FACULTAD DE CIENCIAS BIOLÓGICAS
PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE

"CHARACTERIZATION OF THE EFFECTS INDUCED BY ADAR1 OVER LONG NON-CODING RNAS A-TO-I EDITING AND EXPRESSION LEVELS IN BREAST CANCER"

Thesis submitted to Pontificia Universidad Católica de Chile
in partial fulfilment of the requirements for the degree of
Doctor in Biological Sciences, Cell and Molecular Biology mention

by

PAMELA ROXANA ROJAS DE SANTIAGO

Thesis Advisor
RICARDO ARMISÉN YÁÑEZ, PhD

Thesis Co-advisor
MARCELA SJÖBERG HERRERA PhD

SEPTEMBER, 2020

DEDICATORY

Esta tesis está dedicada a mi familia, a mi madre Patricia y a mis hermanos André, Sebastián y Gonzalo. Ellos han estado siempre a mi lado y me han apoyado de manera incondicional en cada decisión de mi vida.

De forma muy especial quisiera dedicarles esta tesis a mis abuelos, Ana Illanes y Julio de Santiago, quienes fueron padres para mí y mis hermanos. Ellos nos criaron y amaron como si fuésemos sus propios hijos. Nos entregaron valores y una forma de enfrentar la vida que forjó todo lo que somos hoy. La reciente pérdida de mis abuelos ha sido lo más triste que me ha tocado vivir; cómo quisiera haberlos tenido para siempre a mi lado.

Finalmente quisiera dedicarle esta tesis a Benjamín, quien ha sido un pilar fundamental para mí. La amistad y el amor incondicional que nos tenemos no sólo fue primordial para superar cada tropiezo, sino que pavimentó el camino que llevamos juntos y el futuro que tenemos por delante.

ACKNOWLEDGEMENTS

The development of this thesis project would not have been possible without the support of a number of people, to whom I am very grateful.

First of all, special gratitude for my mentors Dr. Ricardo Armisén and Dr. Marcela Sjöberg, for being a constant source of support in this project. I consider myself privileged to have been able to work with them; they taught me and guided me in every step of this long way. I sincerely thanks Ricardo and Marcela because they are not only excellent scientists but also exceptional people; their encouragement was beyond academic extent.

My appreciation also extends to my mates from Dr. Armisén and Dr. Sjöberg laboratories, for their technical assistance and companion and emotional support. The bioinformatician team at the Center of Excellence in Precision Medicine from Pfizer Chile (CEMP-Pfizer Chile) was essential for developing this work.

A debt of gratitude is also owed to Dr. Olivier Harismendy and his team at the University of California San Diego (UCSD) for their support and trust, and for teach and guide me during the internship.

I am also sincerely grateful to my doctoral thesis committee, Dr. María Estela Andrés, Dr. Pilar Carvallo, and Dr. Gonzalo Olivares. I was fortunate to have such a combination of

respectable scientists. Their insightful comments, valuable suggestions, and critical vision of my project and presentations allowed me to improve and become a better scientist.

I would like to acknowledge the financial support of the Chilean Science Ministry through ANID (Agencia Nacional para la Investigación y Desarrollo), the Doctoral Scholarship, and the FONDECYT grants. Our country has limited resources; thus, receiving financial support for this research was a privilege. I also would like to thank the funding provided by CEMP-Pfizer Chile and their effort for developing high-quality science in Chile.

Above all, I want to express my sincere gratitude to my family and all my friends for helping me survive all the stress that implies this process and not let me give up. I cannot thank Benjamín enough for his endless support and for his companion in the hours when the lab's lights were off.

TABLE OF CONTENTS

FIGURE INDEX

INTRODUCTION

RESULTS

CHAPTER 1: Experimental model and RNA-sequencing

CHAPTER 2: ADAR1 and lncRNAs expression levels in breast cancer

CHAPTER 3: ADAR1 and lncRNAs A-to-I editing in breast cancer

DISCUSSION

Discussion: ADAR1 and lncRNAs expression levels in breast cancer

Discussion: ADAR1 and lncRNAs A-to-I editing in breast cancer

TABLE INDEX

INTRODUCTION

MATERIALS

RESULTS

CHAPTER 1: Experimental model and RNA-sequencing

CHAPTER 2: ADAR1 and lncRNAs expression levels in breast cancer

ABBREVIATIONS

ADAR:	Adenosine deaminase acting on RNAs
AJCC:	American Joint Committee on Cancer
A-to-I editing:	Adenine to Inosine editing
BAM:	Binary Alignment Map
BC:	Breast cancer
CIS:	Carcinoma *in situ*
CEMP:	Centro de Excelencia en Medicina de Precisión
DCIS:	Ductal carcinoma *in situ*
dsRBD:	Double stranded RNA-binding domains
dsRNA:	Double-stranded RNA
ER:	Estrogen receptor
FDR:	False discovery rate
fRIP:	Formaldehyde RNA-Immunoprecipitation
GATK:	Global analysis toolkit
GENCODE:	Generalized coding
GFP:	Green-fluorescent protein
GLOBOCAN:	Global cancer observatory
GRCh38:	Genome Reference Consortium Human genome build 38
HER2:	Human epidermal growth factor receptor 2
HR:	Hormone receptor
IBC:	Inflammatory breast cancer

IDC:	Invasive ductal carcinoma
ILC:	Invasive or infiltrating lobular carcinoma
LCIS:	Lobular carcinoma *in situ*
LncRNAs:	Long non-coding RNAs
LincRNAs:	Long intergenic non-coding RNAs
LINC00944:	Long intergenic non-coding RNA 944
mRNAs:	Messenger RNAs
miRNAs:	MicroRNAs
nt:	Nucleotide
OE:	Overexpression
ORF:	Open reading frame
OS:	Overall Survival
PCR:	Polymerase chain reaction
PIC:	Protease inhibitor cocktail
PR:	Progesterone receptor
QC:	Quality control
RESS-qPCR:	RNA Editing Site-Specific-qPCR
RFS:	Relapse-Free Survival
RIPA buffer:	Radioimmunoprecipitation assay buffer
RISC:	RNA-induced silencing complex
RNAi:	RNA interference
RNA-seq:	RNA sequencing
RNP:	Ribonucleoprotein

rRNA:	Ribosomal RNA
RT-qPCR:	Quantitative reverse transcription PCR
SEM:	Standard error of the mean
SDS-PAGE:	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
siRNAs:	Small interfering RNA
SNV:	Single nucleotide variant
ssRNA:	Single-stranded RNA
TCGA:	The Cancer Genome Atlas
TCGA-BRCA:	The Cancer Genome Atlas, Breast invasive carcinoma cohort
TCGA-COAD:	The Cancer Genome Atlas, Colon adenocarcinoma cohort
TCGA-LUAD:	The Cancer Genome Atlas, Lung adenocarcinoma cohort
TCGA-TGCT:	The Cancer Genome Atlas, Testicular Germ Cell Tumors cohort
TILs:	Tumor-infiltrating lymphocytes
TNBC:	Triple-negative breast cancer
TNM:	Tumor-node-metastasis

RESUMEN

La enzima Adenosina desaminasa que actúa sobre ARNs 1 (ADAR1, del inglés Adenosine deaminase acting on RNAs 1) ha sido ampliamente descrita como un factor modulador de la edición, niveles de expresión y función de sus ARNs blancos, pudiendo ser estos codificantes y no-codificantes. Entre los últimos, los ARNs largos no-codificantes (lncRNAs, del inglés long non-coding RNAs) (> 200 nt de largo) se han destacado por ser componentes centrales en procesos celulares tanto fisiológicos como patológicos. En cáncer de mama, ambos ADAR1 y lncRNAs, han sido caracterizados como elementos clave en vías de señalización oncogénicas y supresoras de tumores. Sin embargo, sólo unos pocos reportes en la literatura abordan la relación ADAR1-lncRNAs, quedando mucho por entender a escala transcriptómica. Por este motivo, esta tesis está enfocada principalmente en abordar los efectos inducidos por ADAR1 tanto en los niveles de expresión como en la edición de lncRNAs y cómo esto podría estar relacionado a la progresión del cáncer.

Mediante el uso de datos de secuenciación de ARN (RNA-seq), detectamos que ADAR1 puede modular la expresión de varios lncRNAs, encontrando que el RNA 944 intergénico largo no codificante (LINC00944) respondía de forma consistente a la ganancia y pérdida de función de ADAR1. Al analizar los datos de pacientes de la cohorte de cáncer de mama de The Cancer Genome Atlas (TCGA-BRCA), encontramos que bajos niveles de LINC00944 se correlacionan con fenotipos malignos, como una fracción menor de infiltración linfocitaria en el microambiente tumoral (TILs, del inglés tumor-*i*nfiltrating *l*ymphocytes) y con una disminución en la expresión de marcadores pro-apoptóticos. En la misma línea, encontramos que una baja expresión de LINC00944 se correlaciona con un mal pronóstico en pacientes, ya que la disminución en su expresión se correlacionó con una reducción en la supervivencia general y supervivencia libre de recaídas.

La sobreexpresión de ADAR1 se ha asociado a un mal pronóstico en pacientes con cáncer de mama triple-negativo (TNBC, del inglés triple-negative breast cancer). Al analizar la cohorte de TCGA-BRCA, demostramos que ADAR1 induce la edición A-por-I en aproximadamente el 10% de los lncRNA expresados en tumores TNBC. Asimismo, encontramos que éstos transcritos fueron editados en una alta proporción. En la presente tesis, nosotros ilustramos dos ejemplos de cómo la edición A-por-I podría estar alterando la función de lncRNAs: ya que PVT1 presentó el mayor número de posiciones editadas, hipotetizamos que su función de lncRNA esponja está siendo diversificada por la edición A-por-I y de esta forma permitiendo la progresión del cáncer. Por otro lado, planteamos la hipótesis de que la capacidad de PINK1-AS1 para estabilizar su ARNm sentido PINK1, puede verse alterada mediante la edición del lncRNA en tumores TNBC y que, en este contexto, podría ser maligna.

ABSTRACT

Adenosine deaminase acting on RNAs 1 (ADAR1) protein has been widely described as a modulating factor of RNA editing, expression levels, and function of its interacting coding and non-coding RNAs. Among the latter, long non-coding RNAs (lncRNAs) (≥ 200 nt in length) have emerged as central components in cellular processes in physiological and pathological conditions. In breast cancer, both ADAR1 and lncRNAs have been characterized as key components in oncogenic transformation and tumor-suppressor pathways. However, only a few reports in literature examine ADAR1-lncRNAs interplay, and there is still much to understand on a transcriptomic scale. For this reason, this thesis is focused on the effects induced by ADAR1 in lncRNAs expression levels and A-to-I editing, and how this could be related to cancer progression.

By using RNA-seq data, we detected that ADAR1 could modulate the expression of several lncRNAs, finding that the long intergenic non-coding RNA 944 (LINC00944) was the consistently responsive to ADAR1 gain- and loss-of-function. By analyzing patient's data from the breast cancer cohort of The Cancer Genome Atlas (TCGA-BRCA), we found that low levels of LINC00944 correlated with malignant phenotypes as a lower fraction of tumor-infiltrating lymphocytes (TILs) in the tumor microenvironment and with a decrease in pro-apoptotic markers. In the same line, we found that low LINC00944 expression is correlated to poor prognosis as a reduction in Overall Survival (OS) and Relapse-Free Survival (RFS) were found.

The overexpression of ADAR1 has been correlated to poor prognosis in patients with TNBC. By using data from the TCGA-BRCA cohort, we demonstrated that ADAR1 induces A-to-I editing in about 10% of expressed lncRNAs in TNBC tumors. Likewise, these transcripts were found to be edited in a high proportion. We postulate and illustrated two examples of how A-to-I editing lncRNAs could be altering their function: PVT1 was found as the lncRNA with the highest number of uniquely edited positions, and we hypothesized that the sponge function of PVT1 is being diversified by A-to-I editing and allowing tumorigenesis. On the other hand, we hypothesized that the ability of PINK1-AS1 to stabilize its sense mRNA PINK1 is being altered by editing in TNBC.

INTRODUCTION

1. Breast cancer

1.1 Definition and origin

Cancer is a term for referring diseases in which cells divide aberrantly and without control, leading to the invasion of adjacent tissues (*Definition of Cancer - NCI Dictionary of Cancer Terms - National Cancer Institute*). Cancer occurs as a consequence of genetic abnormalities, epigenetic alterations, and/or environmental factors that provide cells the ability to grow unrestrainedly, evade anti-growth signals, immune surveillance, and apoptosis, sustained angiogenesis, and reprogram metabolic pathways, among others (Um, 2015). In these terms, breast cancer results from of the acquisition of the features mentioned above of breast cells. Breast cancer occurs almost exclusively in women, though men can also be affected (for review, see Zygogianni et al., 2012).

Breast cancer can originate from different breast areas, being the lobules, the glands that produce milk, or the ducts, the conduits that transport milk from the lobules to the nipple, where cancer usually begins (**Fig. I**). Less frequently, breast cancer can initiate from connective tissues (fibrous and fatty tissue) of the breast (**Fig. I**) (*What Is Breast Cancer? | CDC*).

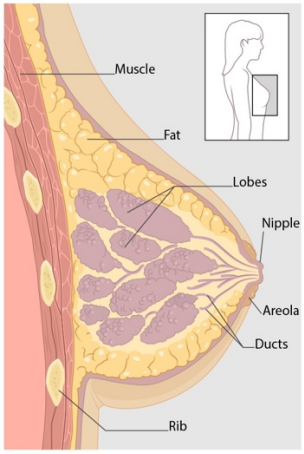


Figure I. Breast diagram. Breasts are located over the pectoral muscles. Breasts are made of fatty tissue and specialized tissue: lobes and ducts. The lobes contain the lobules, which are the structures that produced the milk, and the ducts are the conduits that transport the milk to the nipple. Image source: Centers for Disease Control and Prevention (CDC).

1.2 Breast cancer classification

Since breast cancer is a heterogeneous disease, we can find different morphological characteristics and molecular alterations among breast cancer cells, raising the possibility of different classification systems (Akram et al., 2017).

1.2.1 Immunohistochemical classification of breast cancer

Immunohistochemical (IHC) classification of breast cancer is based on a number of biomarkers that are useful for differentiating benign from malignant lesions, characterize tumor subtypes, and distinguish primary from metastatic tumors, among others (Zaha, 2014). The most

common IHC breast cancer markers used include the estrogen receptor (ER), the progesterone receptor (PR), the human epidermal growth factor receptor-2 (HER2), and the protein Ki-67 (Zaha, 2014).

Receptors of estradiol and progesterone, named estrogen receptor and progesterone receptor, respectively, also known as *hormone receptors*, are nuclear receptors involved in breast cancer development as they can activate many transduction pathways involved in pro-proliferative signaling in breast cells (for review, see Daniel et al., 2012; Duffy, 2006). Normal breast epithelial cells also contain those hormone receptors, but cancer cells contain a higher concentration, which confers them a higher proliferative capability (Carlson & Stockdale, 1988). Breast tumors expressing high levels of ER and/or PR are named ER-positive (ER^+) and/or PR-positive (PR^+), respectively, and are known as *hormone-receptor-positive (HR^+) tumors*.

The human epidermal growth factor receptor 2 (HER2) (*ErbB2*) is a transmembrane receptor tyrosine kinase involved in cell proliferation and survival (for review, see Connell & Doherty, 2017). HER family of receptors constitutes the main driver of tumor growth in approximately 20% of breast tumors (Rimawi et al., 2015). Breast cancer cells with higher than normal levels of HER2, mainly due to gene amplification, are called HER2-positive *($HER2^+$)* (Zaha, 2014).

Triple-negative breast cancer (TNBC) tumors do not express ER, PR, and HER2; thus, its proliferative capability is not driven by the signaling pathways transduced by the mentioned receptors. TNBC has aggressive behavior and poor patient prognosis. TNBC accounts for approximately 15-20% of new cases of breast cancer (for review, see Shen Zhao et al., 2020).

The protein Ki-67 has been extensively studied for its role in mitosis and cell cycle progression (for review, see Sun et al., 2018). Thus, it is used as an indicator of cancer cell proliferation. According to Ki-67 protein levels, breast tumors are categorized as low, intermediate, and highly proliferating (Zaha., 2014).

1.2.2 Breast cancer classification based on the tumor-node-metastasis (TNM) system

The extent of cancer (stage), such as the localization and the size of the tumor, and if it has spread, is a crucial factor that defines prognosis and is critical in determining proper treatment. The most clinically useful staging system is known as the tumor-node-metastasis (TNM) system. It allows the categorization of tumors by the size and magnitude of the primary tumor (T), spreading to regional lymph node (N), and the presence, or not, of distant metastases (M) (American Joint Committee on Cancer, AJCC; 7th edition) (Edge & Compton, 2010). A brief description of the classification of the TNM categories is shown in Table I.

The standardized criteria of the TNM staging system are primarily used to classify solid tumors, comprising the main method for cancer reporting. Importantly, it is regularly used to

make a prognosis and propose individual treatment plans for patients (American Joint Committee on Cancer, AJCC; 7th edition) (Edge & Compton, 2010).

Table I. Tumor Node Metastasis (TNM) staging system. Description of TNM staging system according to the American Joint committee on Cancer, AJCC; 7th edition (Edge & Compton, 2010).

Primary tumor (T)	
T0	Primary tumor cannot be found
T1-T4	Refers to the size of the primary tumor The higher the number, the larger the tumor
TX	Primary tumor cannot be measured
Regional lymph nodes (N)	
N0	There is no cancer in nearby lymph nodes (LN)
N1-N3	Refers to the number and location of LN containing cancer The higher the number, the more LN containing cancer
NX	Cancer cells in nearby lymph nodes cannot be measured
Distant metastasis (M)	
M0	Cancer has not spread to other parts of the body
M1	Cancer has spread to other parts of the body
MX	Metastasis cannot be measured

1.3 Breast cancer and the immune system

The hallmarks of cancer, which include sustaining proliferative signaling, evading growth suppressors, resisting cell death, enabling replicative immortality, inducing angiogenesis, and activating invasion and metastasis, constitute a structured rationale for understanding the extent of neoplastic disease (Hanahan & Weinberg, 2000). Research showed later that additional two key hallmarks should have been included as central for neoplastic transformation: reprogramming of energy metabolism and evading immune destruction. This last hallmark arose from the concept that the immune system is continuously monitoring cells and tissues, detecting and destroying initial cancer cells (immunosurveillance). Accordingly, the growth of tumors implies a failure in the immunosurveillance (Hanahan & Weinberg, 2000).

Several studies have addressed the crosstalk between the immune system and cancer cells. At first sight, contradictory outcomes were found, as, in some context, the immune system seems to act as a pro-tumorigenic factor and in others as anti-tumorigenic. In breast cancer, the current idea indicates that this will depend on the type of immune cell participating in the process. $CD8^+$ T cells, $CD4^+$ T-helper 1 cells, natural killer cells, B cells, macrophages M1, and mature dendritic cells are considered anti-tumorigenic; while $CD4^+$ T-helper 2 cells, regulatory B cells, $CD4^+$ regulatory T cells, myeloid-derived suppressor cells (MDSCs), and macrophages M2 are considered pro-tumorigenic (for review, see Edechi et al., 2019).

A major path for cancer cells to avoid immunosurveillance is inactivating the population of lymphocytes that are infiltrated into the tumor microenvironment (for review, see

Badalamenti et al., 2019). Among the different immune-infiltrated populations, tumor-infiltrating lymphocytes (TILs), which are a selected population of T-cells, have a higher specific immunological reactivity against cancer cells than the non-infiltrating lymphocytes (Badalamenti et al., 2019). Therefore, the infiltration of T lymphocytes into the tumor microenvironment has been considered as an anti-tumorigenic response and has been positively associated with patient prognosis (Badalamenti et al., 2019; Pruneri et al., 2018).

1.4 Breast cancer statistics

Breast cancer represents the second leading cause of cancer-related deaths in women. The Global Cancer Observatory (GLOBOCAN) projected 2,088,849 cases in 2018 worldwide, constituting 11.6% of total cancer cases (Bray et al., 2018). According to the American Cancer Society, in 2020, there will be diagnosed about 276,480 cases of invasive breast cancer and about 48,530 new cases of carcinoma *in situ* (CIS) only in the United States (US), while about 42,170 US women will die from breast cancer (Siegel et al., 2020). In Chile, there were 53,365 new cancer cases in 2018, of which 5,393 corresponded to breast cancer, representing 10.1% of total new cases. The same year, 1,688 women died from breast cancer (GLOBOCAN 2018, Chile) (Bray et al., 2018).

These alarming numbers and the complex biology of breast cancer highlight the extent of this major public health problem and evidence the need for broader knowledge in basic research, diagnosis, and breast cancer treatment.

2. Adenine deaminase acting on RNAs (ADAR)

2.1 Adenine deaminase acting on RNAs (ADAR) proteins and structure

In mammals, the Adenosine deaminase acting on RNAs (ADAR) family of proteins is compose of three members: ADAR1 (official symbol: *ADAR*), which has two ubiquitously expressed isoforms, ADAR1-p110 (constitutive expression) and ADAR1-p150 (interferon-inducible expression) (Kawakubo & Samuel, 2000; U. Kim et al., 1994), ADAR2 (official symbol: *ADARB1*), which is mainly but not exclusively expressed in brain tissue (Melcher et al., 1996), and ADAR3 (official symbol: *ADARB2*), whose expression is restricted to the brain (**Fig. II**) (C. Chen et al., 2000; Melcher et al., 1996).

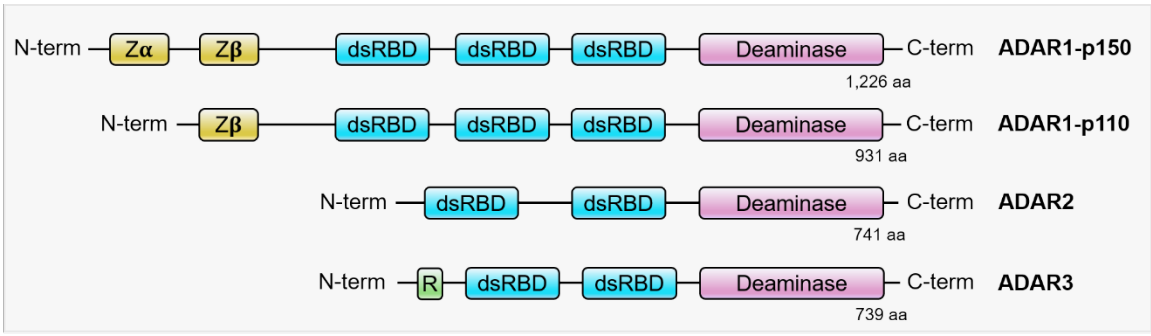


Figure II. ADAR proteins structure. ADAR1 and its major isoforms ADAR1-p150 and ADAR1-p110, ADAR2 and ADAR3. Deaminase domain, double-stranded RNA-binding domain (dsRBD), Z-DNA-binding domains (Zα and Zβ), and the R domain of single-stranded RNA-binding domain. N-term: amino terminal; C-term: carboxyl-terminal; aa: amino acid.

The general structure of ADAR proteins consists of a catalytic *deaminase domain,* followed by a variable number of *double-stranded RNA-binding domains* (dsRBDs) (**Fig. II**). Additionally, two *Z-DNA-binding domains* (Zα and Zβ) are observed in ADAR1-p150, and one

in ADAR1-p110 (Zβ) (**Fig. II**). These domains recognize the left-handed helical variant of DNA, but their biological significance in the protein remains elusive. At its amino-terminal region (N-term), ADAR3 holds a single-stranded RNA (ssRNA)-binding domain called *R domain*, whose functional meaning is also unknown (**Fig. II**) (C. Chen et al., 2000; Nishikura, 2016).

The *deaminase domain* allows ADAR1 and ADAR2 to perform their canonical function known as *A-to-I RNA editing*. This reaction consists of the hydrolytic removal of the amine group of Adenines (A) in double-stranded RNAs (dsRNAs), to convert them into Inosines (I) (**Fig. III**). ADAR homodimerization and the dsRNA-binding domain are necessary for direct contact with their targets and subsequently editing. It has been proposed that ADAR3 lacks this enzymatic activity (C. Chen et al., 2000).

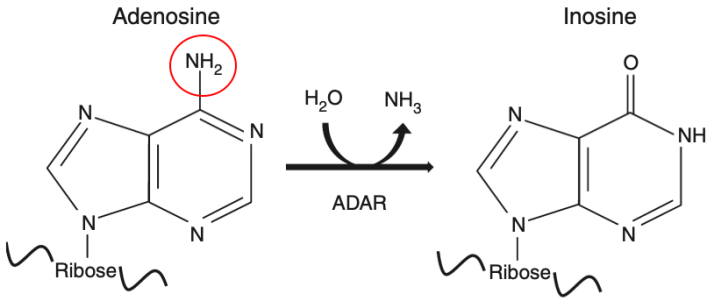


Figure III. Schematic diagram of A-to-I RNA editing reaction. In a hydrolytic reaction, ADARs can remove the amino group of adenines on dsRNAs, converting them into inosines (Deffit et al., 2016).

2.2 RNA A-to-I editing and its biological relevance

Since the high similarity between inosine and guanosine structures (**Fig. IV**), the spliceosome machinery recognizes inosines as guanosines. In this way, RNA editing can generate donor (IU as canonical GU) or acceptor (AI as canonical AG) splice sites or alter splicing regulatory elements (SREs) within exons (**Fig. V**) (Hogg et al., 2011; Rueter et al., 1999; Solomon et al., 2013). Translation machinery also reads inosines as guanosines, leading to modifications in codon sequences known as recoding, or even to introduction of an alternative start or stop codon (**Fig. VI**) (Higuchi et al., 2000; Hogg et al., 2011; Picardi et al., 2015; Pokharel & Beal, 2006). The expression and stability of RNAs also can be affected by ADARs editing, mainly by altering the base-pair recognition of small regulatory RNAs (Borchert et al., 2009) (**Fig. VII**). Besides, RNA editing can affect the subcellular localization of edited RNAs, as nuclear proteins that preferentially bind to dsRNAs containing-inosines (dsRNA-I) can retain them in the nucleus, such as P54NRB, PSF, and PSPC-1 (Cao et al., 2015; Salameh et al., 2015).

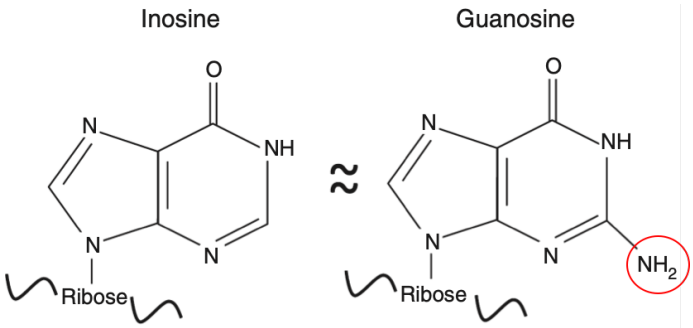


Figure IV. Inosine and Guanosine structures are highly similar. Since the high similitude in their structures, cellular machineries, as spliceosome machinery and translation machinery, recognize inosines as guanosines, having a large impact on different cellular processes. (Deffit et al., 2016)

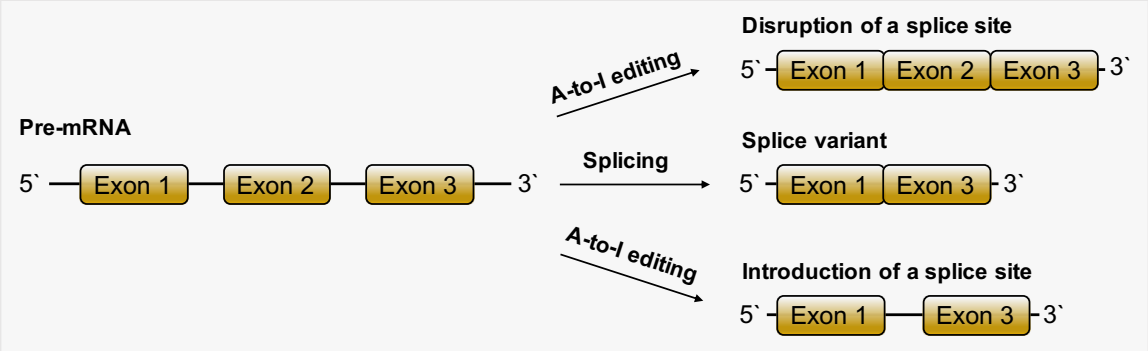


Figure V. A-to-I editing can impair splicing. Editing can disrupt splice sites (upper panel) or introduce new splice sites (bottom panel).

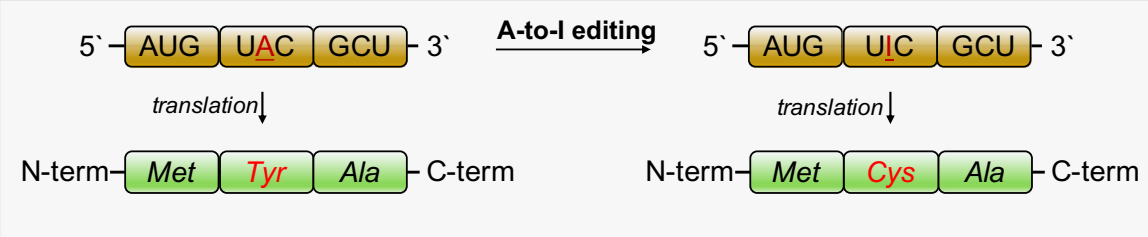


Figure VI. Recoding event induced by A-to-I editing. A-to-I editing in mRNA can alter the exon sequence thereby the encoded amino acid.

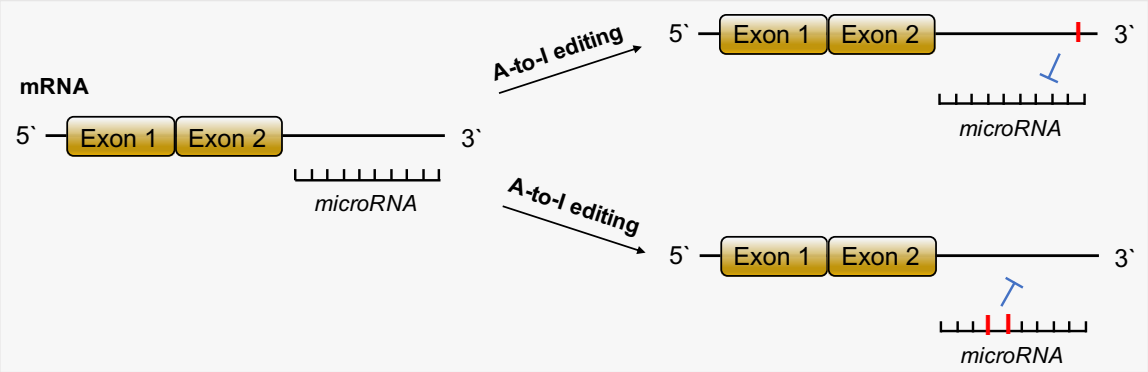


Figure VII. A-to-I editing can impair 3'UTR targeting by microRNAs. RNA editing either in 3'UTR or in miRNA sequence can impair the regulation of miRNA over mRNA expression.

2.3 ADAR1 and its non-canonical functions

ADAR1 function has been characterized beyond the extent of A-to-I editing. The interaction between ADAR1 and protein members of the RNA-induced silencing complex (RISC), such as Dicer, evidenced an important role of ADAR1 in the miRNA processing and RNA interference (RNAi) mechanisms (Heale et al., 2009; Ota et al., 2013). In fact, the lethal embryonic phenotype observed in ADAR1$^{-/-}$ mouse embryos was found tightly related to the global inhibition of miRNAs rather than the editing capability (Ota et al., 2013).

In cancer, the ADAR1/Dicer interaction has also been found to be important. For instance, ADAR1/Dicer interaction can augment the processing of the pre-miR-27a to mature miR-27a, which in turn targets the tumor suppressor gene METTL7A (Methyltransferase Like 7A). In this way, the expression levels of METTL7A are downregulated in hepatocellular carcinoma (HCC), having a key impact on the development of the pathology (Qi et al., 2017).

In an editing-independent manner, ADAR1 can cooperate and/or disrupt mechanisms mediated by other RNA-biding proteins, such as Staufen and the (STAU-1) mediated mRNA decay (SMD), or with the HuR and its role in RNA stability. By these means, ADAR1 has a major impact on transcript stability and protein expression (Sakurai et al., 2017; I. X. Wang et al., 2013).

3. Long non-coding RNAs (lncRNAs)

3.1 Long non-coding RNA (lncRNA) definition

High-throughput sequencing technologies have uncovered that more than 85% of the human genome is transcribed, being only 2-3% protein-coding genes (Hangauer et al., 2013). The vast majority of produced transcripts corresponded to non-protein-coding transcripts, known as non-coding RNAs (ncRNAs) (Hangauer et al., 2013). These ncRNAs comprise a diverse and complex group which has been commonly classified into small non-coding RNAs (small-RNAs) (< 200 nt in length) and long non-coding RNAs (lncRNAs) (> 200 nt in length) (Frankish et al., 2019; K. C. Wang & Chang, 2011).

The GENCODE consortium has been manually annotating a comprehensive set of lncRNAs for several years and has made the data publicly available. Non-coding transcripts are initially classified in this category due to (1) the lack of homology with any protein, (2) no reasonable-sized open reading frame (ORF), and (3) no high conservation in exons. The first catalog of human lncRNAs contained 9,277 lncRNA loci (14,880 transcripts) (Derrien et al., 2012), while the current version contains 17,960 lncRNA genes (GENCODE GRCh38 v.34).

The GENCODE lncRNAs catalog comprises long non-coding transcripts that account for different biological characteristics (biotypes). In Table II, we can find their description.

Table II. Long non-coding RNAs (lncRNAs) biotypes according to GENCODE annotations. Description of biotypes composing lncRNAs according to GENCODE project (Frankish et al., 2019).

Biotype	Description
3' overlapping ncRNA	Transcripts where published experimental data strongly supports the existence of long non-coding transcripts that overlap the 3'UTR of a protein-coding locus on the same strand.
Antisense	Transcripts that overlap the exon or introns of a protein-coding locus on the opposite strand.
lincRNA	Transcripts that are long intergenic non-coding RNA locus with a length > 200bp. Requires lack of coding potential.
Macro lncRNA	Un-spliced lncRNAs that are several kb in size.
Non-coding	Transcripts which are known from the literature to not be protein coding.
Processed transcript	Transcripts that doesn't contain an open reading frame (ORF).
Retained intron	An alternatively spliced transcript believed to contain intronic sequence relative to other, coding, transcripts of the same gene.
Sense intronic	A long non-coding transcript in introns of a coding gene that does not overlap any exons.
Sense overlapping	A long non-coding transcript that contains a coding gene in its intron on the same strand.

3.2 Main characteristics of lncRNAs

Key characteristics defined the group of lncRNAs. LncRNAs show low or no evidence of coding potential, as they do not show ORFs of higher quality than expected of random sequences (Derrien et al., 2012). LncRNAs have general low expression and show more tissue-specific patterns than protein-coding genes (Derrien et al., 2012; Ørom et al., 2010; Ravasi et al., 2006). The vast majority of lncRNAs are spliced and show canonical splices sites (GT/AG) (Derrien et al., 2012). Overall, lncRNA transcripts are shorter than coding transcripts (median 592 bp vs. 2453 bp, respectively), and about a 40% contain poly(A) motifs (Derrien et al., 2012). Their subcellular localization may be either nuclear or cytoplasmic; however, lncRNAs are preferentially localized in chromatin and nuclear RNA fractions (Derrien et al., 2012). Histone profiles of active marks in lncRNAs transcription start sites (TSS), as H3K4me2, H3K4me3, H3K9ac, and H3K27ac, are similar to those of protein-coding genes; and hold a slight excess of other histone marks related to both silencing (H3K27me3) and activation (H3K36me3) (Sati et al., 2012). Interestingly, despite that lncRNAs lack coding potential, a small fraction can generate peptides (Jia et al., 2012), and some lncRNAs may be post-processed into functional smaller RNAs (Derrien et al., 2012).

3.3 Classification of lncRNAs

Annotation and classification of lncRNAs is a developing process; hence it is not exempt from overlapping terminology and ambiguity. Here, we describe the most common and robust classifications used for lncRNAs, which are based on their *location in the genome* and *function*.

3.3.1 Classification of lncRNAs according to their genome location

According to their *genome location* with respect to protein-coding genes (PCG), lncRNAs can be classified as *long* or *large intergenic non-coding RNAs*, *overlapping sense transcripts*, *antisense*, *intronic* and *bidirectional lncRNAs* (**Fig. VIII**).

(i) *Long intergenic non-coding RNAs (lincRNAs),* also known as *large intervening ncRNAs* (**Fig. VIII-A**), do not intersect with any protein-coding gene and ncRNA gene annotation (intergenic lncRNAs). They are usually shorter than protein-coding genes, transcribed by RNA polymerase II, 5´capped, 3´polyadenylated, and undergo splicing. (ii) *Overlapping sense lncRNAs*, also known as *sense lncRNAs* (**Fig. VIII-B**), are intragenic as they encompass exons or whole protein-coding genes within their introns without any sense exon overlap and are transcribed in the same sense direction. (iii) *Antisense lncRNAs* (*asRNAs* or *ancRNAs*) (**Fig. VIII-C**) are transcribed from the opposite strand of a protein-coding gene locus. This classification includes some *natural antisense transcripts* (NATs), which may be subdivided into *cis-NATs* if they impact the expression of the corresponding sense genes, and into *trans-NATs* if they control the expression of genes from other genomic locations. (iv) *Intronic lncRNAs* (**Fig. VIII-D**) are restricted to protein-coding gene introns and could be either expressed independently of the protein-coding gene host or by-products of pre-mRNA processing. (v) *Bidirectional lncRNAs* (**Fig. VIII-E**) are originated from the opposite strand of a protein-coding gene (within 1 kb of promoters). These ncRNAs are underestimated mainly as they are highly unstable and thus difficult to detect (Jarroux et al., 2017).

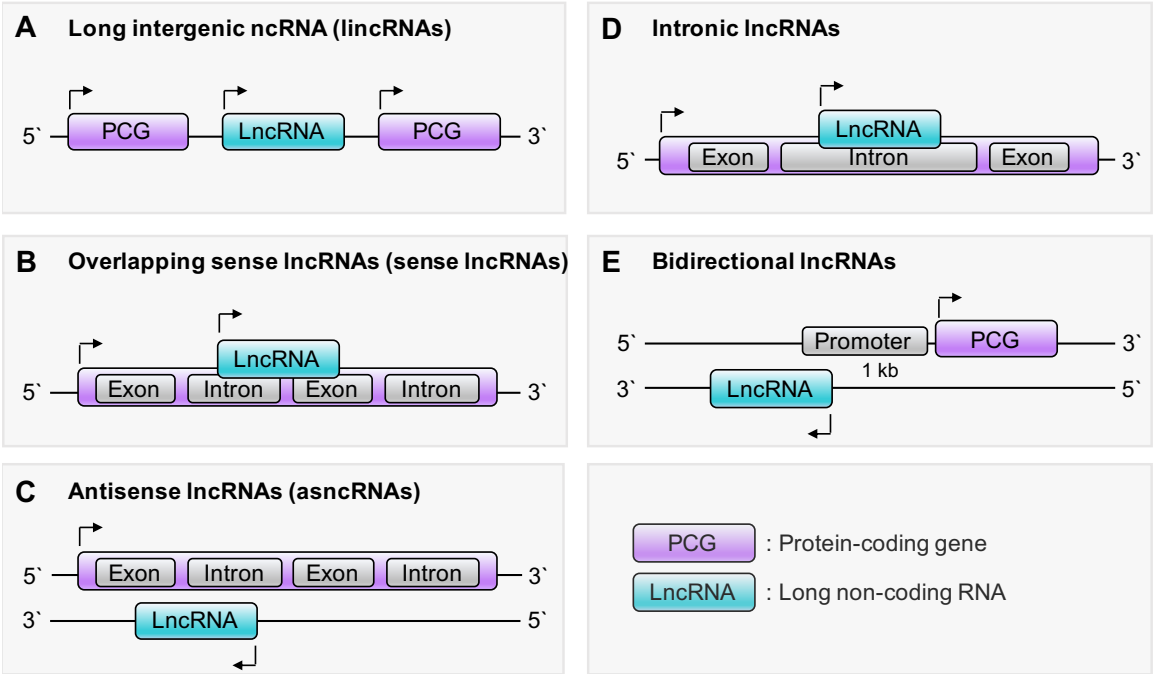


Figure VIII. Long non-coding RNAs classification according to their location in the genome. (A) Long intergenic non-coding RNAs (lincRNAs), **(B)** Overlapping sense lncRNAs, **(C)** Antisense lncRNAs (asncRNAs), **(D)** Intronic lncRNAs, and **(E)** Bidirecional lncRNAs. PCG: protein-coding genes; lncRNAs: long non-coding RNAs.

3.3.2 Classification of lncRNAs according to their functions

To highlight their regulatory role, lncRNAs are classified from a functional perspective. (i) *Scaffolds lncRNAs* act in the assembly of ribonucleoprotein (RNP) complexes, as the structural plasticity allows them to adopt dynamic three-dimensional structures with a high affinity for proteins (**Fig. IX-A**). (ii) *Guide lncRNAs* recruit RNP complexes to specific chromatin loci (**Fig. IX-B**). (iii) *Decoy lncRNAs*, also known as *ribo-repressor lncRNAs*, repress protein activities by inducing allosteric modifications, inhibition of catalytic activity, or blocking their binding sites. *Ribo-activator lncRNAs* act by enhancing protein activities (**Fig. IX-C**). (iv) *Sponge lncRNAs*, also known as *competing endogenous RNAs* (*ceRNAs*), share

partial sequence similarity to protein-coding transcripts and function by competing for microRNA binding (**Fig. IX-D**). Lastly, (v) *precursor lncRNAs* host small RNAs genes (miRNAs, siRNAs, piRNAs) that are mainly involved in the RNAi pathway (**Fig. IX-E**) (Jarroux et al., 2017).

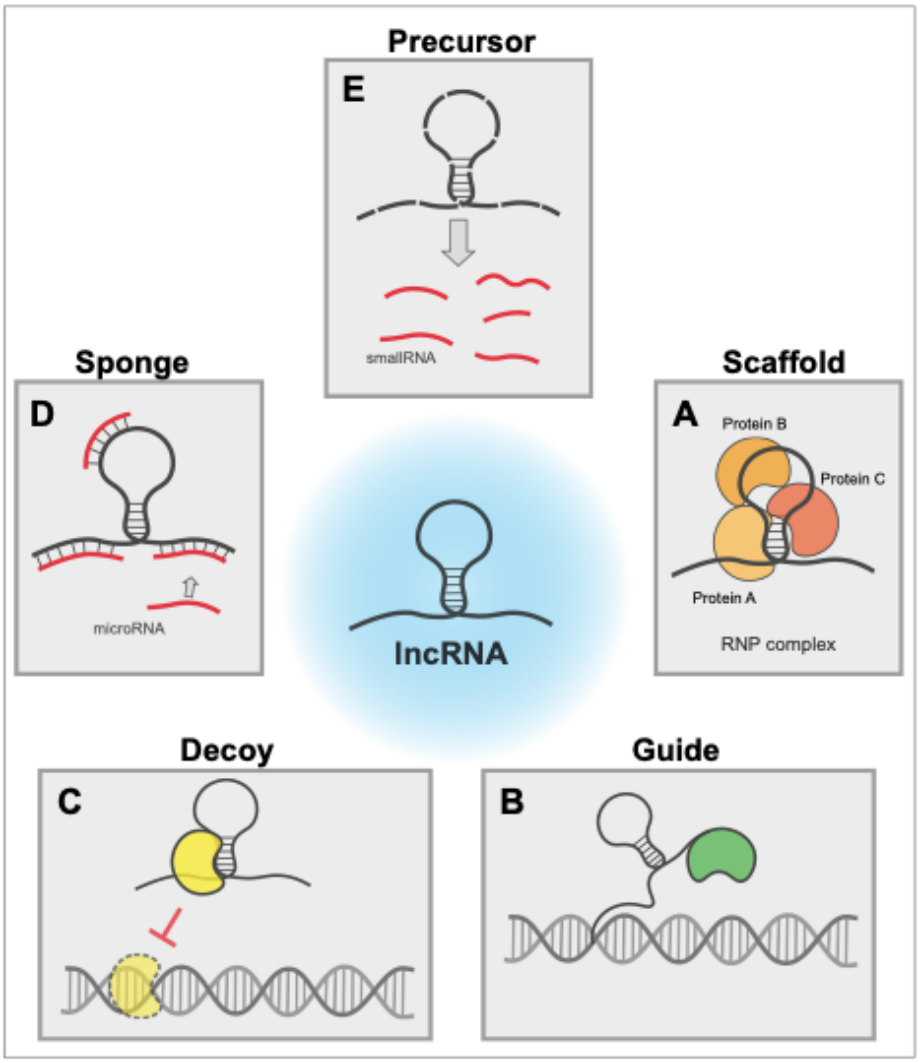


Figure IX. Functional classification of lncRNAs. According to their functions, lncRNAs can be classified in **(A)** Scaffold, **(B)** Guide, **(C)** Decoy, **(D)** Sponge, and **(E)** Precursor.

4. Laboratory findings: ADAR1 and Breast cancer

Our laboratory has been trying to understand how the documented upregulation of ADAR1 in breast cancer could contribute to pathology's development and progression.

4.1 ADAR1 mRNA expression and RNA editing are upregulated in breast cancer

By using public data deposited in the breast cancer cohort from The Cancer Genome Atlas (TCGA-BRCA), we showed that ADAR1 expression is upregulated in tumor samples coming from the principal intrinsic subtypes of breast cancer, Basal-like, Her2-enriched, Luminal B, and Luminal A compared to the normal samples (**Fig. X-A**). Interestingly, tumor samples have a 1.24-fold change (FC) increase ($p < 0.001$) in the number of A-to-I variant counts, which are representative of A-to-I editing (**Fig. X-B**). This increase in variants correlated with ADAR1 expression (*Pearson* $r = 0.7$, $p < 0.0001$), and discretely with ADAR2 expression (*Pearson* $r = -0.08$, $p < 0.01$), but no with ADAR3 levels (*Pearson* $r = -0.03$, $p < 0.38$), suggesting a primary role for ADAR1 in breast cancer (**Fig. X-C**) (Sagredo et al., 2018).

In literature, we can find other studies addressing the upregulation of ADAR1 and aberrant A-to-I editing in tumor samples, not only in breast cancer (Fumagalli et al., 2015; Nakano et al., 2017) but also in lung adenocarcinoma (Anadón et al., 2016), cervical cancer (Y. Chen et al., 2017), hepatocellular carcinoma (L. Chen et al., 2013), esophageal squamous cell carcinoma (Qin et al., 2014) and gastric cancer (Chan et al., 2016), among others (Paz-Yaacov et al., 2015), supporting the central role of ADAR1 in cancer biology.

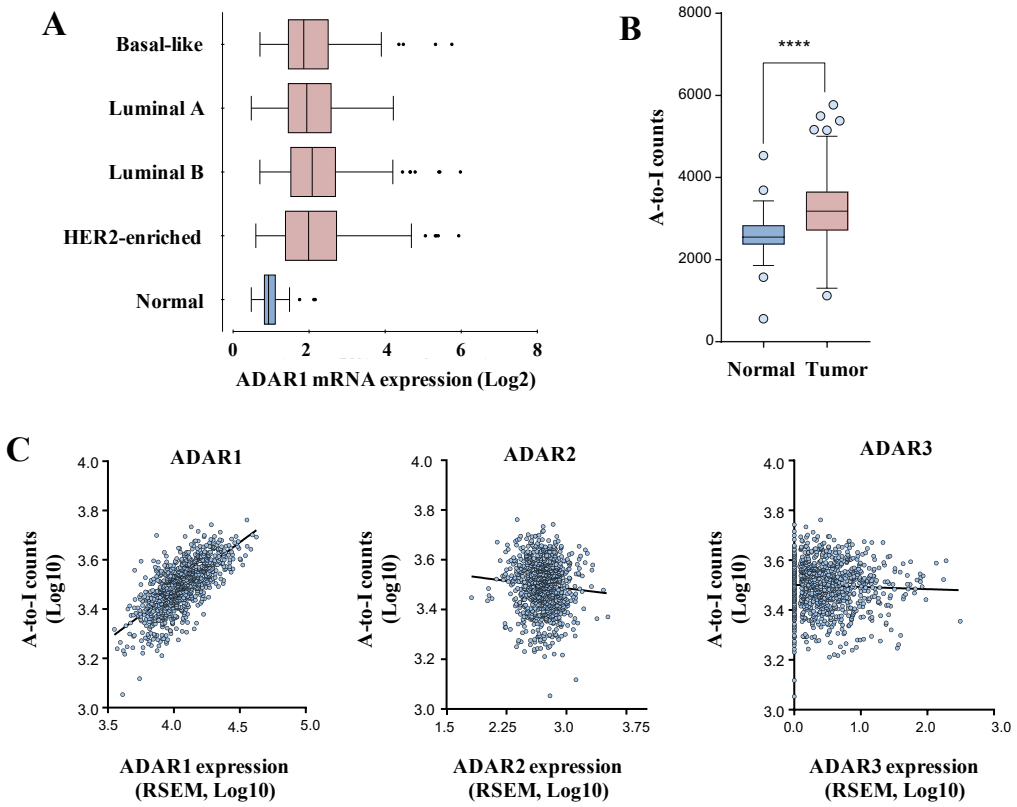


Figure X. ADAR1 mRNA expression and A-to-I editing is upregulated in breast cancer. (A) ADAR1 mRNA expression from microarray data of breast cancer patients from the TCGA-BRCA cohort. **(B)** A-to-I counts which is representative of RNA editing is increased in tumor samples from the TCGA-BRCA cohort. **** *Kolmogorov–Smirnov test*, $p < 0.001$. **(C)** ADARs expression and number of A-to-I counts correlation. A-to-I counts correlates with ADAR1 but no ADAR2 or ADAR3 mRNA expression. Pearson Correlation $r = 0.679$ $p < 0.0001$ (ADAR1); $r = -0.077$ $p < 0.01$ (ADAR2); $r = -0.028$ $p < 0.3765$ (ADAR3). RSEM: RNA-Seq by Expectation-Maximization.

4.2 ADAR1 upregulation correlated to poor survival in breast cancer patients

Our analyses showed that TCGA patients with basal-like breast cancer overexpressing ADAR1 significantly reduced their overall survival rate compared to patients with low or normal ADAR1 expression levels (*Log-rank Mantel-Cox test, p* = 0.02) (**Fig. XI-A**). Noteworthy, censored basal-like cancer patients with higher ADAR1 expression have a significant increase in the number of editing counts, whereas patients with lower ADAR1 expression showed lower edited sites (3683 ± 79.74 versus 2928 ± 103.3 edited sites, mean ± SEM, *Student t-test p* < 0.0001) (**Fig. XI-B**) (Sagredo et al., 2018).

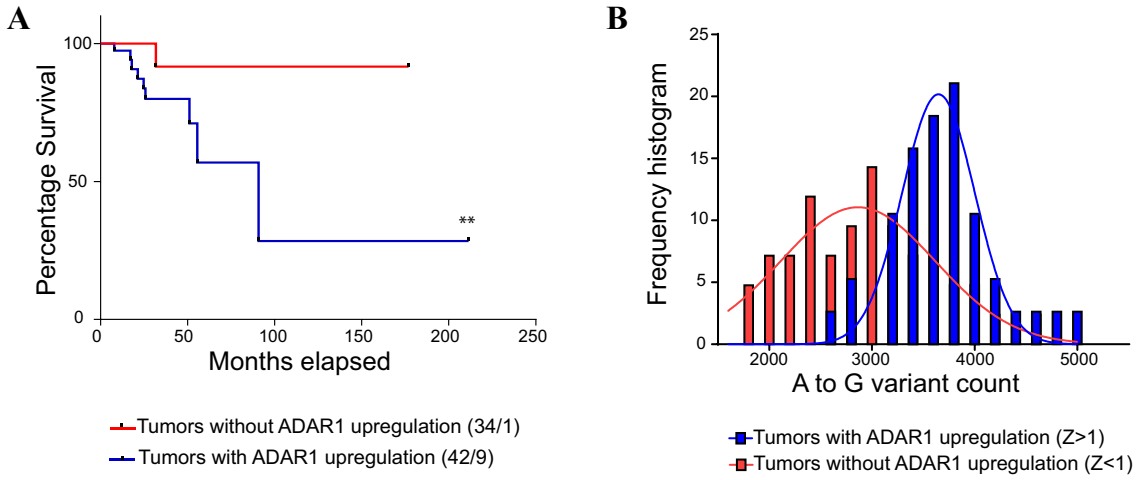


Figure XI. Increase in ADAR1 expression and A-to-I editing is related to breast cancer pathogenesis in basal-like patients. (A) Kaplan-Maier survival proportions for basal-like patients stratified based on ADAR1 mRNA expression levels. ** *Log-rank Mantel-Cox test, p* = 0.02. **(B)** Histogram proportion for basal-like tumors according to their A to G(I) variant count, showing the basal-like patients selected for Kaplan-Maier analysis. Gauss fit distribution for each subgroup is shown with a continuous line. Patients with low ADAR1 expression are depicted in red (Z < 1), and patients with high ADAR1 expression in blue (Z < 1) (Sagredo et al., 2018).

4.3 Role of ADAR1 in cancer biology

Our studies have shown that ADAR1 editing can regulate the expression and stability of a subset of genes related to cell cycle and DNA repair, suggesting that cell proliferation and genomic instability are likely to be associated with the role of ADAR1 in breast cancer progression (Sagredo et al., 2020).

Moreover, by studying ADAR1 overexpression in mice tumors, we have seen that high levels of ADAR1 could promote invasion and neoangiogenesis through activation of the canonical Wnt signaling pathway, suggesting an unknown role of ADAR1 in the aggressiveness of breast cancer tumors (Morales et al., 2020, manuscript under revision).

Taken together, the evidence suggests a role for ADAR1 in cancer biology through a plethora of molecular mechanisms. Nevertheless, our work and the vast majority of published data that address this topic focus on the effects induced by ADAR1 in mRNAs, existing a considerable gap in the understanding of these effects over other important types of RNAs, such as long non-coding RNAs.

5. ADAR1 and lncRNAs

5.1 LncRNAs and the canonical ADAR1 function of RNA editing

Few studies have been published about the regulatory role of ADAR1 editing on lncRNAs in cancer and how this could impact tumor malignancy. Nevertheless, a clear picture is emerging.

In prostate cancer, the antisense lncRNA PCA3 (prostate cancer antigen 3) can promote malignant cell growth by controlling the expression levels of the tumor suppressor gene PRUNE2. PCA3 is transcribed in the antisense direction of gene-encoded PRUNE2, both transcripts base-pair and form an RNA duplex in which ADAR1 performs A-to-I editing. The edited dsRNA is retained in the nucleus following the interaction with nuclear proteins that specifically bind RNAs containing inosines (RNAs-I), such as P54NRB, PSF, and PSPC-1. By this mechanism, ADAR1 regulates PRUNE2 mRNA subcellular distribution and, consequently, its protein translation. Interestingly, the presence of editing induced by ADAR1 in PCA3 and PRUNE2 was confirmed in prostate cancer patients, supporting the medical relevance of this mechanism in tumorigenesis (Salameh et al., 2015).

The nuclear sequestration of edited lncRNAs was also described in the context of viral lncRNAs (vlncRNAs) originated from Epstein Barr Virus (EBV) in Burkitt's lymphoma, supporting a functional role for lncRNA editing induced by ADAR1 in cancer (Cao et al., 2015).

In other models, A-to-I editing of lncRNAs has also been studied. In chickens, T-cell lymphoma induced by the Gallid herpesvirus 2 (GaHV-2), the hyperediting of the ERL lncRNA (edited repeat-long, long non-coding RNA) disrupt its role as the natural antisense transcript of the major transforming oncogene *meq* (*Meq*) (Figueroa et al., 2016).

5.2 LncRNAs and non-canonical functions of ADAR1

ADAR1-lncRNAs interplay has not only been described in terms of A-to-I editing but non-canonical functions of ADAR1. In pancreatic cancer, it has been shown that the antisense lncRNA of glutaminase (GLS-AS) can repress glutaminase (GLS) expression through an ADAR1/Dicer-dependent RNA interference mechanism. It is well documented that cancer cells undergo metabolic reprogramming to support fast proliferation. In this context, GLS is critical for glutamine metabolism. GLS-AS can downregulate GLS expression at the posttranscriptional level, via formation of a dsRNA with GLS pre-mRNA and the ADAR/Dicer-dependent RNA interference processing. In this way, the metabolic reprogramming mediated by GLS in cancer cells is impaired. Remarkably, low expression of GLS-AS in patients was associated with poor clinical outcomes (S. Deng et al., 2018).

It is well reported that cancer cells can overcome hypoxic stress and that HIF-1α (*HIF-1A*) is a pivotal mediator of the cellular hypoxia response. It was found that the natural antisense transcript of the *HIF-1A* gene, the lncRNA HIF1A-AS2, can downregulate its expression levels. In cancer cells, ADAR1 downregulates the levels of HIF1A-AS2, thus antagonizes HIF1A-AS2-dependent suppression of HIF-1A. Remarkable, A-to-I editing was found in both transcripts but

was not found directly involved in the ADAR1-mediated regulation of HIF1A-AS2 and HIF-1A. In the absence of ADAR1, the authors showed a greater association of RNA pol-II with the HIF1A-AS2 promoter, and less association of RNA pol-II with HIF-1A promoter. In this way, they proposed that ADAR1 might act locally in this gene regulation (Ma et al., 2019).

RATIONALE

Although genomic mutations are the main driver of cancer development and progression, they do not account for all the alterations and gene expression changes found in tumor cells. For this reason, and considering the emerging regulatory roles of lncRNAs, we propose that it is relevant to examine cancer biology from this perspective.

Molecular mechanisms and biological functions have been assessed for both lncRNAs and ADAR1-induced changes; however, we are far from understanding how these two components could impact the mechanisms operating in the transition from normal to pathological cells and cancer progression.

In this thesis, we propose to assess the ADAR1-induced changes over lncRNAs A-to-I editing and expression levels and their potential contribution to breast cancer to better understand the complex landscape of cancer.

MATERIALS

The following section lists in detail the reagents, equipment, and software used to perform and analyze the experimental procedures concerning the thesis project. The name of the reagent, source, and catalog number are provided to allow an accurate replication of the results.

Table 1. Reagents and materials. List of all reagents used to perform experiments in this thesis.

CELL CULTURE		
Reagent	**Source**	**Cat. Number**
DMEM-F12	HyClone	Cat # SH30261.01
Fetal Bovine Serum (FBS)	Corning	Cat # 35-010-CV
1x PBS	Gibco	Cat # 10010023
100x Penicillin/Streptomycin	Corning	Cat # 30-002C1
10x Trypsin 10x	Gibco	Cat # 15090046
Cell Lines		
MDA-MB-231	ATCC	Cat # HTB-26
MDA-MB-436	ATCC	Cat # HTB-130
ADENOVIRAL TRANSDUCTION		
Reagent	**Source**	**Cat. Number**
pAV[Exp]-CMV>EGFP	Vector Builder	N/A
pAV[Exp]-CMV>hADAR[NM_001025107.2]*/3xFLAG)	Vector Builder	N/A
SMALL INTERFERING RNAS (siRNAS) TRANSFECTION		
Reagent	**Source**	**Cat. Number**
Transfectargo	Corning	Cat # 40-300-CVR
Lipofectamine 3000 Transfection Reagent	Thermo Scientific	Cat # L300000
siRNAs		
ADAR1, siRNA targeting sequence	Ambion	Cat # AM51331
Control, siRNA targeting sequence	Cell Signaling	Cat # 6568S
E.Z.N.A. Total RNA Kit I	Omega Bio-Tek	Cat # R6834
RNase-free DNase Set I	Omega Bio-Tek	Cat # E1091

RNA ISOLATION AND DNASE TREATMENT		
Reagent	**Source**	**Cat. Number**
E.Z.N.A. Total RNA Kit I	Omega Bio-Tek	Cat # R6834
RNase-free DNase Set I	Omega Bio-Tek	Cat # E1091
cDNA SYNTHESIS		
Reagent	**Source**	**Cat. Number**
AffinityScript QPCR cDNA Synthesis Kit	Agilent Technologies	Cat # 600559
RT-qPCR		
Reagent	**Source**	**Cat. Number**
Brilliant II SYBR Green QPCR Master Mix	Agilent Technologies	Cat # 600828
Oligonucleotides (25 nM)		
ACTB forward	5′-AAC GGC TCC GGC ATG TGC AAG -3′	
ACTB reverse	5′-GCC GTG CTC GAT GGG GTA CTT -3′	
ADAR1 forward	5′-AAG GCA GAA CGC ATG GGT TTC A- 3′	
ADAR1 reverse	5′-AGT GTC TTT GGC TGT GCT TCT GG -3′	
ADAR2 forward	5′-AAT GCG AGC ATC CAA ACG TG -3′	
ADAR2 reverse	5′-AAT GGG CTC CAC GAA AAT GC -3′	
APCDD1L-AS1 forward	5′-ACA AAT GCG CAA GAG CCA TG- 3′	
APCDD1L-AS1 reverse	5′-TGG CAA AAA TGT GGC TGT CG -3′	
H1FX-AS1 forward	5′-TGC TCC ACT TCA CCT TTT GC -3′	
H1FX-AS1 reverse	5′-TGT AGC AAA GCC ACG GAA AG -3′	
LINC00944 forward	5′-AGG GCC TTC AGG AAT CTT CAC -3′	
LINC00944 reverse	5′-ATG CCT TCA ATC TGC AGC TC -3′	
LINC01003 forward	5′-TAC CCA TCC CTT TTC TCC ATG C -3′	
LINC01003 reverse	5′-AAT GCG TCA CCT TGT TAG GG -3′	
RESS-qPCR		
Oligonucleotides (25 nM)		
AZIN1 WT forward	5′-CAT TCA GCT CAG GAA GAA GAC ATC T -3′	
AZIN1 WT reverse	5′-AAT ACA AGG AAG ATG AGC CTC TGT TTA C -3′	
AZIN1 Edited forward	5′-ACT GAA TGA CAT CAT GTA ATA AAT GGC T -3′	
AZIN1 Edited reverse	5′-GAG CTT GAT CAA ATT GTG GCA G -3′	
MDM2 WT forward	5′-ATA GGA CTG AGG TAA TTC TGC ACA GCA -3′	
MDM2 WT reverse	5′-ATA ATG CTT GGA GGA CCT CCA CAT GT -3′	
MDM2 Edited forward	5′-TAA ATG GCC AAA GGG ATT AGT AGT GTG -3′	
MDM2 Edited reverse	5′-AAG AGA TTC TGC TTG GTT GTA GCT GAA G-3′	

PROTEIN EXTRACTION, SDS-PAGE AND WESTERN BLOT		
Reagent	**Source**	**Cat. Number**
BCA Protein Assay kit	Thermo Scientific	Cat # 23227
β-Mercaptoethanol	Merck	Cat # 444203
Bovine serum albumin (BSA)	Winkler	Cat # BM-0150
Halt Phosphatase Inhibitor Cocktail	Thermo Scientific	Cat # 78426
Halt Phosphatase and Protease Inhibitor Cocktail	Thermo Scientific	Cat # 1861284
10% Mini-PROTEAN TGX Precast Protein Gels	Bio-Rad Laboratories	Cat # 4561034
Nitrocellulose membrane (0.45 µm)	Bio-Rad Laboratories	Cat # 162-0115
PageRuler Prestained Protein Ladder	Thermo Scientific	Cat # 26616
Ponceau S	Winkler	Cat # BM-1492
RIPA lysis buffer	Thermo Scientific	Cat # 89900
Sodium Dodecyl Sulfate (SDS)	Calbiochem	Cat # 7910
SuperSignalWest Pico Chemiluminiscent Subs.	Thermo Scientific	Cat# 34579
10x TBS	Bio-Rad Laboratories	Cat # 170-6435
TEMED	Calbiochem	Cat # 8920
10x Tris/Glycine Buffer	Bio-Rad Laboratories	Cat # 161-0734
10x Tris/Glycine/SDS Buffer	Bio-Rad Laboratories	Cat # 161-0732
Tween-20	Winkler	Cat # BM-2031
Antibodies		
Mouse Monoclonal anti- β-actin	Cell Signaling	Cat # 58169
Rabbit Polyclonal anti-ADAR1	Abcam	Cat # ab168809
Rabbit IgG HRP (1:5000)	Jackson ImmunoResearch,	Cat#111-035-045
Mouse IgG HRP (1:5000)	Jackson ImmunoResearch	Cat#115-035-062
RNA IMMUNOPRECIPITATION (RIP)		
Reagent	**Source**	**Cat. Number**
Chloroform	Merck	Cat # 102445
DMSO	Sigma	Cat # D2650
DNase I, RNase-free	QIAGEN	Cat # 79254
1M DTT	Thermo Scientific	Cat # P2325
Dynabeads Protein G	Thermo Scientific	Cat # 10003D
0.5M EDTA Solution	Thermo Scientific	Cat # 1861283
Ethanol absolute	Merck	Cat # 1070172511
Formaldehyde	Thermo Scientific	Cat # 28908
GlycoBlue Blue Coprecipitant	Ambion	Cat # AM9516
N-lauroylsarcosine Sodium Salt	Sigma	Cat # L7414
Proteinase K	Life Technologies	Cat # 25530049
TriZol	Life Technologies	Cat # 15596018
RNaseOUT Ribonuclease Inhibitor	Invitrogen	Cat # 10777-019
RNeasy Mini Kit	Qiagen	Cat # 74104
Tris, Hydrochloride	Merck	Cat # 648313

RNA-SEQUENCING		
Reagent	**Source**	**Cat. Number**
Experion RNA StdSens Reagents and Supplies	Bio-Rad Laboratories	Cat # 7007154
INSTRUMENTS		
Reagent	**Source**	**Cat. Number**
ChemiScope 3500 Mini chemiluminescence imaging system	Clinx Science Instruments	N/A
Cytation 3 Instrument - Take 3	Biotek	N/A
Eco Real-Time PCR System	Illumina	Cat # EC-900-1001
Experion Automated Electrophoresis System	Bio-Rad Laboratories	Cat # 700-7000
LUNA-FL Dual Fluorescence Cell Counter	Logos Biosystems	Cat # L20001
SOFTWARE AND ALGORITHMS		
Software/Algorithms	**Developer**	**Version**
DESeq2	(Love et al., 2014)	N/A
Image Studio Lite	LI-COR	Version 5.2
Prism 8	GraphPad	Version 8
RStudio	R Team Core 2020	Version 3.6

METHODS

The following section describes the methods performed to carry out the experimental procedures concerning the thesis project. Calculations, techniques, statistical analyses, references, and limitations are provided to allow accurate replication and assess the credibility of the results. For details concerning the materials and equipment used, please refer to the *Material section*.

1. Cell culture

MDA-MB-231 and MDA-MB-436 cell lines were obtained from American Type Culture Collection (ATCC). MDA-MB-231 and MDA-MB-436 were cultured in Dulbecco's modified Eagle medium (DMEM)-F12 supplemented with 10% fetal bovine serum (FBS) and 100 U/ml penicillin and streptomycin solution. All cell lines were maintained at sub-confluent densities at 37°C in a 5% CO_2 humidified incubator.

2. Adenovirus transduction

The breast cancer cell line MDA-MB-231 was transduced with an adenovirus vector, which carried either the short isoform of the human ADAR1 DNA sequence (ADAR1-p110) (NM_001025107.2) to overexpress the protein (ADAR1 overexpression condition) (ADAR1 OE), or a GFP sequence (Mock condition) as control. For each condition and biological replicate, 500.000 cells were plated in a 60 mm dish. After 24 hours, adenoviral particles (Mock and ADAR1) were transduced at MOI 200, and expression was allowed for 48 hours.

3. Small interfering RNAs (siRNAs) transfection

siRNAs transfections were carried out using Lipofectamine 3000 Transfection Reagent according to the manufacturer's directions. Briefly, 500.000 cells were plated in a 60 mm dish. After 24 hours, control or ADAR1 siRNAs were transfected at a final concentration of 20 nM and incubated for 48 hours.

4. RNA isolation and DNase treatment

RNA was isolated using the E.Z.N.A. Total RNA Kit I, according to the manufacturer's instructions. Samples were treated with RNase-free DNase Set I for DNA removal, as the manufacturer recommended. RNA concentration and quality were measured in the Take 3 - Cytation 3 Instrument. Only suitable samples were used for downstream applications. All RNAs obtained were stored at -80°C.

5. cDNA synthesis

cDNA was synthesized by using the AffinityScript cDNA Synthesis Kit, according to the manufacturer's protocol. 1 µg of total RNA was used as a template in each reverse transcription reaction, and a mixture of oligo-dT and random primers was used to improve efficiency. All cDNA products were stored at -20°C.

6. RT-qPCR

Real-time PCR reactions were performed using an Eco Real-Time PCR System and Brilliant II SYBR Green QPCR Master Mix. Reactions were prepared and performed according to the manufacturer's instructions. For the detection of mRNAs and lncRNAs, 35 cycles and 40 were

used, respectively. GAPDH and/or β-actin served as internal controls genes. Relative RNA expression was calculated using the comparative cycle threshold (Ct) ($2^{-\Delta\Delta Ct}$) method. Primers were designed using Primer3Plus software (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) and synthesized by Integrated DNA Technologies (IDT).

7. RESS-qPCR

RNA Editing Site-Specific-qPCR (RESS-qPCR) was performed as described by Crews et al., 2015 (Crews et al., 2015). Briefly, for detecting each RNA editing site, two sets of primers are designed, one pair for detecting WT/non-edited transcript (A nucleotide), and one pair detecting the edited transcript (G nucleotide, representing inosine substitution). cDNA synthesis and qPCR were performed as described previously. Relative RNA editing ratios (Relative Edited/WT RNA) were calculated using the comparative cycle threshold (Ct) [$2^{-(Ct\ Edit - Ct\ WT)}$].

8. Protein extraction, SDS-PAGE and Western Blot

Cells were lysed with RIPA lysis buffer and protease inhibitors following the manufacturer's protocol. Then, cell homogenates were separated by a 10% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Primary antibodies were probed overnight at 4°C. HRP-linked secondary antibodies (1:5000 dilution) were incubated 1.5 hours at room temperature, and blots were visualized with SuperSignal West Pico Chemiluminescent Substrate in the ChemiScope3500 Mini chemiluminescence imaging system. Western blot densitometry was performed using Image Studio Lite Software and normalized to β-actin. Primary antibodies used: 1:2000 rabbit polyclonal anti-ADAR1, 1:5000 mouse anti-β-actin.

9. RNA-sequencing

RNA-sequencing was performed by the BGI Genomics Company in China (https://www.bgi.com/global/). The detailed protocol is shown below.

9.1 RNA isolation and DNase treatment

RNA was isolated from transduced cells using the RNeasy Mini Kit, according to manufacturer's instructions. Samples were then incubated with DNase I for 20 minutes at 37°C. Concentration and quality of RNAs were measured in the Take 3 - Cytation 3 Instrument, and RNA integrity was assessed in the Experion Automated Electrophoresis System. Only suitable samples were further used. At this point, aliquots of RNA were separated for sequencing and RT-qPCR validation.

9.2 Library preparation for strand-specific total RNA-sequencing

RNA samples were sent to the BGI's Next Generation Sequencing Service for library preparation and RNA-sequencing. Briefly, for library preparation, the biotin-labeled Ribo-Zero rRNA Removal Kit was used to remove ribosomal (rRNA) from total RNA. Then, the un-purified RNA was fragmented under specific temperature and ion environment. Random primers and the reverse transcriptase of TruSeq Stranded Kit were used to synthesize the first-strand cDNA. Then DNA polymerase I and RNase H were used to synthesize double-stranded cDNA. In the second cDNA strand synthesis process, RNA templates were removed, and dTTPs were replaced by dUTPs. Adenylation and adapter ligation were performed, and adapters were then added to the double-strand cDNA products. Ligation products were amplified and purified to get the final cDNA library. Finally, the library was sequenced using Illumina HiSeq4000.

10. Bioinformatic analysis of differentially expressed lncRNAs

FASTQ files containing clean reads were obtained from BGI. QuickRNASeq workflow (Shanrong Zhao et al., 2016) was followed as developers recommended: individual samples were processed independently in the mapping to GRCh37 Human Reference Genome using STAR software and FeatureCounts software (Liao et al., 2014) for counting reads in genomic features. Next, the counts were used in differential expression analysis between MDA-MB-231 Mock and ADAR1 overexpression conditions using DEseq2 software and following standard recommendations (Love et al., 2014). A False Discovery Rate (FDR) ≤ 0.05 was considered statistically significant, and no fold change cutoff was considered since observed changes in lncRNAs expression were modest. All the bioinformatic steps were applied by the bioinformatics team at CEMP-Pfizer Chile (from Spanish Centro de Excelencia en Medicina de Precisión).

11. Bioinformatic analysis of A-to-I lncRNA editing

A-to-I variants were detected in lncRNAs using the Python REDItoolDenovo.py script, which is part of the REDItools package (https://github.com/BioinfoUNIBA/REDItools) (Picardi & Pesole, 2013). Briefly, BAM files from TNBC patients were obtained from the TCGA-BRCA cohort. Reference genome (GRCh38.p13) was obtained from the National Center for Biotechnology Information (NCBI). Both the BAM file and the human reference genome were the input for running the Python REDItoolDenovo.py script. Key parameters were set in this step: *Per base coverage*, potentially edited sites not supported by ≥ 10 reads were filtered out; *Quality score*, positions with a Phred score < 25 were excluded, and *Mapping quality*, reads with a mapping quality score < 25 were removed. Output tables containing single nucleotide

variants (SNVs) were filtered using RStudio (R Core Team, 2020), and only significant ($p \leq 0.05$, *Fisher test*) variants representative of A-to-I editing (A-to-G in the sense strand and T-to-C in the antisense strand) were kept. SNVs that met all the criteria mentioned above were called edited positions. Annotation of edited sites to gene symbols was performed using the accessory Python *AnnotateTable.py* script provided by REDItools and the lncRNA gene annotation file obtained from GENCODE (Release 34, GRCh38.p13) (https://www.gencodegenes.org/human/). Finally, SNPs were removed using the NCBI database of genetic variation: dbSNP v.138 (Sherry et al., 2001).

12. ADAR1 Formaldehyde RNA immunoprecipitation (fRIP)

ADAR1 RNA immunoprecipitation was performed following the protocol described by Hendrickson et al., 2016 (Hendrickson et al., 2016). The detailed protocol is shown below.

12.1 Cross-linking

$5x10^6$ cells/mL were resuspended with serum-free culture medium, and formaldehyde was added to a final concentration of 0.1%. Crosslinking was performed for 10 minutes at room temperature, and then the reaction was halted with 2.5M glycine to a final concentration of 125 mM for 5 minutes at room temperature. Cells were washed with cold 1x PBS, and pellets of $10x10^6$ cells were stored at -80ºC.

12.2 Cell lysis

Frozen pellets were re-suspended in 1 mL of RIPA lysis buffer with protease inhibitor cocktail and 100 U/mL of RNaseOUT. Cells were lysed, and the lysate was spun at 4ºC at 12000

rpm for 10 minutes. The supernatant was collected and diluted in fresh fRIP binding/wash buffer (150 mM KCl, 25 mM Tris pH 7.5, 5 mM EDTA, 0.5 % NP-40, 0.5 mM DTT, 1× fresh protease inhibitor cocktail (PIC), 100 U/mL RNaseOUT). 50 µl of lysate was removed for the input sample. The lysate was clarified by passage through a 0.45 µM syringe filter. The pre-cleared filtered lysate was then incubated with Dynabeads Protein G at a concentration of 25 µl of beads per 5 million cells for 30 minutes at 4°C with slow rotation.

12.3 ADAR1 Immunoprecipitation

4 µg of ADAR1 antibody were added to the lysate and rotated at 4ºC for 2 hours before adding 50 µl of Dynabeads Protein G. Beads and lysate were rotated at 4ºC for 1 hour before washing twice with 1 mL of fRIP binding/washing buffer + 1xPIC and 100 U/mL RNaseOUT. After the final wash, the supernatant was removed, and the pellet containing beads was collected.

12.4 Reverse crosslinking and RNA purification

Beads were resuspended in 56 µl of RNase-free water, and 33 µL of 3x reverse-crosslinking buffer [3x PBS (without Mg or Ca), 6% N-lauroyl sarcosine, 30 mM EDTA, 15 mM DTT (fresh)), 10 µl of Proteinase K, and 1 µl of RNaseOUT were added to resuspended beads and input samples. Protein degradation and reverse-crosslinking were performed for 1 hour at 42ºC plus 1 hour at 55ºC. Beads and reaction buffer were added to 1 mL of TriZol. After agitation, 200 µl of chloroform was added, and a 20 min spin at 4ºC max speed followed. The aqueous phase was collected and added to 750 µl of ethanol + 1 µl GlycoBlu and ran it over a Qiagen RNeasy minElute column. RNA was eluted in 15 µl of RNase-free water.

13. Kaplan-Meier survival curves

Survival (OS) and Relapse-free Survival (RFS) curves were performed using the web-based Kaplan-Meier Plotter (https://kmplot.com/analysis/) (Györffy et al., 2010) on breast cancer dataset. Gene symbol: LINC00944, Affymetrix ID: 1560573_at. The Log-rank test with $p \leq 0.05$ was considered statistically significant.

14. Gene expression data retrieval

Gene expression (FPKM-UQ values) and clinical data were obtained from The Cancer Genome Atlas (TCGA) Project (https://www.cancer.gov/tcga) through the Genomic Data Commons (GDC) data portal (https://portal.gdc.cancer.gov/) (Grossman et al., 2016). TCGA studies retrieved: Breast invasive carcinoma (BRCA), Colon adenocarcinoma (COAD), Lung adenocarcinoma (LUAD), and Testicular Germ Cell Tumors (TGCT).

15. Guilt-by-Association analysis

LINC00944 expression was correlated to protein-coding genes (GRCh38) expression in normal and tumor samples by Pearson Correlation. The *p*-value was adjusted using the Benjamini-Hochberg method, and an FDR ≤ 0.05 was considered statistically significant. Correlations with Pearson coefficient (r) $\leq |0.3|$ were taken for computing the overlap against Hallmark gene sets collection (Liberzon et al., 2015) using the Gene Set Enrichment Analysis (GSEA) online tool (https://www.gsea-msigdb.org/gsea/index.jsp) (Subramanian et al., 2015). An FDR ≤ 0.05 was considered statistically significant.

16. LINC00944 expression and clinicopathological parameters correlation

LINC00944 expression and clinical data from TCGA-BRCA were obtained as mentioned above. Patients were classified based on the LINC00944 *z*-score. Upper (High LINC00944 expression group) and lower (Low LINC00944 expression group) quartiles were chosen for further comparison (n= 250). Data were arranged in 2x2 contingency tables, and the Fisher's exact test was run for the comparison. A p-value ≤ 0.05 was considered statistically significant.

17. Statistical analyses

R Software v3.6 (R Core Team, 2020) was used for data management. Statistical tests were performed in R Software v3.6 and GraphPad Prism v8.3. Two-tailed Student's t-test was used in the analysis of RT-qPCR to establish differences between two groups in RT-qPCR, and each analysis was performed in 5-6 independent experiments. For non-parametric data, the Mann-Whitney test was used in two-group comparisons.

HYPOTHESIS AND OBJECTIVES

1. HYPOTHESIS

"Adenosine deaminase acting on RNAs 1 (ADAR1) expression induces changes in long non-coding RNAs (lncRNAs) A-to-I editing and/or expression levels, which in turns correlates with malignant phenotypes and poor outcomes in breast cancer"

2. OBJECTIVES

2.1 General objective

Characterize the effects of ADAR1 over lncRNAs expression levels and A-to-I editing in breast cancer, and evaluate if lncRNA dysregulation induced by ADAR1 correlates with malignant phenotypes and poor prognosis in breast cancer.

2.2 Specific objectives

Specific objective 1: Characterize the effects induced by ADAR1 on lncRNAs expression levels in breast cancer.

Specific objective 2: Evaluate if the ADAR1-induced effects on lncRNAs expression levels are involved in a malignant phenotype and/or poor breast cancer outcomes.

Specific objective 3: Characterize lncRNAs A-to-I editing induced by ADAR1 in triple-negative breast cancer (TNBC) and the potential effects over a malignant phenotype.

RESULTS

This chapter encompasses the rationale behind the experimental model used to assess the hypothesis proposed in this thesis. Moreover, here we show a detailed quality control of the RNA-sequencing to ensure that we generated accurate datasets for downstream analysis.

CHAPTER 1: Experimental model and RNA-sequencing

ADAR1 is upregulated in breast cancer, predicting poor overall survival in patients (Kung et al., 2020; Sagredo et al., 2018). It has been proved that many pathologic processes mediated by ADAR1 are through the modulation of A-to-I editing and expression levels of its targeted coding and non-coding RNAs (for review, see C. Wang et al., 2017; Xu et al., 2018). Among the latter, lncRNAs have emerged as central components in oncogenic and tumor suppressor pathways (for review, see Isin & Dalay, 2015; Schmitt & Chang, 2016). Still, there are few reports in the literature that examine the ADAR1-lncRNAs interplay.

Taken that into account, *we examined the effects induced by ADAR1 expression on lncRNAs A-to-I editing and expression levels*. By using MDA-MB-231 breast cancer cells overexpressing ADAR1, we performed an RNA-seq and assessed the effects over lncRNAs on a transcriptomic scale.

Since the RNA-seq was a central experiment for this thesis, we performed a diligent quality control (QC) on samples used in the sequencing and on the generated data before any downstream analysis.

1. **Establishing a cellular model for studying the effects of ADAR1 overexpression on lncRNAs in breast cancer**

Previous analysis from our laboratory shows that breast cancer patients expressing higher mRNA levels of ADAR1 display a lower overall survival in time (**Fig. XIII**), therefore we took a cell line with low levels of ADAR1 (vs. other breast cancer cell lines), MDA-MB-231, and by using an adenoviral vector, we overexpressed ADAR1. To study the impact on lncRNAs, we obtained RNA samples for RNA-sequencing (RNA-seq) and RT-qPCR, and protein samples for western blot controls.

Schemes of adenoviral vectors used in this thesis are presented in Figure 1A. In both vectors, the expression is driven by the cytomegalovirus (CMV) promoter, followed by the green fluorescent protein (GFP) sequence in the control vector (Mock vector) and by the human ADAR1-p110 sequence (NCBI reference sequence: NM_001025107.2) in the ADAR1 vector. Additionally, the ADAR1 vector has three *FLAG*-tags (**Fig. 1A**).

A general scheme of the experimental design is presented in Figure 1B. Viral vectors were expressed in the breast cancer cell line MDA-MB-231. The Mock vector was used as control, and the ADAR1 vector was the experimental condition (ADAR1 OE). Vector

expression was allowed for 48 hours, and then RNA and protein samples were obtained for downstream applications (**Fig. 1B**).

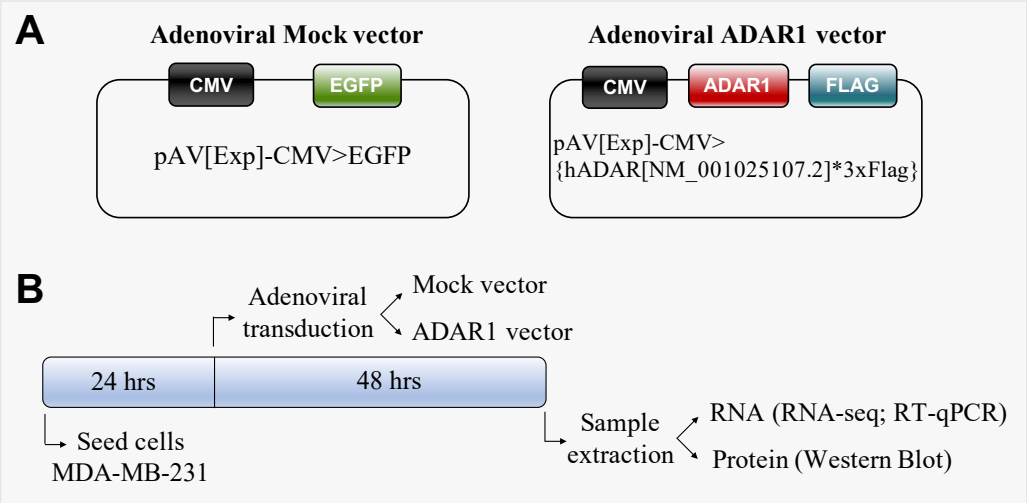


Figure 1. Experimental design. (A) Schemes of adenoviral Mock control (left panel) and ADAR1 (NM_001025107.2, encoding the short isoform) (right panel) vectors. **(B)** Scheme of experimental design.

2. Samples quality control for RNA-sequencing

Isolated total RNA was subjected to quality control in order to be suitable for sequencing. The quality control consisted of the measurement of:

- **RNA concentration**, which is an important parameter for successful downstream RNA sequencing. A minimum of 500 ng per sample was used for sequencing.

- **Ratio 260/280**, which is an accepted parameter of the RNA preparation purity concerning to protein or phenol contaminations. Only samples which met a ratio of 260/280 ≥ 1.8 were used for sequencing.

- **RQI (RNA Quality Indicator)**, which calculates RNA integrity by comparing the electropherogram of RNA samples to a series of standardized degraded RNA samples. The RQI method returns a number between 10 (intact RNA) and 1 (highly degraded RNA) for each eukaryotic RNA sample. Only samples which met a ratio RQI ≥ 7.0 were used for sequencing.

- **Ratio 28S/18S**, which indicates the quantification of the large (28S) and small (18S) subunit ribosomal RNA subunits (rRNAs). A ratio around 2 indicates that the larger subunit is twice the smaller band, which is an indicator of intact RNA. Only samples which met a ratio 28S/18S ≥ 1.0 were used for sequencing.

As shown in Table 1, all samples qualified for sequencing as they met the quality criteria mentioned above.

Table 1. Sample quality control. All samples were subjected to quality controls in order to guarantee proper RNA-sequencing.

Sample	Concentration µg/µl	Ratio 260/280	RQI	Ratio 28S/18S	Test result
Mock N1	399.7	2.15	10.0	2.06	Qualify
Mock N2	190.2	2.16	10.0	1.97	Qualify
Mock N3	286.3	2.17	10.0	1.90	Qualify
ADAR1 OE N1	212.1	2.16	10.0	1.91	Qualify
ADAR1 OE N2	212.1	2.15	10.0	1.93	Qualify
ADAR1 OE N3	206.4	2.15	10.0	2.02	Qualify

Mock correspond to control condition. ADAR1 OE correspond to ADAR1 overexpression condition. N: Biological replicate.

3. RNA-sequencing of MDA-MB-231 cells overexpressing ADAR1

RNA samples were sequenced with the BGI Genomics next-generation sequencing service (BGI Co., Ltd, China). The experimental workflow is described in the *Methods Section*. In Table 2 is presented a brief description of the main RNA-sequencing parameters.

For library preparation, ribosomal rRNA was depleted, and strand-specific reads were obtained using the protocol described in the *Methods Section*. Short inserts of 100 bp in average size were employed for library construction to avoid the formation of secondary structures and ensure an optimal sequencing. The platform used for RNA-sequencing was the Illumina HiSeq4000, and fragments were sequenced for both ends (paired-end). Reads of 100 bp long were generated, and a minimum of 100 million reads was obtained.

Table 2. Summary of RNA-seq parameters.

Library preparation	Ribo-Zero + Strand-Specific total RNA sequencing
Library construction	Short-insert library construction
Platform for sequencing	HiSeq4000
Sequencing strategy	Paired-end
Read length	100 bp
Data output	100M reads

It is worth to note that performing stranded RNA sequencing was key for an accurate lncRNA identification and quantification of expression levels, as many of them are encoded in antisense strands that overlap sense encoded protein-coding genes (Derrien et al., 2012). Moreover, obtaining a high coverage and high sequencing depth were crucial for detecting

lncRNAs, as most lncRNAs display low expression levels (Derrien et al., 2012). Table 3 shows the number of total clean reads per sample used in the downstream bioinformatics analysis.

Table 3. Total clean data per sequenced sample. The average clean reads obtained per sample was 147.837.596.

Sample	Total clean reads
Mock N1	141,437,020
Mock N2	138,719,022
Mock N3	171,906,448
ADAR1 OE N1	145,017,052
ADAR1 OE N2	152,960,002
ADAR1 OE N3	136,986,032

4. Sequencing quality control

The quality control (QC) of datasets obtained from RNA-sequencing is a critical step since indicates if the process of sequencing itself was appropriately performed, if it has a low error rate associated, and if the process was biased or not (Babraham Bioinformatics, FastQC project). Hence, the quality control provides security for using the data/reads obtained in downstream analysis.

4.1 Base percentage composition along reads

Bases composition along reads should resemble the proportions of the four DNA bases of the organism in study, and never should be found an extreme imbalanced between bases. As random primers are used to prime the reverse transcription to produce cDNA and the end of

which are eventually sequenced, it is expected that the reads to start at random locations along with the transcript. In that case, the first base positions should show a typical noise. Consequently, there should not be any base composition bias along with the reads (Babraham Bioinformatics, FastQC project).

In a successful sequencing process is predictable that the Adenine (A) curve overlapped with the Thymine (T) curve and that the Guanine (G) curve overlapped with the Cytosine (C) curve. In the same line, the unknown bases (N) should not be significantly represented in the reads obtained (Babraham Bioinformatics, FastQC project).

The base percentage composition plots of Mock (**Fig. 2A-C**) and ADAR1 OE samples (**Fig. 2D-F**) showed a balanced composition of DNA bases, as the A curve overlapped with the T curve and the C curve with the G curve. Also, the N bases in each sample are not significant (**Fig. 2A-F**). A warning is showed in this test if the difference between the curves of A and T, or G and C is greater than 10% in any position and, a failure is showed if the difference between A and T, or G and C is greater than 20% (Babraham Bioinformatics, FastQC project). As none of the cases just mentioned are observed in the plots, we can conclude that the sequencing was not biased and that there is no contamination of overrepresented sequences such as adapter sequences or rRNA.

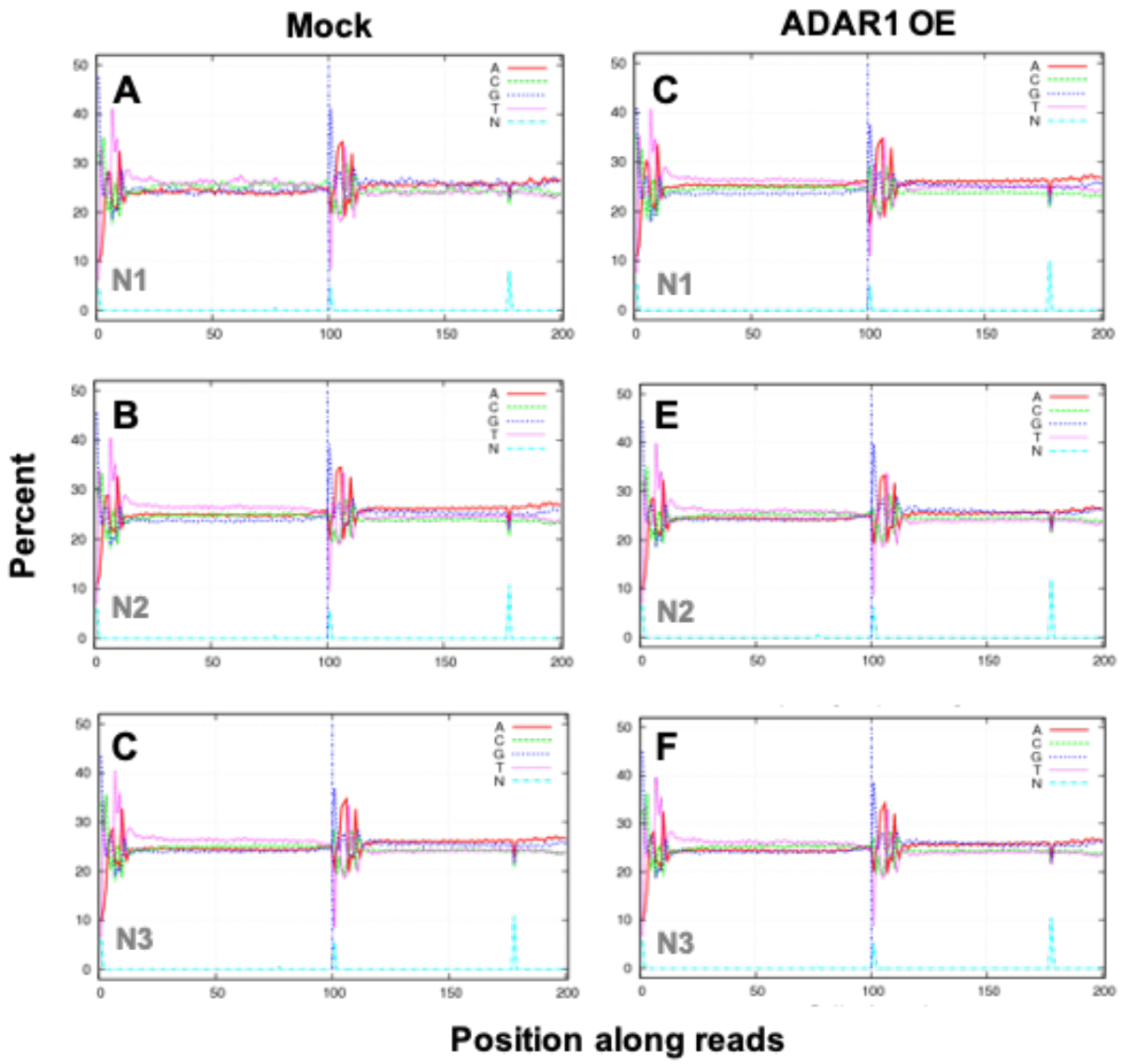


Figure 2. Base composition along reads. Plots of base percentage composition for clean reads of the (**A-C**) Mock and (**D-F**) ADAR1 OE conditions samples. N1-N3 represents the three biological replicates. On the horizontal axis, position 1-99 bp represent read 1, and 100-200 bp represent read 2. The vertical axis represents the percentage of nucleotides. Adenine (A) bases are depicted in red, Cytosine (C) bases in green, Guanine (G) bases in purple, Thymine (T) bases in pink, and N bases are depicted in light blue. OE: Overexpression.

4.2 Base quality distribution along reads

The *Phred* quality score (Q score) indicates the base calling accuracy and represents the most common metric used to evaluate the precision of a sequencing platform. Quality of called bases can vary depending on the position in the read due to the sequencing procedure itself; thus, a common practice of quality control is to determine the average quality score of each base, averaging all reads in a file (Babraham Bioinformatics, FastQC project). The algorithm of Illumina GA Pipeline v1.5 calculates and reports base quality values ranged from 2 to 41. Since the Q score calculation ponders the estimated probability of the base call being wrong, we can infer the error rate of the sequencing process. In this way, we can state that a Q score of 30 is equivalent to the probability of an incorrect base call 1 in 1000 times. The consensus values for base quality scores are as follow: very good base quality is > 28, reasonable base quality is 20 - 28, and poor quality is < 20 (Illumina Inc, Quality Scores for Next-Generation Sequencing).

The base quality distribution along reads of Mock (**Fig. 3A-C**) and ADAR1 OE samples (**Fig. 3D-F**) shows that all samples had Q scores > 30, indicating that the quality of the data is very good. Based on the relationship between quality values and sequencing error rate, we can conclude that the sequencing process had a base call accuracy of 99,9%.

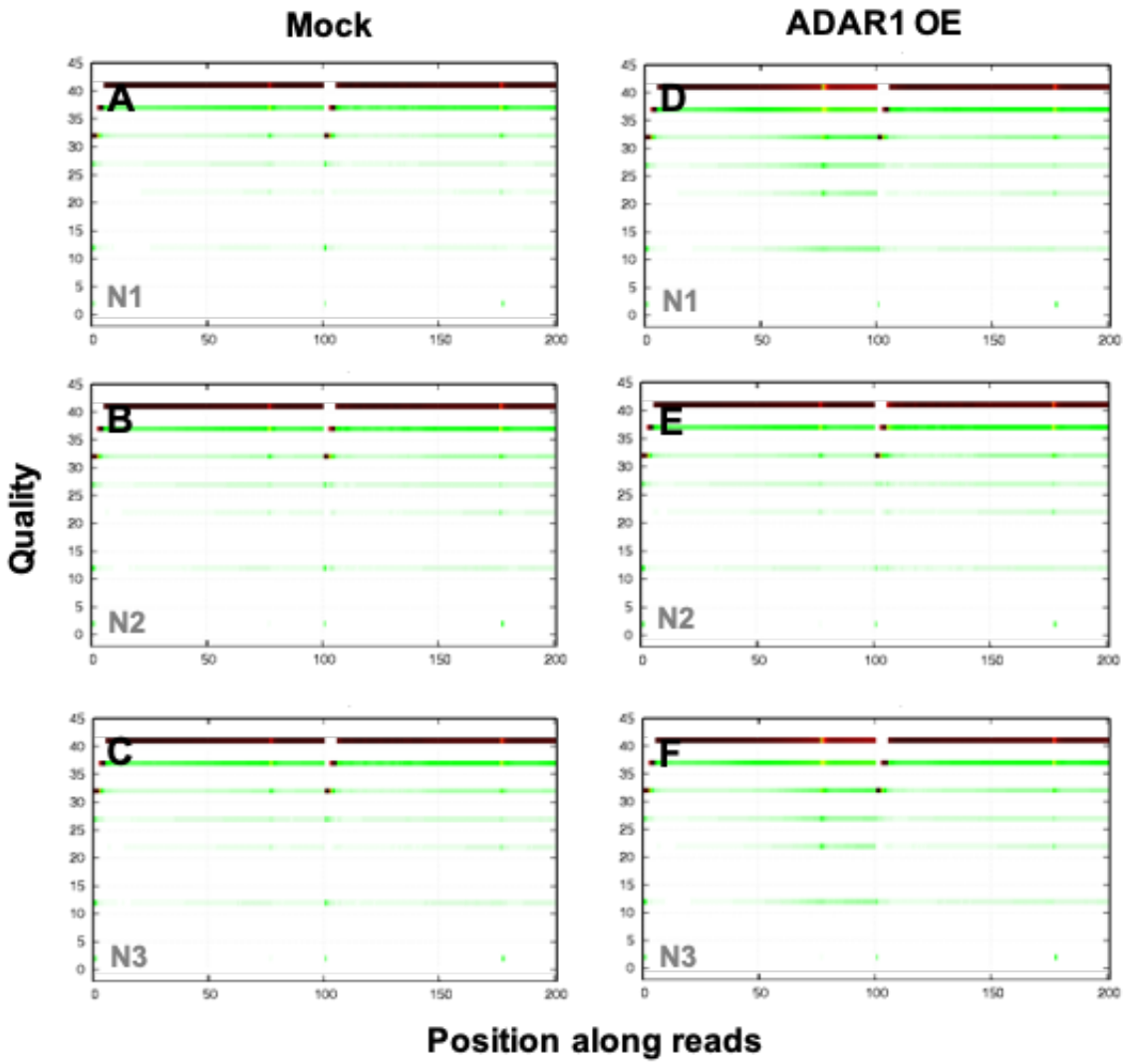


Figure 3. Base quality distribution along reads. Plots of base quality (Q score) distribution along clean reads of the (**A-C**) Mock and (**D-F**) ADAR1 OE conditions samples. N1-N3 represents the three biological replicates. The horizontal axis shows the position along reads, and the vertical axis the base quality score (Q score). On the horizontal axis, position 1-99 bp represent read 1, and 100-200 bp represent read 2. Each green dot in the image represents the Q score of the given position. The more accumulated are the dots (green), the more intense the color is (red and black). OE: Overexpression.

Taken together the QC plots, we can conclude that the RNA-sequencing procedure provided us with good and accurate datasets for downstream analysis.

CHAPTER 2: ADAR1 and lncRNAs expression levels in breast cancer

This chapter encompasses the results obtained in order to evaluate the specific objective 1, "*Characterize the effects induced by ADAR1 on lncRNAs expression levels in breast cancer*", and specific objective 2, "*Evaluate if the ADAR1-induced effects on lncRNAs expression levels are involved in a malignant phenotype and/or poor outcomes in breast cancer*".

As a result of evaluating the hypothesis about the effects induced by ADAR1 expression over lncRNAs expression levels and their involvement in breast cancer cells' malignancy, we generated a scientific publication that is currently under review. The scientific publication manuscript is attached in the following pages.

Scientific Article: ADAR1 and lncRNAs expression in breast cancer

Manuscript submitted

Cover letter

Santiago, July 16th, 2020.

Dear Editor
Life Sciences

Please find enclosed a manuscript entitled "*Low expression of immune-related lncRNA LINC00944 indicates poor outcomes in breast cancer"* which we are submitting for consideration at Life Sciences. For this study, a multidisciplinary group of Researchers from different scientific background (bioinformatics, medical oncology, system biology, and cellular and molecular biology) and from diverse research Institutions (Universidad de Chile, Pontificia Universidad Catolica, University of California San Diego, CORFO Center of Excellence in Precision Medicine and Universidad del Desarrollo) joined their effort and expertise to address a potential novel mechanism of breast cancer progression.

The oncogenic capabilities of the double stranded RNA-specific adenosine 1 (ADAR1) has been widely demonstrated in several cancer models. ADAR1 is able to modify the expression and function of an important number of RNA involved in many hallmarks of cancer. In this context, the effects of ADAR1 over long non-coding RNAs (lncRNAs) is an emerging area, and few reports have been published.

In breast cancer (BC), ADAR1 is upregulated and predicts poor prognosis. Here, we provide new evidence regarding ADAR1 role over lncRNAs expression. We show that the long intergenic non-coding RNA, LINC00944, is responsive to ADAR1 gain- and loss-of function in breast cancer cell lines. By using publicly available transcriptomic data of cancer patients, we predicted LINC00944 function via Guilt-by-Association correlation analysis, finding a strong relationship with immune system. Moreover, we were able to positively correlate LINC00944 expression to tumor-infiltrating lymphocytes in several cancer cohorts. In terms of clinicopathological parameters, we found LINC00944 to correlate to age at diagnosis, tumor size and estrogen and progesterone receptors expression, indicating that LINC00944 may have clinical relevance. Finally, we observed that a low expression of LINC00944 in BC patients is a poor prognosis factor, as a reduction in their overall survival and relapse-free survival was found.

Taken together, these results show that ADAR plays an important role in BC progression, in part through the regulation of LINC00944 expression, likely impacting BC patient outcomes. We believe these results will appeal to the Researchers with close interest in these topics but also to a broader cellular/molecular biology audience.

Thank you very much for your time and effort in handling this manuscript in these difficult times. We hope that you will find that it merits publication in Life Science

Sincerely,

Ricardo Armisén Y.
Professor, Facultad de Medicina Clínica
Alemana Universidad del Desarrollo.

1. Title: Low expression of immune-related lncRNA LINC00944 indicates poor outcomes in breast cancer

Pamela R. de Santiago[1,2], Alejandro Blanco[2], Fernanda Morales[2,3]. Katherine Marcelain[3,4], Olivier Harismendy[5,6], Marcela Sjöberg[1*] & Ricardo Armisen[2,7*].

Affiliations

[1]Departamento de Biología Celular y Molecular, Pontificia Universidad Católica de Chile. Santiago, Chile.
[2]Center of Excellence in Precision Medicine, Pfizer Chile. Santiago, Chile.
[3]Centro de Investigación y Tratamiento del Cáncer, Facultad de Medicina, Universidad de Chile, Santiago, Chile.
[4]Departamento de Oncología Básico Clínica, Facultad de Medicina, Universidad de Chile, Santiago, Chile.
[5]Division of Biomedical Informatics, Department of Medicine, University of California San Diego, California, United States.
[6]Moores Cancer Center, University of California San Diego, California, United States.
[7]Centro de Genética y Genómica, Instituto de Ciencias e Innovación en Medicina, Facultad de Medicina Clínica Alemana Universidad del Desarrollo.

***Correspondence to:**
Dr. Ricardo Armisen
Centro de Genética y Genómica, Instituto de Ciencias e Innovación en Medicina, Facultad de Medicina Clínica Alemana Universidad del Desarrollo.
Av. Las Condes 12461, 7590943, Santiago, Chile.
E-mail: rarmisen@udd.cl
Dr. Marcela Sjöberg
Departamento de Biología Celular y Molecular, Facultad de Ciencias Biológicas, Pontificia Universidad Católica de Chile.
Av. Libertador Bernardo O'Higgins 340, 8320000, Santiago, Chile.
E-mail: msjoberg@bio.puc.cl

2. Abstract

Breast cancer is one of the leading causes of women's deaths worldwide, being a major public health problem. It has been reported that the expression of the RNA-editing enzyme Adenosine Deaminase Acting on RNAs 1 (ADAR1) is upregulated in breast cancer, predicting poor prognosis in patients. A few reports in the literature examined ADAR1 and long non-coding RNAs (lncRNAs) interplay in cancer and suggest key roles in oncogenic and tumor suppressor pathways. Here we show that the lncRNA LINC00944 is responsive to ADAR1 up- and downregulation in breast cancer cell lines. We predicted LINC00944 function via Guilt-by-Association correlation analysis, finding a strong relationship with immune signaling pathways. Further assessment of the TCGA-BRCA cohort showed that LINC00944 expression was positively correlated to tumor-infiltrating T lymphocytes and pro-apoptotic markers. Moreover, we found that LINC00944 expression was correlated to the age at diagnosis, tumor size, and estrogen and progesterone receptor expression. Finally, we showed that the low expression of LINC00944 is correlated to poor prognosis, as a decrease in the Overall Survival (OS) and Relapse-Free Survival (RFS) were observed in patients.

Keywords: Breast cancer, lncRNAs, LINC00944, ADAR1, Prognosis, Tumor-infiltrating lymphocytes

3. Introduction

Breast cancer is the second leading cause of cancer-related deaths in women. The Global Cancer Observatory (GLOBOCAN) projected 2,088,849 cases in 2018 worldwide, which represented 11.6% of total cancer cases. According to the American Cancer Society, in 2020, there will be diagnosed about 276,480 cases of invasive breast cancer and about 48,530 new cases of carcinoma *in situ* (CIS) only in the United States, while about 42,170 US women will die from breast cancer. These alarming numbers highlight the extent of this major public health problem and evidences the need for broader knowledge in basic research, diagnosis, and treatment of breast cancer.

The Adenosine Deaminase Acting on RNAs (ADAR) family of proteins is integrated by three members, ADAR1, ADAR2, and ADAR3, being ADAR1 indispensable for life in mammals[1,2]. The canonical function of ADAR enzymes is the deamination of Adenine-to-Inosine in double-stranded RNAs (dsRNAs), in a process known as A-to-I RNA editing[1]. The introduction of inosines into dsRNAs has a plethora of biological consequences, ranging from biogenesis to expression and function[1,3]. Elevated *ADAR1* expression has been reported in breast cancer[4–6], predicting poor overall survival in patients[7,8]. We and others have shown that modulation of *ADAR1* expression leads to expression changes in coding and non-coding RNAs, being an important factor in cancer biology[7,9,10]. Among non-coding RNAs, long non-coding RNAs (lncRNAs) have emerged as central players in cancer development and progression[11] and proposed as prognostic biomarkers in several cancer types[12–14]. However, few reports regarding ADAR1 effect over lncRNAs expression and/or function are found in the literature[15–18]. For instance, it has been shown in human prostate cancer that the antisense lncRNA PCA3 can promote malignant cell growth by controlling the expression levels of the tumor suppressor PRUNE2, via the formation of an RNA duplex and the nuclear sequestration induced by ADAR1 A-to-I editing[15]. In pancreatic cancer, it has been proved that the antisense lncRNA of glutaminase (GLS-AS) can repress glutaminase (GLS) expression through an ADAR1/Dicer-dependent RNA interference mechanism, thus impairing the metabolic reprogramming mediated by GLS in cancer cells. Accordingly, low expression of GLS-AS was associated with poor clinical outcomes[16].

The emerging picture of ADAR1-lncRNAs interplay in cancer biology prompted us to further investigate the effect of ADAR1 over lncRNAs expression in breast cancer. To this end, we examined RNA-seq data on the MDA-MB-231 cell line overexpressing ADAR1 to look for differentially expressed lncRNAs. Next, ADAR1 gain-of-function and loss-of-function approaches were used to validate expression changes, finding that the long intergenic non-coding RNA LINC00944 presented very consistent expression changes. By using the Guilt-by-Association method, we investigated the LINC00944 function. Interrogating the breast cancer cohort of The Cancer Genome Atlas (TCGA-BRCA) allowed us to correlate LINC00944 expression to tumor-infiltrating lymphocytes (TILs) and apoptotic markers. Finally, we could correlate LINC00944 expression to clinicopathological parameters and survival outcomes in breast cancer patients.

4. Methods

2.1 Cell Culture

MDA-MB-231(ATCC, Cat# HTB-26) and MDA-MB-436 (ATCC, Cat# HTB-130) human cell lines were obtained from American Type Culture Collection (ATCC). Cells were cultured in Dulbecco's modified Eagle medium (DMEM)-F12 (HyClone, Cat# SH30261.01) supplemented with 10% fetal bovine serum (FBS) (Corning, Cat# 35-010-CV) and 100 U/ml penicillin and streptomycin solution (Corning, Cat# 30-002C1). All cell lines were maintained at subconfluent densities at 37°C in a 5% CO_2 humidified incubator.

2.2 Adenoviral transduction

ADAR1 overexpression was induced in breast cell lines by using an adenoviral vector (ADAR1 OE condition) (Vector Builder, pAV[Exp]-CMV>hADAR[NM_001025107.2]*/3xFLAG). A GFP sequence was used in the control condition (Mock condition) (Vector Builder, pAV[Exp]-CMV>EGFP). For each condition and biological replicate, 500.000 cells were plated in a 60 mm dish. After 24 hours, the adenoviral particles were transduced at MOI 200, and expression was allowed for additional 48 hours.

2.3 Small interfering RNAs (siRNAs) transfection

siRNAs transfections were carried out using Lipofectamine 3000 Transfection Reagent (Thermo Scientific, Cat# L300000) according to the manufacturer's directions. Briefly, 500.000 cells were plated in a 60 mm dish. After 24 hours, control (Cell Signaling, Cat# 6568S), or ADAR1 (Ambion, Cat# AM51331) siRNAs were transfected at a final concentration of 20 nM and incubated for additional 48 hours.

2.4 RNA isolation and DNase treatment

Total RNA was isolated using the E.Z.N.A. Total RNA Kit I (Omega Bio-Tek, Cat# R6834), according to the manufacturer's instructions. Samples were treated with RNase-free DNase Set I (Omega Bio-Tek, Cat# E1091) for DNA removal, as the manufacturer recommended. Concentration and quality of RNAs were measured in the Take 3 - Cytation 3 Instrument

(Biotek). Only suitable samples were used for downstream applications. All RNAs obtained were stored at -80°C.

2.5 cDNA synthesis

cDNA was synthesized by using the AffinityScript cDNA Synthesis Kit (Agilent Technologies, Cat# 600559), according to manufacturer's protocol. 1 µg of total RNA was used as a template in each reverse transcription reaction, and a mixture of oligo-dT and random primers was used to improve efficiency. All cDNA products were stored at -20°C.

2.6 RT-qPCR

Real-time PCR reactions were performed using an Eco Real-Time PCR System (Illumina, Cat# EC-900-1001) and Brilliant II SYBR Green QPCR Master Mix (Agilent Technologies, Cat# 600828). Reactions were prepared and performed according to the manufacturer's instructions. For the detection of mRNAs and lncRNAs, 35 cycles and 40 were used, respectively. β-actin served as an internal control. Relative RNA expression was calculated using the comparative cycle threshold (Ct) ($2^{-\Delta\Delta Ct}$) method. Primers were designed using Primer3Plus software (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) and synthesized by Integrated DNA Technologies (IDT).

2.7 RESS-qPCR

RNA Editing Site-Specific-qPCR (RESS-qPCR) was performed as described by Crews et al., 2015[19]. Briefly, for detecting each RNA editing site, two sets of primers were used, one pair for detecting WT transcript, and one pair detecting the edited transcript. cDNA synthesis and RT-qPCR were performed as described previously. Relative RNA editing ratios (Relative Edited/WT RNA) were calculated using the comparative cycle threshold (Ct) [$2^{-(Ct\ Edit - Ct\ WT)}$].

2.8 Protein extraction, SDS-PAGE and Western Blot

Cells were lysed with RIPA lysis buffer (Thermo Scientific, Cat# 89900) and protease inhibitors (Thermo Scientific, Cat# 1861284) following the manufacturer's protocol. Cell homogenates were separated by a 10% SDS-PAGE (Bio-Rad Laboratories, Cat# 4561034). Primary antibodies were probed overnight at 4°C. HRP-linked secondary antibodies (1:5000 dilution)

were incubated 1.5 hours at room temperature, and blots were visualized with SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific, Cat# 34579) in the ChemiScope3500 Mini chemiluminescence imaging system (Clinx Science Instruments). Western blot densitometry was performed using Image Studio Lite Software v5.2 (LI-COR) and normalized to β-actin. Primary antibodies used: 1:2000 rabbit polyclonal anti-ADAR1 (Abcam, Cat# ab168809), 1:5000 mouse anti-β-actin (Sigma-Aldrich, A1978).

2.9 RIP-qPCR

ADAR1 RNA-immunoprecipitation was performed as previously described by Hendrickson et al., 2016[20]. Briefly, $5x10^6$ cells/mL were crosslinked using formaldehyde to a final concentration of 0.1%. Cells were resuspended in 1 mL of RIPA lysis buffer with protease inhibitors and RNaseOUT (Invitrogen, Catalog # 10777019). Cell lysates were incubated with 4 µg of ADAR1 antibody (Abcam, Catalog # ab168809). Dynabeads Protein G beads were added and rotated at 4ºC. Beads were resuspended in RNase-free water and 3x reverse-crosslinking buffer, Proteinase K (Invitrogen, Catalog # 25530049), and RNaseOUT. Protein degradation and reverse-crosslinking were performed for 1 hour at 42ºC plus 1 hour at 55ºC. 1 mL of TRIzol Reagent (Invitrogen, Catalog #15596026) was added. Total RNA was isolated, and RT-qPCR was performed as above mentioned.

2.10 Analysis of differentially expressed lncRNAs

Data were obtained from Sagredo et al., 2020[21]. Differential expression analysis between MDA-MB-231 Mock and ADAR1 overexpression conditions was performed using DEseq2 software following standard recommendations. A False Discovery Rate (FDR) ≤ 0.05 was considered statistically significant, and no fold change cutoff was considered given that observed changes in lncRNAs expression were modest.

2.11 Kaplan-Meier survival curves

Overall Survival (OS) and Relapse-free Survival (RFS) curves were performed using the web-based Kaplan-Meier Plotter (https://kmplot.com/analysis/)[22] on breast cancer dataset. Gene symbol: LINC00944, Affymetrix ID: 1560573_at. *Log-rank* $p \leq 0.05$ was considered statistically significant.

2.12 Gene expression data retrieval

Gene expression (FPKM-UQ values) and clinical data were obtained from The Cancer Genome Atlas (TCGA) Project (https://www.cancer.gov/tcga) through the Genomic Data Commons (GDC) data portal (https://portal.gdc.cancer.gov/)[23]. TCGA studies retrieved: Breast invasive carcinoma (BRCA), Colon adenocarcinoma (COAD), Lung adenocarcinoma (LUAD), and Testicular Germ Cell Tumors (TGCT).

2.13 Guilt-by-Association analysis

LINC00944 expression was correlated to protein-coding genes (GRCh38) expression in normal and tumor samples by Pearson Correlation. The *p*-value was adjusted using the Benjamini-Hochberg method, and an FDR ≤ 0.05 was considered statistically significant. Correlations with Pearson coefficient (r) $\leq |0.3|$ were taken for computing the overlap against Hallmark gene sets collection[24] using the Gene Set Enrichment Analysis (GSEA) online tool (https://www.gsea-msigdb.org/gsea/index.jsp)[25]. An FDR ≤ 0.05 was considered statistically significant.

2.14 LINC00944 expression and clinicopathological parameters correlation

LINC00944 expression and clinical data from TCGA-BRCA were obtained, as mentioned above. Patients were classified based on the LINC00944 *z*-score. Upper (High LINC00944 expression group) and lower (Low LINC00944 expression group) quartiles were chosen for further comparison (n= 250). Data were arranged in 2x2 contingency tables, and the *Fisher's exact test* was run for the comparison. A *p*-value ≤ 0.05 was considered statistically significant.

2.15 Statistical analyses

R Software v3.6 (R Core Team, 2020) was used for data management. Statistical tests were performed in R Software v3.6 and GraphPad Prism v8.3 (GraphPad Software, La Jolla, CA, USA). *Two-tailed Student's t-test* was used in RT-qPCR to establish differences between two groups, and each analysis was performed in 5-6 independent experiments. For non-parametric data, the *Mann-Whitney test* was used in two-group comparisons.

5. Results

5.1 ADAR1 overexpression induces differential expression of lncRNAs

To assess the effect of *ADAR1* overexpression over lncRNAs expression levels, we examined our previously reported RNA-seq data on the breast cancer cell line MDA-MB-231 overexpressing ADAR1[21]. The ADAR1 overexpression resulted in the upregulation of 24 lncRNAs and the downregulation of 17 lncRNAs (*p-adj* ≤ 0.05) (**Fig. 1A and Supplementary Table 1**). According to GENCODE biotype annotation, differentially expressed lncRNAs were composed of 14 antisense RNAs (AS-RNAs), 18 long intergenic non-coding RNAs (lincRNAs), 7 processed transcripts, and 2 sense intronic RNAs (**Fig. 1A**).

In order to validate the observed expression changes in lncRNAs, we performed an adenoviral ADAR1 overexpression on MDA-MB-231 cells and assessed lncRNAs expression using RT-qPCR (**Fig. 1B**). Two downregulated, LINC00944 and APCDD1L-AS1, and two upregulated lncRNAs, LINC01003 and H1FX-AS1, were taken for this purpose. Controls for ADAR1 mRNA expression, protein levels, and function upregulation were performed (**Supplementary Fig. 1**). As is shown in Figure 1B, we validated the expression changes for LINC00944 ($p = 0.0003$, *Student´s t-test*), APCDD1L-AS1 ($p = 0.0327$, *Student´s t-test*) and LINC01003 ($p = 0.0004$, *Student´s t-test*), but H1FX-AS1 did not reach a significant difference ($p = 0.055$, *Student´s t-test*) (**Fig. 1B**).

Aiming to corroborate our findings in a second breast cancer cell line, we attempted to overexpress *ADAR1* in MDA-MB-436 cells. We were not able to obtain viable MDA-MB-436 ADAR1-overexpressing cells, perhaps because MDA-MB-436 has high intrinsic levels of ADAR1 compared to other breast cell lines (data not shown).

5.2 ADAR1 knockdown reverted the effect in LINC00944 expression levels

To further examine whether ADAR1 could alter the expression levels of lncRNAs in breast cancer, we investigated the effect of *ADAR1* loss-of-function in the MDA-MB-231 cell line using a siRNA (siADAR1). Controls for ADAR1 mRNA expression, protein levels, and function downregulation were performed (**Supplementary Fig. 2**). Interestingly, ADAR1 downregulation resulted in the upregulation of LINC00944 ($p < 0.0029$, *Student´s t-test*), and the downregulation of LINC01003 ($p < 0.0441$, *Student´s t-test*). However, we did not observe changes in APCDD1L-AS1 or H1FX-AS1 expression levels ($p < 0.835$ and $p < 0.2212$ respectively, *Student´s t-test*) (**Fig. 1C**).

In addition, we investigated the effect of *ADAR1* loss-of-function over lncRNAs expression levels in a second breast cancer cell line, MDA-MB-436. Controls for ADAR1 mRNA expression, protein levels, and function downregulation were performed (**Supplementary Fig. 3**). In the ADAR1 KD condition, we observed the expected upregulation of LINC00944 ($p = 0.0029$, *Student´s t-test*) and APCDD1L-AS1 ($p = 0.0004$, *Student´s t-test*). Nevertheless, no significant changes were found in LINC01003 ($p = 0.5977$, *Student´s t-test*), and H1FX-AS1 presented a consistent upregulation even after *ADAR1 knockdown* ($p = 0.0031$, *Student´s t-test*) (**Fig. 1D**).

5.3 ADAR1 and LINC00944 are interacting

Since LINC00944 expression levels were responsive to *ADAR1* gain- and loss-of-function, we decided to examine their potential interaction. To this end, we performed an ADAR1 RNA-immunoprecipitation followed by an RT-qPCR (RIP-qPCR). β-actin (*ACTB* gene) was used as a negative control since there is no reported interaction with ADAR1. Remarkably, we could recover ~30% of LINC00944 RNA in the ADAR1 immunoprecipitated fraction compared to the input sample, confirming that ADAR1 and LINC00944 are interacting in the breast cancer cell line MDA-MB-231 (**Fig. 1E**).

The canonical function of ADAR1 is the catalysis of adenosine's deamination reaction of adenosine to inosine in RNAs (A-to-I editing); thus, we next hypothesized that LINC00944 might be an editing target of ADAR1. To test this possibility, we searched for A-to-I variants in LINC00944 RNA on the RNA-seq data that originated from MDA-MB-231 cells overexpressing ADAR1[21]. The bioinformatic pipeline for A-to-I variant detection showed no evidence of editing in LINC00944 RNA, suggesting that ADAR1 may alter LINC00944 expression levels by means of non-canonical functions.

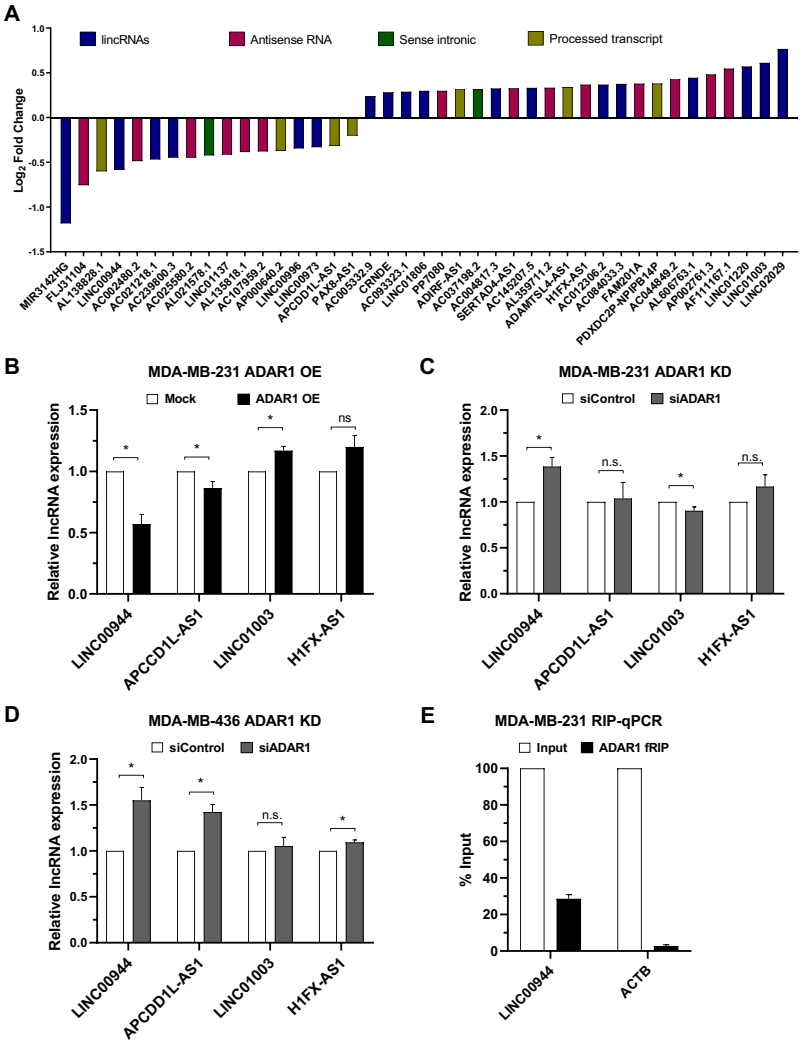


Figure 1. ADAR1 overexpression effects over lncRNAs expression levels in breast cancer cell lines. (A) Differentially expressed lncRNAs in MDA-MB-231 ADAR1 OE. *LincRNAs* are depicted in blue, *antisense* in pink, *sense intronic* in green and *processed transcripts* in yellow. **(B)** RT-qPCR. Experimental validation of expression changes in LINC00944, APCDD1L-AS1, LINC01003, and H1FX-AS1 in MDA-MB-231 ADAR1 OE. RT-qPCR. Assessment of LINC00944, APCDD1L-AS1, LINC01003, and H1FX-expression in **(C)** RT-qPCR. ADAR1 *knockdown* in MDA-MB-231 and **(D)** MDA-MB-436 cell lines. Expression values were calculated relative to the control condition using $2^{-(\Delta\Delta Ct)}$. β-actin was used as an internal control. Data are shown as mean, and error bars represent ± SEM of 5-6 biological replicates. Data were analyzed using unpaired, two-tailed *Student's t-test*. A *p*-value ≤ 0.05 was considered as statistically significant (*). *'ns'* indicates no significant difference between compared groups. **(E)** RIP-qPCR. ADAR1 RNA immunoprecipitation followed by RT-qPCR in MDA-MB-231 cells. Values are relative to the input sample. *ACTB* was used as a negative control. n=2.

5.4 LINC00944 has low expression in human tissues

To gain insight into LINC00944, we examined its expression across 46 human tissues based on data generated by the Genotype-Tissue Expression (GTEx) Project (https://gtexportal.org/home/). As is shown in Figure 2A, LINC00944 has a general low expression in normal tissues, as is expected for most lncRNAs. The highest expression was found in Testis (Median TPM: 9.1, n=361), followed by Spleen (Median TPM: 1.6, n=241) and Small intestine (Median TPM: 1.0, n=187), while Breast showed a moderate expression (Median TPM: 0.11, n=459). The lowest expression values were found in brain and heart tissues (Heart-Left ventricle, Median TPM: 0.013, n= 432, and Brain-Nucleus accumbens, Median TPM: 0.013, n=246) (**Fig. 2A**). Interestingly, no expression (Median TPM: 0) was reported in several brain tissues as the cerebellar hemisphere, cerebellum, caudate (basal ganglia), hypothalamus, putamen (basal ganglia), and substantia nigra.

5.5 LINC00944 is connected to immune signaling pathways in normal and tumor samples

To begin to understand the LINC00944 function, we used the Guilt-by-Association method. To this end, we performed correlation analysis between LINC00944 and all protein-coding genes in normal and tumor datasets from the TCGA-BRCA cohort (*Pearson correlation*, $|r| \geq 0.3$ and p-adj ≤ 0.01), followed by an overlap against Hallmark Collection. The results show a connection between LINC00944 and the immune system-related functions, as 'Interferon-gamma response', 'Inflammatory response', 'IL2 STAT5 signaling' and 'TNFA signaling via NFKB' gene set collections were significantly represented in normal and tumor datasets (**Fig. 2B and C, and Supplementary Tables 2 and 3**). To provide further support to this scenario, we took additional TCGA datasets in which LINC00944 was shown to have high expression, as Testis, Lung (Median TPM: 0.43, n=578), and Colon (Median TPM: 0.29, n=406) (**Fig. 2A**), and performed the Guilt-by-Association analysis. Remarkably, this analysis showed that in normal tissue from the lung, in Lung adenocarcinoma (TCGA-LUAD), in Testicular Germ Cell Tumors (TCGA-TGCT), and in Colon Adenocarcinoma (TCGA-COAD) cohorts, LINC00944 is also significantly related to immune functions (**Supplementary Fig. 4A, B, C,**

and D, respectively). Thus, our results implied that LINC00944 might be participating in immune-related signaling pathways.

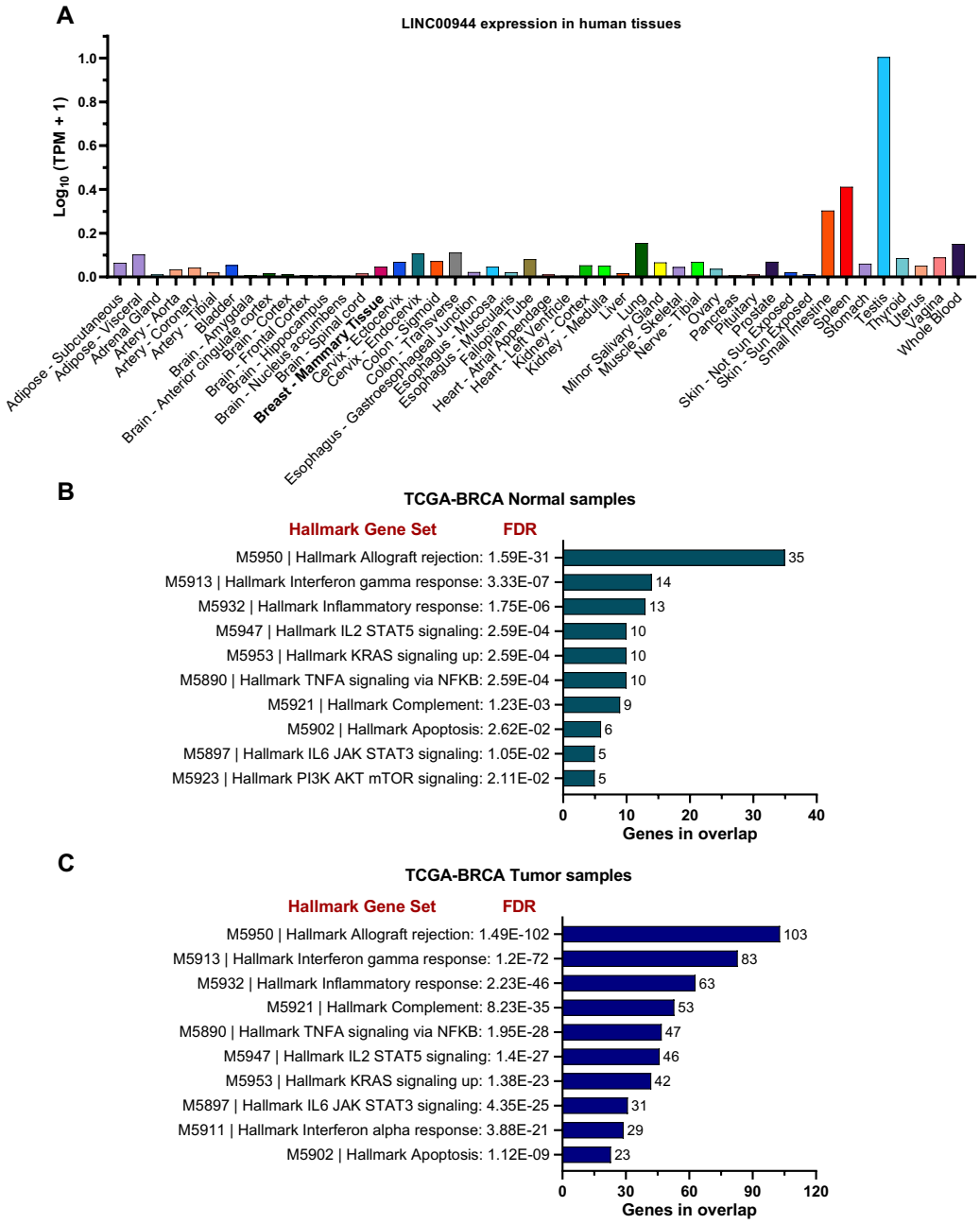


Figure 2. Guilt-by-Association analysis connects LINC00944 to immune signaling pathways. (A) LINC00944 tissue-specific median expression (Log_{10} TPM+1) across 46 human tissues. Data were obtained from the Genotype-Tissue Expression (GTEx) Project Portal (https://www.gtexportal.org/home/gene/LINC00944). **(B)** Guilt-by-Association analysis of LINC00944 co-expressed protein-coding genes followed by an overlap against Hallmark Collection in normal (solid tissue normal) and **(C)** tumor samples from the TCGA-BRCA cohort. Significant terms were ranked based on the number of genes in the overlap, and the top 10 are depicted. TPM: Transcripts Per Million; FDR: False discovery Rate.

5.6 LINC00944 expression serves as a marker for tumor-infiltrating lymphocytes in breast cancer

Since the Guilt-by-Association analysis predicted that LINC00944 might be participating in immune signaling pathways, we next sought to ascertain whether LINC00944 could be a good marker for tumor-infiltrating lymphocytes (TILs). Two different approaches were employed for this purpose. In the first, RNA-seq data was used to assess the expression of high confidence marker genes for T-cell population[26] in tumor samples from TCGA-BRCA. As is shown in Fig. 3A-F, LINC00944 expression was positively correlated to the six confidence marker genes, CD3D ($p < 0.0001$, *Pearson* $r = 0.63$), CD3E ($p < 0.0001$, *Pearson* $r = 0.60$), CD3G ($p < 0.0001$, *Pearson* $r = 0.61$), CD6 ($p < 0.0001$, *Pearson* $r = 0.62$), SH2D1A ($p < 0.0001$, *Pearson* $r = 0.61$) and TRAT1 ($p < 0.0001$, *Pearson* $r = 0.59$), respectively, suggesting that LINC00944 could serve as marker for the presence of T-cells in tumor microenvironment. To provide additional support to this inference, we performed the same correlation in TCGA-TGCT, TCGA-LUAD, and TCGA-COAD datasets, finding significant and strong expression correlations in each case (**Supplementary Table 4**).

In the second approach, we reviewed previously published data in which exome reads mapping was employed to detect tumor-infiltrating T lymphocytes (immune DNA signature, *iDNA score*) in tumor samples from TCGA-BRCA cohort[27]. By using this data, we observed that breast cancer patients expressing low levels of LINC00944 have lower iDNA scores, thus suggesting a lower infiltration of T lymphocytes in those patients (**Fig. 3G**).

Together, our results suggest that LINC00944 expression positively correlates to tumor-infiltrating T lymphocytes, indicating that LINC00944 could serve as a marker for infiltration of T lymphocytes in tumor microenvironment.

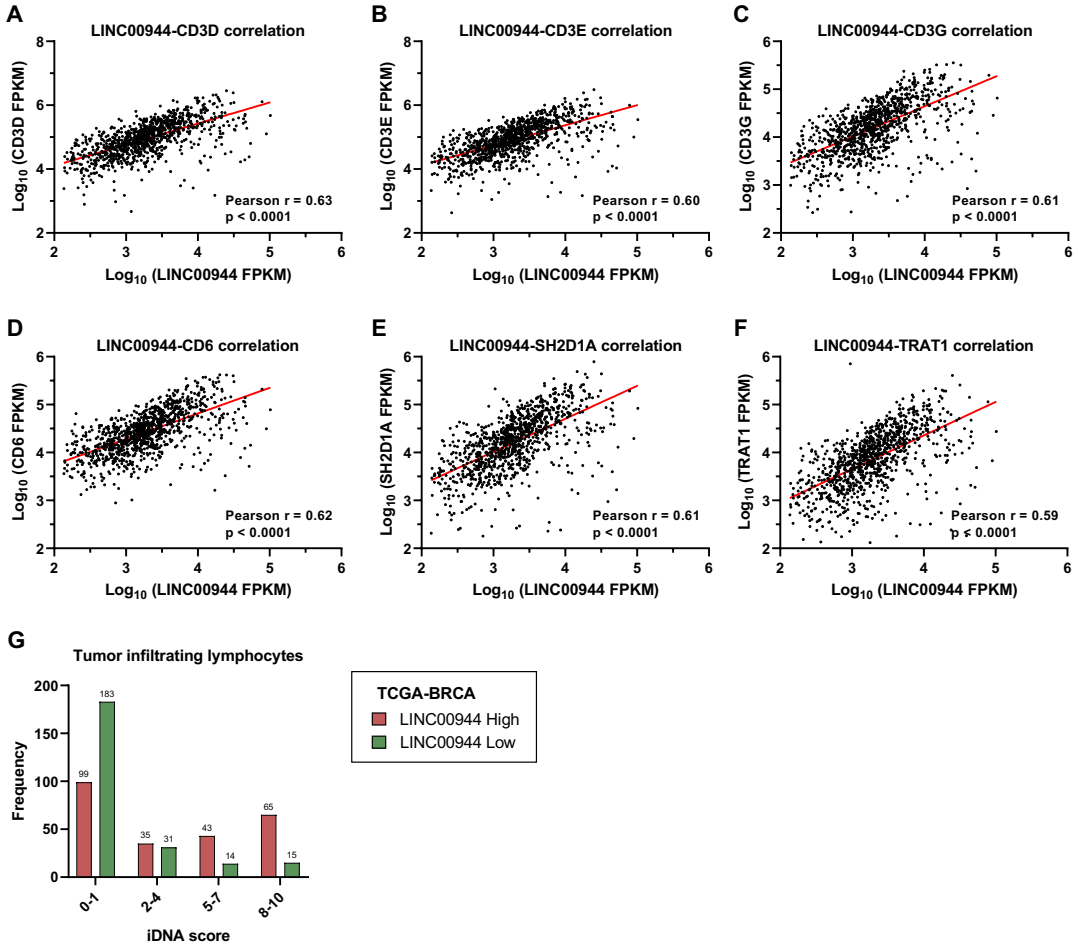


Figure 3. LINC00944 expression positively correlates to tumor-infiltrating T lymphocytes confident markers in breast cancer. Pearson correlation between LINC00944 and **(A)** CD3D, **(B)** CD3E, **(C)** CD3G, **(D)** CD6, **(E)** SH2D1A and **(F)** TRAT1 expression in TCGA-BRCA tumor samples. A p-value ≤ 0.05 was considered as statistically significant. **(G)** iDNA score as an indicator of tumor-infiltrating T lymphocytes in patients expressing high (red) and low (green) levels of LINC00944 in TCGA-BRCA tumor samples. FPKM: Fragment Per Kilobase Million.

5.7 Pro-apoptotic markers are downregulated in breast cancer patients expressing low levels of LINC00944

One major role of the immune system in tumor progression is to drive apoptosis in cancer cells. With that in mind, we then aimed to elucidate if LINC00944 expression was correlated to alterations in programmed cell death. By interrogating the gene expression of apoptotic markers in TCGA-BRCA, we observed that patients expressing low levels of LINC00944 (LINC00944 low group) have a significant downregulation of the pro-apoptotic markers Bak (*BAK1*) ($p < 0.0001$, *Mann-Whitney test*) and Bax (*BAX*) ($p < 0.0001$, *Mann-Whitney test*) (**Fig. 4A**), and an upregulation of the anti-apoptotic markers Bcl2 (*BCL2*) ($p < 0.0001$, *Mann-Whitney test*) and $BclX_L$ (*BCL2L1*) ($p < 0.0001$, *Mann-Whitney test*) when compared to patients expressing high levels of LINC00944 (**Fig. 4B**). In addition, the initiator caspase, caspase-8 (*CASP8*) ($p < 0.0001$, *Mann-Whitney test*) and the executioner caspase, caspase-3 (*CASP3*) ($p = 0.0406$, *Mann-Whitney test*), were significantly downregulated in LINC00944 low group (**Fig. 4C**). Taken together, these results suggest that low expression of LINC00944 in breast cancer patients is correlated to a decrease in the apoptotic program.

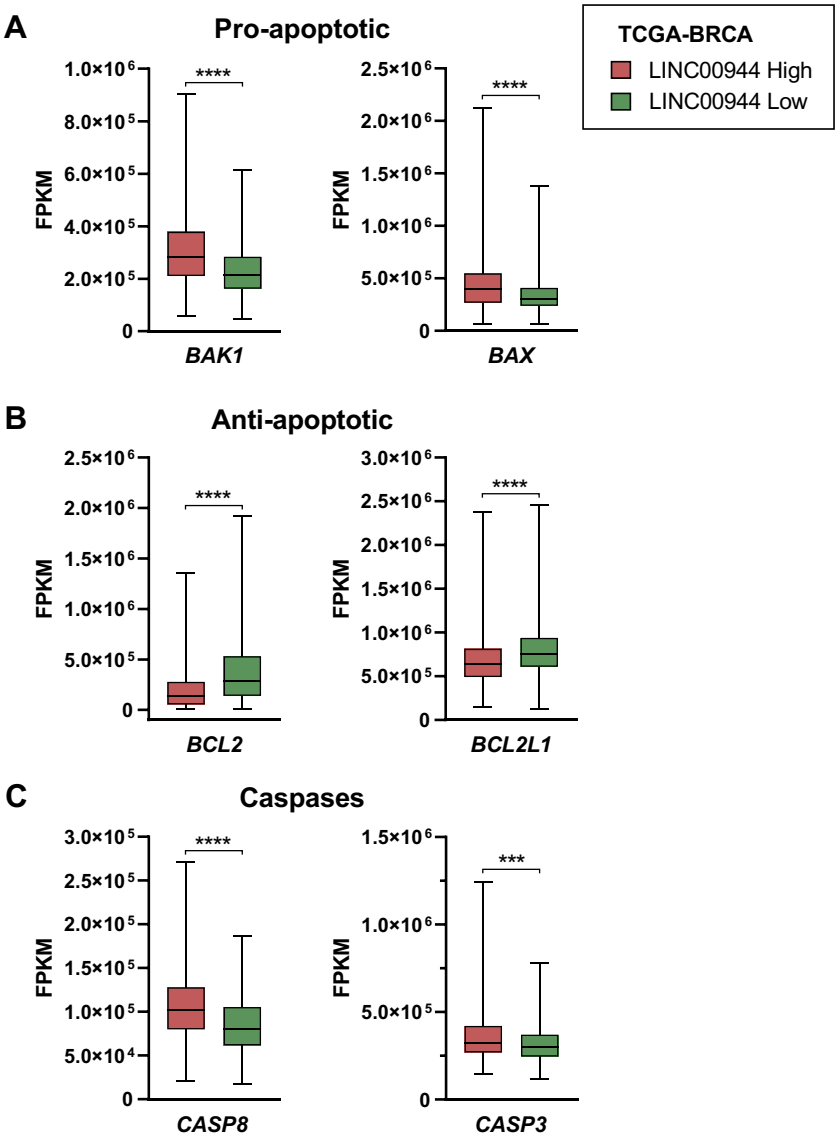


Figure 4. Pro-apoptotic markers are downregulated in breast cancer patients expressing low levels of LINC00944. Expression levels of pro-apoptotic markers *BAK1* and *BAX*, anti-apoptotic markers *BCL2* and *BCL2L1,* and the caspases 8 and 3 (*CASP8, CASP3)* in TCGA-BRCA LINC00944 high (red) and low (green) groups. Data are depicted as Log_2(FPKM), and error bars represent Min to Max values. Data were analyzed using the *Mann-Whitney test*. **** $p < 0.0001$, *** $p = 0.0406$. FPKM: Fragment Per Kilobase Million.

5.8 LINC00944 expression in breast cancer patients is correlated to clinicopathological parameters

Considering that immune pathways are important components in the anti-tumor process, we evaluated whether measuring LINC00944 expression had a clinical value in diagnosing breast cancer. By interrogating RNA-seq expression data from TCGA-BRCA, we detected no significant differences between normal (n=113) and primary tumor (n=1045) samples (p = 0.2014, *Mann-Whitney test*) (**Fig. 5A**). Nevertheless, when we subdivided tumor samples by receptor status, we found that triple-negative breast cancer patients (TNBC, n=113) had significant upregulation of LINC00944 when compared to normal samples ($p < 0.0001$, *Kruskal-Wallis test followed by Dunn's multiple comparison test*) (**Fig. 5B**). This data suggests that LINC00944 high levels would be useful as a biomarker in TNBC patients.

To further explore the clinicopathological value of LINC00944 expression we examined the clinical data provided by TCGA-BRCA, which revealed that LINC00944 expression was correlated to age at diagnosis (p = 0.008, *Fisher's exact test*), tumor size (AJCC Pathologic T; p = 0.041, *Fisher's exact test*) (**Table 1**), estrogen receptor status ($p = 1.84E^{-17}$, *Fisher's exact test*), progesterone receptor status ($p = 5.69E^{-09}$, *Fisher's exact test*) and breast cancer subtype by IHC ($p = 2.92E^{-09}$, *Chi-squared test*) (**Table 2**). However, no significant correlations were observed between LINC00944 expression and other clinicopathological parameters such as pathologic stages (p = 0.329, *Fisher's exact test*), metastasis to lymph nodes (AJCC Pathologic N; p = 0.149, *Fisher's exact test*), distant metastases (AJCC Pathologic M; p = 0.441, *Fisher's exact test*) and HER2 status (p = 0.111, *Fisher's exact test*).

Table 1. Correlation between LINC00944 expression and clinicopathological parameters in breast cancer patients

Parameters	n	LINC00944 expression		*p*-value
		High (n=250)	Low (n=250)	
LINC00944 expression				
FPKM (mean)		12498	473.5	
Age at diagnosis				
≥ 50 years	363	168 (46.3%)	195 (53.7%)	0.008*
< 50 years	132	79 (59.8%)	53 (40.2%)	
Pathologic Stage (AJCC Pathologic Stage)				
Stages I + II	378	194 (51.3%)	184 (48.7%)	0.329
Stages III + IV	110	50 (45.5%)	60 (54.5%)	
Tumor size (AJCC Pathologic T)				
T1 + T2	415	217 (52.3%)	198 (47.7%)	0.041*
T3 + T4	83	33 (39.8%)	50 (60.2%)	
Lymph node metastasis (AJCC Pathologic N)				
N0	249	134 (53.8%)	115 (46.2%)	0.149
N1 + N2 + N3	242	114 (47.1%)	128 (52.9%)	
Distant metastasis (AJCC Pathologic M)				
M0	413	200 (48.4%)	213 (51.6%)	0.441
M1	11	4 (36.4%)	7 (63.6%)	

TCGA-BRCA LINC00944 high and low expression groups were determined based on the LINC00944 *z*-score. Data were analyzed using the *Fisher's exact test*. A *p*-value ≤ 0.05 was considered statistically significant (*). AJCC: American Joint Committee on Cancer.

Table 2. Correlation between LINC00944 expression and receptor status in breast cancer patients

Parameters	n	LINC00944 expression		*p*-value
		High (n=250)	Low (n=250)	
Estrogen receptor status				
Positive	331	125 (37.8%)	206 (62.2%)	$1.84E^{-17}$*
Negative	143	114 (79.7%)	29 (20.3%)	
Progesterone receptor status				
Positive	278	108 (38.8%)	170 (61.2%)	$5.69E^{-09}$ *
Negative	195	129 (66.2%)	66 (33.8%)	
HER2 status				
Positive	73	30 (41.1%)	43 (58.9%)	0.111
Negative	254	133 (52.4%)	121 (47.6%)	
Subtype by IHC				
HR+/HER2-	181	73 (41.3%)	108 (59.7%)	
HER2+	73	30 (41.1%)	43 (58.9%)	$2.92E^{-09}$ * #
Triple-negative	70	58 (82.9%)	12 (17.1%)	

TCGA-BRCA LINC00944 high and low expression groups were determined based on the LINC00944 *z*-score. Data were analyzed using the *Fisher's exact test*. # Data was analyzed using the *Chi-squared test*. A *p*-value ≤ 0.05 was considered statistically significant (*). HR: Hormone receptor; IHC: Immunohistochemistry.

5.9 LINC00944 low expression in breast cancer patients is correlated to poor survival outcomes

Considering that alterations in immune pathways, lower T lymphocytic infiltration, and dysregulation in cell death are key traits for tumor progression, we wanted to further evaluate if LINC00944 expression has a prognostic significance in breast cancer patients' survival. By using the web-based Kaplan Meier-Plotter (https://kmplot.com/analysis/), we assessed survival on the breast cancer dataset, finding that both Overall Survival (OS) ($p = 0.011$, *Log-rank test*) and Relapse-Free Survival (RFS) ($p = 2.1E^{-06}$, *Log-rank test*) are significantly poorer in patients expressing low levels of LINC00944 (**Fig. 5C** and **D**, respectively). The analysis showed that the low expression cohort had a reduction of 40.8 months in OS (106.8 vs. 66 months), while a decrease of 22.4 months in RFS (53.6 vs. 31.2 months). From this data, we can conclude that low expression of LINC00944 in breast cancer patients is correlated to poor overall survival and relapse-free survival.

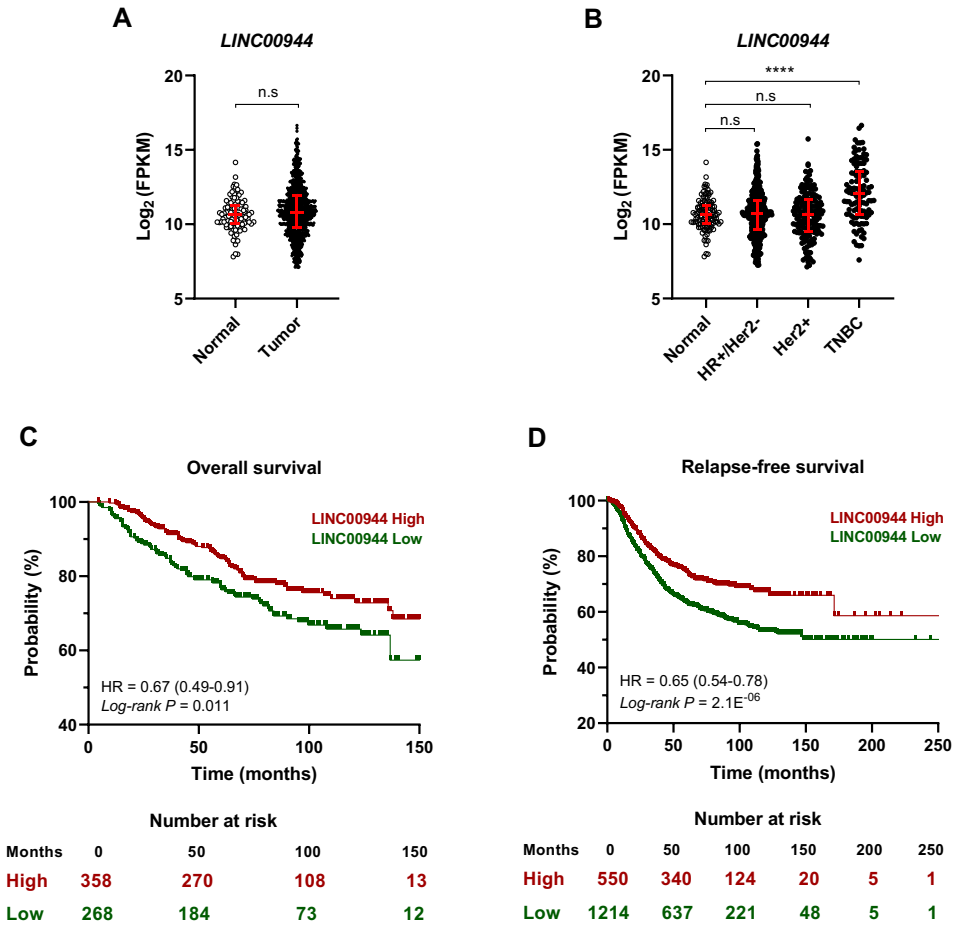


Figure 5. LINC00944 low expression in breast cancer patients is correlated to poor Overall survival and Relapse-free survival. (A) LINC00944 expression from the TCGA-BRCA cohort in solid tissue normal (n=113) and primary tumor (n=1045) samples. Data were analyzed using the *Mann-Whitney test*. **(B)** LINC00944 expression in normal (n=113), hormone receptor-positive (HR+/HER2-), HER2 positive (HER2+), and TNBC samples. Expression data is depicted as Log_2 (FPKM), and error bars represent median with interquartile range. Data were analyzed using the *Kruskal-Wallis test,* followed by *Dunn's multiple comparison test*. **** $p < 0.0001$. **(C)** Kaplan-Meier survival curves and Risk tables for breast cancer patients based on LINC00944 expression. Red and green lines indicate high and low LINC00944 expression, respectively. **(D)** Relapse-free survival curves of 1764 breast cancer patients. The *Log-rank test* was used to analyze data. A p-value ≤ 0.05 was considered statistically significant. *'ns'* indicates no significant difference between compared groups. TNBC: triple-negative breast cancer; HR: Hazard Ratio; FPKM: Fragment Per Kilobase Million.

6. Discussion

The ADAR1 upregulation observed in several cancer types has been largely described as an oncogenic feature[28]. Evidence has shown that ADAR1 has no preference for a specific type of RNA, being extensively described in interactions with mRNAs, microRNAs, and viral RNAs[1]. Despite the sharp rise in lncRNAs studies, only a few have addressed ADAR1-lncRNAs interactions. In this paper, we provided a transcriptomic analysis of the effect of ADAR1 over lncRNAs expression. 41 lncRNAs were differentially expressed after ADAR1 upregulation (p-adj ≤ 0.05), and we could further confirm that the *ADAR1* loss-of-function reverts the effect on expression in some of those in the breast cancer cell lines MDA-MB-231 and MDA-MB-436. A quick exploration of DE lncRNAs in PubMed (https://www.ncbi.nlm.nih.gov/pubmed) showed that several lncRNAs were already linked to cancer development. For instance, APCDD1L-AS1 was found to have an important prognostic value in lung squamous cell carcinoma[29]. Functional enrichment analysis of co-expressed genes revealed that critical pathways in cancer development as 'positive regulation of cell migration' and 'proteinaceous extracellular matrix' were enriched[29]. In our analysis, we showed that ADAR1 gain- and loss-of-function were capable of modulating the expression of APCDD1L-AS1, indicating that this lncRNA may also have a role in the breast cancer malignancy related to ADAR1. Another good example is the upregulation of the lncRNA FAM201A, which has been found mediating metastasis of lung squamous cell cancer[30] and resistant to radiotherapy in non-small cell lung cancer[31] and esophageal squamous cell cancer[32]. In the ADAR1 OE condition, we found a significant upregulation of FAM201A, suggesting that this lncRNA may also have a role in ADAR1-mediated tumorigenesis in breast cancer. On the other hand, the majority of DE lncRNAs have no associated literature, so we postulate them as principal candidates for exploring their function in cancer.

In the present study, we showed that LINC00944 expression levels were susceptible to ADAR1 up- and downregulation in two breast cancer cell lines, and the RNA-immunoprecipitation approach suggested that ADAR1 and LINC00944 may be interacting in breast cancer cells. Mechanisms by which ADAR1 can disrupt RNA expression levels through A-to-I editing range from stability impairment to alteration in splicing processing and nuclear

retention in paraspeckles, among others[1]. In LINC00944, we found no evidence of A-to-I editing, excluding it as an underlying mechanism. On the other hand, it has been shown that ADAR1 can modify RNA expression levels by complexing with other RNA-binding proteins, as Dicer[16] and by an interplay with Staufen[33] and HuR[34]. These mechanisms are worth to be explored for a complete understanding of the ADAR1-LINC00944 interplay.

LINC00944 has been outlined in cancer in previous publications. For instance, LINC00944 expression has been shown as downregulated in colorectal cancer tissues from patients with liver metastasis[35] and reported as an epigenetically activated lncRNA in several cancer types when comparing to their normal tissues[36]. Besides, LINC00944 was recently identified as an immune-related lncRNA in cancer[37]. This data agrees with our findings, in which we reported that LINC00944 has a strong relationship with and may play a role in immune-related signaling pathways. Nonetheless, the molecular mechanism underlying LINC00944 participation should be further investigated.

The immune system plays key roles in cancer initiation and development[38], and the infiltration of immune cells into tumor microenvironment is a major factor in cancer progression[39]. By using publicly available data generated from RNA-sequencing and Exome sequencing, we observed that LINC00944 positively correlates to tumor-infiltrating T lymphocytes (TILs) in the tumor microenvironment, postulating LINC00944 as a useful marker. These findings become relevant when considering that TILs have been related to a favorable clinical outcome in several cancer types, including colorectal cancer[40], non-small cell lung cancer[41], and breast cancer[42].

The execution of an effective immune response in the tumor microenvironment could lead to increased programmed cell death in cancer cells[43]; thus, we further studied a correlation between LINC00944 expression and apoptosis. In the present study, we showed that patients expressing low levels of LINC00944 had downregulation of pro-apoptotic markers, as the gene expression of *BAX* and *BAK1*, and the caspases *CASP8* and *CASP3* are significantly downregulated in breast cancer. Accordingly, the expression of the anti-apoptotic *BCL2* and

BCL2L1 is upregulated. These data may be indicating that cells expressing low levels of LINC00944 have a decrease in the apoptotic process, yet functional assays are needed.

LINC00944 expression between normal and tumor samples from breast cancer was not statistically different, yet we found that LINC00944 was statistically upregulated in triple-negative breast cancer (TNBC) patients. In the same line, assessment of LINC00944 expression and receptor status showed that LINC00944 expression was positively correlated to estrogen and progesterone receptor, all relevant medical concerns regarding therapies.

Regarding clinicopathological parameters, we found that LINC00944 expression is correlated to the age at diagnosis and tumor size. Interestingly, samples coming from larger tumors had low expression levels of LINC00944. Taken the results mentioned above, this may be implying that those tumors have a higher fraction of malignant cells than tumor-infiltrating T lymphocytes and that the apoptotic program may be decreased. Kaplan-Meier survival analysis supported this idea, as Overall survival and Relapse-free survival decreased significantly in the LINC00944 low expression cohort.

7. Conclusion

In summary, LINC00944 expression was responsive to variations in ADAR1 levels, an important player in oncogenic processes. LINC00944 may have a role in immune signaling pathways since a strong connection with the immune system was found. LINC00944 was largely upregulated only in triple-negative breast cancer samples, and lower expression indicated poorer survival outcomes. Moreover, our results suggested that LINC00944 may be used with prognostic value, as its expression was correlated to tumor-infiltrating T lymphocytes and a decrease in pro-apoptotic markers.

8. Funding

This study was financially supported by Government of Chile through Agencia Nacional de Investigación y Desarrollo (ANID), Ph.D. fellowship 21161206 to PRdS, Ph.D. fellowship 21130317 to FM, FONDECYT 1151435 to KM, FONDECYT 1171004 to MS, FONDECYT 1151446 to RA

9. Authors' contributions

PRdS, MS, and RA conceived and designed the project. PRdS and FM performed experiments. PRdS and AB analyzed the data. PRdS, MS, and RA interpreted the data. PRdS wrote the manuscript. OH, KM, MS and RA discussed data, reviewed, and edited the manuscript. All authors read and approved the final manuscript. MS and RA jointly supervised this project.

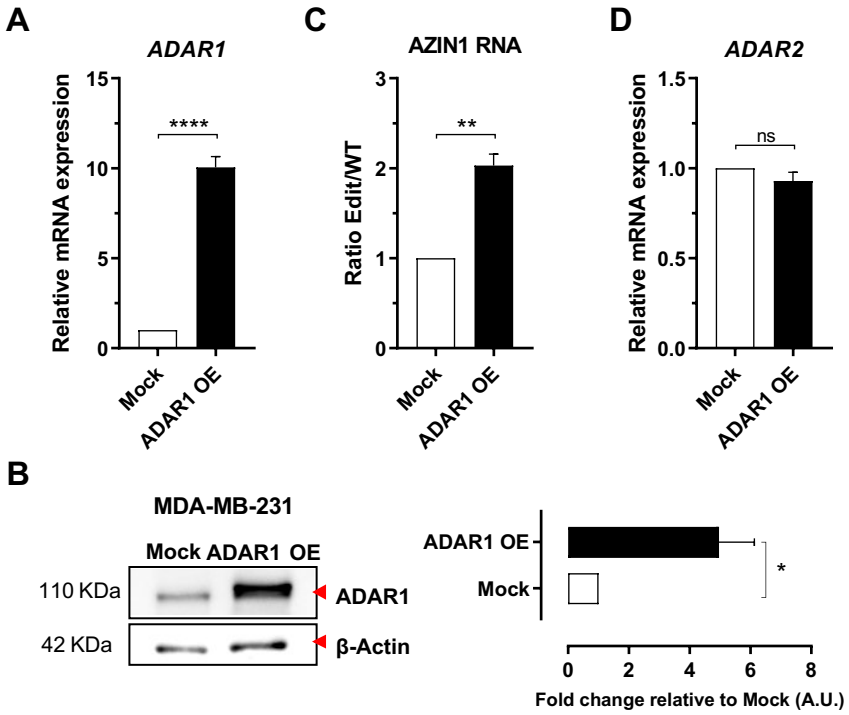
10. Conflict of interest statement

PRdS, AB, FM, and RA were Pfizer Chile employees. The authors have no further financial or non-financial conflicts of interest.

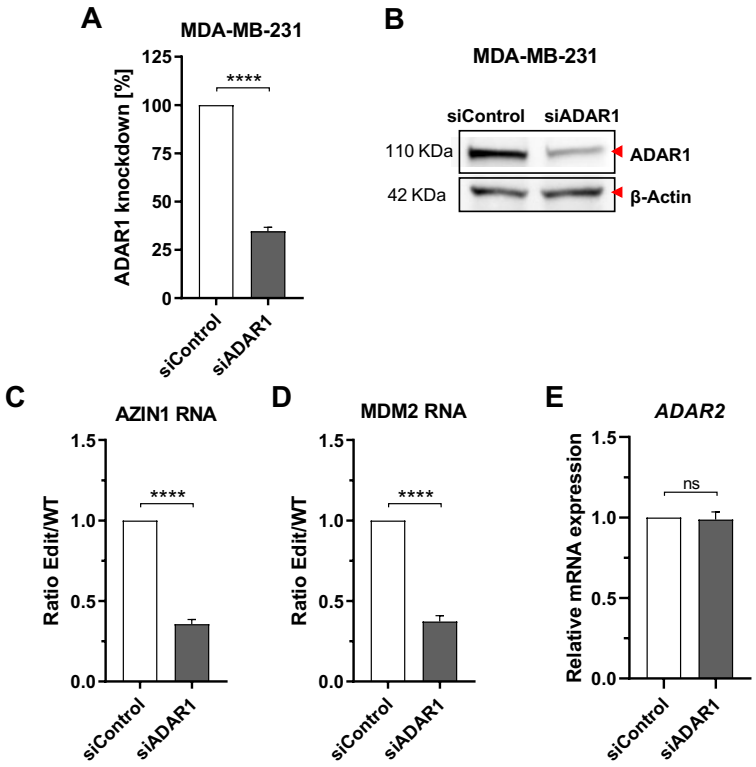
11. Acknowledgments

The results shown here are in part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health,and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from dbGaP accession number phs000424.v8.p2).

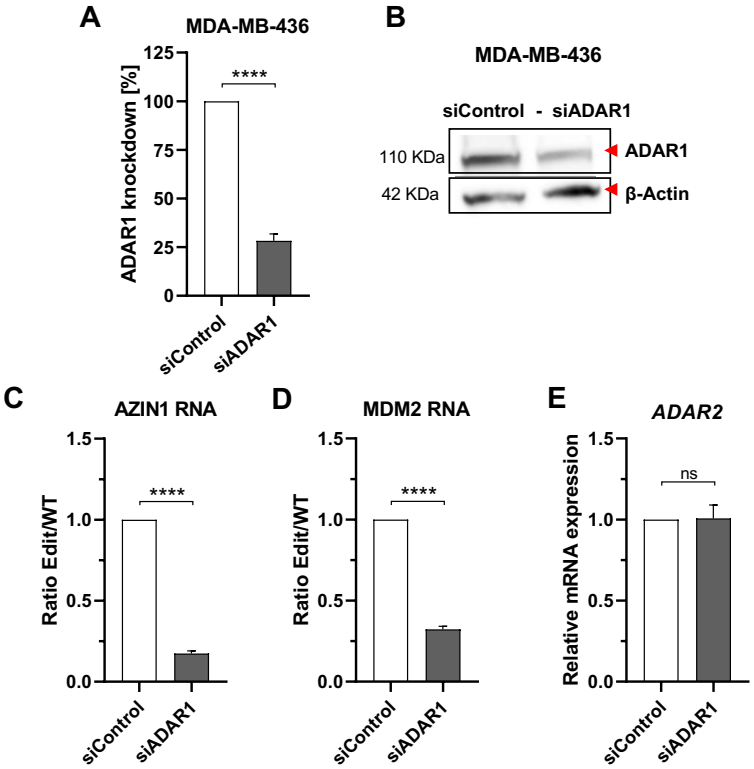
12. Supplementary Figures



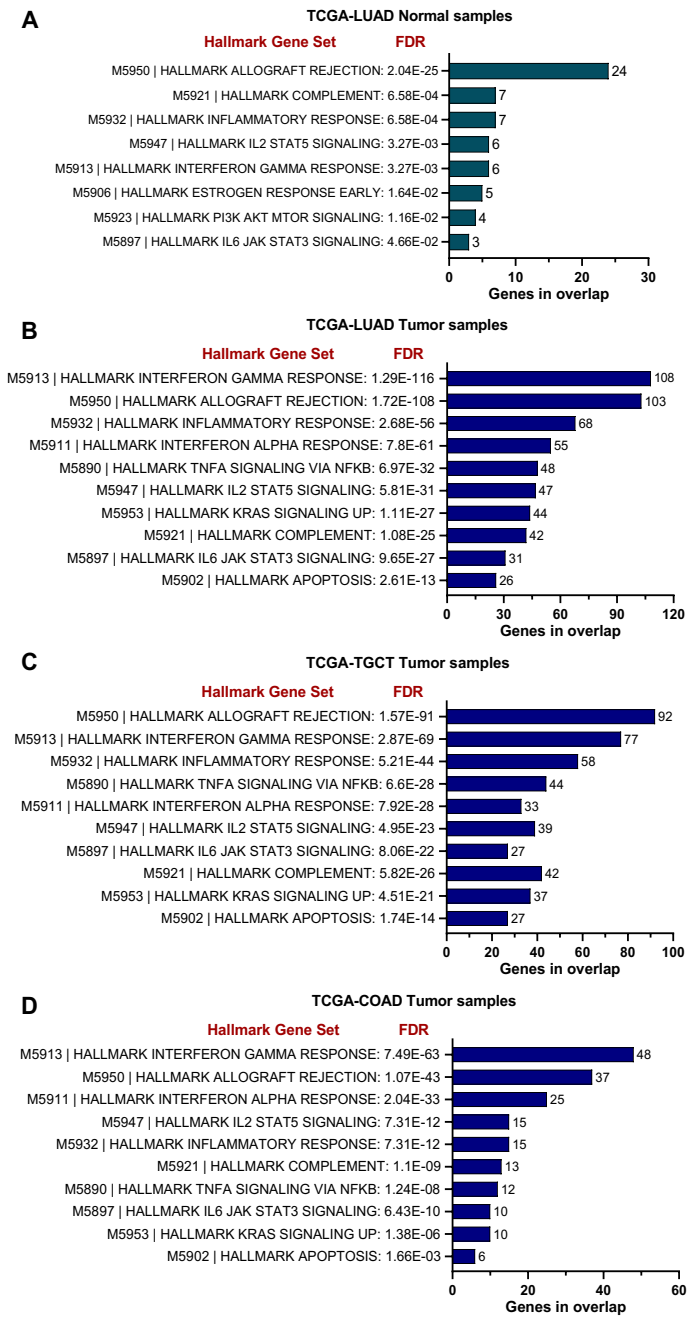
Supplementary Figure 1. Experimental validation of ADAR1 overexpression on the MDA-MB-231 cell line. (A) RT-qPCR. ADAR1 mRNA expression evaluation after adenovirus transduction (MOI 200). Expression values were calculated relative to the Mock condition using $2^{-(\Delta\Delta Ct)}$. **** $p < 0.0001$. **(B)** Western Blot. ADAR1 protein levels evaluation after adenovirus transduction. (Top) Representative image of ADAR1 overexpression. (Bottom) Quantification of band intensities relative to loading control. * $p < 0.0268$. **(C)** RESS-qPCR. Editing assessment of canonical ADAR1 target AZIN1. Editing ratios were calculated relative to Mock condition using $2^{-(Ct\ Edit\ -\ Ct\ WT)}$. ** $p < 0.0024$. **(D)** RT-qPCR. ADAR2 mRNA evaluation after ADAR1 manipulation. β-Actin was used as an internal control. Data are shown as mean, and error bars represent ± SEM of 3 biological replicates. Data were analyzed using unpaired, two-tailed *Student's t-test*. A p-value ≤ 0.05 was considered statistically significant. *'ns'* indicates no significant difference between compared groups. A.U.: Arbitrary Units.



Supplementary Figure 2. Experimental validation of ADAR1 knockdown on the MDA-MB-231 cell line. (A) RT-qPCR. ADAR1 mRNA expression after treatment with siADAR1 (20 nM) in MDA-MB-231 cells. Expression values were calculated relative to control siRNA (siControl) using $2^{-(\Delta\Delta Ct)}$. **(B)** Western Blot. ADAR1 protein levels evaluation after siADAR1 treatment (20 nM). Representative image of ADAR1 KD, n=1. **(C)** RESS-qPCR. Editing assessment of canonical ADAR1 targets AZIN1 and **(D)** MDM2. Editing ratios were calculated relative to the siControl using $2^{-(Ct\ Edit\ -\ Ct\ WT)}$. **(E)** RT-qPCR. ADAR2 mRNA evaluation after ADAR1 KD. Data are shown as mean, and error bars represent ± SEM of 6 biological replicates. Data were analyzed using unpaired, two-tailed *Student's t-test*. **** $p < 0.0001$. A p-value ≤ 0.05 was considered statistically significant. *'ns'* indicates no significant difference between compared groups. The siRNA against ADAR1 reached a decrease of ~70% in both mRNA and protein levels. ADAR1 function also decreased, as its editing targets AZIN1 and MDM2 diminished their edited fraction. As a control, ADAR2 mRNA levels were measure finding no changes after ADAR1 KD ($p = 0.792$).



Supplementary Figure 3. Experimental validation of ADAR1 knockdown on the MDA-MB-436 cell line. (A) RT-qPCR. ADAR1 mRNA expression after treatment with siADAR1 (20 nM) in MDA-MB-436 cells. Expression values were calculated relative to control siRNA (siControl) using $2^{-(\Delta\Delta Ct)}$. **(B)** Western Blot. ADAR1 protein levels evaluation after siADAR1 treatment (20 nM). Representative image of ADAR1 KD, n=1. **(C)** RESS-qPCR. Editing assessment of canonical ADAR1 targets AZIN1 and **(D)** MDM2. Editing ratios were calculated relative to siControl using $2^{-(Ct\ Edit\ -\ Ct\ WT)}$. **(E)** RT-qPCR. ADAR2 mRNA evaluation after ADAR1 KD. Data are shown as mean, and error bars represent ± SEM of 6 biological replicates. Data were analyzed using unpaired, two-tailed *Student's t-test*. **** $p < 0.0001$. A p-value ≤ 0.05 was considered as statistically significant. *'ns'* indicates no significant difference between compared groups. The siRNA against ADAR1 reached a decrease of ~70% in both mRNA and protein levels. ADAR1 function also decreased, as its editing targets AZIN1 and MDM2 diminished their edited fraction. As a control, ADAR2 mRNA levels were measure finding no changes after ADAR1 knockdown ($p = 0.935$).



Supplementary Figure 4. Guilt-by-Association analysis connects LINC00944 to immune signaling pathways in different datasets from TCGA. The analysis was performed in **(A)** normal samples (n=58, *Pearson* $|r| \geq 0.3$) and **(B)** tumor samples from TCGA-LUAD (n=503, *Pearson* $|r| \geq 0.3$), **(C)** tumor samples from the TCGA-TGCT (n=137, *Pearson* $|r| \geq 0.5$) and **(D)** tumor samples from TCGA-COAD (n=409, *Pearson* $|r| \geq 0.3$) cohorts. Significant terms were ranked based on the number of genes in overlap against Hallmark Collection[24], and the top 10 are depicted. FDR: False discovery Rate.

13. Supplementary Tables

Supplementary Table 1. Differentially expressed lncRNAs after ADAR1 overexpression in the breast cancer cell line MDA-MB-231 (Sagredo et al., 2020[21])

ENSEMBL ID	Gene symbol	Log_2 (FoldChange)	*p*-adj	Direction	Gene type
ENSG00000253522	MIR3142HG	-1.183304936	5.29E-19	Downregulated	lincRNA
ENSG00000227908	FLJ31104	-0.755619169	9.83E-08	Downregulated	Antisense RNA
ENSG00000237596	AL138828.1	-0.600826589	6.60E-05	Downregulated	Antisense RNA
ENSG00000256128	LINC00944	-0.583417638	0.000103942	Downregulated	lincRNA
ENSG00000232759	AC002480.2	-0.487173631	3.07E-07	Downregulated	Antisense RNA
ENSG00000204876	AC021218.1	-0.468997229	0.000150168	Downregulated	lincRNA
ENSG00000263590	AC239800.3	-0.449980053	0.000675278	Downregulated	lincRNA
ENSG00000259354	AC025580.2	-0.449099873	0.006173816	Downregulated	Antisense RNA
ENSG00000275894	AL021578.1	-0.423251869	0.001800796	Downregulated	Sense intronic
ENSG00000233621	LINC01137	-0.414660577	0.012772368	Downregulated	Antisense RNA
ENSG00000258875	AL135818.1	-0.385465794	0.026761998	Downregulated	Processed transcript
ENSG00000246130	AC107959.2	-0.377780659	0.020583826	Downregulated	Antisense RNA
ENSG00000255355	AP000640.2	-0.373528260	0.024920440	Downregulated	Processed transcript
ENSG00000242258	LINC00996	-0.345441714	0.044371677	Downregulated	lincRNA
ENSG00000240476	LINC00973	-0.331654491	0.000217902	Downregulated	lincRNA
ENSG00000231290	APCDD1L-AS1	-0.317924898	0.015194891	Downregulated	Processed transcript
ENSG00000189223	PAX8-AS1	-0.202055204	0.007729818	Downregulated	Processed transcript
ENSG00000278730	AC005332.9	0.240168069	0.031444921	Upregulated	lincRNA
ENSG00000245694	CRNDE	0.284452354	0.026847637	Upregulated	lincRNA
ENSG00000170846	AC093323.1	0.289627345	0.025524733	Upregulated	lincRNA
ENSG00000227403	LINC01806	0.299446349	0.023036988	Upregulated	lincRNA
ENSG00000188242	PP7080	0.29964691	0.029485244	Upregulated	Antisense RNA
ENSG00000272734	ADIRF-AS1	0.316893303	0.017323522	Upregulated	Processed transcript
ENSG00000276107	AC037198.2	0.318098808	0.01662948	Upregulated	Sense intronic
ENSG00000269927	AC004817.3	0.325125311	0.043083902	Upregulated	lincRNA
ENSG00000203706	SERTAD4-AS1	0.329114239	0.011809133	Upregulated	Antisense RNA
ENSG00000263731	AC145207.5	0.332655519	0.006068884	Upregulated	lincRNA
ENSG00000260273	AL359711.2	0.334641289	0.040667384	Upregulated	Antisense RNA
ENSG00000203804	ADAMTSL4-AS1	0.341062215	0.012569507	Upregulated	Processed transcript
ENSG00000206417	H1FX-AS1	0.368254632	0.033119549	Upregulated	Antisense RNA
ENSG00000272667	AC012306.2	0.368736515	0.040227617	Upregulated	lincRNA
ENSG00000257698	AC084033.3	0.377478397	0.020136283	Upregulated	lincRNA
ENSG00000204860	FAM201A	0.379917863	0.031811505	Upregulated	Antisense RNA
ENSG00000196696	PDXDC2P-NPIPB14P	0.380748939	0.016490349	Upregulated	Processed transcript
ENSG00000272256	AC044849.2	0.428009661	0.007995402	Upregulated	Antisense RNA
ENSG00000281091	AL606763.1	0.444551507	0.007542919	Upregulated	lincRNA
ENSG00000257038	AP002761.3	0.482191428	0.002799499	Upregulated	Antisense RNA
ENSG00000258820	AF111167.1	0.547561335	0.000193314	Upregulated	Antisense RNA
ENSG00000259687	LINC01220	0.570885631	0.000182563	Upregulated	lincRNA
ENSG00000261455	LINC01003	0.612467281	5.13E-06	Upregulated	lincRNA
ENSG00000241544	LINC02029	0.767795798	8.22E-13	Upregulated	lincRNA

Supplementary Table 2. Guilt-by-Association analysis of LINC00944 in normal samples from TCGA-BRCA. LINC00944 co-expressed protein-coding genes (n=113, *Pearson* | *r* | ≥ 0.3) were taken to overlap against the Hallmark gene set Collection[24].

Systematic name	H: Hallmark Gene Set Name	Description	FDR	Genes
M5950	H: ALLOGRAFT REJECTION	Genes upregulated during transplant rejection	$1.59E^{-31}$	CCL5 CXCL9 IRF4 GZMA IL7 STAT4 LCK IL18RAP IL2RG MAP4K1 CD40LG PRKCB CD2 FYB1 PTPRC CD96 TLR6 CCR5 SIT1 CXCR3 IL12A ITGAL ITK KLRD1 TRAT1 CRTAM CCR2 ZAP70 CD3D CD3E CD3G CD247 CD8A CD8B CD79A
M5913	H: INTERFERON GAMMA RESPONSE	Genes upregulated in response to IFNG	$3.33E^{-07}$	CCL5 CXCL9 IRF4 GZMA IL7 STAT4 CDKN1A SOCS3 ADAR ITGB7 CIITA MARCHF1 SLAMF7 ZBP1
M5932	H: INFLAMMATORY RESPONSE	Genes defining inflammatory response	$1.75E^{-06}$	CCL5 CXCL9 LCK IL18RAP CDKN1A CD48 HBEGF IL7R GP1BA CXCR6 CCR7 KCNA3 SLAMF1
M5947	H: IL2 STAT5 SIGNALING	Genes upregulated by STAT5 in response to IL2 stimulation	$2.59E^{-04}$	IRF4 CD48 TRAF1 RABGAP1L GADD45B CCR4 POU2F1 CDC42SE2 EOMES CST7
M5953	H: KRAS SIGNALING UP	Genes upregulated by KRAS activation	$2.59E^{-04}$	IL2RG MAP4K1 HBEGF IL7R TRAF1 RABGAP1L BIRC3 DOCK2 IKZF1 PRDM1
M5890	H: TNFA SIGNALING VIA NFKB	Genes regulated by NF-kB in response to TNF	$2.59E^{-04}$	CCL5 CDKN1A SOCS3 HBEGF IL7R TRAF1 GADD45B BIRC3 IER3 JUNB
M5921	H: COMPLEMENT	Genes encoding components of the complement system, which is part of the innate immune system	$1.23E^{-03}$	CCL5 GZMA LCK CD40LG GP1BA CDK5R1 CR2 GZMK PIK3CG
M5897	H: IL6 JAK STAT3 SIGNALING	Genes upregulated by IL6 via STAT3, e.g., during acute phase response.	$1.05E^{-02}$	CXCL9 IL7 IL2RG SOCS3 ITGA4
M5923	H: PI3K AKT MTOR SIGNALING	Genes upregulated by activation of the PI3K/AKT/mTOR pathway	$2.11E^{-02}$	LCK IL2RG PRKCB CDKN1A CAMK4
M5902	H: APOPTOSIS	Genes mediating programmed cell death (apoptosis) by activation of caspases	$2.62E^{-02}$	CD2 CDKN1A GADD45B BIRC3 IER3 CYLD

Supplementary Table 3. Guilt-by-Association analysis of LINC00944 in tumor samples from TCGA-BRCA. LINC00944 co-expressed protein-coding genes (n=1045, *Pearson* | *r* | ≥ 0.3) were taken to overlap against the Hallmark gene set Collection[24].

Systematic name	H: Hallmark Gene Set Name	Description	FDR	Genes
M5950	H: ALLOGRAFT REJECTION	Genes upregulated during transplant rejection	$1.49E^{-102}$	IL6 CCL5 LCP2 ICAM1 CCL2 IL2RB CXCL9 IL15 CD40 GZMA TAP1 SOCS1 IRF8 CD86 IRF4 IL7 FAS STAT1 WARS1 B2M CD74 PSMB10 IFNAR2 STAT4 ST8SIA4 LCK LYN TLR2 IL18 IL12B IL10 NLRP3 TLR1 IL18RAP CTSS C2 CD40LG GZMB WAS TNF CD80 CCL4 IL2RA LTB CCND2 GPR65 IL2RG CXCL13 CCR1 IL12RB1 ETS1 ITGB2 MAP4K1 GBP2 FASLG PRF1 CD2 EGFR FYB1 PTPRC NCK1 PRKCB IFNG CFP CD1D HCLS1 CD96 IGSF6 TLR6 CCR5 FCGR2B FGR SIT1 CXCR3 IL2 IL12A IL16 ITGAL ITK KLRD1 NCF4 TRAT1 SRGN CRTAM CCL13 CCL19 SPI1 TAP2 CCR2 ZAP70 CD3D CD3E CD3G CD247 CD4 CD7 CD8A CD8B CD28 NCR1 LY86 IL27RA CD79A
M5913	H: INTERFERON GAMMA RESPONSE	Genes upregulated in response to IFNG	$1.2E^{-72}$	IL6 CCL5 LCP2 ICAM1 CCL2 IL2RB CXCL9 IL15 CD40 GZMA TAP1 SOCS1 IRF8 CD86 IRF4 IL7 FAS STAT1 WARS1 B2M CD74 PSMB10 IFNAR2 STAT4 ST8SIA4 IRF1 CXCL10 CXCL11 IL15RA CD69 IL10RA CMKLR1 NMI FPR1 TNFAIP3 PIM1 PLSCR1

				CFH CASP1 C1S LAP3 PSMB9 CASP4 SERPING1 C1R PLA2G4A SOD2 GBP4 FGL2 STAT2 CD38 CSF2RB IFI44L IFI30 IFI44 BATF2 SAMD9L EPSTI1 VCAM1 CD274 UPP1 PML IL18BP VAMP5 GBP6 FCGR1A KLRK1 LYSMD2 SAMHD1 GPR18 IDO1 ITGB7 CIITA MX2 XAF1 MARCHF1 BANK1 SLAMF7 XCL1 APOL6 ZBP1 NLRC5 P2RY14
M5932	H: INFLAMMATORY RESPONSE	Genes defining inflammatory response	$2.23E^{-46}$	IL6 CCL5 LCP2 ICAM1 CCL2 IL2RB CXCL9 IL15 CD40 LCK LYN TLR2 IL18 IL12B IL10 NLRP3 TLR1 IL18RAP IRF1 CXCL10 CXCL11 IL15RA CD69 IL10RA CMKLR1 NMI FPR1 PIK3R5 GP1BA TNFRSF9 IL7R CCL20 CCRL2 GPR183 PTGER4 TNFRSF1B IL18R1 SELL CD48 EBI3 C3AR1 LAMP3 CHST2 MEFV EMP3 HAS2 SEMA4D CXCR6 CCR7 CYBB ADGRE1 GPR132 KCNA3 LTA P2RX7 PTAFR PTGIR RGS1 CCL17 SLAMF1 C5AR1 MARCO CD70
M5921	H: COMPLEMENT	Genes encoding components of the complement system, which is part of the innate immune system	$8.23E^{-35}$	IL6 CCL5 LCP2 GZMA LCK LYN CTSS C2 CD40LG GZMB WAS IRF1 TNFAIP3 PIM1 PLSCR1 CFH CASP1 C1S LAP3 PSMB9 CASP4 SERPING1 C1R PLA2G4A PIK3R5 GP1BA PLEK FCER1G LGMN DPP4 FYN C3 C1QA CTSC SH2B3 CR1 CR2 CTSL DGKG F5

				FCN1 GNGT2 GZMK ITGAM HPCAL4 PIK3CG GNG2 APOBEC3G C1QC PLA2G7 CASP10 L3MBTL4 SPOCK2
M5890	H: TNFA SIGNALING VIA NFKB	Genes regulated by NF-kB in response to TNF	$1.95E^{-28}$	IL6 CCL5 ICAM1 CCL2 TAP1 TLR2 IL18 IL12B TNF CD80 CCL4 IRF1 CXCL10 CXCL11 IL15RA CD69 TNFAIP3 SOD2 TNFRSF9 IL7R CCL20 CCRL2 GPR183 PTGER4 PLEK BMP2 TRAF1 NFIL3 SLC2A3 CD83 BIRC3 GFPT2 CFLAR SGK1 PTX3 RNF19B DRAM1 TNFAIP8 DUSP2 SLC2A6 FUT4 NFKBIE IL23A BCL2A1 RELB NR4A3 MSC
M5947	H: IL2 STAT5 SIGNALING	Genes upregulated by STAT5 in response to IL2 stimulation	$1.4E^{-27}$	IL2RB SOCS1 IRF8 CD86 IRF4 IL10 IL2RA LTB CCND2 GPR65 CXCL10 IL10RA PIM1 PLSCR1 GBP4 FGL2 TNFRSF9 TNFRSF1B IL18R1 SELL CD48 BMP2 TRAF1 NFIL3 SLC2A3 CD83 IL1R2 TGM2 PRNP CCR4 CTLA4 GLIPR2 SYT11 ICOS ITGAE TLR7 PLAGL1 BATF3 CDC42SE2 TNFRSF4 EOMES LRRC8C CST7 ADAM19 TNFRSF8 CD79B
M5897	H: IL6 JAK STAT3 SIGNALING	Genes upregulated by IL6 via STAT3, e.g., during acute phase response.	$4.35E^{-25}$	IL6 CXCL9 SOCS1 IL7 FAS STAT1 TLR2 TNF IL2RA LTB IL2RG CXCL13 CCR1 IL12RB1 IRF1 CXCL10 CXCL11 IL15RA PIM1 STAT2 CD38 CSF2RB PIK3R5 TNFRSF1B IL18R1 EBI3 IL1R2 CSF2RA IL9R ITGA4 ACVRL1

M5953	H: KRAS SIGNALING UP	Genes upregulated by KRAS activation	$1.38E^{-23}$	IRF8 CTSS CCND2 IL2RG ETS1 ITGB2 MAP4K1 CXCL10 IL10RA CMKLR1 TNFAIP3 CFH IL7R CCL20 TNFRSF1B C3AR1 FCER1G BMP2 TRAF1 BIRC3 GFPT2 CSF2RA CXCR4 TMEM176B HSD11B1 MAFB DOCK2 GYPC TSPAN13 IKZF1 GPNMB EPHB2 LY96 ADAMDEC1 LCP1 CLEC4A TLR8 PRDM1 LAT2 LAPTM5 PDCD1LG2 CD37
M5911	H: INTERFERON ALPHA RESPONSE	Genes upregulated in response to alpha interferon proteins	$3.88E^{-21}$	IL15 TAP1 IL7 WARS1 B2M CD74 GBP2 IRF1 CXCL10 CXCL11 NMI PLSCR1 CASP1 C1S LAP3 PSMB9 GBP4 STAT2 IFI44L IFI30 IFI44 BATF2 SAMD9L EPSTI1 CCRL2 SELL LAMP3 NCOA7 TMEM140
M5902	H: APOPTOSIS	Genes mediating programmed cell death (apoptosis) by activation of caspases	$1.12E^{-09}$	IL6 TAP1 FAS IL18 TNF CCND2 FASLG PRF1 CD2 IRF1 CD69 CASP1 CASP4 SOD2 CD38 BMP2 BIRC3 CFLAR ANXA1 ERBB3 CYLD DPYD PLCB2

Supplementary Table 4. Pearson correlation between LINC00944 and tumor-infiltrating T lymphocytes in tumor samples from TCGA.

Dataset	Marker	LINC00944-marker Pearson *r*	Number of XY pairs	*p*-value
TCGA-TGCT	CD3D	**0.76**	137	<0.0001
	CD3E	**0.76**	137	<0.0001
	CD3G	**0.63**	137	<0.0001
	CD6	**0.77**	137	<0.0001
	SH2D1A	**0.69**	137	<0.0001
	TRAT1	**0.69**	137	<0.0001
TCGA-LUAD	CD3D	**0.49**	503	<0.0001
	CD3E	**0.50**	503	<0.0001
	CD3G	**0.46**	503	<0.0001
	CD6	**0.45**	503	<0.0001
	SH2D1A	**0.50**	502	<0.0001
	TRAT1	**0.45**	502	<0.0001
TCGA-COAD	CD3D	**0.35**	409	<0.0001
	CD3E	**0.36**	409	<0.0001
	CD3G	**0.32**	409	<0.0001
	CD6	**0.34**	409	<0.0001
	SH2D1A	**0.37**	406	<0.0001
	TRAT1	**0.31**	393	<0.0001

TGCT Testicular Germ Cell Tumors, LUAD: Lung adenocarcinoma, and COAD: Colon adenocarcinoma datasets from The Cancer Genome Atlas (TCGA).

14. References

1. Nishikura K. A-to-I editing of coding and non-coding RNAs by ADARs. *Nat Rev Mol Cell Biol*. 2016;17(2):83-96. doi:10.1038/nrm.2015.4
2. Hartner JC, Schmittwolf C, Kispert A, Müller AM, Higuchi M, Seeburg PH. Liver Disintegration in the Mouse Embryo Caused by Deficiency in the RNA-editing Enzyme ADAR1. *J Biol Chem*. 2004;279(6):4894-4902. doi:10.1074/jbc.M311347200
3. Yang Y, Zhou XX, Jin YF. ADAR-mediated RNA editing in non-coding RNA sequences. *Sci China Life Sci*. 2013;56(10):944-952. doi:10.1007/s11427-013-4546-5
4. Paz-Yaacov N, Bazak L, Buchumenski I, et al. Elevated RNA Editing Activity Is a Major Contributor to Transcriptomic Diversity in Tumors. *Cell Rep*. 2015;13(2):267-276. doi:10.1016/j.celrep.2015.08.080
5. Anantharaman A, Gholamalamdari O, Khan A, et al. RNA-editing enzymes ADAR1 and ADAR2 coordinately regulate the editing and expression of Ctn RNA. *FEBS Lett*. 2017;591(18):2890-2904. doi:10.1002/1873-3468.12795
6. Fumagalli D, Gacquer D, Rothé F, et al. Principles Governing A-to-I RNA Editing in the Breast Cancer Transcriptome. *Cell Rep*. 2015;13(2):277-289. doi:10.1016/j.celrep.2015.09.032
7. Sagredo EA, Blanco A, Sagredo AI, et al. ADAR1-mediated RNA-editing of 3′UTRs in breast cancer. *Biol Res*. 2018;51(1):1-10. doi:10.1186/s40659-018-0185-4
8. Kung C-P, Cottrell KA, Ryu S, et al. ADAR1 editing dependency in triple-negative breast cancer. *bioRxiv*. 2020:2020.01.31.928911. doi:10.1101/2020.01.31.928911
9. Wang IX, So E, Devlin JL, Zhao Y, Wu M, Cheung VG. ADAR Regulates RNA Editing, Transcript Stability, and Gene Expression. *Cell Rep*. 2013;5(3):849-860. doi:10.1016/j.celrep.2013.10.002
10. Wang Q, Hui H, Guo Z, et al. ADAR1 regulates ARHGAP26 gene expression through RNA editing by disrupting miR-30b-3p and miR-573 binding. *Rna*. 2013;19(11):1525-1536. doi:10.1261/rna.041533.113
11. Soudyab M, Iranpour M, Ghafouri-Fard S. The role of long non-coding RNAs in breast cancer. *Arch Iran Med*. 2016;19(7):508-517. doi:0161907/AIM.0011
12. Yan X, Hu Z, Feng Y, et al. Comprehensive Genomic Characterization of Long Non-coding RNAs across Human Cancers. *Cancer Cell*. 2015;28(4):529-540. doi:10.1016/j.ccell.2015.09.006
13. Qiu J jun, Yan J bin. Long non-coding RNA LINC01296 is a potential prognostic biomarker in patients with colorectal cancer. *Tumor Biol*. 2015;36(9):7175-7183. doi:10.1007/s13277-015-3448-5
14. Li J, Lian Y, Yan C, et al. Long non-coding RNA FOXP4-AS1 is an unfavourable prognostic factor and regulates proliferation and apoptosis in colorectal cancer. *Cell Prolif*. 2017;50(1):1-11. doi:10.1111/cpr.12312
15. Salameh A, Lee AK, Cardó-Vila M, et al. PRUNE2 is a human prostate cancer suppressor regulated by the intronic long noncoding RNA PCA3. *Proc Natl Acad Sci U S A*. 2015;112(27):8403-8408. doi:10.1073/pnas.1507882112
16. Deng SJ, Chen HY, Zeng Z, et al. Nutrient stress-dysregulated antisense lncRNA GLS-AS impairs GLS-mediated metabolism and represses pancreatic cancer progression.

Cancer Res. 2019;79(7):1398-1412. doi:10.1158/0008-5472.CAN-18-0419
17. Haas R, Ganem NS, Keshet A, Orlov A, Fishman A, Lamm AT. A-to-I RNA editing affects lncRNAs expression after heat shock. *Genes (Basel)*. 2018;9(12):1-13. doi:10.3390/genes9120627
18. Ma C, Liu H, Yi-Feng Chang I, et al. ADAR1 promotes robust hypoxia signaling via distinct regulation of multiple HIF-1α-inhibiting factors. *EMBO Rep*. 2019;20(5):1-16. doi:10.15252/embr.201847107
19. Crews LA, Jiang Q, Zipeto MA, et al. An RNA editing fingerprint of cancer stem cell reprogramming. *J Transl Med*. 2015;13(1):1-12. doi:10.1186/s12967-014-0370-3
20. Hendrickson D, Kelley DR, Tenen D, Bernstein B, Rinn JL. Widespread RNA binding by chromatin-associated proteins. *Genome Biol*. 2016;17(1):1-18. doi:10.1186/s13059-016-0878-3
21. Sagredo EA, Sagredo AI, Blanco A, et al. ADAR1 Transcriptome editing promotes breast cancer progression through the regulation of cell cycle and DNA damage response. *Biochim Biophys Acta - Mol Cell Res*. 2020:118716. doi:10.1016/j.bbamcr.2020.118716
22. Nagy Á, Lánczky A, Menyhárt O, Gyorffy B. Validation of miRNA prognostic power in hepatocellular carcinoma using expression data of independent datasets. *Sci Rep*. 2018;8(1):1-9. doi:10.1038/s41598-018-27521-y
23. Grossman RL, Heath AP, Ferretti V, et al. Toward a Shared Vision for Cancer Genomic Data. *N Engl J Med*. 2016;375(12):1109-1112. doi:10.1056/NEJMp1607591
24. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst*. 2015;1(6):417-425. doi:10.1016/j.cels.2015.12.004
25. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545-15550. doi:10.1073/pnas.0506580102
26. Danaher P, Warren S, Dennis L, et al. Gene expression markers of Tumor Infiltrating Leukocytes. *J Immunother Cancer*. 2017;5(1):1-15. doi:10.1186/s40425-017-0215-8
27. Levy E, Marty R, Gárate Calderón V, et al. Immune DNA signature of T-cell infiltration in breast tumor exomes. *Sci Rep*. 2016;6(July):1-10. doi:10.1038/srep30064
28. Anadón C, Guil S, Simó-Riudalbas L, et al. Gene amplification-associated overexpression of the RNA editing enzyme ADAR1 enhances human lung tumorigenesis. *Oncogene*. 2016;35(33):4407-4413. doi:10.1038/onc.2015.469
29. Luo Y, Xuan Z, Zhu X, Zhan P, Wang Z. Long non-coding RNAs RP5-821D11.7, APCDD1L-AS1 and RP11-277P12.9 were associated with the prognosis of lung squamous cell carcinoma. *Mol Med Rep*. 2018;17(5):7238-7248. doi:10.3892/mmr.2018.8770
30. He W, Qiao ZX, Ma B. Long noncoding RNA FAM201A mediates the metastasis of lung squamous cell cancer via regulating ABCE1 expression. *Eur Rev Med Pharmacol Sci*. 2019;23(23):10343-10353. doi:10.26355/eurrev_201912_19672
31. Liu AM, Zhu Y, Huang ZW, Lei L, Fu SZ, Chen Y. Long noncoding RNA FAM201A involves in radioresistance of non-small-cell lung cancer by enhancing EGFR expression via miR-370. *Eur Rev Med Pharmacol Sci*. 2019;23(13):5802-5814. doi:10.26355/eurrev_201907_18319
32. Chen M, Liu P, Chen Y, et al. Long Noncoding RNA FAM201A Mediates the Radiosensitivity of Esophageal Squamous Cell Cancer by Regulating ATM and mTOR

Expression via miR-101. *Front Genet*. 2018;9(December):1-12. doi:10.3389/fgene.2018.00611

33. Sakurai M, Shiromoto Y, Ota H, et al. ADAR1 controls apoptosis of stressed cells by inhibiting Staufen1-mediated mRNA decay. *Nat Struct Mol Biol*. 2017;24(6):534-543. doi:10.1038/nsmb.3403
34. Stellos K, Gatsiou A, Stamatelopoulos K, et al. Adenosine-to-inosine RNA editing controls cathepsin S expression in atherosclerosis by enabling HuR-mediated post-transcriptional regulation. *Nat Med*. 2016;22(10):1140-1150. doi:10.1038/nm.4172
35. Chen D, Sun Q, Cheng X, et al. Genome-wide analysis of long noncoding RNA (lncRNA) expression in colorectal cancer tissues from patients with liver metastasis. *Cancer Med*. 2016;5(7):1629-1639. doi:10.1002/cam4.738
36. Wang Z, Yang B, Zhang M, et al. lncRNA Epigenetic Landscape Analysis Identifies EPIC1 as an Oncogenic lncRNA that Interacts with MYC and Promotes Cell-Cycle Progression in Cancer. *Cancer Cell*. 2018;33(4):706-720.e9. doi:10.1016/j.ccell.2018.03.006
37. Li Y, Jiang T, Zhou W, et al. Pan-cancer characterization of immune-related lncRNAs identifies potential oncogenic biomarkers. *Nat Commun*. 2020;11(1). doi:10.1038/s41467-020-14802-2
38. Hanahan D, Weinberg RA. Hallmarks of cancer: The next generation. *Cell*. 2011;144(5):646-674. doi:10.1016/j.cell.2011.02.013
39. Gonzalez H, Hagerling C, Werb Z. Roles of the immune system in cancer: From tumor initiation to metastatic progression. *Genes Dev*. 2018;32(19-20):1267-1284. doi:10.1101/GAD.314617.118
40. Pagès F, Kirilovsky A, Mlecnik B, et al. In situ cytotoxic and memory T cells predict outcome in patients with early-stage colorectal cancer. *J Clin Oncol*. 2009;27(35):5944-5951. doi:10.1200/JCO.2008.19.6147
41. Al-Shibli KI, Donnem T, Al-Saad S, Persson M, Bremnes RM, Busund LT. Prognostic effect of epithelial and stromal lymphocyte infiltration in non-small cell lung cancer. *Clin Cancer Res*. 2008;14(16):5220-5227. doi:10.1158/1078-0432.CCR-08-0133
42. Mahmoud SMA, Paish EC, Powe DG, et al. Tumor-infiltrating CD8+ lymphocytes predict clinical outcome in breast cancer. *J Clin Oncol*. 2011;29(15):1949-1955. doi:10.1200/JCO.2010.30.5037
43. Igney FH, Krammer PH. Immune escape of tumors: apoptosis resistance and tumor counterattack. *J Leukoc Biol*. 2002;71(6):907-920. doi:10.1189/jlb.71.6.907

CHAPTER 3: ADAR1 and lncRNAs A-to-I editing in breast cancer

This chapter encompasses the results obtained from evaluating the specific objective 3 "*Characterize lncRNAs A-to-I editing induced by ADAR1 in triple-negative breast cancer and the potential effects over a malignant phenotype*".

1. ADAR1 induces A-to-I editing in a small fraction of expressed lncRNAs in triple-negative breast cancer tumors

Previous studies have shown that ADAR1 is upregulated in all breast cancer subtypes, including TNBC (Anantharaman et al., 2017; Kung et al., 2020; Sagredo et al., 2020). In particular, we have seen that basal-like patients overexpressing ADAR1 have poor overall survival in time (Sagredo et al., 2018). Important efforts have been made in order to identify A-to-I editing in mRNAs and microRNAs induced in patients overexpressing ADAR1, and how this is related to their poor prognosis. In this context, a significant class of RNAs, the lncRNAs, has not been thoroughly examined.

Several A-to-I repositories harbor edited positions in lncRNAs; however, a few publications had characterized the connection with cancer pathogenesis.

In order to expand our understanding of this subject, we evaluated lncRNAs A-to-I editing in TNBC tumor samples from the TCGA-BRCA cohort. In Figure 1, we show the bioinformatic pipeline used for detecting edited positions. Briefly, two files per patient were

used as input for running the REDItoolDenovo.py Python script (Picardi & Pesole, 2013), the BAM file (Binary Alignment Map), containing the aligned reads information, and the reference genome (GRCh38). We performed the pipeline on 105 TNBC samples (n = 105) in which strand-specific RNA-seq was performed (**Fig. 1A).**

Critical parameters were set in the *Running script for variant detection* step, as a minimum of reads coverage of 10, and a minimum quality and mapping scores of 25 per read (**Fig. 1B**). Next, output tables were filtered for selecting only significant variants representing A-to-I editing: A to G (sense strand) and T to C (antisense strand) ($p \leq 0.05$, *Fisher Exact test*) (**Fig. 1C**).

Finally, variants were annotated with the official gene symbol and gene name, while single nucleotide polymorphisms (SNPs) were eliminated for avoiding confusion between DNA and RNA variants (**Fig. 1D**). More details about the pipeline can be found in the *Methods section*.

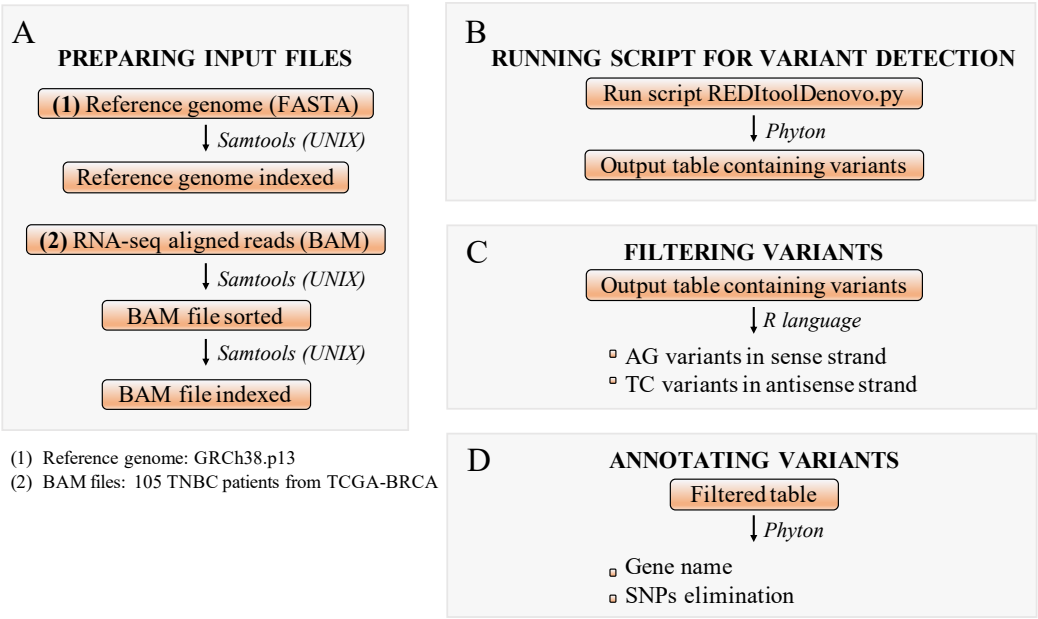


Figure 1. Bioinformatic pipeline for variant detection. (A) Preparing input files. The human reference genome GRCh38.p13 was used in FASTA format and indexed using Samtools. BAM files containing the strand-specific RNA-seq aligned reads were sorted and indexed using Samtools. BAM files were obtained from TNBC patients from the TCGA-BRCA cohort (n = 105). **(B)** Running script for variant detection. The input files were taken for running the variant detection using Python scrips for REDItools. **(C)** Filtering variants. Significant A-to-G (sense strand) and T-to-C (antisense strand) were filtered and considered as edited positions. $p \leq 0.05$, *Fisher Exact test* **(D)** Annotating variants. Edited positions were annotated with gene symbols and gene names. SPNs were annotated using dbSNP v.138. SNPs: Single nucleotide polymorphisms.

By using this bioinformatic pipeline, we detected 2,298 edited lncRNAs, representing 15.2% of the 15,124 expressed lncRNAs in all the 105 TNBC tumors, and a 12.7% of a total of 17,960 lncRNAs annotated in GENCODE (GRCh38, release 34) (**Fig. 2A**). This data suggests that ADAR1 edits a small and specific group of lncRNAs.

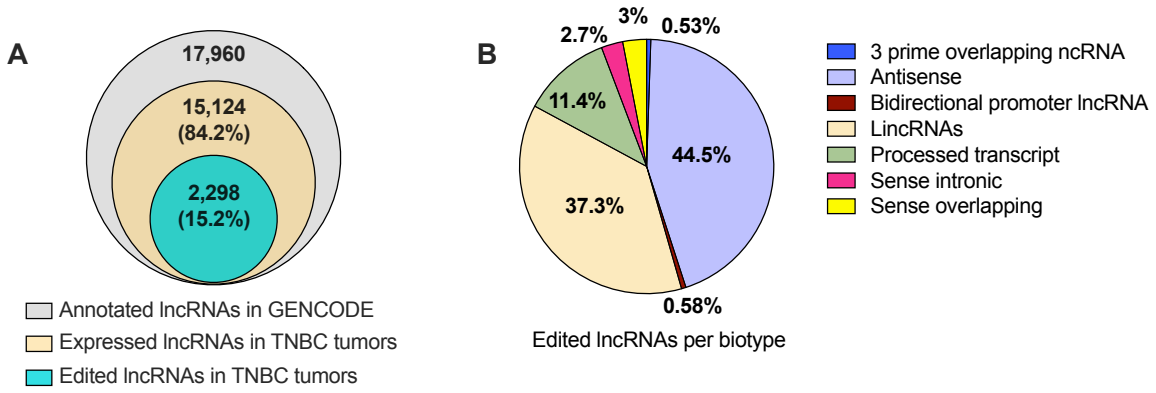


Figure 2. ADAR1 edits a small fraction of expressed lncRNAs. (A) Number of edited lncRNAs found in the 105 analyzed TNBC tumors. **(B)** Classification of edited lncRNAs according to GENCODE biotypes.

According to GENCODE biotypes, the majority of edited lncRNAs corresponded to antisense lncRNAs (44.5%), followed by lincRNAs (37.3%), and processed transcripts (11.4%) (**Fig. 2B**). A minor fraction of edited lncRNAs corresponded to sense overlapping (3%) and sense intronic (2.7%) lncRNAs (**Fig. 2B**). Bidirectional promoter lncRNAs and 3 prime overlapping lncRNAs represented a very limited fraction, -of 0.58% and 0.53%, respectively (**Fig. 2B**).

Next, the genomic location of edited lncRNAs was analyzed, finding that edited lncRNAs have a broad distribution across the genome. In general, about 10% of expressed lncRNAs presented at least one edited position (**Fig. 3A**). These 2,298 edited lncRNAs harbor a total of 31,690 edited positions considering all the 105 patients. These edited positions are edited in a high proportion, as the majority presented high editing levels ranging from 0.4 to 0.6, meaning that on average the 50% (40% - 60%) of the expressed transcripts were edited (**Fig. 3B**).

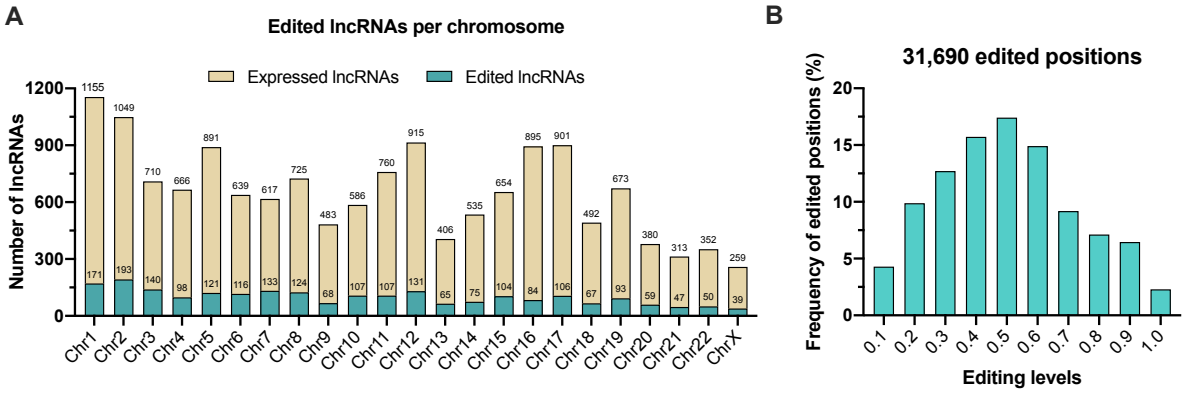


Figure 3. Edited lncRNAs across the genome and frequency of editing levels. (A) Number of edited lncRNAs in each chromosome and the number of expressed lncRNAs per chromosome. **(B)** Frequency distribution of editing levels for the 31,690 uniquely edited positions found. (Editing level = reads supporting the editing [I]/total reads covering the position [A+I]).

2. Edited lncRNAs in TNBC tumors

Aiming at understanding whether A-to-I editing in lncRNAs plays a role in malignant phenotypes and/or poor outcomes in TNBC, edited lncRNAs were grouped according to two criteria. First, *most edited lncRNAs* by considering the number of different edited positions per lncRNA. Secondly, *most frequently edited lncRNAs* by considering the number of patients displaying editing of a particular lncRNA regardless of the edited position.

Most edited lncRNAs were ranked, and the top 30 are depicted in Figure 4A. Despite many lncRNAs exhibited a high number of editing events per gene, PVT1, which has been described as a key mediator of oncogenic progression (for review, see Derderian et al., 2019), appeared to be the most edited, comprising 162 edited positions, following by AC092957.1 (128), FTX (93), CASC15 (87) and LIMD1-AS1 (71) (**Fig. 4A**).

Most frequently edited lncRNAs were ranked, and the top 30 are depicted in Figure 4B. The lncRNAs that showed edited positions in a higher number of patients were LIMD1-AS1 (80 patients), followed by SNAP25.AS1 (78), FTX (76), AC006504.5 (75) and BX322234.1 (71) (**Fig. 4B**).

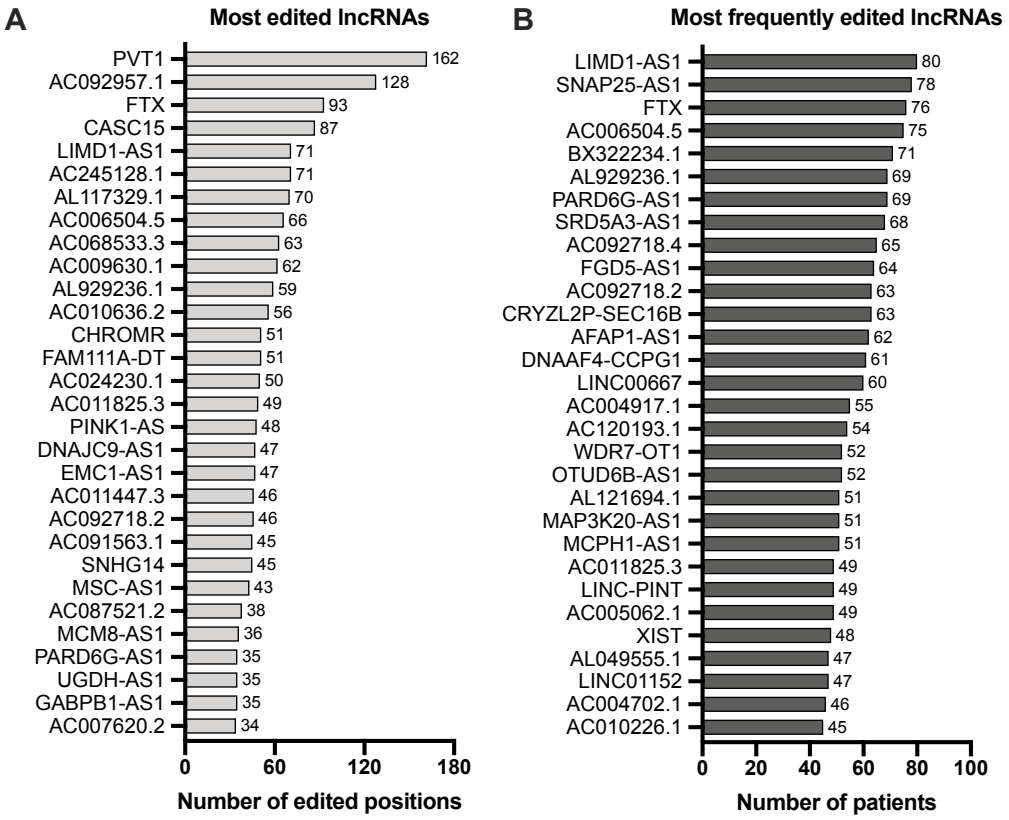


Figure 4. Edited lncRNAs. (A) Most edited lncRNAs. Number of edited positions per lncRNA. **(B)** Most frequently edited lncRNAs. Number of patients harboring edited positions in the same lncRNA.

3. A-to-I editing could modify the sponge function of lncRNAs

In order to understand how editing could be participating in cancer pathogenesis, we considered the function of lncRNAs. The sponge function of lncRNAs, in other words, their capability of binding miRNAs and prevent them from exerting their respective function (**Fig. XI**) (Jarroux et al., 2017), is one of the most studied functions of lncRNAs. In this context, the RNA sequence is crucial, and changes as those introduced by A-to-I editing may modify the interaction with their miRNA targets.

To explore the above-exposed idea, we reviewed the published data on the most edited lncRNA, PVT1 (**Fig. 5A**), finding that it promotes tumorigenesis in several cancer types by acting as a sponge. At least 20 different miRNAs with key roles in the cell cycle, apoptosis, metabolism, autophagy, cell adherens, matrix remodeling, and angiogenesis can be targeted by PVT1 (for review, see W. Wang et al., 2019). In the studied TNBC tumor samples, PVT1 lncRNA was not differentially expressed when comparing patients having high versus low levels of ADAR1 ($p = 0.96$, *Mann-Whitney test*) (**Fig. 5B**), indicating that editing induced by high levels of ADAR1 does not affect PVT1 expression. On the other hand, PVT1 RNA was significantly upregulated in TNBC tumor samples ($p < 0.0001$, *Mann-Whitney test*) (**Fig. 5C**), suggesting a possible role in the pathology. Taken together these observations, we proposed that A-to-I editing does not affect PVT1 expression levels but could be modifying its sponge function by altering its interaction with different miRNAs in breast cancer.

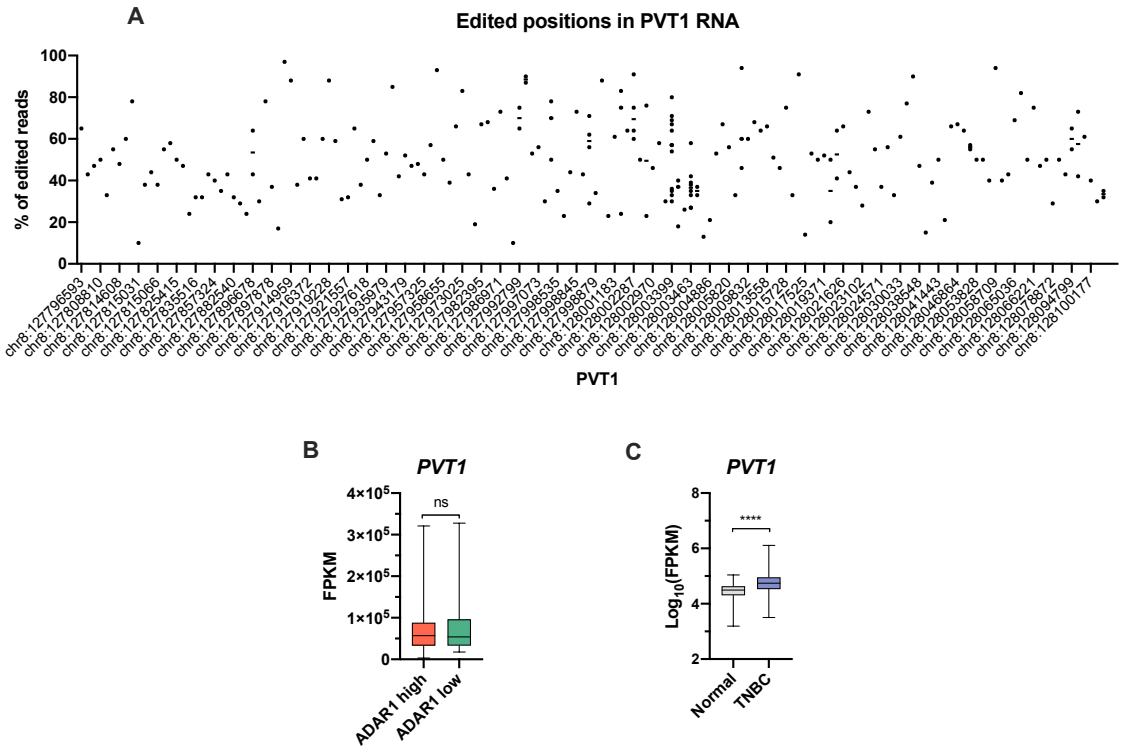


Figure 5. Editing in PVT1 lncRNA could modify its sponge function. (A) Percentage of edited reads in each edited position detected in PVT1 lncRNA. Only the first 54 out of 162 edited positions are depicted. **(B)** PVT1 expression levels in patients expressing high (red) and low (green) ADAR1 levels. **(C)** PVT1 expression levels in normal breast samples (grey) and TNBC samples (blue) from the TCGA-BRCA cohort. *Mann-Whitney test* was used for group comparison. **** $p < 0.0001$. A p-value ≤ 0.05 was considered as statistically significant. 'ns' indicates no significant difference between compared groups.

4. A-to-I editing could alter sense/antisense interaction

An increasing number of studies report that one of the main functions for antisense lncRNAs is to modulate the expression levels of their proximal genes, including their sense protein-coding genes (for review, see Villegas & Zaphiropoulos, 2015). In this context, the sequence complementarity and the base-pairing are fundamental for sense/antisense interaction and the subsequent regulation.

The antisense lncRNA PINK1-AS was one of the most edited lncRNAs (**Fig. 4A**), having both a high number of edited positions (48 uniquely edited positions) and a high proportion of edited transcripts (**Fig. 6A**). Since it was previously reported that PINK1-AS could stabilize its sense protein-coding RNA PINK1 not only in physiological conditions (Scheele et al., 2007) but in breast cancer (Gene et al., 2020), we hypothesized that A-to-I editing could be altering the interaction of PINK1-AS with PINK1 and therefore downregulate its expression. To test this idea, we split TNBC patients by ADAR1 expression and assessed PINK1-AS and PINK1 RNA levels. We found that patients expressing high levels of ADAR1 showed a reduction in PINK1 expression, supporting the idea that the high number of edited positions observed in PINK1-AS alters the interaction with PINK1, and therefore a downregulation in its expression follows ($p = 0.009$, *Mann-Whitney test*) (**Fig. 6C**). On the other hand, PINK1-AS expression was not affected by ADAR1 levels, indicating that the observed downregulation of PINK1 in ADAR1 high patients is independent of PINK1-AS levels ($p = 0.06$, *Mann-Whitney test*) (**Fig. 6B**). Moreover, we found that both PINK1-AS and PINK1 expression levels were significantly downregulated in TNBC tumor samples, indicating that their downregulation may

be necessary for allowing tumor malignancy ($p < 0.0001$, *Mann-Whitney test*) (**Fig. 6D and 6E,** respectively).

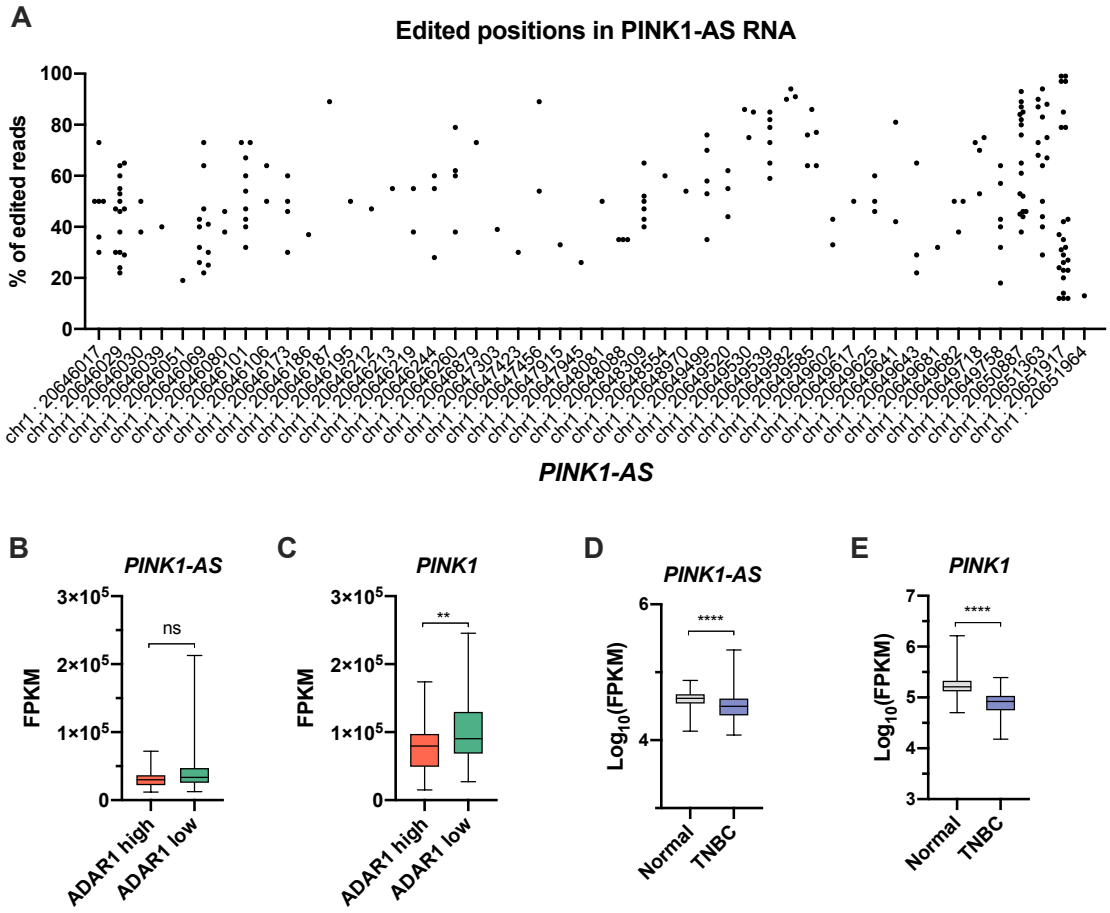


Figure 6. Editing in PINK1-AS could disrupt PINK/PINK1-AS interaction. (A) Percentage of edited reads in each edited position detected in PINK1-AS lncRNA. **(B)** PINK1-AS and **(C)** PINK1 expression levels in patients expressing high (red) and low (green) ADAR1 levels. **(D)** PINK1-AS and **(E)** PINK1 expression levels in normal breast samples (grey) and TNBC samples (blue) from the TCGA-BRCA cohort. *Mann-Whitney test* was used for group comparison. **** $p < 0.0001$. ** $p = 0.009$. A p-value ≤ 0.05 was considered as statistically significant. 'ns' indicates no significant difference between compared groups.

GENERAL DISCUSSION

The meaning and relevance of the main findings of this thesis project are discussed in the following section in two separate chapters. Each of them focuses on the concluding remarks and projections of the two scientific questions addressed in this study.

Discussion: ADAR1 and lncRNAs expression levels in breast cancer

- **ADAR1 overexpression induces changes in lncRNAs expression levels in the breast cancer cell line MDA-MB-231**

It has been extensively described that ADAR1 can modulate the expression levels of several mRNAs, microRNAs, and viral RNAs (Nishikura, 2016). Despite the sharp rise in lncRNAs studies, only few have addressed ADAR1-lncRNAs interactions and how can ADAR1 alter lncRNAs expression levels in cancer (Cao et al., 2015; S. J. Deng et al., 2019; Figueroa et al., 2016; Ma et al., 2019; Salameh et al., 2015; Wei Wang et al., 2019). This thesis provided a transcriptomic analysis of the effect of ADAR1 overexpression over lncRNAs expression in the breast cancer cell line MDA-MB-231. We found that 41 lncRNAs were differentially expressed after ADAR1 upregulation (p-adj ≤ 0.05); and we could further confirm that the ADAR1 loss-of-function reverts the effect on expression in some of those in the breast cancer cell lines MDA-MB-231 and MDA-MB-436, supporting the idea that ADAR1 can exert a regulatory role in lncRNAs expression levels. Eighteen thousand fifty-one lncRNA genes are annotated in the

human reference genome used in this analysis, and the breast cancer cell line MDA-MB-231 express 7,658. According to this, the 41 DE lncRNAs represent only 0.5%, which indicates that the effect induced by ADAR1 upregulation may be target-specific.

The ADAR1 upregulation observed in several cancer types, including breast cancer, has been largely described as an oncogenic feature (Anadón et al., 2016); thus, we hypothesized that the changes induced in lncRNAs expression could have a role in cancer. A quick exploration of differentially expressed lncRNAs in PubMed (https://www.ncbi.nlm.nih.gov/pubmed) showed that several lncRNAs were already linked to cancer development. For instance, APCDD1L-AS1 was found to have an important prognostic value in lung squamous cell carcinoma (Luo et al., 2018). Functional enrichment analysis of APCDD1L-AS1 co-expressed protein-coding genes revealed the enrichment of critical pathways in cancer development as 'positive regulation of cell migration' and 'proteinaceous extracellular matrix' (Luo et al., 2018). In our analysis, we showed that ADAR1 gain- and loss-of-function were capable of modulating the expression of APCDD1L-AS1, indicating that this lncRNA may also have a role in the breast cancer malignancy related to ADAR1.

Another good example is the upregulation of the lncRNA FAM201A, which has been found mediating metastasis of lung squamous cell cancer (He et al., 2019) and resistance to radiotherapy in non-small cell lung cancer (A. M. Liu et al., 2019) and esophageal squamous cell cancer (M. Chen et al., 2018). In the ADAR1 OE condition, we found a significant upregulation of FAM201A, suggesting that this lncRNA may also have a role in ADAR1-mediated breast cancer progression. Lastly, most of differentially expressed lncRNAs have no

associated literature, so we postulate them as principal candidates for exploring their function in cancer.

- **LINC00944 expression levels are responsive to *ADAR1* gain- and loss-of-function**

In the present thesis, we showed that the expression levels of the long intergenic non-coding RNA LINC00944 changed in response to ADAR1 up- and downregulation in two breast cancer cell lines. In addition, the RNA-immunoprecipitation (RIP) experiment shows that ADAR1 and LINC00944 interact in breast cancer cells. Nevertheless, in the RIP assay, we employed formaldehyde as a cross-linker agent; thus, we cannot conclude if the interaction between ADAR1 and LINC00944 is direct or not (Panhale et al., n.d.).

Mechanisms by which ADAR1 can disrupt RNA expression levels through its canonical function of A-to-I editing range from the alteration in splicing processing to stability impairment and nuclear retention in paraspeckles, among others (for review, see Nishikura et al., 2016). In LINC00944, we found no evidence of A-to-I editing, excluding it as an underlying mechanism in breast cancer.

On the other hand, it has been shown that ADAR1 can modify RNA expression levels through non-canonical functions by complexing with other RNA-binding proteins such as Dicer (S. Deng et al., 2018). For instance, upon the heterodimerization of ADAR1 and Dicer, microRNA biogenesis is improved, and thus upregulation of mature microRNAs expression follows (Ota et al., 2013). Importantly, this processing is believed to be essential for life, as it

constitutes the main reason for the death of *ADAR1*$^{-/-}$ null mice in embryonic stages (Ota et al., 2013). Regarding lncRNAs, the interaction with the ADAR1/Dicer complex has been described only in one publication: in pancreatic cancer, the dsRNA formed by the Glutaminase (GLS) mRNA and the lncRNA glutaminase antisense (GLS-AS) is degraded by the ADAR1/Dicer-dependent RNA interference mechanism and comprises an essential factor in metabolism reprogramming of cancer cells (S. Deng et al., 2018).

The interplay between ADAR1 and other RNA binding proteins such as Staufen and HuR has also been described as central for RNA expression regulation. The Staufen1 (STAU-1)-mediated mRNA decay (SMD) is a cellular process in which mammalian cells degrades mRNA by the binding of STAU1 to the 3-UTR of targets transcripts (for review, see Park & Maquat, 2013). In stressed cells, it has been shown that ADAR1 can bind to the 3'UTR of many anti-apoptotic transcripts in detriment of the STAU-1 binding and SMD. In this way, ADAR1 increases the stability and, thus, the expression of those interacting transcripts and suppresses apoptosis (Sakurai et al., 2017). On the other hand, the *Human antigen R* protein (HuR) is an RNA-binding protein that initiates the recognition of the mRNA by the destabilization and degradation machinery (Myer et al., 1997). It has been shown that ADAR1 and HuR cooperate by binding to the same RNA substrates and thus controlling the stability and expression levels of a subset of transcripts (I. X. Wang et al., 2013).

Taken together, we believe that all the mechanisms above mentioned are worth exploring for a complete understanding of how ADAR1 can modulate LINC00944 expression levels in cancer.

- **The emerging role of LINC00944 in cancer**

LINC00944 function has not been elucidated in literature; however, few publications have outlined a possible role in cancer. For instance, LINC00944 (NR_033878) expression was found downregulated in colorectal cancer (CRC) tissues from patients with liver metastasis (vs. patients without metastasis), suggesting an involvement of this lncRNA in the metastatic process (D. Chen et al., 2016). In addition, LINC00944 was reported as an epigenetically activated lncRNA in several cancer types when comparing to their normal tissues (Z. Wang et al., 2018). Interestingly, LINC00944 was recently identified as an immune-related lncRNA in cancer, and associated to antimicrobials, cytokines, interleukins, antigen processing and presentation, natural killer cell cytotoxicity, TCR signaling, cytokine, chemokine, and interleukins receptors pathways in several cancer types, including breast cancer (Li et al., 2020). This data agrees with our findings, in which we reported that LINC00944 has a strong relationship with and may play a role in immune signaling pathways. Nonetheless, the molecular mechanism underlying LINC00944 participation should be further investigated.

- **LINC00944 is related to immune signaling pathways and apoptosis**

By using the Guilt-by-Association approach, we observed that LINC00944 is related to immune signaling pathways in normal and tumoral tissue in different cancer types. Persistent occurrence of immune pathways related to IL-2, IFN-γ, and TNF-α, which are secreted by Th1 cells (for review, see Edechi et al., 2019), prompted us to evaluate the presence of this cells in the tumor microenvironment. By using publicly available data generated from RNA-sequencing

and Exome sequencing, we observed that LINC00944 positively correlates to infiltrating T lymphocytes (TILs) in the tumor microenvironment, postulating LINC00944 expression levels as a useful marker for infiltration of Th1 cells in tumors. These findings become relevant since the infiltration of immune cells into the tumor microenvironment is a major factor in cancer progression (Gonzalez et al., 2018), and because TILs have been related to favorable clinical outcomes in several cancer types, including colorectal cancer (Pagès et al., 2009), non-small cell lung cancer (Al-Shibli et al., 2008) and breast cancer (Mahmoud et al., 2011).

The execution of an effective immune response in the tumor microenvironment could lead to increased programmed cell death in cancer cells (Igney & Krammer, 2002); thus, we further explored a correlation between LINC00944 expression and apoptosis. In the present study, we showed that patients expressing low levels of LINC00944 had downregulation of pro-apoptotic markers, as the gene expression of BAX and BAK1, and the caspases CASP8 and CASP3 are significantly downregulated in breast cancer. Accordingly, the expression of the antiapoptotic BCL2 and BCL2L1 is upregulated. These data suggest that cells expressing low levels of LINC00944 have a decrease in the apoptotic process, yet functional assays are needed.

LINC00944 expression between normal and tumor samples from breast cancer was not statistically different, yet we found that LINC00944 was statistically upregulated in triple-negative breast cancer (TNBC) patients. In the same line, the LINC00944 expression and receptor status assessment showed that LINC00944 expression was positively correlated to estrogen and progesterone receptor, all relevant medical concerns regarding therapies.

- **LINC00944 expression correlated to clinicopathological parameters and survival outcomes**

Regarding clinicopathological parameters, we found that LINC00944 expression is correlated to the age at diagnosis and tumor size. Interestingly, samples coming from larger tumors had low expression levels of LINC00944. Taken the results mentioned above, this may imply that those tumors have a higher fraction of malignant cells compared to tumor-infiltrating lymphocytes and that the apoptotic program may be decreased. Kaplan-Meier survival analysis supported this idea, as Overall survival and Relapse-free survival decreased significantly in the LINC00944 low expression cohort.

Taken together, we proposed a model for summarizing our findings in Figure 1 (**Fig. 1**).

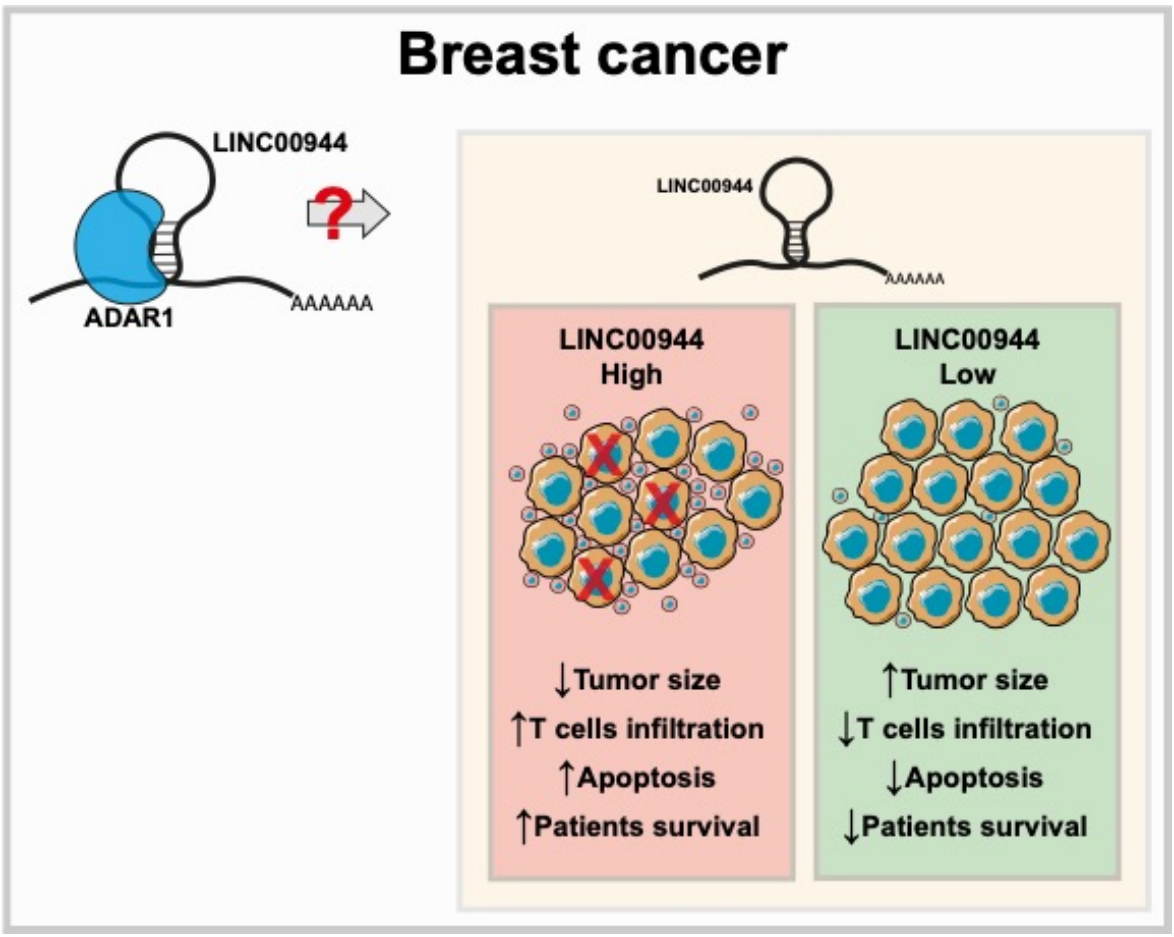


Figure 1. Low expression of the lncRNA LINC00944 indicates poor outcomes in breast cancer. Changes in ADAR1 levels induce differential expression of the lncRNA LINC00944 by an editing-independent mechanism in breast cancer cell lines. In breast cancer patients from The Cancer Genome Atlas (TCGA), LINC00944 expression levels correlated to tumor size, lymphocytic T infiltration in the tumor microenvironment, expression of apoptotic markers, and patient survival.

Discussion: ADAR1 and lncRNAs A-to-I editing in breast cancer

Increasing evidence has proposed ADAR1 as a therapeutic target for TNBC patients. Recently, ADAR1 has been shown to be required for TNBC tumor proliferation, transformation, and tumorigenesis (Kung et al., 2020). Moreover, high expression of ADAR1 has been correlated to poor prognosis in breast cancer (Kung et al., 2020) and in basal-like breast cancer patients (Sagredo et al., 2018), which is highly comparable to TNBC. The A-to-I editing capabilities of ADAR1 has been widely proposed as the underlying mechanism of ADAR1-associated tumorigenesis and tumor progression. In fact, in the literature, multiple examples of editing events can be found in protein-coding genes and miRNAs (for review, see C. Wang et al., 2017; Xu et al., 2018). Nevertheless, to our knowledge, there is no report focused on the A-to-I editing of lncRNAs in TNBC patients.

Here, we report for the first time an RNA A-to-I editing analysis of lncRNAs based on TNBC patients from TCGA data, whose results are discussed below.

- **Detecting A-to-I variants in lncRNAs**

Although several pipelines have been designed for identifying A-to-I editing sites, a consensus approach for detecting single-nucleotide variants in RNA does not exist. Most key parameters, such as genome annotations, read filters, quality filters, cut-off values, among others, will vary mainly depending on RNA-seq depth, and coverage, and RNA expression levels, generating extensive debate in the field. Moreover, the intrinsic complexity of

bioinformatic analysis of high-throughput sequencing data, and the high computational requirement for this type of analysis, increase the difficulties for a non-bioinformatic investigator.

In this thesis, we used a user-friendly pipeline based on a suite of python scripts to investigate RNA A-to-I editing at a transcriptomic scale using massive sequencing data, called REDItools (Picardi & Pesole, 2013). REDItools are freely available at GitHub repository https://github.com/BioinfoUNIBA/REDItools/blob/master/README_1.md and released under the Massachusetts Institute of Technology (MIT) license. In particular, we used the *REDItoolDenovo.py* script, which was designed to detect RNA editing sites using RNA-seq data. The script detects potential RNA editing sites by comparing the aligned reads of RNA-seq data in the standard BAM format with the reference genome. The script then employs the Fisher's exact test to detect statistically significant edited positions by comparing the number of observed bases with the number of expected bases per potential editing site. Moreover, the pipeline allowed us to set up a number of consensus parameters used in editing detection to perform an accurate analysis, including a per-base coverage ≥ 10, a quality score > 25, and mapping quality > 25 per editing site. Finally, REDItools provides accessory scripts for facilitating gene annotation, variant filter, and SNPs elimination for distinguishing genomic variants from the genuine editing sites.

- **Edited lncRNAs in TNBC patients**

We performed the RNA editing analysis on 105 TNBC tumors, detecting a total of 31,690 uniquely edited positions residing in 2,298 lncRNAs. These lncRNAs comprise a small fraction compared to the 15,124-total expressed lncRNAs in those tumors; thus, we suggest that the effect of ADAR1 editing over lncRNA may be specific. Nevertheless, there are technical issues that must be considered. The RNA-seq coverage is key for detecting editing sites, and we set up a minimum of 10 reads for supporting any variant with confidence. If the coverage was below that cut-off, we could be missing authentic variants only due to inadequate coverage. Likewise, detecting variants in RNAs with low expression levels, as is expected for many lncRNAs (Derrien et al., 2012), and/or with low editing levels, would be difficult if the number of reads is insufficient. Although RNA-seq data obtained from TCGA is notable, the sequencing depth and the coverage per potentially edited position are not uniform across all the samples; therefore, a high variability can be found between patients.

The vast majority of edited lncRNAs corresponded to antisense and lincRNAs, which together cover the 81,8% of total edited lncRNAs. This observation was previously reported in Picardi et al., 2014, in which edited lncRNAs found in brain, lung, kidney, liver, heart, and muscle healthy tissues corresponded primarily to antisense RNAs and lincRNAs (Picardi et al., 2014). We could explain these results in part as circumstantial and not causative, due to the fact that most of the lncRNAs annotated in GENCODE are classified as antisense and lincRNAs (Derrien et al., 2012).

In addition, we found that lncRNA editing levels were high, as the mean editing levels for all the 31,690 uniquely edited positions were 0.5, meaning that in general, half of the total reads for a given position (A+I) were supporting the editing (I). Previous studies had shown that editing in lncRNAs is, in general, at low levels (Picardi et al., 2014), but those analyses were performed in normal tissue. This supports our observation that the increased expression of ADAR1 levels in breast cancer tumors increases its editing activity (Sagredo et al., 2018).

- **Potential effects of lncRNAs A-to-I editing in TNBC**

RNA editing has been described as an essential and extensive mechanism for generating RNA diversity in mRNAs and miRNAs (Kume et al., 2014; Pullirsch & Jantsch, 2010). Since RNA sequence is essential for many of the functions described for lncRNAs, we proposed that editing induced by ADAR1 may generate diversity in lncRNAs function. In this context, A-to-I editing events may both promote and inhibit lncRNA-RNA or lncRNA-protein interactions. Moreover, the introduction of inosines could stabilize or destabilize higher-order RNA structures that are indispensable for their function.

LncRNAs ability to bind miRNAs and prevent them from interacting with their target genes is known as *sponge* function (**Fig. IX**) (Jarroux et al., 2017). Here, we hypothesized that editing could allow the lncRNA to interact with new miRNA targets, as adenine base pairs with uracil while inosine with cytosine in RNAs (Alseth et al., 2014). As an example, we evaluated PVT1, whose sponge function has been widely described in tumorigenesis (Wenxi Wang et al., 2019). The high number of edited positions, the expression of PVT1 found in the analyzed

TNBC tumor samples, and the fact that ADAR1 expression did not alter PVT1 expression could be supporting the idea that the main effect of ADAR1 editing over PVT1 is altering its ability to interact with different miRNAs. As the ADAR1 overexpression observed in TNBC tumors has been widely described as an oncogenic feature in mRNAs and miRNAs (for review, see Nishikura et al., 2016), we postulate that the editing in PVT1 lncRNA diversifies its targets; thus, disturb the function of a wide range of miRNAs and allows tumor cells to better adapt. Nevertheless, we cannot overlook the possibility that editing could halt the ability of PVT1 for binding miRNAs.

Other examples of sponge lncRNAs having key functions in cancer and that we found edited are FTX (93 uniquely edited positions), FGD5-AS1 (27 edited positions), and LINC00963 (34 edited positions) (Fan et al., 2020; Yang et al., 2018; Zhang et al., 2019). Thus, we postulate them as major candidates for exploring how RNA editing induced by ADAR1 could alter the interaction with their known targets and consequently their function in TNBC.

Many lncRNAs can regulate the expression of neighboring genes based on the interaction with complementary sequences, forming intermolecular pairs of sense-antisense RNA-RNA, RNA-DNA, and RNA-protein duplexes (for review, see Villegas et al., 2015). Since was previously reported that the antisense lncRNA PINK1-AS could stabilize its sense coding RNA PINK1 (Schelee et al., 2007; Yaghoobi et al., 2020), we tested if RNA editing could modify the ability of PINK1-AS to bind and therefore modulate the expression levels of PINK1. We found a downregulation of PINK1 in patients expressing high levels of ADAR1 and supporting this idea. As the PINK1 function is not fully elucidated in breast cancer and both

anti-apoptotic and cell growth-suppressive effects have been found (Berthier et al., 2011), we cannot categorize the downregulation of PINK1 as tumor suppressor or oncogenic. However, the fact that PINK1-AS and PINK1 were downregulated in TNBC samples may be indicating that this component is necessary for allowing tumor malignancy.

Additional examples of sense/antisense regulation can be further tested based on our analysis: LIMD1-AS1 was the most frequently edited lncRNA, showing A-to-I editing in 80 out of 105 analyzed patients. It has been shown in non-small cell lung cancer that LIMD1-AS1 positively regulates the expression of its sense protein-coding LIMD1, which has a tumor-suppressive impact on cancer cells. LIMD1-AS1 can stabilize the LIMD1 mRNA by allowing the interaction with the mRNA stabilizer protein hnRNP U (Pan et al., 2020); thus, it would be interesting to test if A-to-I editing could alter the mentioned interactions.

On the other hand, A-to-I editing can potentially disrupt the formation of secondary and tertiary structures in lncRNAs, as inosines base pairs with cytosine and adenine with uracil (Alseth et al., 2014). Consequently, we propose that editing could also alter lncRNAs functions that not only depend on their nucleotide sequence but also their structures, as *decoy* and *guides* lncRNAs (**Fig. IX**). Decoy lncRNAs interact with proteins and restrict their binding to other targets (Jarroux et al., 2017). An example of a decoy function has been shown for the OIP5-AS1 lncRNA, which on top of its ability of bind miRNAs, can also bind the protein HuR and repress the HuR-elicited proliferative phenotypes (J. Kim et al., 2016). OIP5-AS1 showed 23 different edited positions in the analyzed TNBC tumors, making it a good candidate for elucidating if the lncRNA-protein interaction can be affected by lncRNA editing.

Finally, A-to-I editing also can potentially affect lncRNAs that act as precursors of small RNAs and their binding (**Fig. XI**) (Jarroux et al., 2017). The introduction of inosines in lncRNAs could alter secondary structures that give rise to small RNAs. Moreover, the presence of inosines in lncRNAs-derived small RNAs can alter the binding capability to their targets. FTX lncRNA has been described as a precursor lncRNA in cancer, as hosts microRNAs in its introns (Z. Liu et al., 2016). In our analysis, we found that FTX not only harbored a high number of uniquely edited positions (93 edited positions) but also was the third lncRNA most frequently edited in patients, showing A-to-I substitutions in 76 out of 105 patients. Considering this, we proposed that the precursor function of FTX is an exciting focus of study in TNBC tumors.

Taken together our data, we proposed a model where lncRNAs A-to-I editing induced by ADAR1 can diversify and/or alter their functions, which in turn may contribute to the establishment of malignant phenotypes required for cancer progression (**Fig. 2**).

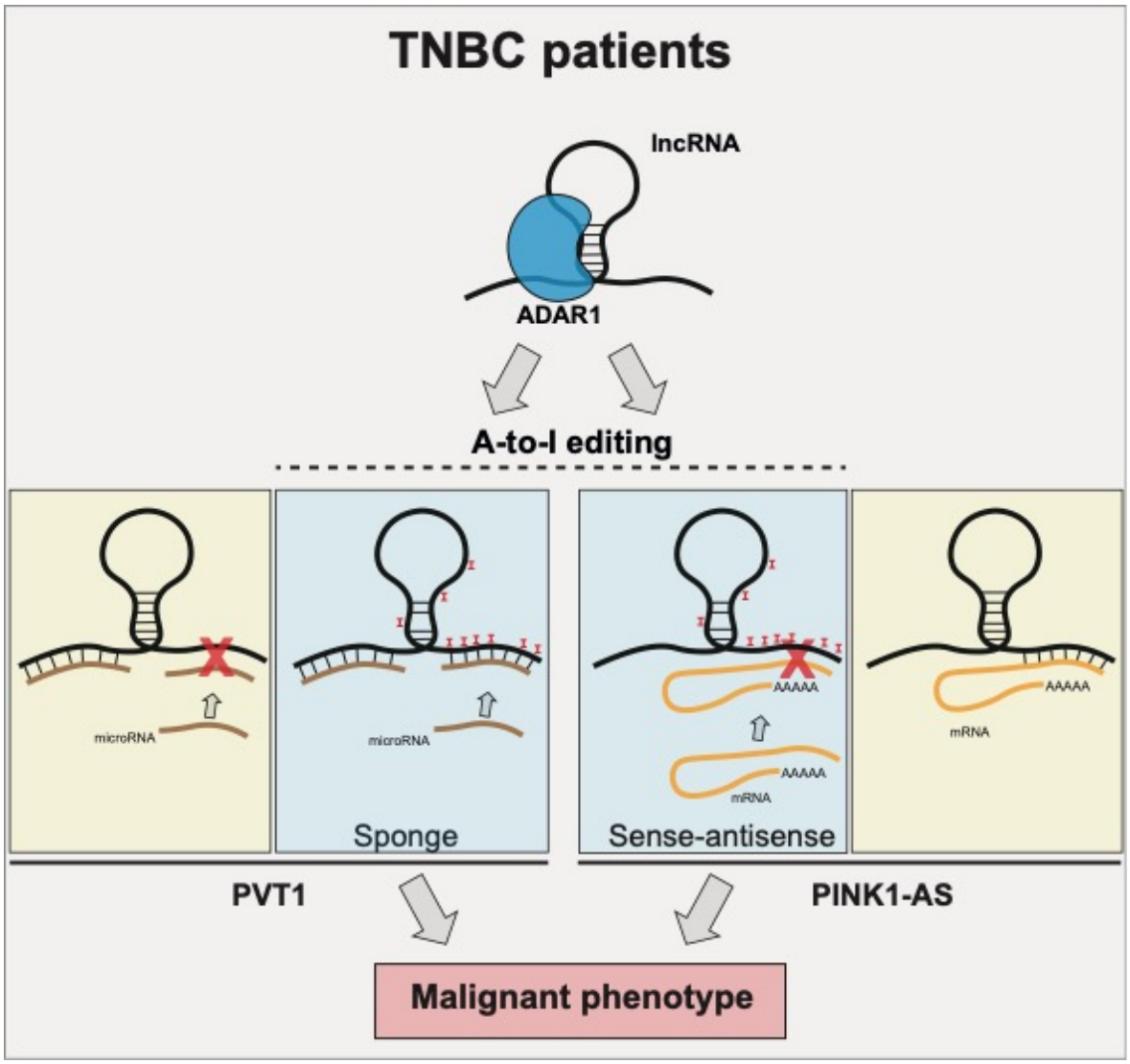


Figure 2. LncRNAs editing induced by ADAR1 in TNBC patients promotes malignant phenotypes. As example, A-to-I editing in PVT1 may broaden its sponge function by binding alternative microRNAs and impacting cellular phenotypes. On the other hand, editing of antisense lncRNAs such as PINK1-AS may disturb the interaction with their respective sense protein-coding gene, PINK1, and therefore alter its expression.

CONCLUSION

In this thesis, we demonstrated that ADAR1 expression induces changes both in lncRNAs A-to-I editing and expression levels in breast cancer. In our model, A-to-I editing and changes in lncRNAs expression were not related processes, since differentially expressed lncRNAs were not edited and vice versa.

The dysregulation of lncRNAs expression after ADAR1 up- and downregulation was related to both malignant phenotypes and poor prognosis. In particular, the overexpression of ADAR1 led to the downregulation of the long intergenic non-coding RNA LINC00944. Accordingly, low LINC00944 expression was correlated to poor outcomes in patients, as lower lymphocytic T infiltration, decrease in pro-apoptotic markers, and decreased overall survival and relapse-free survival.

ADAR1 upregulation in TNBC tumors has been correlated to poor prognosis. Here we demonstrated that ADAR1 induces A-to-I editing in about 10% of expressed lncRNAs in TNBC tumors. Moreover, these transcripts were found to be edited in a high proportion. As lncRNAs sequence is central for their function, we postulate that nearly all functions described for lncRNAs can be either diversified or disrupted by A-to-I RNA editing, which in turn contributes to the ADAR1-associated malignancy. In this context, we illustrated two examples of A-to-I edited lncRNAs and how this could alter their function: the sponge function of PVT1 in tumorigenesis and the ability of PINK1-AS1 to stabilize its sense mRNA PINK1.

REFERENCES

Akram, M., Iqbal, M., Daniyal, M., & Khan, A. U. (2017). Awareness and current knowledge of breast cancer. *Biological Research*, *50*(1), 1–23. https://doi.org/10.1186/s40659-017-0140-9

Al-Shibli, K. I., Donnem, T., Al-Saad, S., Persson, M., Bremnes, R. M., & Busund, L. T. (2008). Prognostic effect of epithelial and stromal lymphocyte infiltration in non-small cell lung cancer. *Clinical Cancer Research*, *14*(16), 5220–5227. https://doi.org/10.1158/1078-0432.CCR-08-0133

Alseth, I., Dalhus, B., & Bjørås, M. (2014). Inosine in DNA and RNA. *Current Opinion in Genetics and Development*, *26*, 116–123. https://doi.org/10.1016/j.gde.2014.07.008

Anadón, C., Guil, S., Simó-Riudalbas, L., Moutinho, C., Setien, F., Martínez-Cardús, A., Moran, S., Villanueva, A., Calaf, M., Vidal, A., Lazo, P. A., Zondervan, I., Savola, S., Kohno, T., Yokota, J., De Pouplana, L. R., & Esteller, M. (2016). Gene amplification-associated overexpression of the RNA editing enzyme ADAR1 enhances human lung tumorigenesis. *Oncogene*, *35*(33), 4407–4413. https://doi.org/10.1038/onc.2015.469

Anantharaman, A., Gholamalamdari, O., Khan, A., Yoon, J. H., Jantsch, M. F., Hartner, J. C., Gorospe, M., Prasanth, S. G., & Prasanth, K. V. (2017). RNA-editing enzymes ADAR1 and ADAR2 coordinately regulate the editing and expression of Ctn RNA. *FEBS Letters*, *591*(18), 2890–2904. https://doi.org/10.1002/1873-3468.12795

Badalamenti, G., Fanale, D., Incorvaia, L., Barraco, N., Listì, A., Maragliano, R., Vincenzi, B., Calò, V., Iovanna, J. L., Bazan, V., & Russo, A. (2019). Role of tumor-infiltrating lymphocytes in patients with solid tumors: Can a drop dig a stone? *Cellular Immunology*, *343*(December 2017). https://doi.org/10.1016/j.cellimm.2018.01.013

Berthier, A., Navarro, S., Jiménez-sáinz, J., Roglá, I., Ripoll, F., Cervera, J., & Pulido, R. (2011). PINK1 displays tissue-specific subcellular location and regulates apoptosis and cell growth in breast cancer cells ☆. *Human Pathology*, *42*(1), 75–87. https://doi.org/10.1016/j.humpath.2010.05.016

Borchert, G. M., Gilmore, B. L., Spengler, R. M., Xing, Y., Lanier, W., Bhattacharya, D., & Davidson, B. L. (2009). *Adenosine deamination in human transcripts generates novel microRNA binding sites*. *18*(24), 4801–4807. https://doi.org/10.1093/hmg/ddp443

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, *68*(6), 394–424. https://doi.org/10.3322/caac.21492

Cao, S., Moss, W., O'Grady, T., Concha, M., Strong, M. J., Wang, X., Yu, Y., Baddoo, M., Zhang, K., Fewell, C., Lin, Z., Dong, Y., & Flemington, E. K. (2015). New Noncoding Lytic Transcripts Derived from the Epstein-Barr Virus Latency Origin of Replication, oriP

, Are Hyperedited, Bind the Paraspeckle Protein, NONO/p54nrb, and Support Viral Lytic Transcription . *Journal of Virology*, *89*(14), 7120–7132. https://doi.org/10.1128/jvi.00608-15

Carlson, R. W., & Stockdale, F. E. (1988). The clinical biology of breast cancer. *Annual Review of Medicine*, *39*, 453–464. https://doi.org/10.1146/annurev.me.39.020188.002321

Chan, T. H. M., Qamra, A., Tan, K. T., Guo, J., Yang, H., Qi, L., Lin, J. S., Ng, V. H. E., Song, Y., Hong, H., Tay, S. T., Liu, Y., Lee, J., Rha, S. Y., Zhu, F., So, J. B. Y., Teh, B. T., Yeoh, K. G., Rozen, S., … Chen, L. (2016). ADAR-Mediated RNA Editing Predicts Progression and Prognosis of Gastric Cancer. *Gastroenterology*, *151*(4), 637-650.e10. https://doi.org/10.1053/j.gastro.2016.06.043

Chen, C., Cho, D. C., Wang, Q., Lai, F., Carter, K. C., & Nishikura, K. (2000). *A third member of the RNA-specific adenosine deaminase gene family , ADAR3 , contains both single- and double-stranded RNA binding domains*. 755–767.

Chen, D., Sun, Q., Cheng, X., Zhang, L., Song, W., Zhou, D., Lin, J., & Wang, W. (2016). Genome-wide analysis of long noncoding RNA (lncRNA) expression in colorectal cancer tissues from patients with liver metastasis. *Cancer Medicine*, *5*(7), 1629–1639. https://doi.org/10.1002/cam4.738

Chen, L., Li, Y., Lin, C. H., Chan, T. H. M., Chow, R. K. K., Song, Y., Liu, M., Yuan, Y. F., Fu, L., Kong, K. L., Qi, L., Li, Y., Zhang, N., Tong, A. H. Y., Kwong, D. L. W., Man, K., Lo, C. M., Lok, S., Tenen, D. G., & Guan, X. Y. (2013). Recoding RNA editing of AZIN1 predisposes to hepatocellular carcinoma. *Nature Medicine*, *19*(2), 209–216. https://doi.org/10.1038/nm.3043

Chen, M., Liu, P., Chen, Y., Chen, Z., Shen, M., Liu, X., Li, X., Li, A., Lin, Y., Yang, R., Ni, W., Zhou, X., Zhang, L., Tian, Y., Li, J., & Chen, J. (2018). Long Noncoding RNA FAM201A Mediates the Radiosensitivity of Esophageal Squamous Cell Cancer by Regulating ATM and mTOR Expression via miR-101. *Frontiers in Genetics*, *9*(December), 1–12. https://doi.org/10.3389/fgene.2018.00611

Chen, Y., Wang, H., Lin, W., & Shuai, P. (2017). ADAR1 overexpression is associated with cervical cancer progression and angiogenesis. *Diagnostic Pathology*, *12*(1), 1–8. https://doi.org/10.1186/s13000-017-0600-0

Connell, C. M., & Doherty, G. J. (2017). *Activating HER2 mutations as emerging targets in multiple solid cancers*. 1–11. https://doi.org/10.1136/esmoopen-2017-000279

Crews, L. A., Jiang, Q., Zipeto, M. A., Lazzari, E., Court, A. C., Ali, S., Barrett, C. L., Frazer, K. A., & Jamieson, C. H. M. (2015). An RNA editing fingerprint of cancer stem cell reprogramming. *Journal of Translational Medicine*, *13*(1), 1–12. https://doi.org/10.1186/s12967-014-0370-3

Daniel, A. R., Hagan, C. R., & Lange, C. A. (2012). *Progesterone receptor action : defining a role in breast cancer*. *6*(3), 359–369. https://doi.org/10.1586/eem.11.25.Progesterone

Definition of cancer - NCI Dictionary of Cancer Terms - National Cancer Institute. (n.d.). Retrieved October 1, 2020, from https://www.cancer.gov/publications/dictionaries/cancer-

terms/def/cancer

Deng, S., Chen, H., Zeng, Z., Deng, S., Zhu, S., Ye, Z., He, C., Liu, M., Huang, K., Zhong, J., Xu, F., & Li, Q. (2018). *Nutrient Stress – Dysregulated Antisense lncRNA GLS-AS Impairs GLS-Mediated Metabolism and Represses Pancreatic Cancer Progression*. *7*, 1–16. https://doi.org/10.1158/0008-5472.CAN-18-0419

Deng, S. J., Chen, H. Y., Zeng, Z., Deng, S., Zhu, S., Ye, Z., He, C., Liu, M. L., Huang, K., Zhong, J. X., Xu, F. Y., Li, Q., Liu, Y., Wang, C., & Zhao, G. (2019). Nutrient stress-dysregulated antisense lncRNA GLS-AS impairs GLS-mediated metabolism and represses pancreatic cancer progression. *Cancer Research*, *79*(7), 1398–1412. https://doi.org/10.1158/0008-5472.CAN-18-0419

Derderian, C., Orunmuyi, A. T., & Olapade-olaopa, E. O. (2019). *PVT1 Signaling Is a Mediator of Cancer Progression*. *9*(June), 1–8. https://doi.org/10.3389/fonc.2019.00502

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D. G., Lagarde, J., Veeravalli, L., Ruan, X., Ruan, Y., Lassmann, T., Carninci, P., Brown, J. B., Lipovich, L., Gonzalez, J. M., … Guigo, R. (2012). *The GENCODE v7 catalog of human long noncoding RNAs : Analysis of their gene structure , evolution , and expression*. 1775–1789. https://doi.org/10.1101/gr.132159.111.

Duffy, M. J. (2006). Estrogen Receptors: Role in Breast Cancer. *Critical Reviews in Clinical Laboratory Sciences*, *43*(4), 325–347. https://doi.org/10.1080/10408360600739218

Edechi, C. A., Ikeogu, N., Uzonna, J. E., & Myal, Y. (2019). *Regulation of Immunity in Breast Cancer*. *Cdc*, 1–18.

Edge, S. B., & Compton, C. C. (2010). The American Joint Committee on Cancer: the 7th Edition of the AJCC Cancer Staging Manual and the Future of TNM. *Annals of Surgical Oncology*, *17*(6), 1471–1474. https://doi.org/10.1245/s10434-010-0985-4

Fan, Y., Li, H., Yu, Z., Dong, W., Cui, X., Ma, J., & Li, S. (2020). *Long non-coding RNA FGD5-AS1 promotes non-small cell lung cancer cell proliferation through sponging hsa-miR-107 to up-regulate FGFRL1*. *0*(January), 1–12.

Figueroa, T., Boumart, I., Coupeau, D., & Rasschaert, D. (2016). Hyperediting by ADAR1 of a new herpesvirus lncRNA during the lytic phase of the oncogenic Marek's disease virus. *Journal of General Virology*, *97*(11), 2973–2988. https://doi.org/10.1099/jgv.0.000606

Frankish, A., Diekhans, M., Ferreira, A., Johnson, R., Jungreis, I., Loveland, J., Mudge, J. M., Sisu, C., Wright, J., Armstrong, J., Barnes, I., Berry, A., Bignell, A., Domenico, D., Sala, S. C., Chrast, J., Cunningham, F., Donaldson, S., Fiddes, I. T., … Flicek, P. (2019). *GENCODE reference annotation for the human and mouse genomes ıa Gir on*. *47*(October 2018), 766–773. https://doi.org/10.1093/nar/gky955

Fumagalli, D., Gacquer, D., Campbell, P. J., Sotiriou, C., Detours, V., Lefort, A., Libert, F., & Brown, D. (2015). *Principles Governing A-to-I RNA Editing in the Breast Cancer Transcriptome Article Principles Governing A-to-I RNA Editing in the Breast Cancer Transcriptome*. 277–289. https://doi.org/10.1016/j.celrep.2015.09.032

Gene, M., Yaghoobi, H., Azizi, H., Kholghi, V., & Taheri, M. (2020). PTEN-induced putative

kinase 1 (PINK1) down-regulation in breast cancer samples in association with mitotic rate. *Meta Gene*, *25*(June), 100748. https://doi.org/10.1016/j.mgene.2020.100748

Gonzalez, H., Hagerling, C., & Werb, Z. (2018). Roles of the immune system in cancer: From tumor initiation to metastatic progression. *Genes and Development*, *32*(19–20), 1267–1284. https://doi.org/10.1101/GAD.314617.118

Grossman, R. L., Heath, A. P., Ferretti, V., Varmus, H. E., Lowy, D. R., Kibbe, W. A., & Staudt, L. M. (2016). Toward a Shared Vision for Cancer Genomic Data. *New England Journal of Medicine*, *375*(12), 1109–1112. https://doi.org/10.1056/NEJMp1607591

Györffy, B., Lanczky, A., Eklund, A. C., Denkert, C., Budczies, J., Li, Q., & Szallasi, Z. (2010). An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast Cancer Research and Treatment*, *123*(3), 725–731. https://doi.org/10.1007/s10549-009-0674-9

Hanahan, D., & Weinberg, R. A. (2000). The Hallmarks of Cancer. *Cell*, *100*(1), 57–70. https://doi.org/10.1016/S0092-8674(00)81683-9

Hangauer, M. J., Vaughn, I. W., & Mcmanus, M. T. (2013). *Pervasive Transcription of the Human Genome Produces Thousands of Previously Unidentified Long Intergenic Noncoding RNAs*. *9*(6). https://doi.org/10.1371/journal.pgen.1003569

He, W., Qiao, Z. X., & Ma, B. (2019). Long noncoding RNA FAM201A mediates the metastasis of lung squamous cell cancer via regulating ABCE1 expression. *European Review for Medical and Pharmacological Sciences*, *23*(23), 10343–10353. https://doi.org/10.26355/eurrev_201912_19672

Heale, B. S. E., Keegan, L. P., McGurk, L., Michlewski, G., Brindle, J., Stanton, C. M., Caceres, J. F., & O'Connell, M. A. (2009). Editing independent effects of ADARs on the miRNA/siRNA pathways. *EMBO Journal*, *28*(20), 3145–3156. https://doi.org/10.1038/emboj.2009.244

Hendrickson, D., Kelley, D. R., Tenen, D., Bernstein, B., & Rinn, J. L. (2016). Widespread RNA binding by chromatin-associated proteins. *Genome Biology*, *17*(1), 1–18. https://doi.org/10.1186/s13059-016-0878-3

Higuchi, M., Maas, S., Single, F. N., Hartner, J., Rozov, A., Burnashev, N., Feldmeyer, D., Sprengel, R., & Seeburg, P. H. (2000). *Point mutation in an AMPA receptor gene rescues lethality in mice de ® cient in the*. *406*(July), 1998–2001.

Hogg, M., Paro, S., Keegan, L. P., & O'Connell, M. A. (2011). RNA Editing by Mammalian ADARs. In *Advances in Genetics* (1st ed., Vol. 73, Issue C, pp. 87–120). Elsevier Inc. https://doi.org/10.1016/B978-0-12-380860-8.00003-3

Igney, F. H., & Krammer, P. H. (2002). Immune escape of tumors: apoptosis resistance and tumor counterattack. *Journal of Leukocyte Biology*, *71*(6), 907–920. https://doi.org/10.1189/jlb.71.6.907

Isin, M., & Dalay, N. (2015). LncRNAs and neoplasia. *Clinica Chimica Acta*, *444*, 280–288. https://doi.org/10.1016/j.cca.2015.02.046

Jarroux, J., Morillon, A., & Pinskaya, M. (2017). History, discovery, and classification of

lncRNAs. In *Advances in Experimental Medicine and Biology*. https://doi.org/10.1007/978-981-10-5203-3_1

Jia, H., Khatun, J., Wood, E., Risk, B., Jr, W. E. G., Kundaje, A., Gunawardena, H. P., Yu, Y., Xie, L., Krajewski, K., Strahl, B. D., Chen, X., Bickel, P., Giddings, M. C., Brown, J. B., & Lipovich, L. (2012). *Long noncoding RNAs are rarely translated in two human cell lines*. 1646–1657. https://doi.org/10.1101/gr.134767.111.

Kawakubo, K., & Samuel, C. E. (2000). *Human RNA-specific adenosine deaminase (ADAR1) gene specifies transcripts that initiate from a constitutively active alternative promoter*. *258*, 165–172.

Kim, J., Abdelmohsen, K., Yang, X., De, S., Grammatikakis, I., Noh, J. H., & Gorospe, M. (2016). LncRNA OIP5-AS1/cyrano sponges RNA-binding protein HuR. *Nucleic Acids Research*, *44*(5), 2378–2392. https://doi.org/10.1093/nar/gkw017

Kim, U., Wang, Y., Sanford, T., Zeng, Y., & Nishikura, K. (1994). Molecular cloning of cDNA for double-stranded RNA adenosine deaminase, a candidate enzyme for nuclear RNA editing. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(24), 11457–11461. https://doi.org/10.1073/pnas.91.24.11457

Kume, H., Hino, K., Galipon, J., & Ui-tei, K. (2014). *A-to-I editing in the miRNA seed region regulates target mRNA selection and silencing efficiency containing either C : G or I : C pairs may account for*. 3–8. https://doi.org/10.1093/nar/gku662

Kung, C.-P., Cottrell, K. A., Ryu, S., Bramel, E. R., Kladney, R. D., Bross, E. A., Maggi, L., & Weber, J. D. (2020). ADAR1 editing dependency in triple-negative breast cancer. *BioRxiv*, 2020.01.31.928911. https://doi.org/10.1101/2020.01.31.928911

Li, Y., Jiang, T., Zhou, W., Li, J., Li, X., Wang, Q., Jin, X., Yin, J., Chen, L., Zhang, Y., Xu, J., & Li, X. (2020). Pan-cancer characterization of immune-related lncRNAs identifies potential oncogenic biomarkers. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-14802-2

Liao, Y., Smyth, G. K., & Shi, W. (2014). *Sequence analysis featureCounts : an efficient general purpose program for assigning sequence reads to genomic features*. *30*(7), 923–930. https://doi.org/10.1093/bioinformatics/btt656

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., & Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Systems*, *1*(6), 417–425. https://doi.org/10.1016/j.cels.2015.12.004

Liu, A. M., Zhu, Y., Huang, Z. W., Lei, L., Fu, S. Z., & Chen, Y. (2019). Long noncoding RNA FAM201A involves in radioresistance of non-small-cell lung cancer by enhancing EGFR expression via miR-370. *European Review for Medical and Pharmacological Sciences*, *23*(13), 5802–5814. https://doi.org/10.26355/eurrev_201907_18319

Liu, Z., Dou, C., Yao, B., Xu, M., Ding, L., Wang, Y., Jia, Y., Li, Q., Zhang, H., Tu, K., Song, T., & Liu, Q. (2016). Ftx non coding RNA-derived miR-545 promotes cell proliferation by targeting RIG-I in hepatocellular carcinoma. *Oncotarget*, *7*(18), 25350–25365. https://doi.org/10.18632/oncotarget.8129

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 1–21. https://doi.org/10.1186/s13059-014-0550-8

Luo, Y., Xuan, Z., Zhu, X., Zhan, P., & Wang, Z. (2018). Long non-coding RNAs RP5-821D11.7, APCDD1L-AS1 and RP11-277P12.9 were associated with the prognosis of lung squamous cell carcinoma. *Molecular Medicine Reports*, *17*(5), 7238–7248. https://doi.org/10.3892/mmr.2018.8770

Ma, C., Liu, H., Yi-Feng Chang, I., Wang, W., Chen, Y., Wu, S., Chen, H., Kuo, Y., Shih, C., Li, C., & Tan, B. C. (2019). ADAR1 promotes robust hypoxia signaling via distinct regulation of multiple HIF-1α-inhibiting factors. *EMBO Reports*, *20*(5), 1–16. https://doi.org/10.15252/embr.201847107

Mahmoud, S. M. A., Paish, E. C., Powe, D. G., Macmillan, R. D., Grainge, M. J., Lee, A. H. S., Ellis, I. O., & Green, A. R. (2011). Tumor-infiltrating CD8+ lymphocytes predict clinical outcome in breast cancer. *Journal of Clinical Oncology*, *29*(15), 1949–1955. https://doi.org/10.1200/JCO.2010.30.5037

Melcher, T., Maas, S., Herb, A., Sprengel, R., Higuchi, M., & Seeburg, P. H. (1996). *of the RNA-specific Adenosine Deaminase Family *. 15*, 31795–31799.

Myer, V. E., Fan, X. C., & Steitz, J. A. (1997). *Identification of HuR as a protein implicated in AUUUA-mediated mRNA decay*. *16*(8), 2130–2139.

Nakano, M., Fukami, T., Gotoh, S., & Nakajima, M. (2017). A-to-I RNA editing up-regulates human dihydrofolate reductase in breast cancer. *Journal of Biological Chemistry*, *292*(12), 4873–4884. https://doi.org/10.1074/jbc.M117.775684

Nishikura, K. (2016). A-to-I editing of coding and non-coding RNAs by ADARs. *Nature Reviews Molecular Cell Biology*, *17*(2), 83–96. https://doi.org/10.1038/nrm.2015.4

Ørom, U. A., Derrien, T., Beringer, M., Gumireddy, K., Gardini, A., Bussotti, G., Lai, F., Zytnicki, M., Notredame, C., Huang, Q., Guigo, R., & Shiekhattar, R. (2010). Long Noncoding RNAs with Enhancer-like Function in Human Cells. *Cell*, *143*(1), 46–58. https://doi.org/10.1016/j.cell.2010.09.001

Ota, H., Sakurai, M., Gupta, R., Valente, L., Wulff, B.-E., Ariyoshi, K., Iizasa, H., Davuluri, R. V., & Nishikura, K. (2013). ADAR1 Forms a Complex with Dicer to Promote MicroRNA Processing and RNA-Induced Gene Silencing. *Cell*, *153*(3), 575–589. https://doi.org/10.1016/j.cell.2013.03.024

Pagès, F., Kirilovsky, A., Mlecnik, B., Asslaber, M., Tosolini, M., Bindea, G., Lagorce, C., Wind, P., Marliot, F., Bruneval, P., Zatloukal, K., Trajanoski, Z., Berger, A., Fridman, W. H., & Galon, J. (2009). In situ cytotoxic and memory T cells predict outcome in patients with early-stage colorectal cancer. *Journal of Clinical Oncology*, *27*(35), 5944–5951. https://doi.org/10.1200/JCO.2008.19.6147

Pan, J., Lu, Y., Tang, Y., Wang, Y., & Li, L. (2020). *LIMD1-AS1 suppressed non-small cell lung cancer progression through stabilizing LIMD1 mRNA via hnRNP U*. *January*, 3829–3839. https://doi.org/10.1002/cam4.2898

Panhale, A., Richter, F. M., Ramírez, F., Mittler, G., Akhtar, A., & Manke, T. (n.d.). CAPRI enables comparison of evolutionarily conserved RNA interacting regions. *Nature Communications*, *2019*. https://doi.org/10.1038/s41467-019-10585-3

Park, E., & Maquat, L. E. (2014). *NIH Public Access*. *4*(4), 423–435. https://doi.org/10.1002/wrna.1168.Staufen-mediated

Paz-Yaacov, N., Bazak, L., Buchumenski, I., Porath, H. T., Danan-Gotthold, M., Knisbacher, B. A., Eisenberg, E., & Levanon, E. Y. (2015). Elevated RNA Editing Activity Is a Major Contributor to Transcriptomic Diversity in Tumors. *Cell Reports*, *13*(2), 267–276. https://doi.org/10.1016/j.celrep.2015.08.080

Picardi, E., D'Erchia, A. M., Gallo, A., Montalvo, A., & Pesole, G. (2014). Uncovering RNA editing sites in long non-coding RNAs. *Frontiers in Bioengineering and Biotechnology*, *2*(DEC), 1–6. https://doi.org/10.3389/fbioe.2014.00064

Picardi, E., Manzari, C., Mastropasqua, F., Aiello, I., D'Erchia, A. M., & Pesole, G. (2015). Profiling RNA editing in human tissues: Towards the inosinome Atlas. *Scientific Reports*, *5*, 1–17. https://doi.org/10.1038/srep14941

Picardi, E., & Pesole, G. (2013). REDItools: High-throughput RNA editing detection made easy. *Bioinformatics*, *29*(14), 1813–1814. https://doi.org/10.1093/bioinformatics/btt287

Pokharel, S., & Beal, P. A. (2006). High-throughput screening for functional adenosine to inosine RNA editing systems. *ACS Chemical Biology*, *1*(12), 761–765. https://doi.org/10.1021/cb6003838

Pruneri, G., Vingiani, A., & Denkert, C. (2018). Tumor infiltrating lymphocytes in early breast cancer. *Breast*, *37*, 207–214. https://doi.org/10.1016/j.breast.2017.03.010

Pullirsch, D., & Jantsch, M. F. (2010). Proteome diversification by adenosine to inosine RNA-editing. *RNA Biology*, *7*(2), 205–212. https://doi.org/10.4161/rna.7.2.11286

Qi, L., Song, Y., Chan, T. H. M., Yang, H., Lin, C. H., Tay, D. J. T., Hong, H. Q., Tang, S. J., Tan, K. T., Huang, X. X., Lin, J. S., Ng, V. H. E., Maury, J. J. P., Tenen, D. G., & Chen, L. (2017). An RNA editing/dsRNA binding-independent gene regulatory mechanism of ADARs and its clinical implication in cancer. *Nucleic Acids Research*, *45*(18), 10436–10451. https://doi.org/10.1093/nar/gkx667

Qin, Y., Qiao, J., Hon, T., Chan, M., Zhu, Y., Li, F., & Liu, H. (2014). *Adenosine-to-Inosine RNA Editing Mediated by ADARs in Esophageal Squamous Cell Carcinoma*. *2*(8), 840–852. https://doi.org/10.1158/0008-5472.CAN-13-2545

Ravasi, T., Suzuki, H., Pang, K. C., Katayama, S., Furuno, M., Okunishi, R., Fukuda, S., Ru, K., Frith, M. C., Gongora, M. M., Grimmond, S. M., Hume, D. A., Hayashizaki, Y., & Mattick, J. S. (2006). *Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome*. 11–19. https://doi.org/10.1101/gr.4200206.1

Rimawi, M. F., Schiff, R., & Osborne, C. K. (2015). Targeting HER2 for the treatment of breast cancer. *Annual Review of Medicine*, *66*, 111–128. https://doi.org/10.1146/annurev-med-042513-015127

Rueter, S. M., Dawson, T. R., & Emeson, R. B. (1999). Regulation of alternative splicing by

RNA editing. *Nature*, *399*(6731), 75–80. https://doi.org/10.1038/19992

Sagredo, E. A., Blanco, A., Sagredo, A. I., Pérez, P., Sepúlveda-Hermosilla, G., Morales, F., Müller, B., Verdugo, R., Marcelain, K., Harismendy, O., & Armisén, R. (2018). ADAR1-mediated RNA-editing of 3′UTRs in breast cancer. *Biological Research*, *51*(1), 1–10. https://doi.org/10.1186/s40659-018-0185-4

Sagredo, E. A., Sagredo, A. I., Blanco, A., De Santiago, P. R., Rivas, S., Assar, R., Pérez, P., Marcelain, K., & Armisén, R. (2020). ADAR1 Transcriptome editing promotes breast cancer progression through the regulation of cell cycle and DNA damage response. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 118716. https://doi.org/10.1016/j.bbamcr.2020.118716

Sakurai, M., Shiromoto, Y., Ota, H., Song, C., Kossenkov, A. V, Wickramasinghe, J., Showe, L. C., Skordalakes, E., Tang, H.-Y., Speicher, D. W., & Nishikura, K. (2017). ADAR1 controls apoptosis of stressed cells by inhibiting Staufen1-mediated mRNA decay. *Nature Structural & Molecular Biology*, *24*(6), 534–543. https://doi.org/10.1038/nsmb.3403

Salameh, A., Lee, A. K., Cardó-Vila, M., Nunes, D. N., Efstathiou, E., Staquicini, F. I., Dobroff, A. S., Marchiò, S., Navone, N. M., Hosoya, H., Lauer, R. C., Wen, S., Salmeron, C. C., Hoang, A., Newsham, I., Lima, L. A., Carraro, D. M., Oliviero, S., Kolonin, M. G., … Arap, W. (2015). PRUNE2 is a human prostate cancer suppressor regulated by the intronic long noncoding RNA PCA3. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(27), 8403–8408. https://doi.org/10.1073/pnas.1507882112

Sati, S., Ghosh, S., Jain, V., Scaria, V., & Sengupta, S. (2012). *Genome-wide analysis reveals distinct patterns of epigenetic features in long non-coding RNA loci*. *40*(20), 10018–10031. https://doi.org/10.1093/nar/gks776

Scheele, C., Petrovic, N., Faghihi, M. A., Lassmann, T., Fredriksson, K., Rooyackers, O., Wahlestedt, C., Good, L., & Timmons, J. A. (2007). *The human PINK1 locus is regulated in vivo by a non-coding natural antisense RNA during modulation of mitochondrial function*. *13*. https://doi.org/10.1186/1471-2164-8-74

Schmitt, A. M., & Chang, H. Y. (2016). Long Noncoding RNAs in Cancer Pathways. *Cancer Cell*, *29*(4), 452–463. https://doi.org/10.1016/j.ccell.2016.03.010

Sherry, S. T., Ward, M., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., Sirotkin, K., & Res, K. S. G. (2001). *dbSNP : the NCBI database of genetic variation*. *29*(1), 308–311.

Siegel, R. L., Miller, K. D., & Jemal, A. (2020). Cancer statistics, 2020. *CA: A Cancer Journal for Clinicians*, *70*(1), 7–30. https://doi.org/10.3322/caac.21590

Solomon, O., Safran, M., Unger, N. D., Akiva, P., Jacob-, J., Cesarkas, K., Kabesa, R., Amariglio, N., Unger, R., Eyal, E., & Jacob-hirsch, J. (2013). *Global Regulation of Alternative Splicing By Adenosine To*. 591–604. https://doi.org/10.1261/rna.038042.112.4

Um, P. (2015). Cancer, Definition. In S. K. Highlander, F. Rodriguez-Valera, & B. A. White (Eds.), *Encyclopedia of Metagenomics* (pp. 65–65). Springer US. https://doi.org/10.1007/978-1-4899-7475-4_106

Villegas, V. E., & Zaphiropoulos, P. G. (2015). Neighboring gene regulation by antisense long

Non-Coding RNAs. *International Journal of Molecular Sciences*, *16*(2), 3251–3266. https://doi.org/10.3390/ijms16023251

Wang, C., Zou, J., Ma, X., Wang, E., & Peng, G. (2017). Mechanisms and implications of ADAR-mediated RNA editing in cancer. *Cancer Letters*, *411*, 27–34. https://doi.org/10.1016/j.canlet.2017.09.036

Wang, I. X., So, E., Devlin, J. L., Zhao, Y., Wu, M., & Cheung, V. G. (2013). ADAR Regulates RNA Editing, Transcript Stability, and Gene Expression. *Cell Reports*, *5*(3), 849–860. https://doi.org/10.1016/j.celrep.2013.10.002

Wang, K. C., & Chang, H. Y. (2011). Molecular mechanisms of long noncoding RNAs. *Molecular Cell*, *43*(6), 904–914. https://doi.org/10.1016/j.molcel.2011.08.01

Wang, Wei, Xie, Y., Chen, F., Liu, X., Zhong, L. L., Wang, H. Q., & Li, Q. C. (2019). LncRNA MEG3 acts a biomarker and regulates cell functions by targeting ADAR1 in colorectal cancer. *World Journal of Gastroenterology*, *25*(29), 3972–3984. https://doi.org/10.3748/wjg.v25.i29.3972

Wang, Wenxi, Zhou, R., Wu, Y., Liu, Y., Su, W., & Xiong, W. (2019). *PVT1 Promotes Cancer Progression via MicroRNAs*. *9*(July), 1–8. https://doi.org/10.3389/fonc.2019.00609

Wang, Z., Yang, B., Zhang, M., Guo, W., Wu, Z., Wang, Y., Jia, L., Li, S., Caesar-Johnson, S. J., Demchok, J. A., Felau, I., Kasapi, M., Ferguson, M. L., Hutter, C. M., Sofia, H. J., Tarnuzzer, R., Wang, Z., Yang, L., Zenklusen, J. C., … Yang, D. (2018). lncRNA Epigenetic Landscape Analysis Identifies EPIC1 as an Oncogenic lncRNA that Interacts with MYC and Promotes Cell-Cycle Progression in Cancer. *Cancer Cell*, *33*(4), 706-720.e9. https://doi.org/10.1016/j.ccell.2018.03.006

What Is Breast Cancer? | CDC. (n.d.). Retrieved October 1, 2020, from https://www.cdc.gov/cancer/breast/basic_info/what-is-breast-cancer.htm

Xu, X., Wang, Y., & Liang, H. (2018). The role of A-to-I RNA editing in cancer development. *Current Opinion in Genetics and Development*, *48*, 51–56. https://doi.org/10.1016/j.gde.2017.10.009

Yang, Y., Zhang, J., Chen, X., Xu, X., Cao, G., Li, H., & Wu, T. (2018). LncRNA FTX sponges miR-215 and inhibits phosphorylation of vimentin for promoting colorectal cancer progression. *Gene Therapy*. https://doi.org/10.1038/s41434-018-0026-7

Zaha, D. C. (2014). Significance of immunohistochemistry in breast cancer. In *World Journal of Clinical Oncology* (Vol. 5, Issue 3, pp. 382–392). https://doi.org/10.5306/wjco.v5.i3.382

Zhang, N., Zeng, X., Sun, C., Guo, H., Wang, T., Wei, L., Zhang, Y., & Zhao, J. (2019). LncRNA LINC00963 Promotes Tumorigenesis and Radioresistance in Breast Cancer by Sponging miR-324-3p and Inducing ACK1 Expression. *Molecular Therapy: Nucleic Acid*, *18*(44), 871–881. https://doi.org/10.1016/j.omtn.2019.09.033

Zhao, Shanrong, Xi, L., Quan, J., Xi, H., Zhang, Y., Schack, D. Von, & Vincent, M. (2016). *QuickRNASeq lifts large-scale RNA-seq data analyses to the next level of automation and interactive visualization*. 1–15. https://doi.org/10.1186/s12864-015-2356-9

Zhao, Shen, Zuo, W.-J., Shao, Z.-M., & Jiang, Y.-Z. (2020). Molecular subtypes and precision

treatment of triple-negative breast cancer. *Annals of Translational Medicine*, *8*(7), 499–499. https://doi.org/10.21037/atm.2020.03.194

Zygogianni, A. G., Kyrgias, G., Gennatas, C., Ilknur, A., Armonis, V., Tolia, M., Papaloukas, C., Pistevou, G., Kouvaris, J., & Kouloulias, V. (2012). Male Breast Carcinoma: Epidemiology, Risk Factors and Current Therapeutic Approaches. *Asian Pacific Journal of Cancer Prevention*, *13*(1), 15–19. https://doi.org/10.7314/APJCP.2012.13.1.015