729970, those results will have to be further investigated in the future.

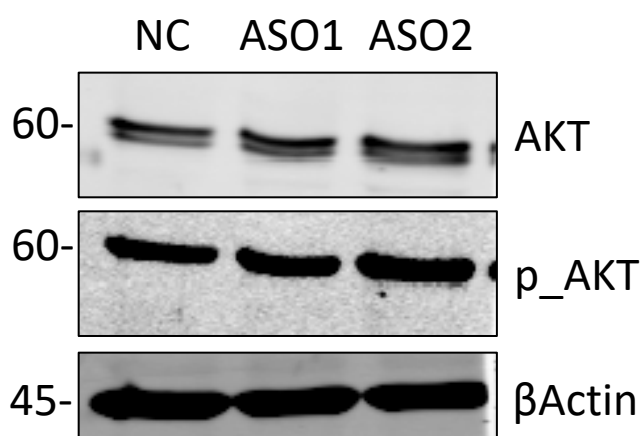


Figure 5.9: Protein level of AKT and p-AKT measured by Western blotting after LOC729970 knockdown using ASOs. For the immunoblot, 40 µg of total protein was loaded per lane onto the gel.

5.2.5 Expression of the six lncRNA candidates under stress conditions

The transcription and resulting levels of lncRNAs can be induced under stress conditions (363). This allows the activation of a cellular response for cells to adapt to the stressful factor. Since the six lncRNA candidates do not seem to be involved in the control of cell proliferation, cell motility or the regulation of EMT markers, we investigated if they might be important in stress response. First, we compared expression of the lncRNA candidates under low (50%) and high confluency conditions

(100%). Cells were harvested 24 hours after plating and lncRNA expression was quantified by qRT-PCR (Figure 5.10 A). Only LOC100134040 seemed to be induced in high confluency conditions. LINC00707 expression seemed to be completely downregulated.

Under serum starvation, we previously observed a loss of p-AKT due to a reduction in RTK signaling (Figure 3.9). ETCC-006 cells were put under serum starvation conditions for 24 hours and expression of the six candidates was compared to the complete media condition. Following harvesting, p-AKT levels were compared in complete media and serum starved conditions. In Figure 5.10 B, we show loss of phosphorylated AKT, indicating that the serum starvation was successful. After performing the qRT-PCR, relative expression of the lncRNA candidates was calculated compared to actin and UBC (ubiquitin C). Actin and UBC were used as reference genes because GAPDH expression changes in starvation conditions, which could lead to a false statistically significant upregulation of LOC729970, MIR210HG and LINC00313 (364). Indeed, when compared to actin or UBC, the six lncRNAs, while showing a trend suggestive of an upregulation in serum starvation conditions, did not reach significance. LINC00707 expression was downregulated under serum starvation conditions. This supports our observation in high confluency conditions.

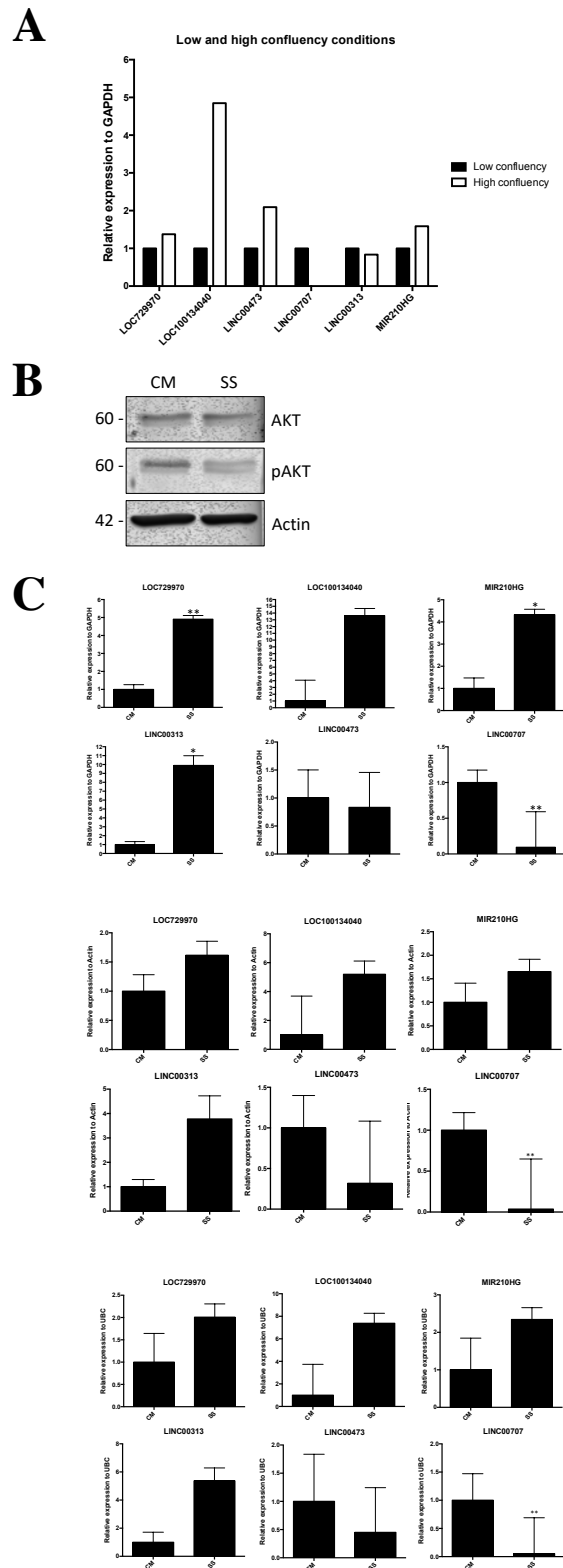
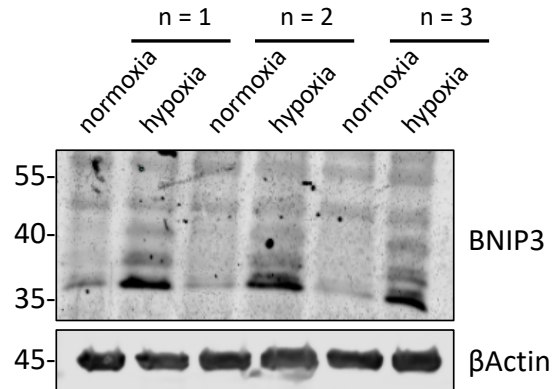


Figure 5.10: Expression of the six lncRNA candidates in stress conditions. (A) Comparison of the expression of the six lncRNA candidates in high and low confluency conditions by qRT-PCR (n = 1) (B) Western blotting of AKT and p-AKT

levels in complete media and serum starved ETCC-006 cells (n = 3). (C) Comparison of the expression of the six lncRNA candidates in complete media and serum starved conditions by qRT-PCR. The data was normalized against GAPDH, β actin and UBC (n = 3).

Next, we analyzed expression of the six candidates under hypoxic conditions. Indeed, hypoxia condition is a key regulator of cancer progression in tumors. A class of lncRNAs induced under hypoxic conditions has started to emerge recently (365). For example, MIR210HG has been shown to be induced under those conditions (343). BNIP3 (BCL2 interacting protein 3) is induced under hypoxic conditions and is used as a control (366). BNIP3 levels were increased after 24 hours in hypoxia conditions (Figure 5.11 A) so we analyzed the expression of the six candidates in those samples by qRT-PCR (Figure 5.11 B). Relative expression was calculated compared to actin and UBC. We confirmed an increase in MIR210HG under hypoxic conditions relative to actin expression; while LOC72970, LINC00473 and LINC00707 (relative to UBC) levels were reduced under the same conditions. LOC100134040 was slightly reduced in hypoxia when expression was analyzed in relation to UBC. On the opposite, we observed higher expression of LINC00313 in hypoxic conditions. This might suggest that LINC00313 is important for the cells to respond to hypoxia.

A



B

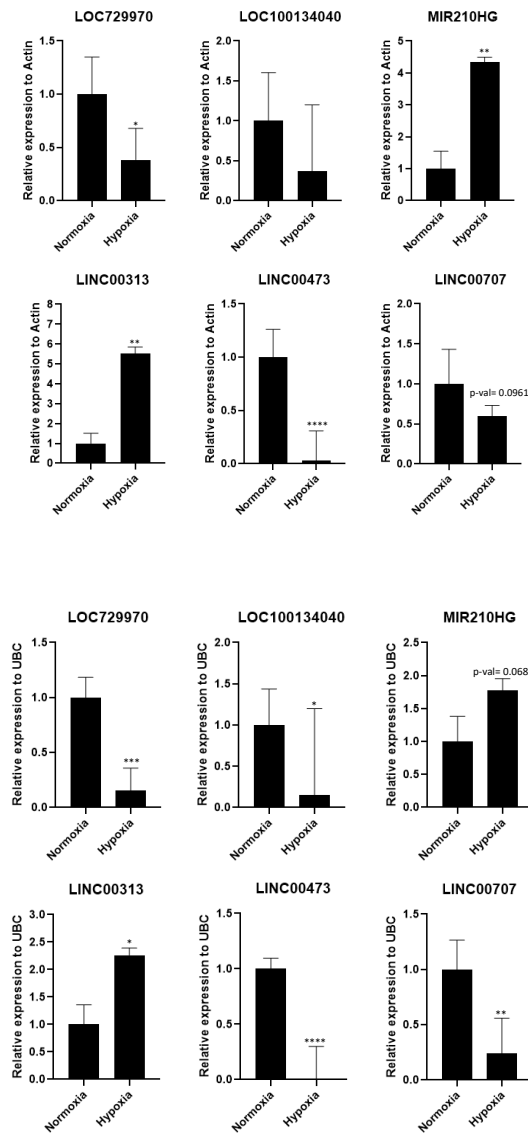


Figure 5.11: Expression of the six lncRNA candidates in hypoxic conditions. (A)

Western blotting comparing BNIP3 levels in normoxia (about 20% oxygen) and

hypoxia conditions (under 5% oxygen). (B) Comparison of the expression of the six lncRNA candidates in normal and hypoxic conditions by qRT-PCR (n = 3).

5.2.6 Identification of LOC729970 protein interactors

We have shown that changes in LOC729970 did not seem to have a global effect on cell proliferation or cell motility (Figure 5.2, 5.3 and 5.8). Moreover, it does not seem to regulate the expression of EMT markers. However, we have seen that LOC729970 expression is lower in IBC compared to normal (Figure 4.6) and that it also seems to be lower in DCIS compared to normal, although this remains to be formally shown (Figure 4.24). Moreover, we have seen that in IBC, lower expression of LOC729970 correlates with better overall survival which suggests that it might have a protective role against progression to IBC. Since we had no evidence of a functional role for LOC729970, we decided to investigate the function in which it could be implicated by identifying its protein interactors. This would give us an indication of the mechanism in which it is involved. To do this, we performed an RNA-affinity purification using the tRSA system (Chapter 2, Figure 2.1), followed by mass spectrometry. LOC729970 and its antisense version were synthesized in an *in vitro* transcription reaction, then bound to streptavidin beads. ETCC-006 cell lysates were prepared and incubated with the beads to allow binding of proteins to the RNA. The antisense version of LOC729970 was used as a negative control to identify proteins binding specifically to LOC729970.

Figure 5.12 A shows an image of a urea SDS-polyacrylamide gel. About 15 µg of RNA was used to analyze the binding of LOC729970 and LOC729970 antisense to

the streptavidin beads. The RNA was detected using SYBR gold nucleic acid stain, a stain 25 to 100 times more sensitive than ethidium bromide. While single bands corresponding to the synthesized RNAs are hard to pinpoint, we can see material in the “bound to beads” lane suggestive that some RNA was bound to the beads. RNA-bound beads were then incubated with ETCC-006 protein lysate to allow protein binding to LOC729970 (sense/antisense) lncRNA. The lncRNA/protein complexes were released from the beads by addition of 2X SDS loading buffer and incubated at 95°C for 10 minutes. The presence of protein in the sample was verified by running an SDS-PAGE gel followed by silver staining. Figure 5.12.B is a silver stain gel of the RNA-affinity purification, showing that some protein was bound with both LOC729970 and its antisense version. This was repeated in triplicate and RNA-bound protein samples (six in total) were sent for mass spectrometry analysis for protein identification at the Center of Excellence for Mass Spectrometry in King’s College, London. Overall, 194 proteins were identified with 94 proteins binding specifically to LOC729970 (48.5%), 32 proteins binding specifically to its antisense version (16.5%) and 68 proteins binding both LOC729970 and its antisense version. (35%) The list of proteins binding specifically to LOC729970 are presented in Table 5.1.

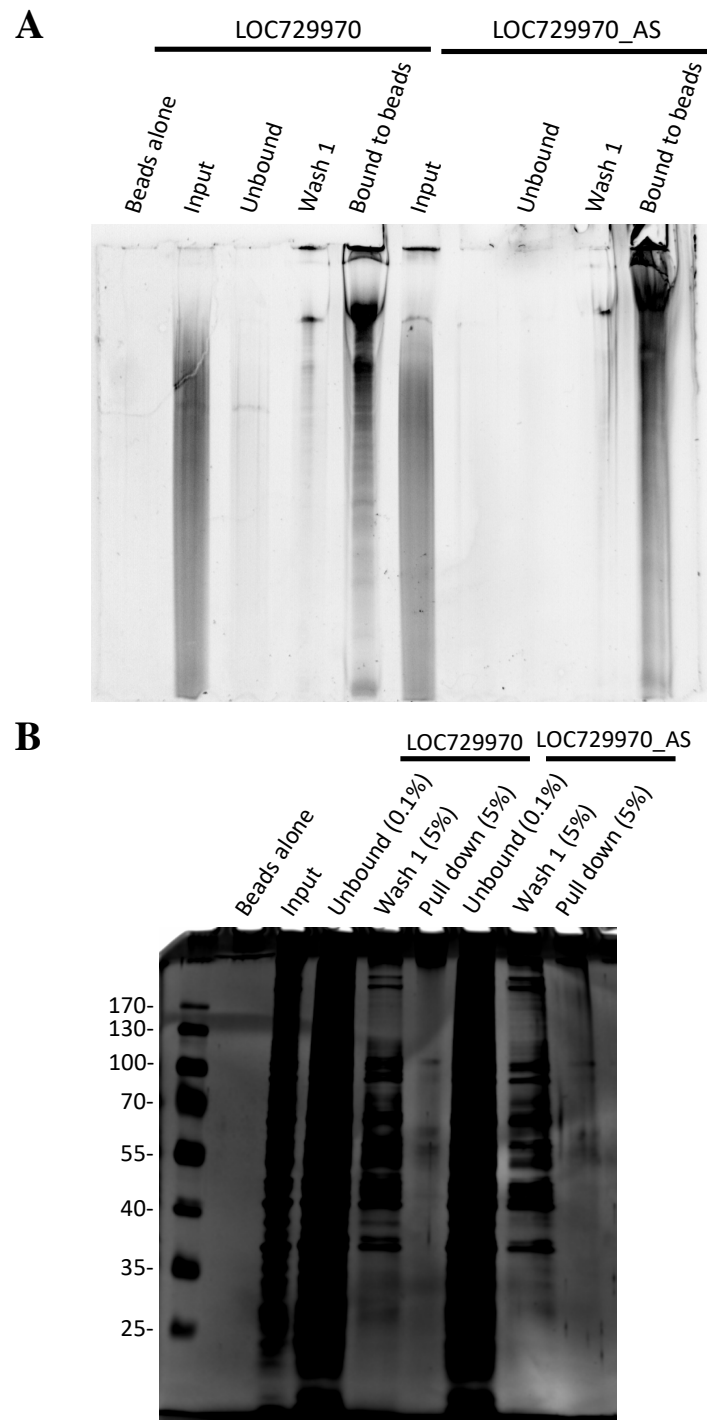


Figure 5.12: Controls to monitor the successful binding of the RNA to the beads and binding of proteins. (A) Verification of the binding of the LOC729970 (sense/antisense) RNA to the streptavidin beads in an urea-SDS PAGE gel followed by staining with SYBG gold. (B) Verification of the presence of proteins in the RNA-affinity purification samples by SDS-PAGE gel followed by silver stain.

Identified Proteins	Accession Number	Alternate ID	Molecular Weight	T-Test (p-value) * (p < 0.00026)	Taxonomy	Sense			Antisense		
60S ribosomal protein L22	RL22_HUMAN	RPL22	15 kDa	0,0022	Homo sapiens	n = 1	n = 2	n = 3	n = 1	n = 2	n = 3
Glyceraldehyde-3-phosphate dehydrogenase	GAP_HUMAN	GAPDH	36 kDa	0,0075	Homo sapiens	1	2	2	0	0	0
Protein CYP61	CYP61_HUMAN	CYP61	42 kDa	0,041	unknown	4	7	1	0	0	0
ATP-dependent RNA helicase DDX18	DDX18_HUMAN	DDX18	75 kDa	0,12	Homo sapiens	0	1	2	0	0	0
Fibroblast growth factor 2	FGF2_HUMAN	FGF2	31 kDa	0,12	Homo sapiens	0	2	0	0	0	0
Lysine-rich nucleolar protein 1	KNOP1_HUMAN	KNOP1	52 kDa	0,12	Homo sapiens	1	1	1	0	0	0
Microtubule-actin cross-linking factor 1, isoforms 1/2/3/5	MACF1_HUMAN	MACF1	838 kDa	0,12	Homo sapiens	0	1	0	0	0	0
Probable 28S rRNA (cytosine(4447)-C(5))-methyltransferase	NOP2_HUMAN	NOP2	89 kDa	0,12	Homo sapiens	2	2	0	0	0	0
Interferon-inducible double-stranded RNA-dependent protein kinase activator A	PRKRA_HUMAN	PRKRA	34 kDa	0,12	Homo sapiens	1	1	1	0	0	0
UV excision repair protein RAD23 homolog B	RD23B_HUMAN	RAD23B	43 kDa	0,12	Homo sapiens	0	1	2	0	0	0
60S ribosomal protein L11	RL11_HUMAN	RPL11	20 kDa	0,12	Homo sapiens	0	2	1	0	0	0
Heterogeneous nuclear ribonucleoprotein A/B	ROAA_HUMAN	HNRNPAB	36 kDa	0,12	Homo sapiens	0	1	1	0	0	0
40S ribosomal protein S30	RS30_HUMAN	FAU	7 kDa	0,12	Homo sapiens	0	1	2	0	0	0
Small nuclear ribonucleoprotein F	RUXF_HUMAN	SNRPF	10 kDa	0,12	Homo sapiens	0	2	1	0	0	0
Small nuclear ribonucleoprotein Sm D2	SMD2_HUMAN	SNRPD2	14 kDa	0,12	Homo sapiens	0	1	0	0	0	0
SLT-ROBO Rho GTPase-activating protein 1	SRGP1_HUMAN	SRGAP1	124 kDa	0,12	Homo sapiens	1	1	0	0	0	0
Cluster of DNA topoisomerase 1	TOP1_HUMAN	TOP1	91 kDa	0,13	Homo sapiens	2	3	0	0	0	0
Histone H1X	H1X_HUMAN	H1FX	22 kDa	0,16	Homo sapiens	1	2	1	0	0	0
Replication factor C subunit 3	RFC3_HUMAN	RFC3	41 kDa	0,16	Homo sapiens	0	2	2	0	0	0
60S ribosomal protein L13	RL13_HUMAN	RPL13	24 kDa	0,16	Homo sapiens	0	1	2	0	0	0
60S ribosomal protein L24	RL24_HUMAN	RPL24	18 kDa	0,16	Homo sapiens	0	1	2	0	0	0
40S ribosomal protein S9	RS9_HUMAN	RPS9	23 kDa	0,16	Homo sapiens	0	1	0	0	0	0
RNA transcription, translation and transport factor protein	RTRAF_HUMAN	RTRAF	28 kDa	0,16	Homo sapiens	1	2	2	0	0	0
Activated RNA polymerase II transcriptional coactivator p15	TCP4_HUMAN	SUB1	14 kDa	0,16	Homo sapiens	0	1	1	0	0	0
Nuclear fragile X mental retardation-interacting protein 2	NUFIP2_HUMAN	NUFIP2	76 kDa	0,21	Homo sapiens	0	3	0	0	0	0
E3 ubiquitin-protein ligase TRIP12	TRIP1_HUMAN	TRIP12	220 kDa	0,21	Homo sapiens	1	3	0	0	0	0

Table 5.1: List of proteins binding specifically to LOC729970, identified by mass spectrometry.

Using the list of specific-LOC729970 protein interactors, we conducted a Gene Ontology (GO) analysis that is presented in Table 5.2. Interestingly, we observe a fold enrichment over 100 for supercoiled DNA binding proteins, a strong enrichment of translation associated proteins and a 4.5 enrichment for cell adhesion molecule binding. This suggests that LOC729970 might be involved in the regulation of replication or transcription or that it might be translated into a protein. Indeed, we have mentioned previously that some lncRNAs are coding for peptides or small proteins. Since LOC729970 was a complete novel lncRNA, we decided to use LNCipedia, a database specific to lncRNAs. Indeed, LOC729970 is characterized as a lncRNA with a CPAT (coding potential assessment tool) of 12.81% in that database. CPAT uses an alignment-free logistic regression model to recognize coding from non-coding transcripts (367). It integrates the length of the open reading frame (ORF), the longer it is the more likely the transcript is coding; the ORF coverage, ORF from mRNAs showing higher coverage than ORF from lncRNAs; the Fickett TESTCODE score that distinguishes mRNAs from non-coding RNAs based on the combinational effect of nucleotide composition and codon usage bias. Last criteria is the relative degree of hexamer usage bias, a positive value associating with mRNAs whereas a negative value associates with lncRNAs. Using GWIPS-Viz, a Genome Browser allowing the direct visualization of ribosome profiling data (368), we noticed ribosome occupation on the LOC729970 transcript in MDA-MB-231 cells (369) (data not shown). However, this does not necessarily mean that LOC729970 is translated as some lncRNAs are sent to ribosomes to be degraded (169).

Homo sapiens	Reference	Input	Expected	Fold Enrichment	+/-	raw P value	FDR
supercoiled DNA binding	5	3	.02	> 100	+	4.72E-06	1.58E-03
nucleic acid binding	4116	72	18.43	3.91	+	2.43E-32	5.67E-29
organic cyclic compound binding	6133	75	27.46	2.73	+	6.06E-24	7.08E-21
binding	15183	91	67.97	1.34	+	8.11E-10	4.21E-07
heterocyclic compound binding	6045	75	27.06	2.77	+	2.30E-24	3.59E-21
large ribosomal subunit rRNA binding	6	3	.03	> 100	+	7.06E-06	2.06E-03
rRNA binding	63	10	.28	35.45	+	9.34E-13	6.23E-10
RNA binding	1673	64	7.49	8.54	+	3.83E-47	1.79E-43
small ribosomal subunit rRNA binding	11	4	.05	81.22	+	4.86E-07	2.07E-04
structural constituent of ribosome	170	19	.76	24.96	+	9.96E-21	9.30E-18
structural molecule activity	739	28	3.31	8.46	+	2.09E-18	1.63E-15
mRNA 3'-UTR binding	91	6	.41	14.73	+	4.76E-06	1.48E-03
mRNA binding	257	15	1.15	13.04	+	1.16E-12	6.74E-10
single-stranded RNA binding	91	5	.41	12.27	+	7.01E-05	1.82E-02
translation regulator activity	151	6	.68	8.88	+	7.34E-05	1.80E-02
cell adhesion molecule binding	495	10	2.22	4.51	+	8.48E-05	1.98E-02
protein binding	11958	78	53.54	1.46	+	1.56E-07	7.28E-05
protein-containing complex binding	1140	17	5.10	3.33	+	1.25E-05	3.45E-03
Unclassified	3362	1	15.05	.07	-	2.88E-06	1.12E-03

Table 5.2: Gene ontology analysis of the proteins binding specifically to LOC729970.

The GO results were confirmed by the PANTHER analysis of the class of proteins present on the list (Table 5.3), indicating that the proteins are mostly translation-related proteins and DNA-binding proteins. In Table 5.4, we present the results of a reactome pathway analysis, suggesting that LOC729970 is mostly associated with translation-related processes.

PANTHER Protein Class	Reference	Input	Expected	Fold Enrichment	+/-	raw P value	FDR
ribosomal protein	160	17	.72	23.73	+	2.80E-18	2.01E-16
RNA binding protein	632	32	2.83	11.31	+	8.75E-25	1.88E-22
nucleic acid binding	1567	41	7.02	5.84	+	1.28E-21	1.38E-19
ribonucleoprotein	47	4	.21	19.01	+	7.88E-05	2.82E-03
mRNA splicing factor	111	7	.50	14.09	+	9.82E-07	4.22E-05
mRNA processing factor	153	8	.68	11.68	+	6.18E-07	3.32E-05
DNA binding protein	457	9	2.05	4.40	+	2.35E-04	7.21E-03
Unclassified	12714	40	56.92	.70	-	4.52E-04	1.22E-02

Table 5.3: PANTHER protein class analysis of the proteins binding specifically to LOC729970.

Reactome pathways	Reference	Input	Expected	Fold Enrichment	+/-	raw P value	FDR
Viral mRNA Translation	89	16	.40	40.15	+	1.45E-20	1.06E-17
Influenza Viral RNA Transcription and Replication	131	16	.59	27.28	+	3.89E-18	5.01E-16
Influenza Life Cycle	140	16	.63	25.53	+	1.02E-17	1.25E-15
Influenza infection	151	16	.68	23.67	+	3.10E-17	3.57E-15
Infectious disease	370	16	1.66	9.66	+	1.46E-11	1.14E-09
Peptide chain elongation	89	16	.40	40.15	+	1.45E-20	7.92E-18
Eukaryotic Translation Elongation	94	16	.42	38.02	+	3.17E-20	9.92E-18
Translation	293	18	1.31	13.72	+	2.18E-15	1.91E-13
Selenocysteine synthesis	93	16	.42	38.43	+	2.72E-20	1.19E-17
Selenoamino acid metabolism	116	16	.52	30.81	+	6.62E-19	1.04E-16
Metabolism of amino acids and derivatives	365	16	1.63	9.79	+	1.20E-11	9.73E-10
Eukaryotic Translation Termination	93	16	.42	38.43	+	2.72E-20	9.92E-18
Nonsense Mediated Decay (NMD) independent of the Exon Junction Complex (EJC)	95	16	.43	37.62	+	3.69E-20	1.01E-17
Nonsense-Mediated Decay (NMD)	115	16	.51	31.08	+	5.84E-19	1.07E-16
Formation of a pool of free 40S subunits	101	16	.45	35.38	+	8.92E-20	2.17E-17
Cap-dependent Translation Initiation	119	16	.53	30.03	+	9.59E-19	1.31E-16
Eukaryotic Translation Initiation	119	16	.53	30.03	+	9.59E-19	1.40E-16
SRP-dependent cotranslational protein targeting to membrane	112	17	.50	33.90	+	1.12E-20	1.23E-17
L13a-mediated translational silencing of Centrioplasmin expression	111	16	.50	32.20	+	3.49E-19	7.65E-17
GTP hydrolysis and joining of the 60S ribosomal subunit	112	16	.50	31.91	+	3.98E-19	7.92E-17
Nonsense Mediated Decay (NMD) enhanced by the Exon Junction Complex (EJC)	115	16	.51	31.08	+	5.84E-19	9.84E-17
Formation of the ternary complex, and subsequently, the 43S complex	51	6	.23	26.28	+	2.02E-07	1.34E-05
Ribosomal scanning and start codon recognition	58	6	.26	23.11	+	4.06E-07	2.54E-05
Translation initiation complex formation	58	6	.26	23.11	+	4.06E-07	2.47E-05
Activation of the mRNA upon binding of the cap-binding complex and eIFs, and subsequent binding to 43S	59	6	.26	22.71	+	4.46E-07	2.64E-05
Regulation of expression of SLTs and ROBOS	170	16	.76	21.02	+	1.76E-16	1.75E-14
Signaling by ROBO receptors	216	17	.97	17.58	+	3.00E-16	2.86E-14
Axon guidance	550	20	2.46	8.12	+	6.54E-13	5.51E-11
Major pathway of RNA processing in the nucleolus and cytosol	181	16	.81	19.74	+	4.40E-16	4.02E-14
rRNA processing in the nucleus and cytosol	191	17	.86	19.88	+	4.41E-17	4.83E-15
rRNA processing	201	17	.90	18.89	+	9.77E-17	1.02E-14
mRNA Splicing - Major Pathway	180	9	.81	11.17	+	1.70E-07	1.16E-05
mRNA splicing	188	9	.84	10.69	+	2.42E-07	1.56E-05
Processing of Capped Intron-Containing Pre-mRNA	237	9	1.06	8.48	+	1.55E-06	8.94E-05

Table 5.4: Reactome pathway analysis of the proteins binding specifically to LOC729970.

Next, we chose to focus on proteins identified across all replicates to increase our chances of identifying specific protein interactions. This reduced the protein interactors list to RPL22 (60s ribosomal protein L22), GAPDH (glyceraldehyde 3-

phosphate dehydrogenase), CYR61 (cysteine-rich angiogenic inducer 61), KNOP1 (lysine rich nucleolar protein 1), PRKRA (protein activator of interferon induced protein kinase EIF2AK2), H1X (histone H1x) and RTRAF (RNA transcription, translation and transport factor). RPL22 is a ribosomal protein, part of the 60S subunit of the ribosome, whose absence induces a disruption in development (370). KNOP1 is a nucleolar protein that might be involved in testis development (provided by RefSeq, Feb 2017 on NCBI). PRKRA is a known activator of the protein kinase R (PKR) (371), a kinase involved in the regulation of mRNA translation, transcription, apoptosis and cell proliferation (372). H1X is a histone linker that accumulates in the nucleoli during the G1 phase (373). RTRAF is a protein involved in the modulation of several phases of RNA metabolism (374).

GAPDH is a glycolytic enzyme important for ATP and pyruvate production (375). It participates in several cellular functions including DNA replication and repair, carcinogenesis and cell death. Hence, an increased level of GAPDH is associated with poor prognosis in patients with cancer (376). GAPDH overexpression has been shown to induce an increase in phosphorylated AKT (377), ultimately resulting in an escape from caspase-independent cell death. In breast cancer specifically, AKT signaling prevents GAPDH nuclear accumulation, thus blocking apoptosis by GAPDH interaction with p53 (376). Interestingly, GAPDH is known to bind to mRNA containing AU-rich sequences, including *COX-2* and *HIF-1 α* mRNA, to facilitate its translation (378).

CYR61 is an angiogenic, matricellular factor involved in the regulation of cell proliferation, survival and angiogenesis (379). CYR61 binds to integrins to induce

downstream signaling pathways including MAPK/ERK, PI3K and AKT, ultimately inducing cell growth and repressing apoptosis (380). Interestingly, CYR61 expression was detected by immunohistochemistry staining in high-grade DCIS lesions where it correlates positively with Cyclin D1 expression (381). To our knowledge, CYR61 has not been identified as an RNA-binding protein.

Since we were unable to confirm the interaction between LOC729970 and all the proteins interactors we identified, we chose to focus on CYR61 because of its interesting association with high-grade DCIS. Moreover, since the association between LOC729970 and AKT remains unclear, we chose to focus on GAPDH in addition to CYR61 in the next experiments.

5.2.7 Attempt at confirming LOC729970 interaction with CYR61 and GAPDH

Using mass spectrometry analysis, we identified several proteins that interact specifically with LOC729970. Investigating those protein interactors in the literature in relation to DCIS and what we have shown about LOC729970, we chose to focus on CYR61 and GAPDH.

In order to confirm the interaction between LOC729970 and CYR61/GAPDH, the tRSA affinity purification was repeated in combination with Western blotting, using antibodies against CYR61 and GAPDH (Figure 5.13 A). Since *FGFR2* has been linked with DCIS in the genome-wide study comparing DCIS to normal and IDC (56) and it was found to bind to LOC729970 in one replicate of the RNA-affinity purification

experiment, we decided to include FGF2 in the first step of the confirmation of the interaction with LOC729970. No expression of CYR61, GAPDH or FGF2 was detected in the bound sample. This strongly suggest that the interaction evidenced previously was, if not unspecific, weak and/or transient. In this case, the affinity purification might not reach detection levels sufficient for Western blotting analysis. Alternatively, we suspected that this interaction might be indirect. To address this point, we identified CYR61 and GAPDH proteins interactors (Table 5.5 and 5.6) and re-examined proteins identified in the mass spectrometry experiment as binding to LOC729970 or LOC729970_AS and whether they are known as RNA-binding proteins. We selected some interactors and analyzed their expression in the RNA-affinity experiment (Figure 5.13 B). We selected TUT7 as a protein interactor of CYR61, known to be an RNA-binding protein, and β actin, a protein interactor of GAPDH. Unfortunately, we could not detect expression of TUT7 and β actin in the RNA-bound sample. While we could not dismiss the hypothesis that the interaction between LOC729970 with GAPDH or CYR61 is indirect after only testing one known protein interactor, we decided to use different approaches to confirm the interaction between LOC729970 and GAPDH/CYR61.

Interactors	Binds to Sense	Binds to Antisense	RNA Binding Protein
14335	No	No	No
APOD	No	No	No
B3A2	No	No	No
BCAR1	No	No	No
CAV1	No	No	No
CNTRB	No	No	No
CNTROB	No	No	No
COG2	No	No	No
CREB1	No	No	No
CRIM1	No	No	No
CTGF	No	No	No
CUL7	No	No	No
ELMO3	No	No	No
ESR1	No	No	No
HIVEP1	No	No	No
IPIL2	No	No	No
ITAV	No	No	No
ITB5	No	No	No
ITGAD	No	No	No
ITGAM	No	No	No
ITGAV	No	No	No
ITGB2	No	No	No
ITGB5	No	No	No
ITPRIPL2	No	No	No
KI26B	No	No	No
KIF26B	No	No	No
LAMC1	No	No	No
MYC	No	No	No
NOV	No	No	No
OPA1	No	No	No
PAR16	No	No	No
PARP16	No	No	No
PLAU	No	No	No
PLXA3	No	No	No
PLXNA3	No	No	No
S18I2	No	No	No
SFN	No	No	No
SLC4A2	No	No	No
SNMA	No	No	No
SOST	No	No	No
SOX2	No	No	No
SPICE	No	No	No
SPICE1	No	No	No
SRGAP3	No	No	No
SRGP3	No	No	No
SS18L2	No	No	No
TBC1D24	No	No	No
TBC24	No	No	No
THBS1	No	No	No
TUT7	No	No	Yes
UBQL2	No	No	No
ZEP1	No	No	No
ZNF207	No	No	Yes
ZWINT	No	No	No

Table 5.5: List of CYR61 protein interactors identified in the database IntAct.

IntAct			
Interactors	Binds to Sense	Binds to Antisense	RNA Binding Protein
ACTB	Yes	Yes	No
ALDOA	No	No	Yes
ANXA7	No	No	Yes
ATP5F1A	No	No	Yes
CDC5L	No	No	Yes
CDKN2A	No	No	Yes
EIF1B	No	No	Yes
G3BP1	No	No	Yes
GOT2	No	No	Yes
HIST1H4L	No	No	Yes
JUN	No	No	Yes
KPNA2	No	No	Yes
LIN28A	No	No	Yes
MAP3K3	No	Yes	No
MAPT	No	No	Yes
PRDX1	No	No	Yes
RBM5	No	No	Yes
SMG1	No	No	Yes
SMN1	No	No	Yes
TPT1	No	No	Yes
TXN	No	No	Yes
YWHAЕ	No	No	Yes
YWHAZ	No	No	Yes

Table 5.6: List of GAPDH protein interactors identified in the database IntAct.

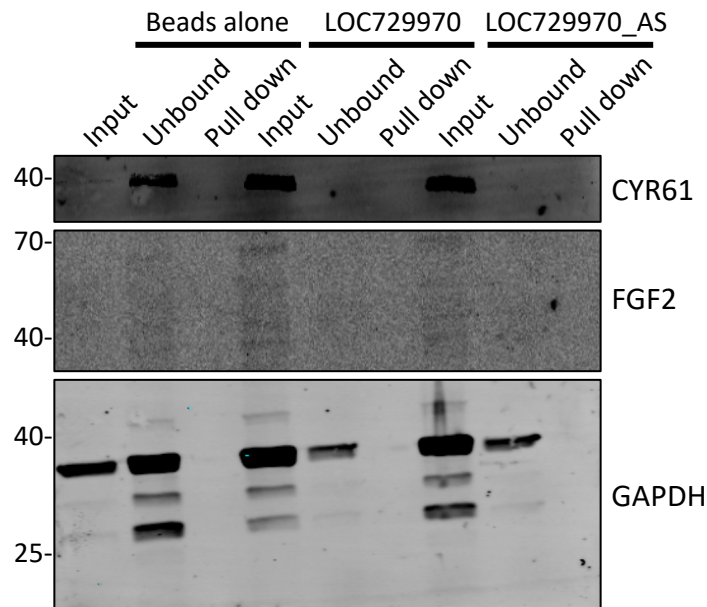
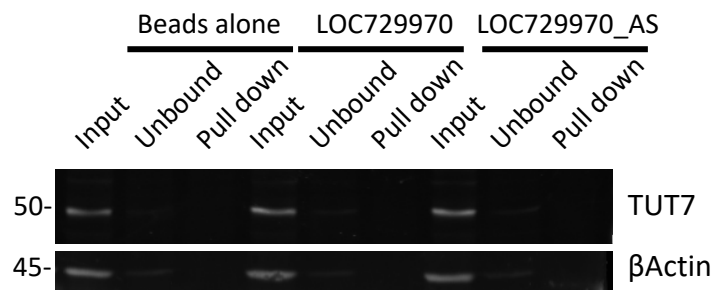
A**B**

Figure 5.13: Attempt at confirming the interaction between LOC729970 and CYR61 or GAPDH using the tRSA-system. (A) Western blotting of CYR61, FGF2 and GAPDH to verify the interaction with LOC729970. (B) Western blotting of TUT7 and β actin, some protein interactors of CYR61 and GAPDH, in the “RNA affinity purification samples” (pull down in the figure).

Instead of using immobilized RNA LOC729970 as a bait, we performed an RNA immunoprecipitation (RIP) after overexpression of myc-tagged constructs of GAPDH and CYR61, using an antibody directed against the myc tag. Presence of LOC729970 in the precipitated complex was analyzed by qRT-PCR. Myc-tagged constructs of

GAPDH and CYR61 were used to perform the IP with the myc-trap system (Chromotek). Since this was completely new in the laboratory, using the myc-trap system allowed us to perform the IP using a validated antibody for an IP. Moreover, using an antibody directly targeting GAPDH, for example, would have required additional optimization steps.

In the first trial, ETCC-006 cells were transfected with the GAPDH/CYR61-myc-tag constructs and then used for an IP using the myc-trap system. The IP sample was divided in two fractions, one fraction was used to verify the success of the IP and the other was used for qRT-PCR to measure the presence of LOC729970 RNA. This trial was performed without crosslinking as shown in Figure 5.14. PABP is an RNA binding protein that binds to the polyadenylated tail present on mRNA transcripts. It was used here as a positive control. Figure 5.14 A shows the result of the immunoprecipitation; however, no PABP was detected in the input or IP lane. We hypothesized that this is explained by the insufficient cell numbers that went into the IP. Indeed, due to the transfection with lipofectamine, a large proportion of cells died and only 1.5 million ETCC-006 cells were used for the experiment. GAPDH-myc and CYR61-myc IP were successful, indeed GAPDH-myc was detected in the input and IP fractions specifically, CYR61 was only detected in the IP fraction. We hypothesize that the lack of CYR61 detection in the input lane is due to insufficient amount of total protein as only 1% of the total protein sample is loaded into the gel for Western blotting. Actin expression was analyzed in all IPs fractions. If the RIP experiment was successful, we would expect to detect an actin mRNA enrichment in the PABP IP fraction and no actin mRNA in the GAPDH and CYR61 IP fractions. Regarding LOC729970, we expected no LOC729970 RNA in the PABP IP fraction and an enrichment in the GAPDH and

CYR61 IP fractions. No enrichment was identified in any IP fractions compared to the control, this suggests again that the interaction between LOC729970 and GAPDH or CYR61 is either weak, transient or nonspecific.

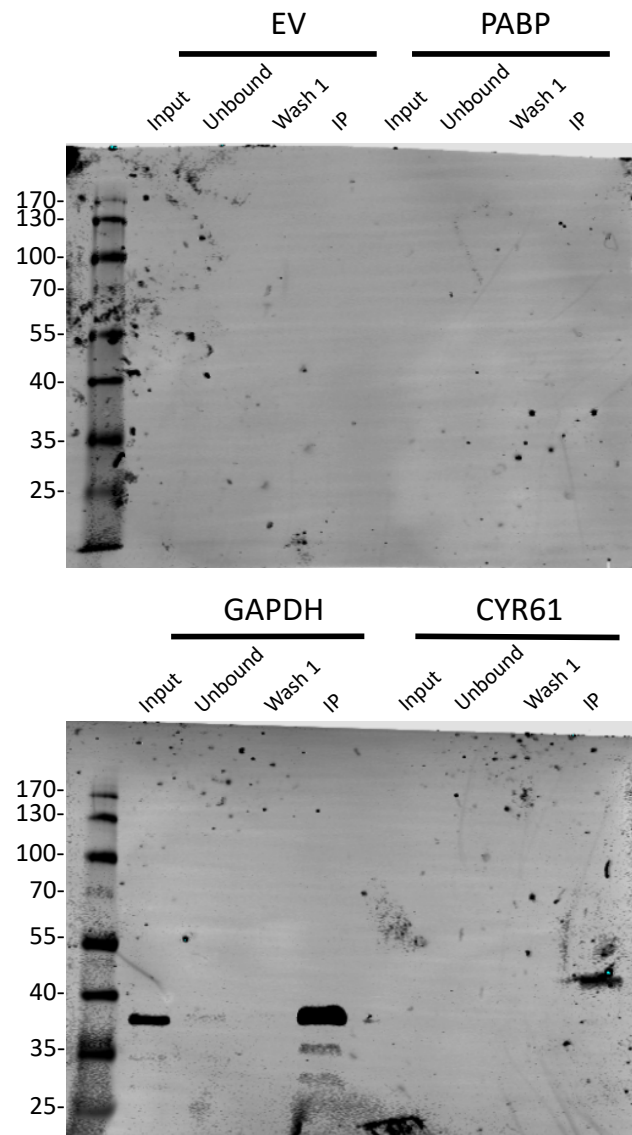
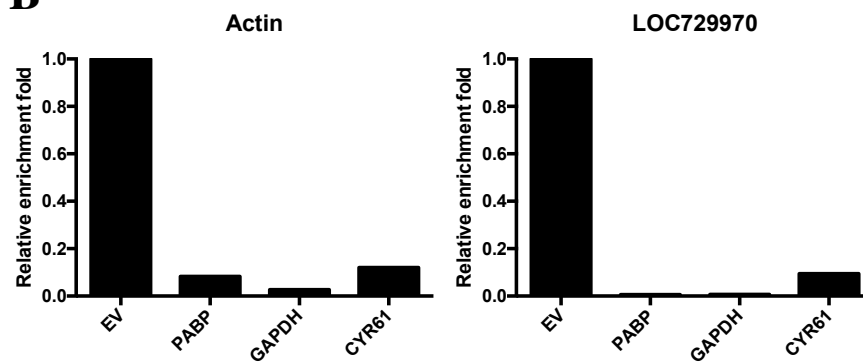
A**B**

Figure 5.14: RNA immunoprecipitation of myc-tagged version of PABBP, GAPDH and CYR61 to confirm GAPDH/CYR61 interaction with LOC729970.

(A) Western blotting using a myc antibody to verify the success of the IP. (B) Detection of actin mRNA and LOC729970 in the IP fraction by qRT-PCR.

To determine if LOC729970 interaction with GAPDH or CYR61 is a weak and transient interaction, we aimed to stabilize the RNA-protein interaction. To do that, crosslinks are induced, either chemically with the use of formaldehyde or with UV (382). We chose to stabilize the RNA-protein complexes chemically using formaldehyde since UV crosslinking is highly inefficient and requires lengthy optimization steps. We repeated this RNA IP after crosslinking the cells with formaldehyde to stabilize the RNA/protein complexes as seen in Figure 5.15. GAPDH and CYR61 IP were successful (Figure 5.15 A) and actin and LOC729970 expression was measured in each IP fractions. No enrichment was detected for either actin in the PABP IP fraction or LOC729970 in the GAPDH and CYR61 IP fractions (Figure 5.15 B).

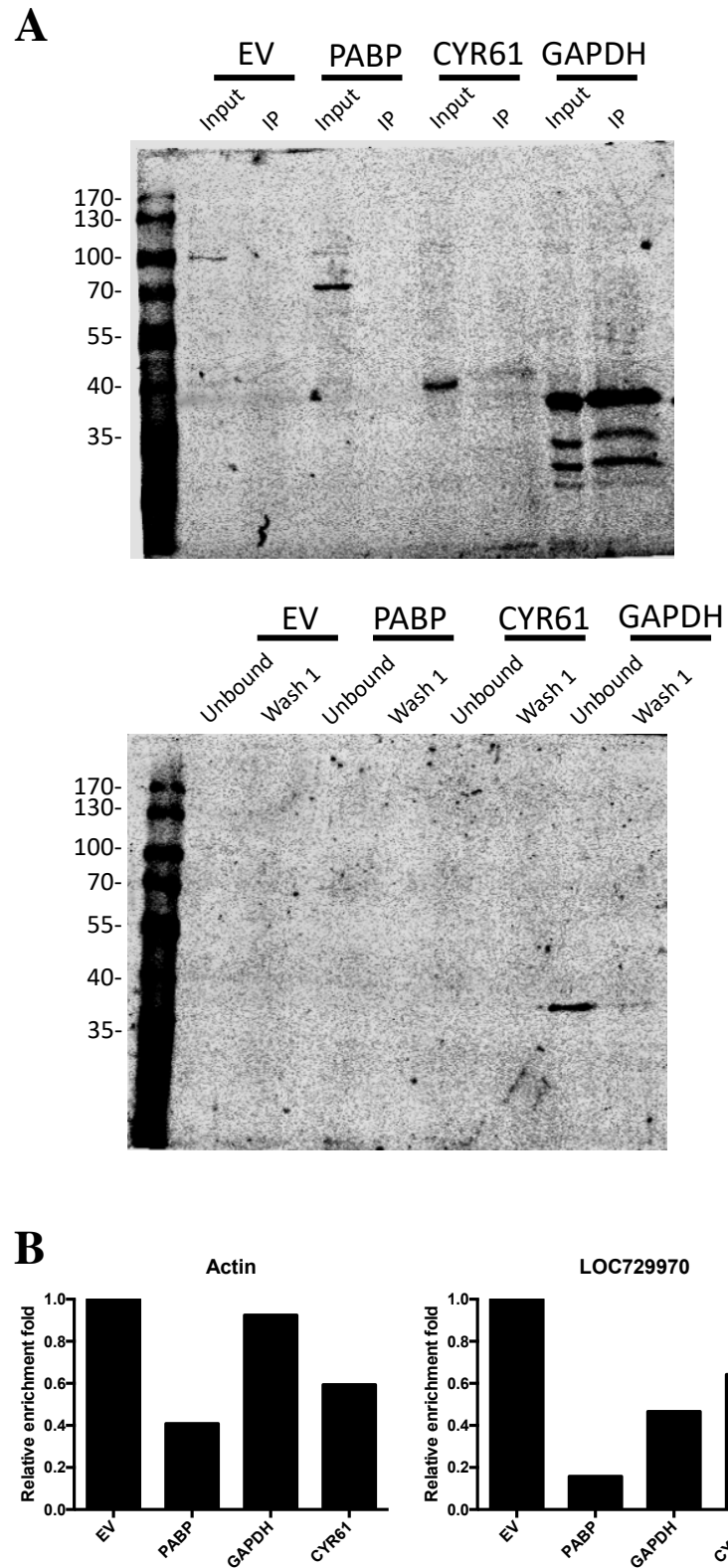


Figure 5.15: RNA immunoprecipitation of myc-tagged version of PAPBP, GAPDH and CYR61 after formaldehyde crosslinking to confirm GAPDH/CYR61 interaction with LOC729970. (A) Western blotting using a myc

antibody to verify the success of the IP. (B) Detection of *actin* mRNA and LOC729970 in the IP fraction by qRT-PCR.

Finally, since the previous attempts were unsuccessful, we decided to re-use the tRSA affinity purification system in ETCC-006 cells transfected with the GAPDH-myc and CYR61-myc constructs. LOC729970 antisense version was processed in parallel as a negative control (Figure 5.16). Similar to previously, we could not detect CYR61 protein most likely due to a large proportion of cells dying. GAPDH protein was detected in the input and the unbound lane. However, no protein was detected in the LOC729970 and LOC729970 antisense pulldown fractions with GAPDH-myc and CYR61-myc lysate.

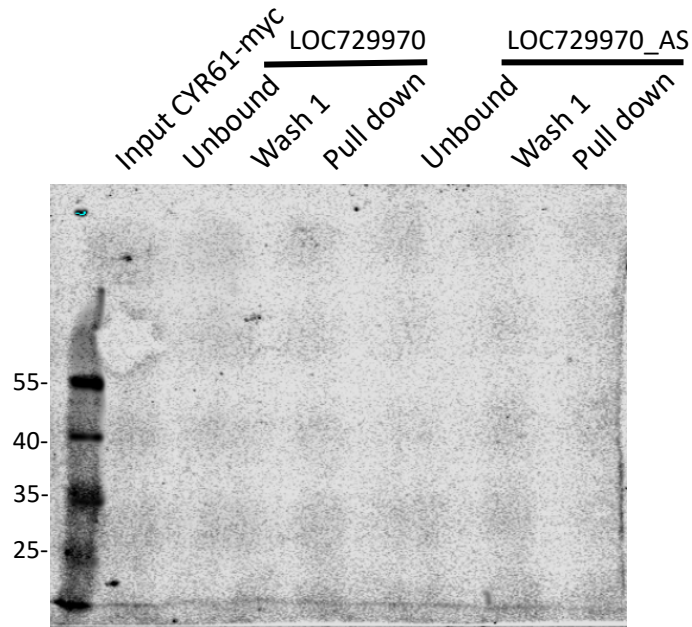
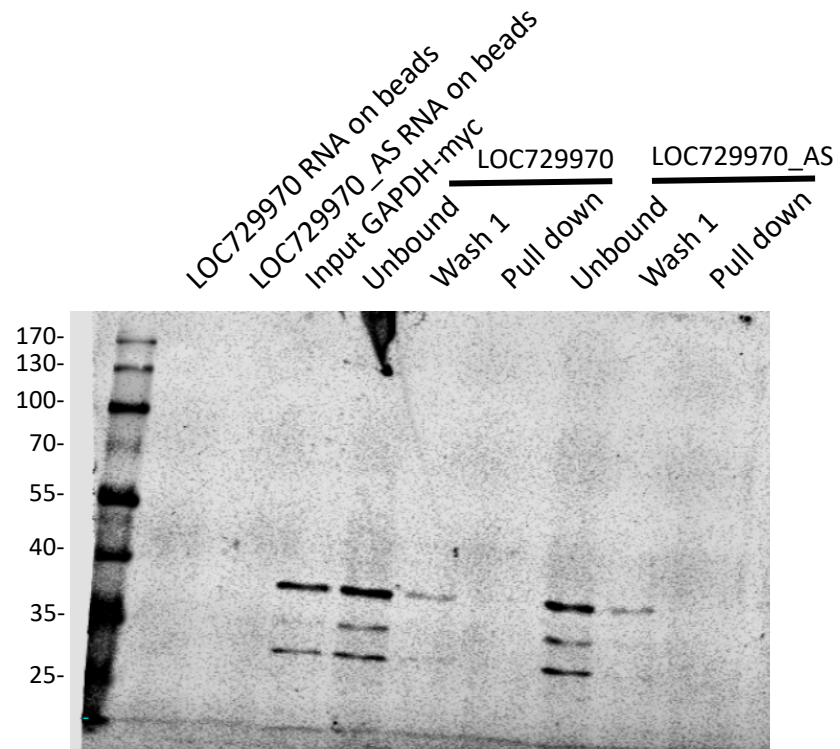


Figure 5.16: Attempt at confirming the interaction between LOC729970 and CYR61/GAPDH using the tRSA-system with myc-tagged version of GAPDH and CYR61. Western blotting with a myc antibody of the “RNA affinity purification samples” (pull down in the figure) to verify the interaction of GAPDH and CYR61 with LOC729970.

Taken together, all of our data strongly suggests that LOC729970's interaction with GAPDH or CYR61 is either unspecific or extremely transient. This presents a challenge to confirm these interactions with the panel of experimental techniques that we have available.

Because the experimental confirmation of an interaction between LOC729970 and its protein interactors GAPDH and CYR61 proved to be a challenge, we decided to use RNA-Protein Interaction seq (RPIseq), a bioinformatic tool that only utilizes the protein and RNA sequence to predict the probability of interaction (383). Two variants are used, one using a support vector machine (SVM) classifier and one using a random forest (RF) classifier. For non-coding RNA and protein interactions in *H. sapiens*, RPIseq has shown a specificity of 74.7% using RF classifiers and 57.3% using SVM classifiers (383).

In Figure 5.17.A, we show a prediction of an interaction calculated using CYR61 protein sequence and LOC729970 RNA sequence. A probability over 0.5 is considered positive and the protein and RNA are likely to interact. We observe probabilities of 0.95 (RF) and 0.94 (SVM). We also analyzed the probability of an interaction between LOC729970 and GAPDH and found a probability of 0.95 (RF) and 0.97 (SVM). In comparison, we analyzed the probability of interaction with the protein Tyrosine protein kinase Fer (FER) or beta-integrin which should not interact with LOC729970 based on our mass spectrometry analysis data. FER and LOC729970 have a probability of interaction of 0.75 (RF) and 0.99 (SVM) ; while beta-integrin and LO729970 have a probability of interaction of 0.85 (RF) and 0.86 (SVM). Those are high probabilities; however, since predictions using RF classifiers have shown to be more accurate for

non-coding RNA and proteins interaction in *H. sapiens*, we mostly focused on those. Hence, while the probabilities of interaction between LOC729970 and FER or beta-integrin were high, they remained lower than the probabilities observed with CYR61 or GAPDH. We also analyzed the probability of interaction between LOC729970 and other proteins identified in the mass spectrometry experiment, DDX18 for example showed probabilities close to those observed with FER. TCP4 however showed higher probabilities than those observed for GAPDH and CYR61. Those results seem to suggest that the lncRNA LOC729970 interact with a large number of proteins and by extension enhance the need to identify those interactions experimentally.

In addition, we analyzed the co-expression between LOC729970 and some protein interactors identified in the mass spectrometry experiment using RNAseq data from the BC TCGA cohort (figure 5.17 B). We show a positive correlation between CYR61 and LOC729970 ($r = 0.305$).

A

	Interaction probabilities	
	Prediction using RF classifier	Prediction using SVM classifier
LOC729970 and CYR61	0.95	0.94
LOC729970 and GAPDH	0.95	0.97
LOC729970 and DDX18	0.8	0.99
LOC729970 and FGF2	0.9	0.97
LOC729970 and TCP4	1	0.98
LOC729970 and NUFP2	0.85	0.95
LOC729970 and FER	0.75	0.99
LOC729970 and beta-integrin	0.85	0.86

B



Figure 5.17: Bioinformatic analysis to support the data identified in the mass spectrometry experiment. (A) Table recapitulating the interaction probabilities between LOC729970 and CYR61/GAPDH/DDX18/FGF2/TCP4/NUFP2/FER/beta-integrin with the database RPIseq. (B) Co-expression analysis of LOC729970 with CYR61, GAPDH, KNOP1, PRKRA, H1FX and RTFAF with RNAseq data from a BC cohort of TGCA.

5.3 Discussion

In this chapter, we investigated the function of lncRNAs LINC00313, LINC00473, LINC00707, MIR210HG, LOC729970 and LOC100134040 in DCIS. By examining how altering their expression affected proliferation, migration and how cells respond to stress, we intended to determine if and how those lncRNAs might affect DCIS transition into IDC.

By performing subcellular fractionation in the cell lines where they are most highly expressed, we showed that LINC00707, LINC00473, LINC00313, MIR210HG and LOC100134040 are nuclear whereas LOC729970 is mostly cytoplasmic. Interestingly, this supports what is known about lncRNAs, with most lncRNAs being nuclear.

Alteration of LINC00313 expression does not seem to induce a change in cell proliferation and cell motility. Under serum starvation or hypoxic conditions, LINC00313 is upregulated. Those results suggest that LINC00313 might be important in cells for stress response although it remains to be fully confirmed. This would, however, explain why no effect on proliferation or migration was observed after altering LINC00313 expression under normal conditions. Recently, Wu et al. showed that in papillary thyroid cancer, LINC00313 knockdown induces apoptosis and cell cycle arrest of cells in the G2/M phase (289). However, after overexpressing LINC00313, we did not see an effect on the cell number (measured by MTT). This means that LINC00313 might have a function different in our DCIS cell lines.

Regarding LINC00473, we showed no effect on proliferation and migration. Under serum starvation and hypoxic conditions, LINC00473 levels were reduced. This might just be explained by the fact that LINC00473 is expressed in conditions normal for cell growth, so conditions in which oxygen and serum with growth factors are present. In Wilms tumors, LINC00473 knockdown inhibits cell viability but we did not see an effect on proliferation after altering LINC00473 levels (302). Since LINC00473 is associated with the cAMP signaling pathway to induce CREB-mediated transcripts in lung cancer, it might be involved in the same pathway in DCIS (300). An easy way to test it would be to measure the level of CREB-mediated transcripts after alteration of LINC00473 in our DCIS cell lines.

LINC00707 change in expression induced no change in proliferation or migration. We observe a downregulation under hypoxic and serum starvation conditions. As with LINC00473, we suspect that downregulation in hypoxia and serum starvation is explained by the lack of growth factors that might regulate LINC00707 expression. In lung adenocarcinoma, it promotes proliferation through CDC42 regulation (308), however we did not see any effect on proliferation. The function of LINC00707 in DCIS remains elusive.

As with the previous lncRNAs, MIR210HG expression does not induce a change in proliferation or migration. It has been shown to be induced under hypoxia (343) and we confirmed increased levels of MIR210HG under hypoxic conditions in ETCC-006 cells. Interestingly, MIR210HG seemed slightly elevated under serum starvation too, suggesting that this lncRNA might be involved in stress response. In invasive breast cancer, MIR210HG promotes metastasis by acting as a competing endogenous RNA

for miR-1226-3p (317). However, we did not see an effect on migration when we induced the overexpression of MIR210HG. Performing a soft agar colony assay that simulates growth in a mouse and tests the capability of cells to invade a tissue after altering MIR210HG levels might allow us to determine if MIR210HG has a similar function in DCIS cells.

LOC100134040 does not seem to influence proliferation and migration. Under serum starvation, LOC100134040 is slightly induced which suggests that it might be important for stress response. Under hypoxia, however, its expression looks slightly reduced. Since LOC100134040 has not been characterized before, it is difficult to estimate how LOC100134040 might modulates stress response in cells. In the future, a project will have to focus on the confirmation of LOC100134040 role in stress response. Hence, cell response to stress will have to be monitored after altering LOC100134040 levels. Next, a large-scale study should be conducted, for example an RNA sequencing experiment, in cells with altered LOC100134040 expression to identify the targets of this lncRNA.

A change in LOC729970 expression, as with previous lncRNA candidates, did not induce an observable effect in proliferation or migration. We did notice that under serum starvation conditions, LOC729970 expression seemed slightly induced. This might mean that LOC729970 participates in the cellular response put in place to react to serum starvation stress conditions. However, LOC729970 expression is reduced in hypoxic conditions. This suggests that LOC729970 expression is modulated differently in response to different stress stimuli.

In order to identify the pathway in which LOC729970 is implicated, we analyzed LOC729970 proteins interactors and identified a short list of protein mostly involved in protein production. We chose to focus on CYR61 and GAPDH. Indeed, CYR61 has been linked with high-grade DCIS (381). GAPDH has been shown to induce an escape from caspase-independent cell death through AKT activation (377). However, while the probability of an interaction between LOC729970 and CYR61 or GAPDH is high, we could not confirm the interaction experimentally. This probably means that the interaction is transient. It could also be explained by the fact that the conditions in which we are examining the interaction between LOC729970 and GAPDH or CYR61 are not the conditions in which this interaction is optimal. Indeed, we have seen that LOC729970 is slightly induced under serum starvation conditions. Hence the interaction between LOC729970 and CYR61 or GAPDH might occur under similar conditions to activate the cellular response to stress. In consequence, further experiments will have to be conducted to confirm those interactions. In addition, imaging techniques for LOC729970 and CYR61 or GAPDH would also allow us to directly visualize a co-localization in cells which would support our data.

While our investigation of the function of the lncRNAs LINC00313, LINC00473, LINC00707, MIR210HG, LOC100134040 and LOC729970 has been unsuccessful overall, our work has brought clues of the possible functions of LINC00473, MIR210HG and LOC100134040 in DCIS for further investigations in the future. In addition, in the last part of this Chapter, we chose to focus specifically on LOC729970 in DCIS. We hypothesized previously that LOC729970 has a protective function against breast cancer progression. While none of our data has given us evidence yet to draw any conclusions about the role of LOC729970 in DCIS transition to IDC, we

hope that confirming the interaction of LOC729970 with CYR61 or GAPDH will help decipher its function in DCIS.

Chapter 6

General discussion

In this thesis, we aimed to study DCIS and to identify factors that are involved in the transition toward IDC. In the general introduction, we have highlighted the lack of overall research in DCIS and this is, especially, demonstrated by the lack of cell lines available to study DCIS.

In the third chapter of this thesis we have characterized two DCIS cell lines, ETCC-006, ETCC-010, in comparison to MCF10DCIS.com, a well-characterized cell line and model of DCIS. We have highlighted some differences in ETCC-006 and ETCC-010 cells compared to the characteristics presented in the original publication. Hence while those cells were published as hormone receptor positive breast cancer cell lines (245), in our hands they behaved like triple negative breast cancer cell lines based on expression of the hormone receptors. That study allowed us to establish a baseline for the behavior of those cells that was largely uncharacterized. Indeed, ETCC-006 and ETCC-010 have not been published in another publication since 2014 (245,247).

In the fourth chapter of this thesis, we aimed to identify novel lncRNAs associated with DCIS. To date, DCIS studies have mostly focused on the investigation of mutations associated with IDC that are present at the DCIS stage. Immunohistochemistry analyses have also allowed researchers to identify proteins expressed at the DCIS stage. For example, high levels of p16, a checkpoint regulator of the transition from G1 to S phase of the cell cycle, are associated with a higher risk

to develop IDC (112). Because many lncRNAs have been identified as dysregulated in invasive breast cancer but very few studies have focused on those in the context of DCIS specifically, we aimed to identify novel lncRNAs specific to DCIS. We performed a targeted, lncRNA-capture, RNAseq experiment on our DCIS cell lines and identified 92 lncRNAs dysregulated in all the DCIS cell lines compared to MCF-10A (Table 4.2). However, since our final aim is to identify lncRNAs with clinical implications, we restricted our candidates list to six lncRNAs whose expression is associated with adverse patient overall survival. Of those six, we identified LOC100134040 and LOC729970 whose expression is lower in the DCIS lesion compared to normal adjacent tissue in our DCIS patient cohort. Interestingly, high LOC729970 expression is associated with better overall survival. This might suggest that LOC729970 has a protective function against breast cancer progression. Moreover, lower expression in DCIS or IDC compared to normal supports that hypothesis.

Next, we tried to investigate the function of LOC729970 in DCIS transition to IDC. If LOC729970 is protective against breast cancer progression, we would expect that altering its expression would affect cell proliferation and/or migration. However, we saw no difference. With LOC729970 being slightly induced under serum starvation conditions, we suspect that this lncRNA could be involved in modulating cellular stress response. We have identified LOC729970 protein interactors. Our gene ontology analysis has shown that a large proportion of LOC729970 protein interactors is associated with ribosomal proteins, for example RPL22 but also ribosomal proteins RPL11, RPL13, RPL24 and RPS59 (Table 5.1). Moreover, the analysis of ribosome profiling data from MDA-MB-231 cells on GWIPS-Viz has shown ribosome

occupation on LOC729970 RNA (368,369). This suggests that either LOC729970 is translated onto a small protein or it could get sent onto the ribosome for degradation (170). Further experiments will have to be conducted to address those questions. Measuring LOC729970 expression after drug-induced ribosome stalling would allow us to determine if LOC729970 is degraded in the ribosome (170). Alternatively, many lncRNAs have been shown to have open reading frames and produce proteins (384). In the case of LOC729970, there is a possible open reading frame that would produce a 128 amino acid peptide. However, this open reading frame does not correspond to the region occupied by ribosomes in the ribosome profiling data obtained from MDA-MB-231 cells. By tagging possible open reading frames of lncRNAs with small protein tags, 'coding lncRNAs' can be identified by Western blotting or directly in cells with imaging techniques including immunofluorescence.

Interestingly, we identified CYR61 and GAPDH as possible LOC729970 protein interactors through RNA-affinity purification, combined with mass spectrometry. However, we could not re-confirm the interaction most likely due to the interaction being too transient. The difference in sensitivity between mass spectrometry and Western blotting analysis might also explain this lack of reproducibility, mass spectrometry being more sensitive than Western blotting (385). To overcome this difficulty, the use of techniques allowing the study of interactions between specific RNAs and proteins will have to be used. Crosslinking and immunoprecipitation (CLIP) is a technique where RNA-protein complexes are crosslinked in cells with UV light, then RNA-protein complexes are immunoprecipitated using antibodies specific to the protein of interest. The use of CLIP will most likely be necessary to stabilize the complexes formed between LOC729970 and GAPDH or CYR61. Moreover, as

mentioned earlier, the conditions in which we have analyzed those complexes may not be the physiological conditions in which those interactions happen. Being able to directly visualize LOC729970 in cells and investigating if it co-localizes with CYR61 or GAPDH would support the mass spectrometry data. We have considered a novel system to tag non-coding RNA, the RNA mango aptamer (386). This small sequence of 40 nucleotides forms a three-tiered G-quadruplex structure with mixed parallel and anti-parallel connectivity that will trap a TO1-Biotin dye and allow direct in-cell visualization.

LOC729970 has been associated with chemoresistance in pancreatic cancer (349). Indeed, in pancreatic cancer cells, Li et al. (349) have identified a hub that includes MIR210HG and LOC729970 that could possibly mediate drug resistance. We have not been able to compare LOC729970 expression between chemoresistant and non-chemoresistant cells due to time constraints. However, since MIR210HG and LOC729970 are both expressed at high levels in our DCIS cell lines, it would be interesting to first examine their level of expression in chemoresistance cells and then assess MIR210HG and LOC729970 function in chemoresistance. Interestingly, one study has suggested that CYR61 could be a target for chemosensitization (387). Could LOC729970 modulate cells chemoresistance with CYR61?

We have tried to identify markers that would allow a better stratification of DCIS patients that will either develop IDC and thus require treatment or will not develop IDC and thus do not require treatment. The research of biomarkers for early detection in cancer, in particular breast cancer have been considered for years (388). Some lncRNAs have been detected in human blood. For example, BC200 an lncRNA

detected in high-grade DCIS has been detected in blood of patients with IBC (224). However, to our knowledge no lncRNAs have been detected in serum of DCIS patients. Ultimately, our research could be extended to identify lncRNA biomarkers that could be used for early detection of DCIS and we are, particularly, considering serum biomarkers for an easy and non-invasive detection technique for patients.

DCIS patients from our cohort have donated blood prior to their surgical procedures. The blood samples will be used, in the future, to compare the expression of our lncRNA candidates in DCIS patients to healthy individuals. If an lncRNA is detected specifically in the serum of a DCIS patient, we are aware that this study will have to be repeated in a larger group of DCIS patients for validation. However, this thesis lays the foundation for a focused study that might have important clinical implications.

In parallel, since some lncRNAs have been found in blood of IDC patients (205, 274), we propose to include those in a study analyzing their expression in serum from DCIS patients. This work has been started by Yu. et al (390) that conducted a meta-analysis with 835 patients with BC and 725 controls and analyzed the sensitivity, specificity and odds ratio of several known lncRNAs as markers for early detection of breast cancer. However, this preliminary study will need additional proof to support its results, in order to be considered for implementation in clinical settings.

With this thesis, we have tried to fill the gap in research between lncRNAs and DCIS. The list of promising lncRNAs we identified, including LOC729970, will need to be further characterized. However, this work lays the foundation for future studies that

will hopefully lead to a significant improvement in early breast cancer detection and stratification of at-risk patients.

BIBLIOGRAPHY

1. Bernard WS, Christopher PW. World cancer report 2014. World Heal Organ. 2014;630.
2. Schiffman JD, Fisher PG, Gibbs P. Early Detection of Cancer: Past, Present, and Future. Am Soc Clin Oncol Educ B. 2015 May 26;35:57–65.
3. Li G, Hu J, Hu G. Biomarker studies in early detection and prognosis of breast cancer. In: *Advances in Experimental Medicine and Biology*. Springer New York LLC; 2017. p. 27–39.
4. Coleman C. Early Detection and Screening for Breast Cancer. Semin Oncol Nurs. 2017 May 1;33(2):141–55.
5. Hassiotou F, Geddes D. Anatomy of the human mammary gland: Current status of knowledge. Clin Anat. 2013;26(1):29–48.
6. Peaker M. The mammary gland in mammalian evolution: A brief commentary on some of the concepts. J Mammary Gland Biol Neoplasia. 2002;7(3):347–353.
7. Love SM, Barsky SH. Anatomy of the nipple and breast ducts revisited. Cancer. 2004;101(9):1947–1957.
8. Visvader JE. Keeping abreast of the mammary epithelial hierarchy and breast tumorigenesis. Genes Dev. 2009;23(22):2563–2577.
9. Musumeci G, Castrogiovanni P, Szychlinska MA, Aiello FC, Vecchio GM, Salvatorelli L, et al. Mammary gland: From embryogenesis to adult life. Acta Histochem. 2015;117(4–5):379–385.
10. Stute P, Wood CE, Kaplan JR, Cline JM. Cyclic changes in the mammary gland of cynomolgus macaques. Fertil Steril. 2004;82(Suppl 3):1160–1170.
11. Need EF, Atashgaran V, Ingman W V., Dasari P. Hormonal regulation of the immune microenvironment in the mammary gland. J Mammary Gland Biol Neoplasia. 2014;19(2):229–239.
12. Chapter: Essentials of Anatomy and Physiology: Reproductive System. brainkart.com. p. Mammary Glands and Anatomy of the Breast-Female.
13. Griffiths AJ, Miller JH, Suzuki DT, Lewontin RC, Gelbart WM. Mutation and cancer. An introduction to Genetic Analysis, 7th Edition. W. H. Freeman; 2000. Chapter 15 Gene mutation.
14. Bertram JS. The molecular biology of cancer. Mol Aspects Med. 2000 Dec 1;21(6):167–223.

15. Hanahan D, Weinberg RA. Hallmarks of cancer: The next generation. *Cell*. 2011 Mar 4;144(5):646–74.
16. Lodish H, Berk A, Zipursky SL, Matsudaira P, Baltimore D, Darnell J. Tumor Cells and the Onset of Cancer. In: *Molecular Cell Biology* 4th edition. W. H. Freeman; 2000. p. Section 24.1.
17. Lowe SW, Cepero E, Evan G. Intrinsic tumour suppression. *Nature*. 2004;432(7015):307–15.
18. Muenst S, Läubli H, Soysal SD, Zippelius A, Tzankov A, Hoeller S. The immune system and cancer evasion strategies: Therapeutic concepts. *J Intern Med*. 2016;279(6):541–62.
19. Witsch E, Sela M, Yarden Y. Roles for Growth Factors in Cancer Progression. *Physiology*. 2010;25(2):85–101.
20. Blasco MA. Telomeres and human disease: Ageing, cancer and beyond. *Nat Rev Genet*. 2005;6(8):611–22.
21. Jones RG, Thompson CB. Tumor suppressors and cell metabolism: A recipe for cancer growth. *Genes Dev*. 2009;23(5):537–48.
22. Hanahan D, Folkman J. Patterns and emerging mechanisms of the angiogenic switch during tumorigenesis. *Cell*. 1996;86(3):353–64.
23. Grivennikov SI, Greten FR, Karin M. Immunity, Inflammation, and Cancer. *Cell*. 2010;140(6):883–99.
24. Talmadge JE, Fidler IJ. AACR centennial series: The biology of cancer metastasis: Historical perspective. *Cancer Res*. 2010;70(14):5649–69.
25. Negrini S, Gorgoulis VG, Halazonetis TD. Genomic instability an evolving hallmark of cancer. *Nat Rev Mol Cell Biol*. 2010;11(3):220–8.
26. Kolak A, Kamińska M, Sygit K, Budny A, Surdyka D, Kukiełka-Budny B, et al. Primary and secondary prevention of breast cancer. *Ann Agric Environ Med*. 2017;24(4):549–53.
27. Yoshida K, Miki Y. Role of BRCA1 and BRCA2 as regulators of DNA repair, transcription, and cell cycle in response to DNA damage. *Cancer Sci*. 2004;95(11):866–71.
28. Lalloo F, Evans DG. Familial Breast Cancer. *Clin Genet*. 2012 Aug;82(2):105–14.
29. Economopoulou P, Dimitriadis G, Psyrri A. Beyond BRCA: New hereditary breast cancer susceptibility genes. *Cancer Treat Rev*. 2015 Jan 1;41(1):1–8.
30. Crabtree JS, Miele L. Breast Cancer Stem Cells. *Biomedicines*. 2018;6(3):77.

31. Savarese T, Low HP, Baik I, Strohsnitter W, Hsieh C-C. Normal breast stem cells, malignant breast stem cells, and the perinatal origin of breast cancer. *Stem Cell Rev.* 2006;2(2):103–9.
32. Berkel H, Birdsell DC, Jenkins H. Breast Augmentation: A Risk Factor for Breast Cancer? *N Engl J Med.* 1992 Jun 18;326(25):1649–53.
33. Ekblom A, Hsieh CC, Lipworth L, Adami HO, Trichopoulos D. Intrauterine environment and breast cancer risk in women: A population- based study. *J Natl Cancer Inst.* 1997 Jan 1;89(1):71–6.
34. Strohsnitter WC, Noller KL, Titus-Ernstoff L, Troisi R, Hatch EE, Poole C, et al. Breast cancer incidence in women prenatally exposed to maternal cigarette smoke. *Epidemiology.* 2005 May;16(3):342–5.
35. Yamasaki H, Loktionov A, Tomatis L. Perinatal and Multigenerational Effect of Carcinogens: Possible Contribution to Determination of Cancer Susceptibility. *Environ Health Perspect.* 1992 Nov;98:39–43.
36. Zhang M, Lee A V., Rosen JM. The cellular origin and evolution of breast cancer. *Cold Spring Harb Perspect Med.* 2017 Mar 1;7(3):a027128.
37. Greaves M, Maley CC. Clonal evolution in cancer. *Nature.* 2012 Jan 19;481(7381):306–13.
38. Shibata M, Shen MM. The roots of cancer: Stem cells and the basis for tumor heterogeneity. *BioEssays.* 2013 Mar;35(3):253–60.
39. Taherian-Fard A, Srihari S, Ragan MA. Breast cancer classification: linking molecular mechanisms to disease prognosis. *Brief Bioinform.* 2015 May 1;16(3):461–74.
40. ELSTON CW, ELLIS IO. pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology.* 1991;19(5):403–10.
41. Barroso-Sousa R, Metzger-Filho O. Differences between invasive lobular and invasive ductal carcinoma of the breast: results and therapeutic implications. *Ther Adv Med Oncol.* 2016;8(4):261–6.
42. Tang Y, Wang Y, Kiani MF, Wang B. Classification, Treatment Strategy, and Associated Drug Resistance in Breast Cancer. *Clin Breast Cancer.* 2016 Oct 1;16(5):335–43.
43. Sporn MB, Lippman SM. Agents for Chemoprevention and Their Mechanism of Action. In: Holland-Frei *Cancer Medicine* 6th edition. 2003. p. Chapter 30 Chemoprevention of Cancer.
44. Fabian CJ. The what, why and how of aromatase inhibitors: Hormonal agents for treatment and prevention of breast cancer. *Int J Clin Pract.*

2007;61(12):2051–63.

45. Cao L, Niu Y. Triple negative breast cancer: special histological types and emerging therapeutic methods. *Cancer Biol Med*. 2020;17(2):293–306.
46. Siddharth Kunte, Jame Abraham AJM. Novel HER2-targeted therapies for HER2-positive metastatic breast cancer. *Cancer*. 2020;Online ahe.
47. Yeo SK, Guan J-L. Breast Cancer: Multiple Subtypes within a Tumor? Breast cancer stratification and its role in guiding therapeutic decisions. *Trends Cancer*. 2017;3(11):753–760.
48. 2019 Annual Report - National Cancer Registry Ireland. 2019.
49. Fu FM, Gilmore RC, Jacobs LK. Ductal Carcinoma In Situ. Vol. 98, *Surgical Clinics of North America*. W.B. Saunders; 2018. p. 725–45.
50. Kader T, Hill P, Rakha EA, Campbell IG, Gorringer KL. Atypical ductal hyperplasia: Update on diagnosis, management, and molecular landscape. *Breast Cancer Res*. 2018 May 2;20(1):39.
51. Casasent AK, Edgerton M, Navin NE. Genome evolution in ductal carcinoma *in situ*: invasion of the clones. *J Pathol*. 2017 Jan;241(2):208–18.
52. Hong YK, McMasters KM, Egger ME, Ajkay N. Ductal carcinoma in situ current trends, controversies, and review of literature. *Am J Surg*. 2018 Nov 1;216(5):998–1003.
53. Yeong J, Thike AA, Tan PH, Iqbal J. Identifying progression predictors of breast ductal carcinoma in situ. *J Clin Pathol*. 2017;70(2):102–8.
54. Reeves GK, Pirie K, Green J, Bull D, Beral V. Comparison of the effects of genetic and environmental risk factors on in situ and invasive ductal breast cancer. *Int J Cancer*. 2012;15(131):930–7.
55. Claus EB, Petruzella S, Matloff E, Carter D. Prevalence of BRCA1 and BRCA2 mutations in women diagnosed with ductal carcinoma in situ. *J Am Med Assoc*. 2005;293(8):964–9.
56. Petridis C, Brook MN, Shah V, Kohut K, Gorman P, Caneppele M, et al. Genetic predisposition to ductal carcinoma in situ of the breast. *Breast Cancer Res*. 2016 Dec 17;18(1):22.
57. Jaffer S, Bleiweiss IJ. Histologic classification of ductal carcinoma in situ. *Microsc Res Tech*. 2002 Oct 15;59(2):92–101.
58. Mayr NA, Staples JJ, Hussey DH, Robinson RA, Vanmetre JE. Morphometric studies in intraductal breast carcinoma using computerized image analysis. *Cancer*. 1991;67(11):2805–12.

59. Shoker BS, Sloane JP. DCIS grading schemes and clinical implications. *Histopathology*. 1999;35(5):393–400.
60. Lagios MD. Heterogeneity of duct carcinoma in situ (DCIS): Relationship of grade and subtype analysis to local recurrence and risk of invasive transformation. *Cancer Lett*. 1995;90(1):97–102.
61. Allred DC. Ductal carcinoma in situ: Terminology, classification, and natural history. *J Natl Cancer Inst - Monogr*. 2010;2010(41):134–8.
62. Fu FM, Gilmore RC, Jacobs LK. Ductal Carcinoma In Situ. *Surg Clin North Am*. 2018 Aug 1;98(4):725–45.
63. Wapnir IL, Dignam JJ, Fisher B, Mamounas EP, Anderson SJ, Julian TB, et al. Long-term outcomes of invasive ipsilateral breast tumor recurrences after lumpectomy in NSABP B-17 and B-24 randomized clinical trials for DCIS. *J Natl Cancer Inst*. 2011;103(6):478–88.
64. Cuzick J, Sestak I, Pinder SE, Ellis IO, Forsyth S, Bundred NJ, et al. Effect of tamoxifen and radiotherapy in women with locally excised ductal carcinoma in situ: Long-term results from the UK/ANZ DCIS trial. *Lancet Oncol*. 2011;12(1):21–9.
65. Lo AC, Truong PT, Wai ES, Nichol A, Weir L, Speers C, et al. Population-based analysis of the impact and generalizability of the NSABP-B24 study on endocrine therapy for patients with ductal carcinoma in situ of the breast. *Ann Oncol*. 2015;26(9):1898–903.
66. Allred DC, Anderson SJ, Paik S, Wickerham DL, Nagtegaal ID, Swain SM, et al. Adjuvant tamoxifen reduces subsequent breast cancer in women with estrogen receptor-positive ductal carcinoma in situ: A study based on NSABP protocol B-24. *J Clin Oncol*. 2012;30(12):1268–73.
67. Forbes JF, Sestak I, Howell A, Bonanni B, Bundred N, Levy C, et al. Anastrozole versus tamoxifen for the prevention of locoregional and contralateral breast cancer in postmenopausal women with locally excised ductal carcinoma in situ (IBIS-II DCIS): A double-blind, randomised controlled trial. *Lancet*. 2016;387(10021):866–73.
68. Margolese RG, Cecchini RS, Julian TB, Ganz PA, Costantino JP, Vallow LA, et al. Anastrozole versus tamoxifen in postmenopausal women with ductal carcinoma in situ undergoing lumpectomy plus radiotherapy (NSABP B-35): A randomised, double-blind, phase 3 clinical trial. *Lancet*. 2016;387(10021):849–56.
69. Siziopikou KP, Anderson SJ, Cobleigh MA, Julian TB, Arthur DW, Zheng P, et al. Preliminary results of centralized HER2 testing in ductal carcinoma in situ (DCIS): NSABP B-43. *Breast Cancer Res Treat*. 2013 Nov;142(2):415–21.
70. Barrio A V., Van Zee KJ. Controversies in the Treatment of Ductal Carcinoma

in Situ. *Annu Rev Med.* 2017 Jan 14;68(1):197–211.

71. Groen EJ, Elshof LE, Visser LL, Rutgers EJT, Winter-Warnars HAO, Lips EH, et al. Finding the balance between over- and under-treatment of ductal carcinoma in situ (DCIS). *Breast.* 2017 Feb 1;31:274–83.
72. Virnig BA, Tuttle TM, Shamliyan T, Kane RL. Ductal carcinoma in Situ of the breast: A systematic review of incidence, treatment, and outcomes. *J Natl Cancer Inst.* 2010;102(3):170–8.
73. Elshof LE, Tryfonidis K, Slaets L, Van Leeuwen-Stok AE, Skinner VP, Dif N, et al. Feasibility of a prospective, randomised, open-label, international multicentre, phase III, non-inferiority trial to assess the safety of active surveillance for low risk ductal carcinoma in situ - The LORD study. *Eur J Cancer.* 2015;51(12):1497–510.
74. Lakhani SR, Audretsch W, Cleton-Jensen AM, Cutuli B, Ellis I, Eusebi V, et al. The management of lobular carcinoma in situ (LCIS). Is LCIS the same as ductal carcinoma in situ (DCIS)? *Eur J Cancer.* 2006;42(14):2205–11.
75. Klauschen F, Müller KR, Binder A, Bockmayr M, Hägele M, Seegerer P, et al. Scoring of tumor-infiltrating lymphocytes: From visual estimation to machine learning. *Semin Cancer Biol.* 2018;52(Pt 2):151–7.
76. Chen XY, Yeong J, Thike AA, Bay BH, Tan PH. Prognostic role of immune infiltrates in breast ductal carcinoma in situ. *Breast Cancer Res Treat.* 2019 Aug 30;177(1):17–27.
77. Abba MC, Gong T, Lu Y, Lee J, Zhong Y, Lacunza E, et al. A molecular portrait of high-grade ductal carcinoma in situ. *Cancer Res.* 2015 Sep 15;75(18):3980–90.
78. Pang J-MB, Gorringe KL, Wong SQ, Dobrovic A, Campbell IG, Fox SB. Appraisal of the technologies and review of the genomic landscape of ductal carcinoma in situ of the breast. *Breast Cancer Res.* 2015 Dec 16;17(1):80.
79. Gao Y, Niu Y, Wang X, Wei L, Lu S. Genetic changes at specific stages of breast cancer progression detected by comparative genomic hybridization. *J Mol Med.* 2009;87(2):145–52.
80. Liao S, Desouki MM, Gaile DP, Shepherd L, Nowak NJ, Conroy J, et al. Differential copy number aberrations in novel candidate genes associated with progression from in situ to invasive ductal carcinoma of the breast. *Genes Chromosom Cancer.* 2012;51(12):1067–78.
81. Moore E, Magee H, Coyne J, Gorey T, Dervan PA. Widespread chromosomal abnormalities in high-grade ductal carcinoma in situ of the breast. Comparative genomic hybridization study of pure high- grade DCIS. *J Pathol.* 1999;187(4):403–9.

82. Vos CBJ, Ter Haar NT, Rosenberg C, Peterse JL, Cleton-Jansen AM, Cornelisse CJ, et al. Genetic alterations on chromosome 16 and 17 are important features of ductal carcinoma in situ of the breast and are associated with histologic type. *Br J Cancer*. 1999;81(8):1410–8.
83. Hwang ES, DeVries S, Chew K Lo, Moore DH, Kerlikowske K, Thor A, et al. Patterns of chromosomal alterations in breast ductal carcinoma in situ. *Clin Cancer Res*. 2004;10(15):5160–7.
84. Ellsworth RE, Ellsworth DL, Love B, Patney HL, Hoffman LR, Kane J, et al. Correlation of levels and patterns of genomic instability with histological grading of DCIS. *Ann Surg Oncol*. 2007;14(11):3070–7.
85. Sakr RA, Weigelt B, Chandarlapaty S, Andrade VP, Guerini-Rocco E, Giri D, et al. PI3K pathway activation in high-grade ductal carcinoma in situ-implications for progression to invasive breast carcinoma. *Clin Cancer Res*. 2014;20(9):2326–37.
86. Miron A, Varadi M, Carrasco D, Li H, Luongo L, Kim HJ, et al. PIK3CA mutations in in situ and invasive breast carcinomas. *Cancer Res*. 2010;70(14):5674–8.
87. DeVaux RS, Herschkowitz JJ. Beyond DNA: the Role of Epigenetics in the Premalignant Progression of Breast Cancer. *J Mammary Gland Biol Neoplasia*. 2018;23(4):223–35.
88. Tommasi S, Karm DL, Wu X, Yen Y, Pfeifer GP. Methylation of homeobox genes is a frequent and early epigenetic event in breast cancer. *Breast Cancer Res*. 2009;11(1):R14.
89. Johnson KC, Koestler DC, Fleischer T, Chen P, Jenson EG, Marotti JD, et al. DNA methylation in ductal carcinoma in situ related with future development of invasive breast cancer. *Clin Epigenetics*. 2015;7(1):75.
90. Teschendorff AE, Gao Y, Jones A, Ruebner M, Beckmann MW, Wachter DL, et al. DNA methylation outliers in normal breast tissue identify field defects that are enriched in cancer. *Nat Commun*. 2016;29(7):10478.
91. Yan PS, Venkataramu C, Ibrahim A, Liu JC, Shen RZ, Diaz NM, et al. Mapping geographic zones of cancer risk with epigenetic biomarkers in normal breast tissue. *Clin Cancer Res*. 2006;12(22):6626–36.
92. Munshi A, Shafi G, Aliya N, Jyothy A. Histone modifications dictate specific biological readouts. *J Genet Genomics*. 2009;36(2):75–88.
93. Strahl BD, Allis CD. The language of covalent histone modifications. *Nature*. 2000;403(6765):41–5.
94. Kleer CG, Cao Q, Varambally S, Shen R, Ota I, Tomlins SA, et al. EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of

breast epithelial cells. *Proc Natl Acad Sci U S A*. 2003;100(20):11606–11.

95. Ding L, Erdmann C, Chinnaiyan AM, Merajver SD, Kleer CG. Identification of EZH2 as a molecular marker for a precancerous state in morphologically normal breast tissues. *Cancer Res*. 2006;66(8):4095–9.
96. Serce N, Gnatzy A, Steiner S, Lorenzen H, Kirfel J, Buettner R. Elevated expression of LSD1 (Lysine-specific demethylase 1) during tumour progression from pre-invasive to invasive ductal carcinoma of the breast. *BMC Clin Pathol*. 2012;12:13.
97. Suzuki J, Chen YY, Scott GK, DeVries S, Chin K, Benz CC, et al. Protein acetylation and histone deacetylase expression associated with malignant breast cancer progression. *Clin Cancer Res*. 2009;15(9):3163–71.
98. Shen H, Xu W, Guo R, Rong B, Gu L, Wang Z, et al. Suppression of Enhancer Overactivation by a RACK7-Histone Demethylase Complex. *Cell*. 2016;165(2):331–42.
99. Knopfmacher A, Fox J, Lo Y, Shapiro N, Fineberg S. Correlation of histopathologic features of ductal carcinoma in situ of the breast with the oncotype DX DCIS score. *Mod Pathol*. 2015 Sep 3;28(9):1167–73.
100. Oncotype DX DCIS score predicts recurrence. *Cancer Discov*. 2015;5(2):OF3.
101. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med*. 2004;351(27):2817–26.
102. Solin LJ, Gray R, Baehner FL, Butler SM, Hughes LL, Yoshizawa C, et al. A Multigene expression Assay to Predict local recurrence risk for Ductal carcinoma in Situ of the Breast. *J Natl Cancer Inst*. 2013;105(10):701–10.
103. Farrugia DJ, Landmann A, Zhu L, Diego EJ, Johnson RR, Bonaventura M, et al. Magee Equation 3 predicts pathologic response to neoadjuvant systemic chemotherapy in estrogen receptor positive, HER2 negative/equivocal breast tumors. *Mod Pathol*. 2017 Aug 1;30(8):1078–85.
104. Nofech-Mozes S, Hanna W, Rakovitch E. Molecular Evaluation of Breast Ductal Carcinoma in Situ with Oncotype DX DCIS. *Am J Pathol*. 2019 May 1;189(5):975–80.
105. Rakovitch E, Gray R, Baehner FL, Sutradhar R, Crager M, Gu S, et al. Refined estimates of local recurrence risks by DCIS score adjusting for clinicopathological features: a combined analysis of ECOG-ACRIN E5194 and Ontario DCIS cohort studies. *Breast Cancer Res Treat*. 2018;169(2):359–69.
106. Rakovitch E, Nofech-Mozes S, Narod SA, Hanna W, Thiruchelvam D, Saskin R, et al. Can we select individuals with low risk ductal carcinoma in situ (DCIS)? A population-based outcomes analysis. *Breast Cancer Res Treat*.

2013;138(2):581–90.

107. McNamara KM, Yoda T, Nurani AM, Shibahara Y, Miki Y, Wang L, et al. Androgenic pathways in the progression of triple-negative breast carcinoma: A comparison between aggressive and non-aggressive subtypes. *Breast Cancer Res Treat.* 2014;145(2):281–93.
108. Davey RA, Grossmann M. Androgen Receptor Structure, Function and Biology: From Bench to Bedside. *Clin Biochem Rev.* 2016;37(1):3–15.
109. Williams AB, Schumacher B. p53 in the DNA-damage-repair process. *Cold Spring Harb Perspect Med.* 2016;6(5):a026070.
110. Giono LE, Manfredi JJ. The p53 tumor suppressor participates in multiple cell cycle checkpoints. *J Cell Physiol.* 2006;209(1):13–20.
111. Kasthuber ER, Lowe SW. Putting p53 in Context. *Cell.* 2017;170(6):1062–78.
112. Pape-Zambito D, Jiang Z, Wu H, Devarajan K, Slater CM, Cai KQ, et al. Identifying a highly-aggressive DCIS subgroup by studying intra-individual DCIS heterogeneity among invasive breast cancer patients. *PLoS One.* 2014 Jun 30;9(6):e100488.
113. Sherr CJ, Beach D, Shapiro GI. Targeting CDK4 and CDK6: From discovery to therapy. *Cancer Discov.* 2016;6(4):353–67.
114. Shan M, Zhang X, Liu X, Qin Y, Liu T, Liu Y, et al. P16 and P53 Play Distinct Roles in Different Subtypes of Breast Cancer. Gao J-X, editor. *PLoS One.* 2013 Oct 11;8(10):e76408.
115. Generali D, Buffa FM, Deb S, Cummings M, Reid LE, Taylor M, et al. COX-2 expression is predictive for early relapse and aromatase inhibitor resistance in patients with ductal carcinoma in situ of the breast, and is a target for treatment. *Br J Cancer.* 2014;111(1):46–54.
116. Karimian A, Ahmadi Y, Yousefi B. Multiple functions of p21 in cell cycle, apoptosis and transcriptional regulation after DNA damage. *DNA Repair (Amst).* 2016;42:63–71.
117. Wiechmann L, Kuerer HM. The molecular journey from ductal carcinoma in situ to invasive breast cancer. *Cancer.* 2008 May 15;112(10):2130–42.
118. Alao JP. The regulation of cyclin D1 degradation: Roles in cancer development and the potential for therapeutic invention. *Mol Cancer.* 2007;6:24.
119. Turner BC, Gumbs AA, Carter D, Glazer PM, Haffty BG. Cyclin D1 expression and early breast cancer recurrence following lumpectomy and radiation. *Int J Radiat Oncol Biol Phys.* 2000;47(5):1169–76.

120. Kuwana T, Newmeyer DD. Bcl-2-family proteins and the role of mitochondria in apoptosis. *Curr Opin Cell Biol.* 2003;500(1):26–34.
121. Kerlikowske K, Molinaro AM, Gauthier ML, Berman HK, Waldman F, Bennington J, et al. Biomarker Expression and Risk of Subsequent Tumors After Initial Ductal Carcinoma In Situ Diagnosis. *JNCI J Natl Cancer Inst.* 2010 May 5;102(9):627–37.
122. Jonas S, Izaurralde E. Towards a molecular understanding of microRNA-mediated gene silencing. Vol. 16, *Nature Reviews Genetics*. Nature Publishing Group; 2015. p. 421–33.
123. Hannafon BN, Ding WQ. miRNAs as Biomarkers for Predicting the Progression of Ductal Carcinoma in Situ. *Am J Pathol.* 2018;188(3):542–9.
124. Hannafon BN, Sebastiani P, de las Morenas A, Lu J, Rosenberg CL. Expression of microRNA and their gene targets are dysregulated in preinvasive breast cancer. *Breast Cancer Res.* 2011;13(2):R24.
125. Iorio M V., Ferracin M, Liu CG, Veronese A, Spizzo R, Sabbioni S, et al. MicroRNA gene expression deregulation in human breast cancer. *Cancer Res.* 2005;1711:55–81.
126. Farazi TA, Horlings HM, Ten Hoeve JJ, Mihailovic A, Halfwerk H, Morozov P, et al. MicroRNA sequence and expression analysis in breast tumors by deep sequencing. *Cancer Res.* 2011 Jul 1;71(13):4443–53.
127. Chen L, Li Y, Fu Y, Peng J, Mo MH, Stamatakis M, et al. Role of Deregulated microRNAs in Breast Cancer Progression Using FFPE Tissue. *PLoS One.* 2013 Jan 29;8(1).
128. Iyer MK, Niknafs YS, Malik R, Singhal U, Sahu A, Hosono Y, et al. The landscape of long noncoding RNAs in the human transcriptome. *Nat Genet.* 2015;47(3):199–208.
129. Jensen TH, Jacquier A, Libri D. Dealing with pervasive transcription. *Mol Cell.* 2013;52(4):473–84.
130. Jarroux J, Morillon A, Pinskaya M. History, discovery, and classification of lncRNAs. *Adv Exp Med Biol.* 2017;1008:1–46.
131. Choudhuri S. Small noncoding RNAs: Biogenesis, function, and emerging significance in toxicology. *J Biochem Mol Toxicol.* 2010;24(3):195–216.
132. Hombach S, Kretz M. Non-coding RNAs: Classification, biology and functioning. *Adv Exp Med Biol.* 2016;937:3–17.
133. Wu T, Du Y. LncRNAs: From basic research to medical application. *Int J Biol Sci.* 2017;13(3):295–307.

134. St.Laurent G, Wahlestedt C, Kapranov P. The Landscape of long noncoding RNA classification. *Trends Genet.* 2015 May 1;31(5):239–51.
135. Hackermüller J, Reiche K, Otto C, Hösler N, Blumert C, Brocke-Heidrich K, et al. Cell cycle, oncogenic and tumor suppressor pathways regulate numerous long and macro non-protein-coding RNAs. *Genome Biol.* 2014;15(3):R48.
136. Kapranov P, St Laurent G, Raz T, Ozsolak F, Reynolds CP, Sorensen PHB, et al. The majority of total nuclear-encoded non-ribosomal RNA in a human cell is “dark matter” un-annotated RNA. *BMC Biol.* 2010;8:149.
137. St Laurent G, Shtokalo D, Dong B, Tackett MR, Fan X, Lazorthes S, et al. VlincRNAs controlled by retroviral elements are a hallmark of pluripotency and cancer. *Genome Biol.* 2013;14(7):R73.
138. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, et al. GENCODE: The reference human genome annotation for the ENCODE project. *Genome Res.* 2012;22(9):1760–74.
139. Nakaya HI, Amaral PP, Louro R, Lopes A, Fachel AA, Moreira YB, et al. Genome mapping and expression analyses of human intronic noncoding RNAs reveal tissue-specific patterns and enrichment in genes related to regulation of transcription. *Genome Biol.* 2007;8(3):R43.
140. St Laurent G, Shtokalo D, Tackett MR, Yang Z, Eremina T, Wahlestedt C, et al. Intronic RNAs constitute the major fraction of the non-coding RNA in mammalian cells. *BMC Genomics.* 2012;13:504.
141. Katayama S, Tomaru Y, Kasukawa T, Waki K, Nakanishi M, Nakamura M, et al. Antisense transcription in the mammalian transcriptome. *Science (80-).* 2005;309(5740):1564–6.
142. Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, et al. An atlas of active enhancers across human cell types and tissues. *Nature.* 2014;507(7493):455–61.
143. Sell S. On the stem cell origin of cancer. *Am J Pathol.* 2010;176(6):2584–494.
144. Mariner PD, Walters RD, Espinoza CA, Drullinger LF, Wagner SD, Kugel JF, et al. Human Alu RNA Is a Modular Transacting Repressor of mRNA Transcription during Heat Shock. *Mol Cell.* 2008;29(4):499–509.
145. Poliseno L, Salmena L, Zhang J, Carver B, Haveman WJ, Pandolfi PP. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature.* 2010;465(7301):1033–8.
146. Wilusz JE, Freier SM, Spector DL. 3' End Processing of a Long Nuclear-Retained Noncoding RNA Yields a tRNA-like Cytoplasmic RNA. *Cell.* 2008;135(5):919–32.

147. Prabakaran S, Hemberg M, Chauhan R, Winter D, Tweedie-Cullen RY, Dittrich C, et al. Quantitative profiling of peptides from RNAs classified as noncoding. *Nat Commun.* 2014;18(5):5429.
148. Nam JW, Choi SW, You BH. Incredible RNA: Dual functions of coding and noncoding. *Mol Cells.* 2016;39(5):367–74.
149. Dykes IM, Emanuelli C. Transcriptional and Post-transcriptional Gene Regulation by Long Non-coding RNA. *Genomics, Proteomics Bioinforma.* 2017 Jun 1;15(3):177–86.
150. Chen LL, Carmichael GG. Decoding the function of nuclear long non-coding RNAs. *Curr Opin Cell Biol.* 2010;22(3):357–64.
151. Cabili MN, Dunagin MC, McClanahan PD, Biaesch A, Padovan-Merhar O, Regev A, et al. Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome Biol.* 2015;16(1):20.
152. Batista PJ, Chang HY. Long noncoding RNAs: Cellular address codes in development and disease. *Cell.* 2013;152(6):1298–307.
153. Mousavi K, Zare H, Dell'Orso S, Grontved L, Gutierrez-Cruz G, Derfoul A, et al. ERNAs Promote Transcription by Establishing Chromatin Accessibility at Defined Genomic Loci. *Mol Cell.* 2013;51(5):606–17.
154. Kung JTY, Colognori D, Lee JT. Long noncoding RNAs: Past, present, and future. *Genetics.* 2013;193(3):651–69.
155. Davidovich C, Cech TR. The recruitment of chromatin modifiers by long noncoding RNAs: Lessons from PRC2. *RNA.* 2015;21(12):2007–22.
156. Hacisuleyman E, Goff LA, Trapnell C, Williams A, Henao-Mejia J, Sun L, et al. Topological Organization of Multi-chromosomal Regions by Firre. *Nat Struct Mol Biol.* 2014;21(2):198–206.
157. Ng SY, Bogu GK, Soh BS, Stanton LW. The long noncoding RNA RMST interacts with SOX2 to regulate neurogenesis. *Mol Cell.* 2013;51(3):349–59.
158. Hung T, Wang Y, Lin MF, Koegel AK, Kotake Y, Grant GD, et al. Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nat Genet.* 2011;43(7):621–9.
159. Long Y, Wang X, Youmans DT, Cech TR. How do lncRNAs regulate transcription? *Sci Adv.* 2017;3(9):eaao2110.
160. Park E, Maquat LE. Staufen-mediated mRNA decay. *Wiley Interdiscip Rev RNA.* 2013;4(4):423–35.
161. Noh JH, Kim KM, McClusky WG, Abdelmohsen K, Gorospe M. Cytoplasmic functions of long noncoding RNAs. *Wiley Interdiscip Rev RNA.*

2018;9(3):e1471.

162. Yoon JH, Abdelmohsen K, Srikantan S, Yang X, Martindale JL, De S, et al. LincRNA-p21 Suppresses Target mRNA Translation. *Mol Cell*. 2012;47(4):648–55.
163. Yang F, Zhang H, Mei Y, Wu M. Reciprocal Regulation of HIF-1 α and LincRNA-p21 Modulates the Warburg Effect. *Mol Cell*. 2014;53(1):88–100.
164. Lu G, Li Y, Ma Y, Lu J, Chen Y, Jiang Q, et al. Long noncoding RNA LINC00511 contributes to breast cancer tumorigenesis and stemness by inducing the miR-185-3p/E2F1/Nanog axis. *J Exp Clin Cancer Res*. 2018;37(1):289.
165. Liu AN, Qu HJ, Gong WJ, Xiang JY, Yang MM, Zhang W. LncRNA AWPPH and miRNA-21 regulates cancer cell proliferation and chemosensitivity in triple-negative breast cancer by interacting with each other. *J Cell Biochem*. 2019;120(9):14860–6.
166. Dong H, Hu J, Zou K, Ye M, Chen Y, Wu C, et al. Activation of LncRNA TINCR by H3K27 acetylation promotes Trastuzumab resistance and epithelial-mesenchymal transition by targeting MicroRNA-125b in breast Cancer. *Mol Cancer*. 2019;18(1):3.
167. Wang P, Xue Y, Han Y, Lin L, Wu C, Xu S, et al. The STAT3-binding long noncoding RNA Inc-DC controls human dendritic cell differentiation. *Science* (80-). 2014;344(6181):310–3.
168. Lin A, Hu Q, Li C, Xing Z, Ma G, Wang C, et al. The LINK-A lncRNA interacts with PtdIns(3,4,5)P3 to hyperactivate AKT and confer resistance to AKT inhibitors. *Nat Cell Biol*. 2017;19(3):238–51.
169. Zeng C, Fukunaga T, Hamada M. Identification and analysis of ribosome-associated lncRNAs using ribosome profiling data. *BMC Genomics*. 2018;19(1):414.
170. Carlevaro-Fita J, Rahim A, Guigó R, Vardy LA, Johnson R. Cytoplasmic long noncoding RNAs are frequently bound to and degraded at ribosomes in human cells. *RNA*. 2016;22(6):867–82.
171. Noh JH, Kim KM, Abdelmohsen K, Yoon JH, Panda AC, Munk R, et al. HuR and GRSF1 modulate the nuclear export and mitochondrial localization of the lncRNA RMRP. *Genes Dev*. 2016;30(10):1224–39.
172. Sarropoulos I, Marin R, Cardoso-Moreira M, Kaessmann H. Developmental dynamics of lncRNAs across mammalian organs and species. *Nature*. 2019 Jul 25;571(7766):510–4.
173. Bunch H. Gene regulation of mammalian long non-coding RNA. *Mol Genet Genomics*. 2018 Feb 1;293(1):1–15.

174. Patrat C, Ouimette J-F, Rougeulle C. X chromosome inactivation in human development. *Development*. 2020 Jan 3;147(1):dev183095.
175. Alvarez-Dominguez JR, Lodish HF. Emerging mechanisms of long noncoding RNA function during normal and malignant hematopoiesis. Vol. 130, *Blood*. American Society of Hematology; 2017. p. 1965–75.
176. Li L, Zhuang Y, Zhao X, Li X. Long Non-coding RNA in Neuronal Development and Neurological Disorders. *Front Genet*. 2019;10(JAN):9–744.
177. Dinger ME, Amara PP, Mercer TR, Pang KC, Bruce SJ, Gardiner BB, et al. Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res*. 2008 Sep;18(9):1433–45.
178. Yu CY, Chuang CY, Kuo HC. Trans-spliced long non-coding RNA: an emerging regulator of pluripotency. Vol. 75, *Cellular and Molecular Life Sciences*. Birkhauser Verlag AG; 2018. p. 3339–51.
179. Shi G, Jin Y. Role of Oct4 in maintaining and regaining stem cell pluripotency. *Stem Cell Res Ther*. 2010;1(5):39.
180. Mohamed JS, Gaughwin PM, Lim B, Robson P, Lipovich L. Conserved long noncoding RNAs transcriptionally regulated by Oct4 and Nanog modulate pluripotency in mouse embryonic stem cells. *RNA*. 2010 Feb;16(2):324–37.
181. Bao X, Wu H, Zhu X, Guo X, Hutchins AP, Luo Z, et al. The p53-induced lincRNA-p21 derails somatic cell reprogramming by sustaining H3K9me3 and CpG methylation at pluripotency gene promoters. *Cell Res*. 2015 Jan 5;25(1):80–92.
182. Zhang H na, Xu Q qiao, Thakur A, Alfred MO, Chakraborty M, Ghosh A, et al. Endothelial dysfunction in diabetes and hypertension: Role of microRNAs and long non-coding RNAs. Vol. 213, *Life Sciences*. Elsevier Inc.; 2018. p. 258–68.
183. Lekka E, Hall J. Noncoding RNAs in disease. *FEBS Lett*. 2018 Sep;592(17):2884–900.
184. Findeis MA. The role of amyloid β peptide 42 in Alzheimer's disease. *Pharmacol Ther*. 2007 Nov;116(2):266–86.
185. Luo Q, Chen Y. Long noncoding RNAs and Alzheimer's disease. *Clin Interv Aging*. 2016 Jun 29;11:867–72.
186. Fields CR, Bengoa-Vergniory N, Wade-Martins R. Targeting Alpha-Synuclein as a Therapy for Parkinson's Disease. *Front Mol Neurosci*. 2019;12:299.
187. Oe S, Kimura T, Yamada H. Regulatory non-coding RNAs in nervous system development and disease. Vol. 24, *Frontiers in Bioscience - Landmark*. Frontiers in Bioscience; 2019. p. 1203–40.

188. Wang S, Zhang X, Guo Y, Rong H, Liu T. The long noncoding RNA HOTAIR promotes parkinson's disease by upregulating LRRK2 expression. *Oncotarget*. 2017;8(15):24449–56.
189. Liu M, Jia J, Wang X, Liu Y, Wang C, Fan R. Long non-coding RNA HOTAIR promotes cervical cancer progression through regulating BCL2 via targeting miR-143-3p. *Cancer Biol Ther*. 2018;19(5):391–9.
190. Carrick WT, Burks B, Cairns MJ, Kocerha J. Noncoding RNA Regulation of Dopamine Signaling in Diseases of the Central Nervous System. *Front Mol Biosci*. 2016 Oct 25;3:69.
191. Bannon MJ, Savonen CL, Jia H, Dachet F, Halter SD, Schmidt CJ, et al. Identification of long noncoding RNAs dysregulated in the midbrain of human cocaine abusers. *J Neurochem*. 2015 Oct 1;135(1):50–9.
192. Ip JY, Sone M, Nashiki C, Pan Q, Kitaichi K, Yanaka K, et al. Gomafu lncRNA knockout mice exhibit mild hyperactivity with enhanced responsiveness to the psychostimulant methamphetamine. *Sci Rep*. 2016 Jun 2;6:27204.
193. Huarte M. The emerging role of lncRNAs in cancer. *Nat Med*. 2015 Oct 1;21(11):1253–61.
194. Bussemakers MJG, Van Bokhoven A, Verhaegh GW, Smit FP, Karthaus HFM, Schalken JA, et al. DD3: A new prostate-specific gene, highly overexpressed in prostate cancer. *Cancer Res*. 1999 Dec 1;59(23):5975–9.
195. Srikantan V, Zou Z, Petrovics G, Xu L, Augustus M, Davis L, et al. PCGEM1, a prostate-specific gene, is overexpressed in prostate cancer. *Proc Natl Acad Sci U S A*. 2000 Oct 24;97(22):12216–21.
196. Washietl S, Kellis M, Garber M. Evolutionary dynamics and tissue specificity of human long noncoding RNAs in six mammals. *Genome Res*. 2014;24(4):616–28.
197. Dinescu S, Ignat S, Lazar AD, Constantin C, Neagu M, Costache M. Epitranscriptomic signatures in lncRNAs and their possible roles in cancer. *Genes (Basel)*. 2019 Jan 1;10(1):52.
198. Gutschner T, Diederichs S. The hallmarks of cancer: A long non-coding RNA point of view. *RNA Biol*. 2012;9(6):703–19.
199. Peng WX, Koirala P, Mo YY. LncRNA-mediated regulation of cell signaling in cancer. *Oncogene*. 2017 Oct 12;36(41):5661–7.
200. Gugnoni M, Ciarrocchi A. Long noncoding RNA and epithelial mesenchymal transition in cancer. *Int J Mol Sci*. 2019 Apr 2;20(8):1924.
201. Weidle UH, Birzele F, Kollmorgen G, Rüger R. Long Non-coding RNAs and their Role in Metastasis. *Cancer Genomics Proteomics*. 2017;14(3):143–60.

202. Fan Q, Yang L, Zhang X, Peng X, Wei S, Su D, et al. The emerging role of exosome-derived non-coding RNAs in cancer biology. *Cancer Lett.* 2018 Feb 1;414:107–15.
203. Xu C-G, Yang M-F, Ren Y-Q, Wu C-H, Wang L-Q. Exosomes mediated transfer of lncRNA UCA1 results in increased tamoxifen resistance in breast cancer cells. *Eur Rev Med Pharmacol Sci.* 2016;20(20):4362–8.
204. Zhang R, Xia Y, Wang Z, Zheng J, Chen Y, Li X, et al. Serum long non coding RNA MALAT-1 protected by exosomes is up-regulated and promotes cell proliferation and migration in non-small cell lung cancer. *Biochem Biophys Res Commun.* 2017 Aug 19;490(2):406–14.
205. Conigliaro A, Costa V, Lo Dico A, Saieva L, Buccheri S, Dieli F, et al. CD90+ liver cancer cells modulate endothelial cell phenotype through the release of exosomes containing H19 lncRNA. *Mol Cancer.* 2015 Aug 14;14(1):155.
206. Pan Y, Chen H, Shen X, Wang X, Ju S, Lu M, et al. Serum level of long noncoding RNA H19 as a diagnostic biomarker of multiple myeloma. *Clin Chim Acta.* 2018;480:199–205.
207. Wu Y, Wang YQ, Weng WW, Zhang QY, Yang XQ, Gan HL, et al. A serum-circulating long noncoding RNA signature can discriminate between patients with clear cell renal cell carcinoma and healthy controls. *Oncogenesis.* 2016;5(2):e192.
208. Matsui M, Corey DR. Non-coding RNAs as drug targets. *Nat Rev Drug Discov.* 2017 Mar 1;16(3):167–79.
209. Leucci E, Vendramin R, Spinazzi M, Laurette P, Fiers M, Wouters J, et al. Melanoma addiction to the long non-coding RNA SAMMSON. *Nature.* 2016 Mar 23;531(7595):518–22.
210. Berteaux N, Lottin S, Monté D, Pinte S, Quatannens B, Coll J, et al. H19 mRNA-like noncoding RNA promotes breast cancer cell proliferation through positive control by E2F1. *J Biol Chem.* 2005;280(33):29625–36.
211. Silva JM, Boczek NJ, Berres MW, Ma X, Smith DI. LSINCT5 is overexpressed in breast and ovarian cancer and affects cellular proliferation. *RNA Biol.* 2011;8(3):496–505.
212. Lanz RB, Chua SS, Barron N, Soder BM, DeMayo F, O'Malley BW. Steroid Receptor RNA Activator Stimulates Proliferation as Well as Apoptosis In Vivo. *Mol Cell Biol.* 2003;23(20):7163–76.
213. Bin X, Hongjian Y, Xiping Z, Bo C, Shifeng Y, Binbin T. Research progresses in roles of LncRNA and its relationships with breast cancer. *Cancer Cell Int.* 2018 Dec 12;18(1):179.
214. Li X, Wu Y, Liu A, Tang X. Long non-coding RNA UCA1 enhances tamoxifen

resistance in breast cancer cells through a miR-18a-HIF1 α feedback regulatory loop. *Tumor Biol.* 2016;37(11):14733–43.

215. Li Y, Wang B, Lai H, Li S, You Q, Fang Y, et al. Long non-coding RNA CRALA is associated with poor response to chemotherapy in primary breast cancer. *Thorac Cancer.* 2017 Nov 1;8(6):582–91.
216. Mamun M Al, Mannoor K, Cao J, Qadri F, Song X. SOX2 in cancer stemness: tumor malignancy and therapeutic potentials. *J Mol Cell Biol.* 2018 Dec 5;12(2):85–98.
217. Askarian-Amiri ME, Seyfoddin V, Smart CE, Wang J, Kim JE, Hansji H, et al. Emerging role of long non-coding RNA SOX2OT in SOX2 regulation in breast cancer. *PLoS One.* 2014;9(7):e102140.
218. Park IK, Morrison SJ, Clarke MF. Bmi1, stem cells, and senescence regulation. Vol. 113, *Journal of Clinical Investigation.* 2004. p. 175–9.
219. Liu Y, Sharma S, Watabe K. Roles of lncRNA in breast cancer. *Front Biosci (Schol Ed).* 2015 Jun 1;7:94–108.
220. Özgür E, Mert U, Isin M, Okutan M, Dalay N, Gezer U. Differential expression of long non-coding RNAs during genotoxic stress-induced apoptosis in HeLa and MCF-7 cells. *Clin Exp Med.* 2013;13(2):119–26.
221. Piao HL, Ma L. Non-coding RNAs as regulators of mammary development and breast cancer. *J Mammary Gland Biol Neoplasia.* 2012;17(1):33–42.
222. Iacoangeli A, Lin Y, Morley EJ, Muslimov IA, Bianchi R, Reilly J, et al. BC200 RNA in invasive and preinvasive breast cancer. *Carcinogenesis.* 2004 Nov;25(11):2125–33.
223. Samson J, Cronin S, Dean K. BC200 (BCYRN1) – The shortest, long, non-coding RNA associated with cancer. *Non-coding RNA Res.* 2018;3(3):131–43.
224. Iacoangeli A, Adzovic L, Chen EQ, Latif R, Soff GA, Tiedge H. Regulatory BC200 RNA in peripheral blood of patients with invasive breast cancer. *J Invest Med.* 2018 Oct 1;66(7):1055–63.
225. Zhang Z, Weaver DL, Olsen D, DeKay J, Peng Z, Ashikaga T, et al. Long non-coding RNA chromogenic in situ hybridisation signal pattern correlation with breast tumour pathology. *J Clin Pathol.* 2016 Jan 1;69(1):76–81.
226. Deva Magendhra Rao AK, Patel K, Korivi Jyothiraj S, Meenakumari B, Sundersingh S, Sridevi V, et al. Identification of lncRNA s associated with early-stage breast cancer and their prognostic implications. *Mol Oncol.* 2019 May 8;13(6):1342–55.
227. DeVaux RS, Ropri AS, Grimm SL, Hall PA, Herrera EO, Chittur S V., et al. Long noncoding RNA BHLHE40-AS1 promotes early breast cancer

- progression through modulating IL-6/STAT3 signaling. *J Cell Biochem.* 2020 Jan 7;doi: 10.1002/jcb.29621
228. Volders P-J, Anckaert J, Verheggen K, Nuytens J, Martens L, Mestdagh P, et al. LNCipedia 5: towards a reference set of human long non-coding RNAs. *Nucleic Acids Res.* 2019 Jan 8;47(D1):D135–9.
 229. Jordan M, Wurm F. Transfection of adherent and suspended cells by calcium phosphate. *Methods.* 2004;33(2):136–43.
 230. Afgan E, Baker D, van den Beek M, Blankenberg D, Bouvier D, Čech M, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res.* 2016;44(W1):W3–10.
 231. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30(15):2114–20.
 232. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29(1):15–21.
 233. Liao Y, Smyth GK, Shi W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics.* 2014;30(7):923–30.
 234. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550.
 235. Yu G, He QY. ReactomePA: An R/Bioconductor package for reactome pathway analysis and visualization. *Mol Biosyst.* 2016;12(2):477–9.
 236. Feng Y. Long Non-Coding RNAs. *Methods and Protocols.* In: *Anticancer research.* 2016. p. 73–100.
 237. Trapnell C, Pachter L, Salzberg SL. TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics.* 2009;25(9):1105–11.
 238. Ghosh S, Chan CKK. Analysis of RNA-seq data using TopHat and cufflinks. *Methods Mol Biol.* 2016;1374:339–61.
 239. Iioka H, Loiselle D, Haystead TA, MacAra IG. Efficient detection of RNA-protein interactions using tethered RNAs. *Nucleic Acids Res.* 2011 Apr;39(8):e53.
 240. Iioka H, Macara IG. Detection of RNA–protein interactions using tethered RNA affinity capture. *Methods Mol Biol.* 2015 May 16;1316:67–73.
 241. Dai X, Cheng H, Bai Z, Li J. Breast cancer cell line classification and Its relevance with breast tumor subtyping. *J Cancer.* 2017;8(16):3131–41.
 242. Holliday DL, Speirs V. Choosing the right cell line for breast cancer research.

Breast Cancer Res. 2011;13(4):215.

243. Miller FR, Santner SJ, Tait L, Dawson PJ. MCF10DCIS.com Xenograft Model of Human Comedo Ductal Carcinoma In Situ. *JNCI J Natl Cancer Inst.* 2000;92(14):1185–6.
244. Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, Fevr T, et al. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell.* 2006;10(6):515–27.
245. Yong JWY, Choong ML, Wang SF, Wang Y, Lim SQY, Lee MA. Characterization of ductal carcinoma in situ cell lines established from breast tumor of a Singapore Chinese patient. *Cancer Cell Int.* 2014;14(1):94.
246. Lee KM, Choi KH, Ouellette MM. Use of exogenous hTERT to immortalize primary human cells. *Cytotechnology.* 2004;45(1–2):33–8.
247. Brock EJ, Ji K, Shah S, Mattingly RR, Sloane BF. In Vitro Models for Studying Invasive Transitions of Ductal Carcinoma In Situ. *J Mammary Gland Biol Neoplasia.* 2019;24(1):1–15.
248. Bunnell TM, Burbach BJ, Shimizu Y, Ervasti JM. β -Actin specifically controls cell growth, migration, and the G-actin pool. *Mol Biol Cell.* 2011;22(21):4047–58.
249. Naylor MJ, Ormandy CJ. Mouse strain-specific patterns of mammary epithelial ductal side branching are elicited by stromal factors. *Dev Dyn.* 2002;225(1):100–5.
250. Varner VD, Nelson CM. Cellular and physical mechanisms of branching morphogenesis. *Dev.* 2014;141(14):2750–9.
251. Allred DC, Wu Y, Mao S, Nagtegaal ID, Lee S, Perou CM, et al. Ductal carcinoma in situ and the emergence of diversity during breast cancer evolution. *Clin Cancer Res.* 2008;14(2):370–8.
252. Subik K, Lee JF, Baxter L, Strzepek T, Costello D, Crowley P, et al. The expression patterns of ER, PR, HER2, CK5/6, EGFR, KI-67 and AR by immunohistochemical analysis in breast cancer cell lines. *Breast Cancer Basic Clin Res.* 2010;4:35–41.
253. Dawson PJ, Wolman SR, Tait L, Heppner GH, Miller FR. MCF10AT: A model for the evolution of cancer from proliferate breast disease. *Am J Pathol.* 1996;148(1):313–9.
254. Kadota M, Yang HH, Gomez B, Sato M, Clifford RJ, Meerzaman D, et al. Delineating genetic alterations for tumor progression in the MCF10A series of breast cancer cell lines. *PLoS One.* 2010;5(2):e9201.
255. Worsham MJ, Pals G, Schouten JP, Miller F, Tiwari N, Van Spaendonk R, et

- al. High-resolution mapping of molecular events associated with immortalization, transformation, and progression to breast cancer in the MCF10 model. *Breast Cancer Res Treat*. 2006;96(2):177–86.
256. van Seijen M, Lips EH, Thompson AM, Nik-Zainal S, Futreal A, Hwang ES, et al. Ductal carcinoma in situ: to treat or not to treat, that is the question. *Br J Cancer*. 2019;121(4):285–92.
 257. Baldin V, Lukas J, Marcote MJ, Pagano M, Draetta G. Cyclin D1 is a nuclear protein required for cell cycle progression in G1. *Genes Dev*. 1993;7(5):812–21.
 258. Tye BK. MCM Proteins in DNA Replication. *Annu Rev Biochem*. 1999;68:649–86.
 259. Musa J, Aynaud MM, Mirabeau O, Delattre O, Grünwald TG. MYBL2 (B-Myb): a central regulator of cell proliferation, cell survival and differentiation involved in tumorigenesis. *Cell Death Dis*. 2017;8(6):e2895.
 260. Brown LF, Guidi AJ, Schnitt SJ, Van De Water L, Iruela-Arispe ML, Yeo TK, et al. Vascular stroma formation in carcinoma in situ, invasive carcinoma, and metastatic carcinoma of the breast. *Clin Cancer Res*. 1999;5(5):1041–56.
 261. Franken NAP, Rodermond HM, Stap J, Haveman J, van Bree C. Clonogenic assay of cells in vitro. *Nat Protoc*. 2006;1(5):2315–9.
 262. PUCK TT, MARCUS PI, CIECIURA SJ. Clonal growth of mammalian cells in vitro; growth characteristics of colonies from single HeLa cells with and without a feeder layer. *J Exp Med*. 1956;103(2):273–83.
 263. Roberts AB, Roche NS, Sporn MB. Selective inhibition of the anchorage-independent growth of myc-transfected fibroblasts by retinoic acid. *Nature*. 1985;315:237–9.
 264. Kalluri R, Weinberg RA. The basics of epithelial-mesenchymal transition. *J Clin Invest*. 2009;119(6):1420–8.
 265. Seyfried TN, Huysentruyt LC. On the origin of cancer metastasis. *Crit Rev Oncog*. 2013;18(1–2):43–73.
 266. Yang Y, Fujita J, Bando S, Ohtsuki Y, Yamadori I, Yoshinouchi T, et al. Detection of antivimentin antibody in sera of patients with idiopathic pulmonary fibrosis and non-specific interstitial pneumonia. *Clin Exp Immunol*. 2002;128(1):169–74.
 267. Gilles C, Polette M, Mestdagt M, Nawrocki-Raby B, Ruggeri P, Birembaut P, et al. Transactivation of vimentin by β -catenin in human breast cancer cells. *Cancer Res*. 2003;63(10):2658–64.
 268. Wang Z, Zhang H, Hou J, Niu J, Ma Z, Zhao H, et al. Clinical implications of

- β -catenin protein expression in breast cancer. *Int J Clin Exp Pathol*. 2015;8(11):14989–94.
269. Pan B, Guo J, Liao Q, Zhao Y. β 1 and β 3 integrins in breast, prostate and pancreatic cancer: A novel implication. *Oncol Lett*. 2018;15(4):5412–6.
 270. Howe GA, Addison CL. β 1 integrin: An emerging player in the modulation of tumorigenesis and response to therapy. *Cell Adhes Migr*. 2012;6(2):71–7.
 271. Wee P, Wang Z. Epidermal growth factor receptor cell proliferation signaling pathways. *Cancers (Basel)*. 2017;9(5):52.
 272. Denduluri SK, Idowu O, Wang Z, Liao Z, Yan Z, Mohammed MK, et al. Insulin-like growth factor (IGF) signaling intumorigenesis and the development of cancer drug resistance. *Genes Dis*. 2015;2(1):13–25.
 273. Vos CBJ, Ter Haar NT, Peterse JL, Cornelisse CJ, Van De Vijver MJ. Cyclin D1 gene amplification and overexpression are present in ductal carcinoma in situ of the breast. *J Pathol*. 1999;187(3):279–84.
 274. Barnabas N, Cohen D. Phenotypic and Molecular Characterization of MCF10DCIS and SUM Breast Cancer Cell Lines. *Int J Breast Cancer*. 2013;2013:872743.
 275. Vogt PK, Hart JR, Gymnopoulos M, Jiang H, Kang S, Bader AG, et al. Phosphatidylinositol 3-kinase: The oncoprotein. *Curr Top Microbiol Immunol*. 2010;347:9–104.
 276. Filipów S, Łaczmański Ł. Blood circulating miRNAs as cancer biomarkers for diagnosis and surgical treatment response. *Front Neurosci*. 2019;10:169.
 277. Guo X, Hua Y. CCAT1: an oncogenic long noncoding RNA in human cancers. *J Cancer Res Clin Oncol*. 2017;143(4):555–62.
 278. Liu Z, Chen QJ, Hann SS. The functions and oncogenic roles of CCAT1 in human cancer. *Biomed Pharmacother*. 2019;115:108943.
 279. Han C, Li X, Fan Q, Liu G, Yin J. CCAT1 promotes triple-negative breast cancer progression by suppressing miR-218/ZFX signaling. *Aging (Albany NY)*. 2019;11(14):4858–75.
 280. Lai Y, Chen Y, Lin Y, Ye L. Down-regulation of LncRNA CCAT1 enhances radiosensitivity via regulating miR-148b in breast cancer. *Cell Biol Int*. 2018;42(2):227–36.
 281. Zhang XF, Liu T, Li Y, Li S. Overexpression of long non-coding RNA CCAT1 is a novel biomarker of poor prognosis in patients with breast cancer. *Int J Clin Exp Pathol*. 2015;8(8):9440–5.
 282. Li ZT, Zhang X, Wang DW, Xu J, Kou KJ, Wang ZW, et al. Overexpressed

- lncRNA GATA6-AS1 Inhibits LNM and EMT via FZD4 through the Wnt/ β -Catenin Signaling Pathway in GC. *Mol Ther - Nucleic Acids*. 2020;19:827–40.
283. Xiong Y, Zhang X, Lin Z, Xiong A, Xie S, Liang J, et al. SFTA1P, LINC00968, GATA6-AS1, TBX5-AS1, and FEZF1-AS1 are crucial long non-coding RNAs associated with the prognosis of lung squamous cell carcinoma. *Oncol Lett*. 2019;18(4):3985–93.
 284. Zhang J, Piao H yan, Guo S, Wang Y, Zhang T, Zheng Z chao, et al. LINC00163 inhibits the invasion and metastasis of gastric cancer cells as a ceRNA by sponging miR-183 to regulate the expression of AKAP12. *Int J Clin Oncol*. 2020;25(4):570–83.
 285. Guo X, Wei Y, Wang Z, Liu W, Yang Y, Yu X, et al. LncRNA LINC00163 upregulation suppresses lung cancer development though transcriptionally increasing TCF21 expression. *Am J Cancer Res*. 2018;8(12):2494–506.
 286. Fan H, Yuan J, Li X, Ma Y, Wang X, Xu B, et al. LncRNA LINC00173 enhances triple-negative breast cancer progression by suppressing miR-490-3p expression. *Biomed Pharmacother*. 2020;125:109987.
 287. Yang Q, Tang Y, Tang C, Cong H, Wang X, Shen X, et al. Diminished LINC00173 expression induced miR-182-5p accumulation promotes cell proliferation, migration and apoptosis inhibition via AGER/NF- κ B pathway in non-small-cell lung cancer. *Am J Transl Res*. 2019;11(7):4248–62.
 288. Zeng F, Wang Q, Wang S, Liang S, Huang W, Guo Y, et al. Linc00173 promotes chemoresistance and progression of small cell lung cancer by sponging miR-218 to regulate Etk expression. *Oncogene*. 2020;39(2):293–307.
 289. Wu WJ, Yin H, Hu JJ, Wei XZ. Long noncoding RNA LINC00313 modulates papillary thyroid cancer tumorigenesis via sponging miR-4429. *Neoplasma*. 2018;65(6):933–42.
 290. Yan DG, Liu N, Chao M, Tu YY, Liu WS. SP1-induced upregulation of long noncoding RNA LINC00313 contributes to papillary thyroid cancer progression via the miR-422a. *Eur Rev Med Pharmacol Sci*. 2019;23(3):1134–44.
 291. Zhao X, Hu X. Downregulated long noncoding RNA LINC00313 inhibits the epithelial–mesenchymal transition, invasion, and migration of thyroid cancer cells through inhibiting the methylation of ALX4. *J Cell Physiol*. 2019;234(11):20992–1004.
 292. Du W, Feng Z, Sun Q. LncRNA LINC00319 accelerates ovarian cancer progression through miR-423-5p/NACC1 pathway. *Biochem Biophys Res Commun*. 2018;507(1–4):198–202.
 293. Li F, Liao J, Duan X, He Y, Liao Y. Upregulation of LINC00319 indicates a poor prognosis and promotes cell proliferation and invasion in cutaneous

squamous cell carcinoma. *J Cell Biochem.* 2018;119(12):10393–405.

294. Zhou B, Yuan W, Li X. Long Intergenic Noncoding RNA 319 (linc00319) Promotes Cell Proliferation and Invasion in Lung Cancer Cells by Directly Downregulating the Tumor Suppressor MiR-32. *Oncol Res Featur Preclin Clin Cancer Ther.* 2017;10.3727/096504017X15016337254650. Online ahead of.
295. Zhang Y, Huang Z, Sheng F, Yin Z. MYC upregulated LINC00319 promotes human acute myeloid leukemia (AML) cells growth through stabilizing SIRT6. *Biochem Biophys Res Commun.* 2019;509(1):314–21.
296. Shen Y, Wang Z, Loo LWM, Ni Y, Jia W, Fei P, et al. LINC00472 expression is regulated by promoter methylation and associated with disease-free survival in patients with grade 2 breast cancer. *Breast Cancer Res Treat.* 2015;154(3):473–82.
297. Wang Z, Katsaros D, Biglia N, Shen Y, Loo L, Yu X, et al. ER α upregulates the expression of long non-coding RNA LINC00472 which suppresses the phosphorylation of NF- κ B in breast cancer. *Breast Cancer Res Treat.* 2019;175(2):353–68.
298. Zou A, Liu X, Mai Z, Zhang J, Liu Z, Huang Q, et al. LINC00472 Acts as a Tumor Suppressor in NSCLC through KLLN-Mediated p53-Signaling Pathway via MicroRNA-149-3p and MicroRNA-4270. *Mol Ther - Nucleic Acids.* 2019;17:563–77.
299. Mao X, Zhou X, Liu J, Mao Y, Zhou H. Up-regulated Linc00472 suppresses development of lung cancer cell via inhibition of MiR-196b-5p. *Biosci Biotechnol Biochem.* 2019;1-13 Online ahead of print.
300. Chen Z, Li JL, Lin S, Cao C, Gimbrone NT, Yang R, et al. cAMP/CREB-regulated LINC00473 marks LKB1-inactivated lung cancer and mediates tumor growth. *J Clin Invest.* 2016 Jun 1;126(6):2267–79.
301. Han PB, Ji XJ, Zhang M, Gao LY. Upregulation of lncRNA LINC00473 promotes radioresistance of HNSCC cells through activating Wnt/ γ -catenin signaling pathway. *Eur Rev Med Pharmacol Sci.* 2018;22(21):7305–13.
302. Zhu S, Fu W, Zhang L, Fu K, Hu J, Jia W, et al. LINC00473 antagonizes the tumour suppressor miR-195 to mediate the pathogenesis of Wilms tumour via IKK α . *Cell Prolif.* 2018;51(1):e12416.
303. Shi C, Yang Y, Yu J, Meng F, Zhang T, Gao Y. The long noncoding RNA LINC00473, a target of microRNA 34a, promotes tumorigenesis by inhibiting ILF2 degradation in cervical cancer. *Am J Cancer Res.* 2017;7(11):2157–68.
304. Liu X, Lu X, Zhen F, Jin S, Yu T, Zhu Q, et al. LINC00665 Induces Acquired Resistance to Gefitinib through Recruiting EZH2 and Activating PI3K/AKT Pathway in NSCLC. *Mol Ther - Nucleic Acids.* 2019;16:155–61.

305. Guo B, Wu S, Zhu X, Zhang L, Deng J, Li F, et al. Micropeptide CIP 2A- BP encoded by LINC 00665 inhibits triple-negative breast cancer progression. *EMBO J.* 2020;39(1):e102190.
306. Cong Z, Diao Y, Xu Y, Li X, Jiang Z, Shao C, et al. Long non-coding RNA linc00665 promotes lung adenocarcinoma progression and functions as ceRNA to regulate AKR1B10-ERK signaling by sponging miR-98. *Cell Death Dis.* 2019;10(2):84.
307. Wang J, Luo Z, Yao T, Li W, Pu J. LINC00707 promotes hepatocellular carcinoma progression through activating ERK/JNK/AKT pathway signaling pathway. *J Cell Physiol.* 2019;234(5):6908–16.
308. Ma T, Ma H, Zou Z, He X, Liu Y, Shuai Y, et al. The Long Intergenic Noncoding RNA 00707 Promotes Lung Adenocarcinoma Cell Proliferation and Migration by Regulating Cdc42. *Cell Physiol Biochem.* 2018;45(4):1566–80.
309. Li T, Li Y, Sun H. MicroRNA-876 is sponged by long noncoding RNA LINC00707 and directly targets metadherin to inhibit breast cancer malignancy. *Cancer Manag Res.* 2019;11:5255–69.
310. Zinovieva OL, Grineva EN, Prokofjeva MM, Karpov DS, Zheltukhin AO, Krasnov GS, et al. Expression of long non-coding RNA LINC00973 is consistently increased upon treatment of colon cancer cells with different chemotherapeutic drugs. *Biochimie.* 2018;151:67–72.
311. Jing C, Ma R, Cao H, Wang Z, Liu S, Chen D, et al. Long noncoding RNA and mRNA profiling in cetuximab-resistant colorectal cancer cells by RNA sequencing analysis. *Cancer Med.* 2019;8(4):1641–51.
312. Yu J, Zhang B, Zhang H, Qi Y, Wang Y, Wang W, et al. E2F1-induced upregulation of long non-coding RNA LMCD1-AS1 facilitates cholangiocarcinoma cell progression by regulating miR-345-5p/COL6A3 pathway. *Biochem Biophys Res Commun.* 2019;512(2):150–5.
313. Wang T, Fu X, Jin T, Zhang L, Liu B, Wu Y, et al. Aspirin targets P4HA2 through inhibiting NF- κ B and LMCD1-AS1/let-7g to inhibit tumour growth and collagen deposition in hepatocellular carcinoma. *EBioMedicine.* 2019;45:168–80.
314. Larach DB, Engoren MC, Schmidt EM, Heung M. Genetic variants and acute kidney injury: A review of the literature. *J Crit Care.* 2018;44:203–11.
315. Stafford-Smith M, Li YJ, Mathew JP, Li YW, Ji Y, Phillips-Bute BG, et al. Genome-wide association study of acute kidney injury after coronary bypass graft surgery identifies susceptibility loci. *Kidney Int.* 2015;88(4):823–32.
316. Wang Y, Li W, Chen X, Li Y, Wen P, Xu F. MIR210HG predicts poor prognosis and functions as an oncogenic lncRNA in hepatocellular carcinoma. *Biomed Pharmacother.* 2019;111:1297–301.

317. Li XY, Zhou LY, Luo H, Zhu Q, Zuo L, Liu GY, et al. The long noncoding RNA MIR210HG promotes tumor metastasis by acting as a ceRNA of miR-1226-3p to regulate mucin-1c expression in invasive breast cancer. *Aging (Albany NY)*. 2019 Aug 15;11(15):5646–65.
318. Wang AH, Jin CH, Cui GY, Li HY, Wang Y, Yu JJ, et al. MIR210HG promotes cell proliferation and invasion by regulating miR-503-5p/TRAF4 axis in cervical cancer. *Aging (Albany NY)*. 2020;12(4):3205–17.
319. Kang X, Kong F, Huang K, Li L, Li Z, Wang X, et al. LncRNA MIR210HG promotes proliferation and invasion of non-small cell lung cancer by upregulating methylation of CACNA2D2 promoter via binding to DNMT1. *Onco Targets Ther*. 2019;12:3779–90.
320. Huang H, Chen J, Ding CM, Jin X, Jia ZM, Peng J. LncRNA NR2F1-AS1 regulates hepatocellular carcinoma oxaliplatin resistance by targeting ABCC1 via miR-363. *J Cell Mol Med*. 2018;22(6):3238–45.
321. Guo F, Fu Q, Wang Y, Sui G. Long non-coding RNA NR2F1-AS1 promoted proliferation and migration yet suppressed apoptosis of thyroid cancer cells through regulating miRNA-338-3p/CCND1 axis. *J Cell Mol Med*. 2019;23(9):5907–19.
322. Qu CX, Shi XC, Zai LQ, Bi H, Yang Q. LncRNA CASC19 promotes the proliferation, migration and invasion of non-small cell lung carcinoma via regulating miRNA-130b-3p. *Eur Rev Med Pharmacol Sci*. 2019;23(2 Suppl):247–55.
323. Wang XD, Lu J, Lin YS, Gao C, Qi F. Functional role of long non-coding RNA CASC19/miR-140-5p/CEMIP axis in colorectal cancer progression in vitro. *World J Gastroenterol*. 2019;25(14):1697–714.
324. Guo K, Chen L, Wang Y, Qian K, Zheng X, Sun W, et al. Long noncoding RNA RP11-547D24.1 regulates proliferation and migration in papillary thyroid carcinoma: Identification and validation of a novel long noncoding RNA through integrated analysis of TCGA database. *Cancer Med*. 2019;8(6):3105–19.
325. Yang F, Deng X, Ma W, Berletch JB, Rabaia N, Wei G, et al. The lncRNA Firre anchors the inactive X chromosome to the nucleolus by binding CTCF and maintains H3K27me3 methylation. *Genome Biol*. 2015;16(1):52.
326. Shi X, Cui Z, Liu X, Wu S, Wu Y, Fang F, et al. LncRNA FIRRE is activated by MYC and promotes the development of diffuse large B-cell lymphoma via Wnt/ β -catenin signaling pathway. *Biochem Biophys Res Commun*. 2019;510(4):594–600.
327. Lu Y, Liu X, Xie M, Liu M, Ye M, Li M, et al. The NF- κ B-Responsive Long Noncoding RNA FIRRE Regulates Posttranscriptional Regulation of Inflammatory Gene Expression through Interacting with hnRNPU. *J Immunol*.

2017;199(10):3571–82.

- 328. Zang Y, Zhou X, Wang Q, Li X, Huang H. LncRNA FIRRE/NF- κ B feedback loop contributes to OGD/R injury of cerebral microglial cells. *Biochem Biophys Res Commun*. 2018;501(1):131–8.
- 329. Jafarzadeh M, Soltani BM, Soleimani M, Hosseinkhani S. Epigenetically silenced LINC02381 functions as a tumor suppressor by regulating PI3K-Akt signaling pathway. *Biochimie*. 2020;(171–172):63–71.
- 330. Yang B, Zhang L, Cao Y, Chen S, Cao J, Wu D, et al. Overexpression of lncRNA IGFBP4-1 reprograms energy metabolism to promote lung cancer progression. *Mol Cancer*. 2017;16(1):154.
- 331. Masson JJR, Cherry CL, Murphy NM, Sada-Ovalle I, Hussain T, Palchaudhuri R, et al. Polymorphism rs1385129 within glut1 gene SLC2A1 is linked to poor CD4+ T cell recovery in antiretroviral-treated HIV+ individuals. *Front Immunol*. 2018;9:900.
- 332. Yu S, Wang D, Shao Y, Zhang T, Xie H, Jiang X, et al. SP1-induced lncRNA TINCR overexpression contributes to colorectal cancer progression by sponging miR-7-5p. *Aging (Albany NY)*. 2019;11(5):1389–403.
- 333. Shao M, Chen G, Lv F, Liu Y, Tian H, Tao R, et al. LncRNA TINCR attenuates cardiac hypertrophy by epigenetically silencing CaMKII. *Oncotarget*. 2017;8(29):47565–73.
- 334. Liu Y, Du Y, Hu X, Zhao L, Xia W. Up-regulation of ceRNA TINCR by SP1 contributes to tumorigenesis in breast cancer. *BMC Cancer*. 2018;18(1):367.
- 335. Ding H, Liu J, Zou R, Cheng P, Su Y. Long non-coding RNA TPTEP1 inhibits hepatocellular carcinoma progression by suppressing STAT3 phosphorylation. *J Exp Clin Cancer Res*. 2019;38(1):189.
- 336. Ding H, Zhang X, Su Y, Jia C, Dai C. GNAS promotes inflammation-related hepatocellular carcinoma progression by promoting STAT3 activation. *Cell Mol Biol Lett*. 2020;25:8.
- 337. Dinh TA, Vitucci ECM, Wauthier E, Graham RP, Pitman WA, Oikawa T, et al. Comprehensive analysis of the Cancer Genome Atlas reveals a unique gene and non-coding RNA signature of fibrolamellar carcinoma. *Sci Rep*. 2017;7:44653.
- 338. Chen Z, Lin S, Li JL, Ni W, Guo R, Lu J, et al. CRTC1-MAML2 fusion-induced lncRNA LINC00473 expression maintains the growth and survival of human mucoepidermoid carcinoma cells. *Oncogene*. 2018;37(14):1885–95.
- 339. Zhang L, Wang Y, Li X, Xia X, Li N, He R, et al. ZBTB7A Enhances Osteosarcoma Chemoresistance by Transcriptionally Repressing lncRNALINC00473-IL24 Activity. *Neoplasia (United States)*. 2017;19(11):908–18.

340. Shi X, Wang X. LINC00473 mediates cyclin D1 expression through a balance between activation and repression signals in breast cancer cells. *FEBS Lett.* 2019;593(7):751–9.
341. Bai J, Zhao WY, Li WJ, Ying ZW, Jiang DQ. Long noncoding RNA LINC00473 indicates a poor prognosis of breast cancer and accelerates tumor carcinogenesis by competing endogenous sponging miR-497. *Eur Rev Med Pharmacol Sci.* 2019;23(8):3410–20.
342. Niu L, Zhou Y, Zhang W, Ren Y. Long noncoding RNA LINC00473 functions as a competing endogenous RNA to regulate MAPK1 expression by sponging miR-198 in breast cancer. *Pathol Res Pract.* 2019;215(8):152470.
343. Voellenkle C, Garcia-Manteiga JM, Pedrotti S, Perfetti A, De Toma I, Da Silva D, et al. Implication of Long noncoding RNAs in the endothelial cell response to hypoxia revealed by RNA-sequencing. *Sci Rep.* 2016;6:24141.
344. Min W, Dai D, Wang J, Zhang D, Zhang Y, Han G, et al. Long noncoding RNA miR210HG as a potential biomarker for the diagnosis of glioma. *PLoS One.* 2016;11(9):e0160451.
345. Li J, Wu QM, Wang XQ, Zhang CQ. Long noncoding RNA miR210HG sponges miR-503 to facilitate osteosarcoma cell invasion and metastasis. *DNA Cell Biol.* 2017;36(12):1117–25.
346. Li XY, Zhou LY, Luo H, Zhu Q, Zuo L, Liu GY, et al. The long noncoding RNA MIR210HG promotes tumor metastasis by acting as a ceRNA of miR-1226-3p to regulate mucin-1c expression in invasive breast cancer. *Aging (Albany NY).* 2019;11(15):5646–65.
347. Li M, Qiu M, Xu Y, Mao Q, Wang J, Dong G, et al. Differentially expressed protein-coding genes and long noncoding RNA in early-stage lung cancer. *Tumor Biol.* 2015;36(12):9969–78.
348. Yang S, Xu J, Zeng X. A six-long non-coding RNA signature predicts prognosis in melanoma patients. *Int J Oncol.* 2018;52(4):1178–88.
349. Li D, Qian X, Xu P, Wang X, Li Z, Qian J, et al. Identification of lncRNAs and Their Functional Network Associated with Chemoresistance in SW1990/GZ Pancreatic Cancer Cells by RNA Sequencing. *DNA Cell Biol.* 2018 Oct;37(10):839–49.
350. Haeussler M, Zweig AS, Tyner C, Speir ML, Rosenbloom KR, Raney BJ, et al. The UCSC Genome Browser database: 2019 update. *Nucleic Acids Res.* 2019;47(D1):D853–8.
351. Paul A, Paul S. The breast cancer susceptibility genes (BRCA) in breast and ovarian cancers. *Front Biosci - Landmark.* 2014;19:605–18.
352. Aso T, Tsai P, Kawaguchi T, Menninger JC, Kitajima S, Yasukochi Y, et al.

Assignment of the human GTF2F1 gene to chromosome 19p13.3. *Genomics*. 1993;16(1):252–3.

353. Wang F, Flanagan J, Su N, Wang LC, Bui S, Nielson A, et al. RNAscope: A novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues. *J Mol Diagnostics*. 2012 Jan;14(1):22–9.
354. Jin S, Yang X, Li J, Yang W, Ma H, Zhang Z. p53-targeted lincRNA-p21 acts as a tumor suppressor by inhibiting JAK2/STAT3 signaling pathways in head and neck squamous cell carcinoma. *Mol Cancer*. 2019 Dec 11;18(1):38.
355. Tan J, Yang L. Long noncoding RNA VPS9D1-AS1 overexpression predicts a poor prognosis in non-small cell lung cancer. *Biomed Pharmacother*. 2018 Oct 1;106:1600–6.
356. Sauvageau M. Diverging RNPs: Toward Understanding lncRNA-Protein Interactions and Functions. *Adv Exp Med Biol*. 2019;1203:285–312.
357. Noh JH, Kim KM, McClusky WG, Abdelmohsen K, Gorospe M. Cytoplasmic functions of long noncoding RNAs. *Wiley Interdiscip Rev RNA*. 2018 May;9(3):e1471.
358. Sun Q, Hao Q, Prasanth K V. Nuclear Long Noncoding RNAs: Key Regulators of Gene Expression. *Trends Genet*. 2018;34(2):142–57.
359. Wu K, Xu K, Liu K, Huang J, Chen J, Zhang J, et al. Long noncoding RNA BC200 regulates cell growth and invasion in colon cancer. *Int J Biochem Cell Biol*. 2018;99:219–25.
360. Gao Y, Feng B, Han S, Lu L, Chen Y, Chu X, et al. MicroRNA-129 in Human Cancers: From Tumorigenesis to Clinical Treatment. *Cell Physiol Biochem*. 2016;39(6):2186–202.
361. Behlke M. Mini-review on current strategies to knockdown long non-coding RNAs. *J Rare Dis Res Treat*. 2016;1(3):66–70.
362. Lee JS, Mendell JT. Antisense-Mediated Transcript Knockdown Triggers Premature Transcription Termination. *Mol Cell*. 2020;77(5):1044–54.
363. Valadkhan S, Valencia-Hipólito A. lncRNAs in stress response. *Curr Top Microbiol Immunol*. 2016;5:41.
364. Krzystek-Korpacka M, Hotowy K, Czapinska E, Podkowik M, Bania J, Gamian A, et al. Serum availability affects expression of common house-keeping genes in colon adenocarcinoma cell lines: implications for quantitative real-time PCR studies. *Cytotechnology*. 2016;68(6):2503–17.
365. Shih JW, Kung HJ. Long non-coding RNA and tumor hypoxia: New players ushered toward an old arena. *J Biomed Sci*. 2017;24(1):53.

366. Bellot G, Garcia-Medina R, Gounon P, Chiche J, Roux D, Pouyssegur J, et al. Hypoxia-Induced Autophagy Is Mediated through Hypoxia-Inducible Factor Induction of BNIP3 and BNIP3L via Their BH3 Domains. *Mol Cell Biol*. 2009;29(10):2570–81.
367. Wang L, Park HJ, Dasari S, Wang S, Kocher JP, Li W. CPAT: Coding-potential assessment tool using an alignment-free logistic regression model. *Nucleic Acids Res*. 2013;41(6):e74.
368. Michel AM, Fox G, M. Kiran A, De Bo C, O'Connor PBF, Heaphy SM, et al. GWIPS-viz: Development of a ribo-seq genome browser. *Nucleic Acids Res*. 2014;42(Database issue):D859-64.
369. Rubio CA, Weisburd B, Holderfield M, Arias C, Fang E, DeRisi JL, et al. Transcriptome-wide characterization of the eIF4A signature highlights plasticity in translation regulation. *Genome Biol*. 2014;15(10):476.
370. Fahl SP, Wang M, Zhang Y, Duc ACE, Wiest DL. Regulatory roles of Rpl22 in hematopoiesis: An old dog with new tricks. *Crit Rev Immunol*. 2015;35(5):379–400.
371. Hisamatsu T, McGuire M, Wu SY, Rupaimoole R, Pradeep S, Bayraktar E, et al. PRKRA/PACT expression promotes chemoresistance of mucinous ovarian cancer. *Mol Cancer Ther*. 2019;18(1):162–72.
372. Gal-Ben-Ari S, Barrera I, Ehrlich M, Rosenblum K. PKR: A kinase to remember. *Front Mol Neurosci*. 2019;11:480.
373. Ichihara-Tanaka K, Kadomatsu K, Kishida S. Temporally and Spatially Regulated Expression of the Linker Histone H1fx During Mouse Development. *J Histochem Cytochem*. 2017;65(9):513–30.
374. Pazo A, Pérez-González A, Oliveros JC, Huarte M, Chavez JP, Nieto A. HCLE/RTRAF-HSPC117-DDx1-FAM98B: A new cap-binding complex that activates mRNA translation. *Front Physiol*. 2019;10:92.
375. Nicholls C, Li H, Liu JP. GAPDH: A common enzyme with uncommon functions. *Clin Exp Pharmacol Physiol*. 2012;39(8):674–9.
376. Zhang JY, Zhang F, Hong CQ, Giuliano AE, Cui XJ, Zhou GJ, et al. Critical protein GAPDH and its regulatory mechanisms in cancer cells. *Cancer Biol Med*. 2015;39(8):674–9.
377. Jacquin MA, Chiche J, Zunino B, Bénétteau M, Meynet O, Pradelli LA, et al. GAPDH binds to active Akt, leading to Bcl-xL increase and escape from caspase-independent cell death. *Cell Death Differ*. 2013;39(8):674–9.
378. Garcin ED. GAPDH as a model non-canonical AU-rich RNA binding protein. *Semin Cell Dev Biol*. 2019;39(8):674–9.

379. Lau LF. CCN1/CYR61: The very model of a modern matricellular protein. *Cell Mol Life Sci.* 2011;68(19):3149–63.
380. Chen Y, Du XY. Functional properties and intracellular signaling of CCN1/Cyr61. *J Cell Biochem.* 2007;100(6):1337–45.
381. Saglam O, Dai F, Husain S, Zhan Y, Toruner G, Haines GK. Matricellular protein CCN1 (CYR61) expression is associated with high-grade ductal carcinoma in situ. *Hum Pathol.* 2014;45(6):1269–75.
382. Harris ME, Christian EL. RNA crosslinking methods. *Methods Enzymol.* 2009;468:127–46.
383. Muppirla UK, Honavar VG, Dobbs D. Predicting RNA-Protein Interactions Using Only Sequence Information. *BMC Bioinformatics.* 2011;12:489.
384. Wu P, Mo Y, Peng M, Tang T, Zhong Y, Deng X, et al. Emerging role of tumor-related functional peptides encoded by lncRNA and circRNA. *Mol Cancer.* 2020;19(1):22.
385. Aebersold R, Burlingame AL, Bradshaw RA. Western blots versus selected reaction monitoring assays: Time to turn the tables? *Mol Cell Proteomics.* 2013;12(9):2381–2.
386. Autour A, Jeng SCY, Cawte AD, Abdolazadeh A, Galli A, Panchapakesan SSS, et al. Fluorogenic RNA Mango aptamers for imaging small non-coding RNAs in mammalian cells. *Nat Commun.* 2018;9(1):656.
387. Banerjee S, Ghosh A, VonHoff DD, Banerjee SK. Cyr61/CCN1 targets for chemosensitization in pancreatic cancer. *Oncotarget.* 2019;10(38):3579–80.
388. Kazarian A, Blyuss O, Metodieva G, Gentry-Maharaj A, Ryan A, Kiseleva EM, et al. Testing breast cancer serum biomarkers for early detection and prognosis in pre-diagnosis samples. *Br J Cancer.* 2017;116(4):501–8.
389. Zhang K, Luo Z, Zhang Y, Wang Y, Cui M, Liu L, et al. Detection and analysis of circulating large intergenic non-coding RNA regulator of reprogramming in plasma for breast cancer. *Thorac Cancer.* 2018;9(1):66–74.
390. Yu G, Zhang W, Zhu L, Xia L. Upregulated long non-coding RNAs demonstrate promising efficacy for breast cancer detection: A meta-analysis. *Onco Targets Ther.* 2018;11:1491–9.