UNIVERSITAT ROVIRA i VIRGILI

EXPERIMENTAL APPROACHES FOR DETECTING LONG NONCODING RNAs IN MALIGNANT PLEURAL EFFUSION

FINAL DEGREE PROJECT IN BIOCHEMISTRY AND MOLECULAR BIOLOGY

Double Degree in Biotechnology and Biochemistry and Molecular Biology

Núria Serrano i Ribes

Tutor: Dra. Helena Torrell Galceran

Supervisors: Dra. M. Alba Sorolla Bardaji and Dra. Eva Parisi Capdevila

Tarragona, June 2023

This project is based on the results obtained over the curricular practices carried out in the Research group of Cancer Biomarkers (GReBiC) under the mentorship of Dra. M. Alba Sorolla Bardaji and Dra. Eva Parisi Capdevila.

ACKNOWLEDGEMENTS

I would like to start this project by thanking all those people who, in some way, have made this work possible and, above all, those who have accompanied me along the path that has led me here.

Firstly, I would like to thank the Research group of Cancer Biomarkers (GReBiC) for allowing me to carry out the internship and final degree project, especially Dra. Eva Parisi and Dra. M. Alba Sorolla, who have guided me throughout the entire process. Also, to the rest of the group, who made me feel at home at all times. As well, I would like to thank the Translational Research in Respiratory Medicine group, particularly Thalía Belmonte.

I also thank my tutor, Dra. Helena Torrell, who has been very attentive and helped me whenever I needed it.

Finally, I would like to thank my family for their unconditional support at every step of this journey that has just begun.

This work would not have been possible without any of you. Thank you all.

INDEX

ABSTRACT

Malignant pleural effusion (MPE) is observed in 15% of lung cancer (LC) patients during diagnosis. However, the current biomarkers tested in pleural fluid (PF) demonstrate limited sensitivity and specificity. This study focuses on establishing the optimal experimental condition for long noncoding RNAs (lncRNAs) which are tumour-derived products present in PF. The aim is to enable further research in this area. A qPCR assay was conducted to evaluate the impact of different initial sample volumes (50 µL, 100 µL, 200 µL), the inclusion of an RNA spike-in (cel-miR-39) during the extraction process and the implementation of a preamplification step. The results suggest that the optimal conditions for lncRNA extraction from PF involve an initial volume of 200 µL, the addition of cel-miR-39 and carrying out of a preamplification procedure. Furthermore, *ACTB* and *GAPDH* were validated as reference genes for data normalisation in PF samples. Considering the active role of lncRNAs within the oncogenic functions, it is conceivable that the identification of specific lncRNAs in PF could potentially contribute to the development of an algorithm for the diagnosis and prognosis of patients with LC and associated MPE.

Key words: Lung cancer, biomarker, lncRNA, malignant pleural effusion.

INDEX OF ABBREVIATIONS

ACTB	Actin Beta	**miRNA**	MicroRNA
ADC	Adenocarcinoma	**MPE**	Malignant Pleural Effusion
BAX	BCL2 Associated X	**mRNA**	Messenger RNA
CDKN2A	Cyclin Dependent Kinase Inhibitor 2A	**mTOR**	Mechanistic Target of Rapamycin Kinase
cDNA	Complementary DNA	***NEAT1***	Nuclear Paraspeckle Assembly Transcript 1
CEA	Carcinoembryonic Antigen	***NEDD9***	Neural Precursor Cell Expressed, Developmentally Down-Regulated 9
cfDNA	Cell-Free DNA	**NSCLC**	Non-Small Cell Lung Cancer
cGAS	Cyclic GMP-AMP Synthase	**PBS**	Phosphate-Buffered Saline
$\mathbf{C_q}$	Quantification Cycle	**PCR**	Polymerase Chain Reaction
$\mathbf{C_t}$	Threshold Cycle	**PF**	Pleural Fluid
CTC	Circulating Tumour Cells	**PI3K**	Phosphatidylinositol-4,5-Bisphosphate 3-Kinase
ctDNA	Circulating Tumour DNA	**piRNA**	Piwi-Interacting RNA
DNA	Deoxyribonucleic Acid	**PRC1/2**	Polycomb Repressive Complex 1/2
DNMT1	DNA Methyltransferase 1	***PTEN***	Phosphatase and Tensin Homolog
dNTP	Deoxynucleotide Triphosphate	**qPCR**	Quantitative PCR
EGFR	Epidermal Growth Factor Receptor	**REST**	RE1 Silencing Transcription Factor
EMT	Epithelial-Mesenchymal Transition	**RIN**	RNA Integrity Number

ERK	Extracellular Signal-Regulated Kinase
EZH2	Enhancer Of Zeste 2 Polycomb Repressive Complex 2 Subunit
FUT4	Fucosyltransferase 4
GAPDH	Glyceraldehyde-3-Phosphate Dehydrogenase
GAS5	Growth Arrest Specific 5
GR	Glucocorticoid Receptors
HOTAIR	Hox Transcript Antisense RNA
HPRT1	Hypoxanthine Phosphoribosyltransferase 1
HUAV	Hospital Universitari Arnau De Vilanova
LCLC	Large-Cell Lung Carcinoma
lncRNA	Long Noncoding RNA
LSD1	Lysine Demethylase 1a
MALAT1	Metastasis Associated Lung Adenocarcinoma Transcript 1
MAPK6	Mitogen-Activated Protein Kinase 6
MEG3	Maternally Expressed 3
MEK	Mitogen-Activated ERK-Regulating Kinase
RNA	Ribonucleic Acid
RT-PCR	Real-Time PCR
SCLC	Small Cell Lung Cancer
SLC34A2	Solute Carrier Family 34 Member 2
SLUG	Snail Family Transcriptional Repressor 2
SNP	Single-Nucleotide Polymorphism
SqCC	Squamous Cell Carcinoma
STING	Stimulator Of Interferon Response CGAMP Interactor
TEAD	Tea Domain Transcription Factor
TGF-β	Transforming Growth Factor Beta
TP53	Tumour Protein P53
TWIST	Twist Family BHLH Transcription Factor
VEGF	Vascular Endothelial Growth Factor
WIF-1	Wnt Inhibitory Factor 1
ZEB1/2	Zinc Finger E-Box Binding Homeobox 1
ZBED5-AS1	ZBED5 Antisense RNA 1

1. INTRODUCTION

In 2019, cancer accounted for approximately 18% of total deaths globally, making it the second leading cause of dead worldwide. Among the most frequent types, lung cancer is one of the most prevalent, ranking second in terms of incidence. In 2020 alone, it is estimated to have resulted in the deaths of 1.8 million individuals (*Figure 1*).

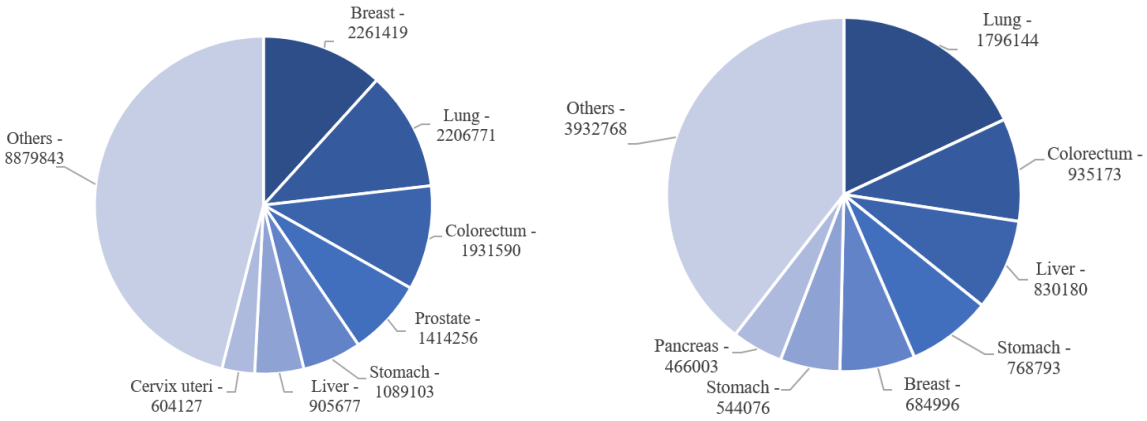


***Figure 1**. Estimated number of new cases (A) and deaths (B) worldwide in 2020, both sexes and all ages*[1].

The high incidence of lung cancer can be attributed to various factors, including its aetiology of lung cancer, as well as several associated risk factors. Well-established risk factors include exposure to occupational carcinogens such asbestos or radon, environmental air pollution, and a family history of lung cancer[2–4]. However, smoking remains the predominant threat, as more than 80% of lung cancers occur in individuals with a history of tobacco use. Mainstream smoke contains many potential carcinogens, including polycyclic aromatic hydrocarbons, aromatic amines, *N*-nitrosamines and other organic and inorganic compounds, such as benzene, vinyl chloride, arsenic and chromium[3].

Certain genetic alterations have been recognised as driver mutations in the development of lung cancer. Among them, the EGFR-RAS-RAF-MEK-ERK signalling pathway (*Figure 2*), a key regulator of cell growth and transformation, frequently harbours mutations in genes such as, the *EGFR*, *KRAS* and *BRAF*, being these mutations all mutually exclusive. Constitutive activation of this pathway results in uncontrolled cell proliferation, increased cell survival and the initiation and progression of tumour[5–7].

The *Epidermal Growth Factor Receptor* or *EGFR* gene encodes a transmembrane glycoprotein, which is a member of the protein kinase superfamily[6]. Mutations in *EGFR* gene have been shown to play a critical role not in tumour development, but also in its stability and maintenance.

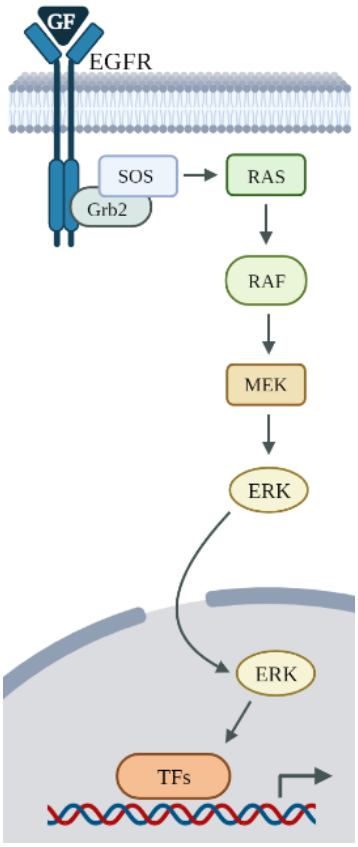


***Figure 2**. Schematic representation of the EGFR-RAS-RAF-MEK-ERK pathway. GF, growth factor; TFs, transcription factors.*

On the contrary, mutations in both *KRAS* and *BRAF* induce cell proliferation and cellular atypia, leading to preinvasive lesions. However, these mutation rarely progress to invasive lesions unless additional genomic events occur, such as the inactivation of *TP53* or *CDKN2A*[7].

The presence of different mutations, combined with varying exposure to various risk factors result, give rise to two primary histotypes that classify lung cancers: small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC).

SCLC accounts for 15% of all lung cancers and it is preferentially located in the central area of the lung. Strongly associated with tobacco smoking, SCLC is characterised by its high aggressiveness and rapid growth.

Otherwise, advances in molecular profiling have facilitated the subdivision of NSCLC into three main subtypes: adenocarcinoma (ADC) and its variants, squamous cell carcinoma (SqCC) and large-cell lung carcinoma (LCLC). While there are other types, such as salivary gland-type tumours, sarcomatoid carcinomas, and others, the constitute a small fraction of the overall NSCLC cases.

ADC accounts for around 40% of all lung cancer cases, making it the most common subtype of NSCLC. It is predominantly associated with non-smokers and EGFR mutations; it is characterised as a malignant epithelial tumour with glandular differentiation. Typically, ADC is located at the periphery of the lung, affecting the pleura and the thoracic wall. On the contrary, SqCC, which accounts for nearly 20% of lung cancers and tends to manifest in a central location, arising in a main or lobar

bronchus, and it often exhibits a necrotic interior. Lastly, LCLC is defined as an undifferentiated NSCLC that represents less than 10% of the total cases. It receives its name because of the size of the cells constituting it [6,8].

Amongst the different conditions that can derive from the cancer disease, the malignant pleural effusion (MPE) is a recurrent complication in oncogenic patients associated to many types of tumours, including lung cancer[9]. The pleural fluid (PF) is a plasma ultrafiltrate enclosed between the pleura membrane coating the lungs and the thoracic cavity, allowing both to slide with ease during pulmonary respiration. Under physiological conditions, the estimated volume of pleural fluid is approximately 0.15 mL/kg of bodyweight per hemithorax[10].

Any accumulation of clinically detectable amounts is considered abnormal and may be indicative of a disease, either pleural, pulmonary or extrapulmonary. Within the most frequent sources of pleural effusion (PE) are heart failure, pneumonia or tuberculosis. However, PE becomes malignant (MPE) when it is due to cancer, which is the leading cause, being lung cancer the most frequent primary tumour (37% of the cases)[11].

MPE often leads to dyspnoea, shortness of breath, in patients, serving as a significant factor in diagnosing the cancer. In fact, 15% of lung cancer patients present with PE at the time of diagnosis, and this rate can increase up to 50% over time[12]. Unfortunately, the presence of MPE in lung cancer indicated an advanced disease, leading to an upstaging of the cancer to stage 4, which is associated with poor survival outcomes.

This relatively late-stage diagnosis in lung cancer, with the majority of cases being diagnosed at an advanced stage, combined with its aggressive nature, significantly impacts overall survival rates: the 5-year relative survival rate is only about 8% within the distant stage (cancer with metastasis) patients, which account for more than 50% of the cases at the time of diagnosis[4,13]. These statistics highlight the urgent need to improve and expand diagnostic techniques in this field.

Currently, to ensure accurate diagnosis and, a precise histological classification, a combination of thoracic imaging (usually computed tomography) and immunohistochemical assays is performed. The latter requires invasive testing techniques such as biopsies[9,14].

Cytological examination of sputum is another diagnostic procedure for lung cancer. It is particularly useful for detecting central tumours, like SqCC or SCLC, but it has limitations in detecting adenocarcinomas with a diameter smaller than two centimetres located in bronchial ramifications. In fact, sputum cytology is able to diagnose only 20-30% of early lung cancer cases[14].

This exam can also be performed on PF in case of the patient presenting pleural effusion. Pleural fluid is obtained through a medical procedure called thoracocentesis, in which a transthoracic needle is used to extract the fluid. Nevertheless, PF cytology has a sensitivity of only around 60% and can to 25% for certain specific subtypes such as squamous cell carcinoma[11].

Given the critical importance of an accurate diagnosis for the optimal treatment of lung cancer patients, there is an urgent need to identify and implement novel techniques that offer safer and more precise options. In this context, liquid biopsy is emerging as a promising non-invasive diagnostic tool for malignant pleural effusions, as they provide a comprehensive representation state of the tumour's status and enables real-time monitoring.

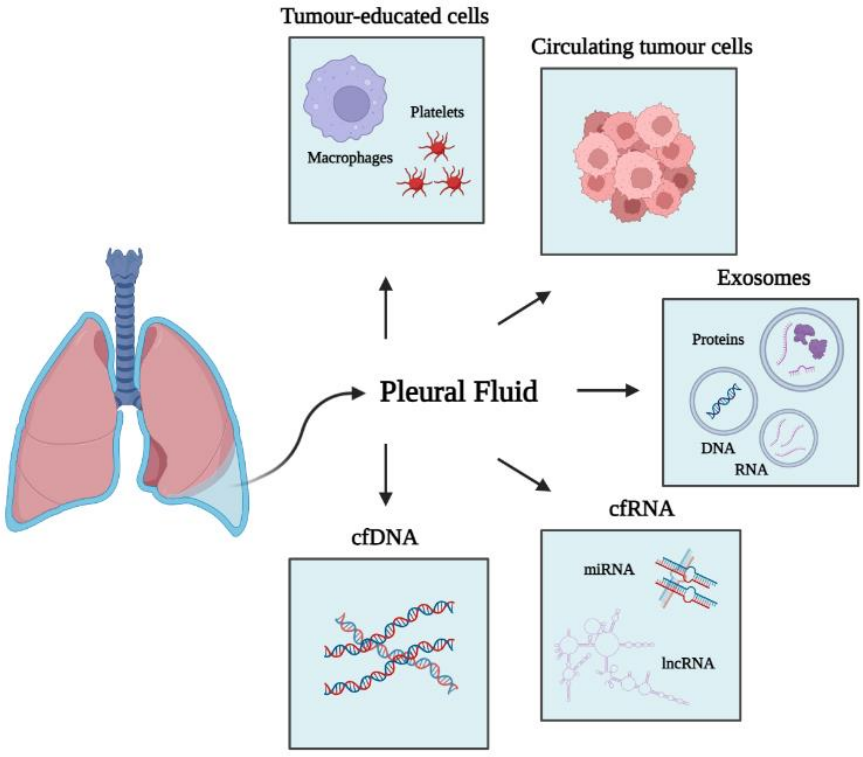


***Figure 3**. Schematic representation of the main tumour-derived products in PF: cfDNA, cell-free DNA; cfRNA, cell-free RNA. Adapted from Sorolla* et al. *(2021).*

Currently, liquid biopsy is mainly focused on analysing circulating tumour DNA (ctDNA) to detect of driver mutations and guide therapeutic interventions. Moreover, malignant pleural fluid (MPF) has been found to contain abundant tumour-derived products with clinical potential (*Figure 3*), including circulating tumour cells (CTC), tumour-educated cells, exosomes, cfDNA, micro-RNAs (miRNA) or long noncoding RNAs (lncRNA). This project specifically focuses on exploring the potential of lncRNA[15,16].

Long noncoding RNAs are transcripts commonly defined as greater than 200 nucleotides in length, generally ranging from 1,000 to 10,000. Despite their limited protein-coding potential, lncRNA have been found to have significant regulatory roles in gene expression, chromatin dynamics and cellular processed such as growth, development and differentiation [17,18]. Their large size enables them to adopt complex secondary or tertiary structures, allowing interactions with proteins, mRNA and DNA sequences.

The enormous number of newly discovered lncRNAs, along with their diverse biological functions, has posed a major challenge classifying them into different categories. However, they have been broadly classified on the basis of their function into three not mutually exclusive groups (*Figure 4*): guides, scaffolds and molecular decoys[17,18].

LncRNA guides serve as regulators that bind to proteins with either regulatory or enzymatic function, including transcription factors or chromatin modifiers, in order to lead them to exact sites in the genome, either in close proximity (*cis*) or at distant locations (*trans*) locations[17]. One example is *HOTAIR*, a transcript that recruits PRC2 and LSD1 in *trans* at the HOXD locus, leading to gene silencing through histone methylation. Its expression has been associated with tumour invasiveness and metastasis[18].

Scaffold lncRNAs play a structural role by providing a central platform for the assembly of multiple enzymatic complexes and/or other regulatory cofactors. Additionally, these ribonucleoprotein (RNP) complexes, often transient, can target a particular gene in order to regulate its expression. The lncRNA *MALAT1*, among its various functions, acts as a dynamic scaffold that plays a crucial role in stem cell

differentiation and early embryonic development by linking chromatin modifying complexes PRC1 and PRC2[17].

Molecular decoys lncRNA function by modulating gene expression by binding to particular regulatory factors, such as transcription factors, miRNAs or other catalytic proteins. The transcripts act as negative regulators by preventing effector factors from interacting with their primary targets[18]. Two notable examples in this category *GAS5* and *MEG3*. *GAS5* sequesters miR-106, thus inducing *PTEN*, and growth arrest via binding the glucocorticoid receptor and inhibiting its transcriptional induction. *MEG3* represses *MDM2* resulting in p53 accumulation and, hence, apoptosis and inhibition of proliferation. *MEG3* can also negatively regulate the expression of gens associated with the TGF-β pathway, involved in immune regulation and cell invasion[17].

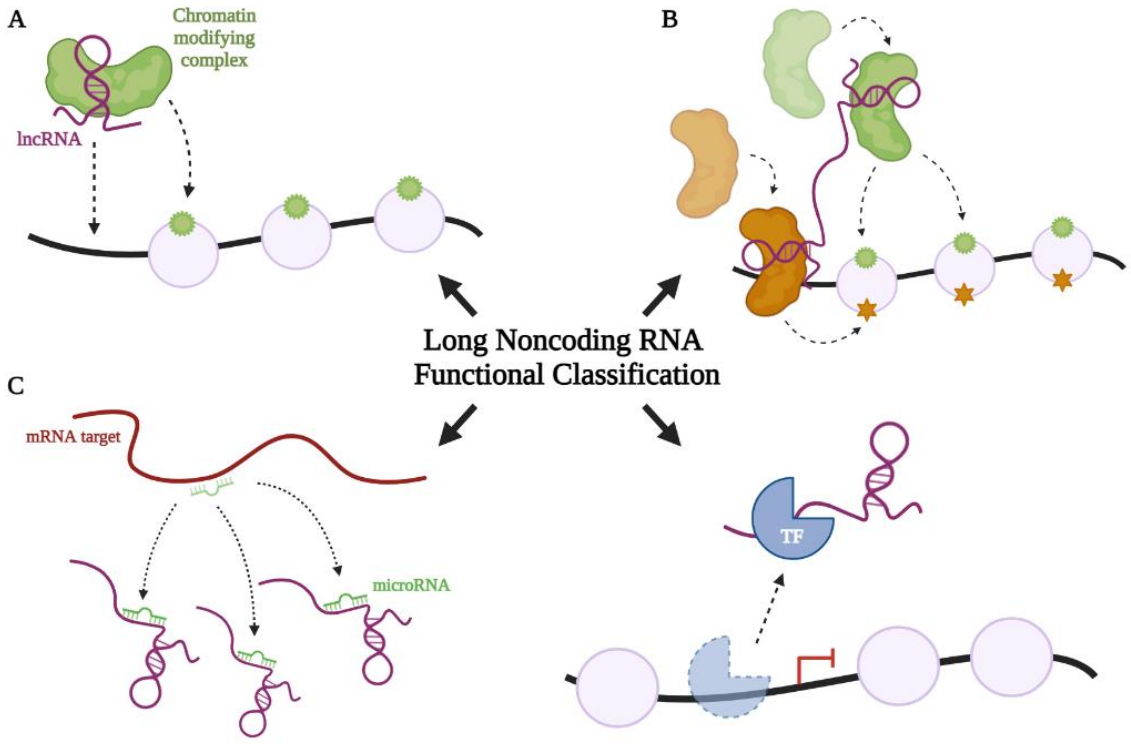


***Figure 4**. General mechanisms for lncRNA classification. LncRNA as (A) guide, (B) dynamic scaffold and (C) decoy. TF, transcription factor. Adapted from Balas* et al. *(2018).*

LncRNAs are highly expressed and widely implicated in various type of cancers. In fact, it has been established by several genome-wide association studies that more than 80% of the SNPs associated to cancer occur in noncoding regions of the genome. Some of these loci are sometimes transcribed to lncRNAs with a role in tumorigenesis, metastasis and tumour stage, in either tumour-suppression or tumour-promoting fashion.

Their expression is closely correlated with the transformed phenotype of cancer cells, impacting processes such as cell cycle regulation, cell mobility and pluripotency [17,18].

Numerous lncRNAs have shown specificity for particular types of cancer. In the case of lung cancer, studies utilising sequencing data and microarray profiling have proven significant alterations in lncRNA expression during cancer initiation and progression. While the precise biological functions and molecular mechanisms of these lncRNA remain complex and not fully elucidated, their dysregulation seem to play a crucial role in regulating essential processes such as proliferation, survival, invasion, metastasis as well as drug and radiation resistance[19]. Among the numerous lncRNAs associated with lung cancer, four transcripts have been widely detected: *HOTAIR*, *MALAT1*, *GAS5* and *MEG3*. *HOTAIR* and *MALAT1* are frequently upregulated and considered oncogenic, while *GAS5* and *MEG3* exhibit tumour-suppressive properties and are typically downregulated[18,19].

Although through largely unknown mechanisms, lncRNAs can be released into body fluids like plasma or pleural fluid, suggesting their promising role as biomarkers. However, limited studies have investigated the content of lncRNA in pleural fluid in order to exploit its diagnostic potential[19]. A notable example is the study of Wang *et al*. (2018) which determined that *MALAT1* in pleural effusion, in combination with carcinoembryonic antigen (CEA), shows higher sensitivity in discriminating between malignant and benign PE than CEA alone[20].

The development of novel diagnostic strategies based on lncRNA holds promise and signifies a new paradigm in cancer research, potentially emerging as a significant diagnostic approach in the near future.

2. HYPOTHESIS AND OBJECTIVE

As stated in the previous section, the significant number of lung cancer-related deaths, primarily caused by late diagnosis, underscores the importance of ongoing research into novel elements, including biomarkers, to improve disease detection systems.

With this objective in mind, the project revolves around validating the following hypothesis:

> Analysis of the differential noncoding RNA expression, such as lncRNAs, in pleural liquid samples may have the potential to serve as a tool to determine the optimal methodology to analyse those lncRNA species present in patients with malignant pleural effusion secondary to lung cancer.

Based on this premise, the following has been established as the main objective of this research:

1. To establish the optimal experimental conditions for lncRNAs extraction and its expression level analysis: initial volume, use of RNA spike-in and need of a preamplification step.

3. MATERIALS AND METHODS

3.1. Sample selection

Being one of the main aims of the project to set the optimal experimental conditions, a limited number of samples were chosen from the collection of the Pleural Medicine Unit of the Hospital Arnau de Vilanova (HUAV) in Lleida, led by Dr. José Manuel Porcel. These samples were obtained from patients diagnosed with malignant pleural effusion secondary to lung cancer, and their clinical details are listed below (*Table 1*).

***Table 1**. Clinical data from the selected patients.*

ID	Sex	Age	Date of collection	Histotype
523	Male	48	05/04/2023	Adenocarcinoma
521	Male	68	31/03/2023	Adenocarcinoma
465	Female	87	03/11/2022	Adenocarcinoma
425	Female	65	14/07/2022	Adenocarcinoma

The study was conducted according to the Declaration of Helsinki and approved by the ethics committee of HUAV. Informed consent was obtained from all the patients involved. It must also be noted that all samples come from the diagnostic surplus and its extraction does not alter the course of the disease in any way.

3.2. Initial sample processing

For each patient, around 100 mL of pleural fluid is obtained and later processed and stored by the personnel in the Biobank of IRBLleida.

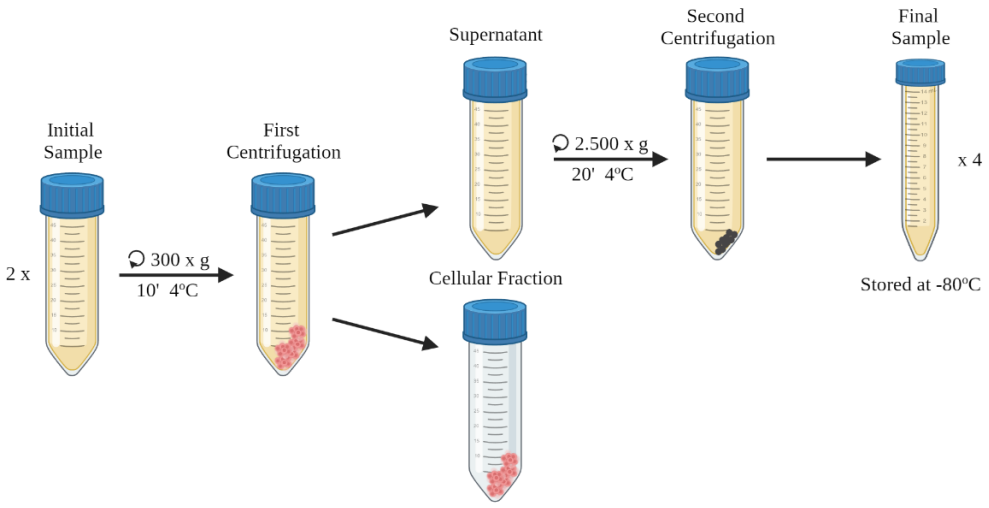


***Figure 5**. Diagram of the pleural fluid processing.*

The protocol consists of two consecutive centrifugation steps to separate the cellular fraction and other impurities. The resulting sample is then transferred to 15 mL Falcon tubes and stored at -80ºC until the moment of the analysis (*Figure 5*).

3.3. RNA isolation from the pleural fluid

***Figure 6**. Personal protective equipment used to enter the biosafety level II lab.*

Pleural fluids, being human biological samples, necessitate manipulation under the appropriated biosafety conditions. Consequently, the initial stages of the protocol were performed within a biosafety level II laboratory, inside a biosecurity II hood.

The entrance to this room requires a specific training provided by the Technical Scientific Cell Culture Service in IRBLleida, as well as a previously approved protocol by the Biosecurity Committee of the institution. Additionally, personal protective equipment (cap, security glasses, FFP2 mask, extra-safe lab coat, gloves and plastic shoe covers) must be worn (*Figure 6*).

To ensure the integrity of RNA, the frequent use of RNase AWAY® (VWR, PA, USA) was employed for decontaminating apparatus, benchtops and plastic ware due to its capability to eliminate both RNases and DNA from various surfaces.

The isolation of RNA was performed with the miRNeasy Mini Kit (Qiagen, NLD) which enables the purification of total RNA from a wide range of animal tissues and cells, including difficult-to-lyse samples. Its efficient enrichment of RNA over 200 nucleotides, along with the high purity of the resulting genetic material, makes the isolated RNA suitable for various downstream applications, even when limited amounts of starting material are available.

Firstly, 50, 150 and 200 µL aliquots of each pleural fluid sample were placed into 2 mL tubes followed by the addition of up to 200 µL of PBS. Later, 1000 µL of QIAzol Lysis Reagent was added to the tubes, and the mixture was vortexed for 15 seconds.

QIAzol is a monophasic solution consisting of phenol and guanidine thiocyanate. It is specially formulated to facilitate the tissue lysis, inhibit RNases and remove most of the cellular DNA and proteins from the lysate through organic extraction. QIAzol's ability to inactivate the sample allows the protocol to be continued in a standard laboratory

setting, however it is important to note that it should be handled under a fume hood due to its toxic nature, as well as that of some of the remaining reagents.

After a 5-minute incubation at room temperature, the samples were spined, 200 µL of chloroform were added (Sigma-Aldrich, MA, USA) and the tubes were vortexed again for 15 seconds. Continuing with a centrifugation at 12,000 x g for 15 minutes at 4ºC before transferring the upper aqueous phase (300 µL) to a new collection tube, avoiding any of the interphase material (*Figure 7*). Next, 450 µL of 100% ethanol (PanReac, ESP) were added and the mixture was thoroughly mixed to facilitate RNA precipitation. The entire volume was then transferred to a RNeasy Mini column in a 2 mL collection tube and centrifuged at 8,000 x g for 15 seconds at room temperature, discarding the flow-through afterwards.

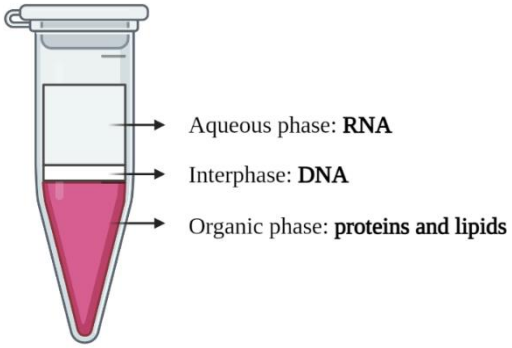


***Figure 7**. Different phases inside the tube after adding QIAzol and chloroform and its content.*

The following process involved several washing steps to remove impurities and excess salts from the column-bound RNA. The composition of the buffers used is undisclosed but requires ethanol for their preparation. Ethanol aids in dissolving and removing residual salts from the RNA pellet. Here are the detailed washing steps.

1. Add 700 µL of Buffer RWT to the column and centrifuge at 8,000 x g for 15 seconds at room temperature. Discard the flow-through.
2. Add 500 µL of Buffer RPE to the column and centrifuge at 8,000 x g for 15 seconds at room temperature. Discard the flow-through.
3. Repeat step 2 by adding another 500 µL of Buffer RPE to the column. Centrifuge at 8,000 x g for 2 minutes at room temperature. Discard the flow-through.

After the washing steps, the column was then transferred to a new 2 mL collection tube, and centrifuged at 16,000 x g for 1 minute at room temperature to dry the membrane. This is a key step to ensure that no residual ethanol is carried over during RNA elution, as it could interfere with downstream reactions.

Finally, the column was transferred to a new collection tube (1.5 mL) for the RNA elution: 30 µL of RNase-free water were added directly onto the column membrane

which was then centrifuged at 16,000 x g for 1 minute at room temperature. The eluted RNA was immediately placed on ice and stored at -80ºC until further use.

3.3. RNA quantification

RNA quantification was carried out through two different methods: the NanoPhotometer® microvolume spectrophotometer (Implen, DEU) and the 2100 Bioanalyzer system (Agilent Technologies Inc., CA, USA).

3.3.1. NanoPhotometer® microvolume spectrophotometer[21]

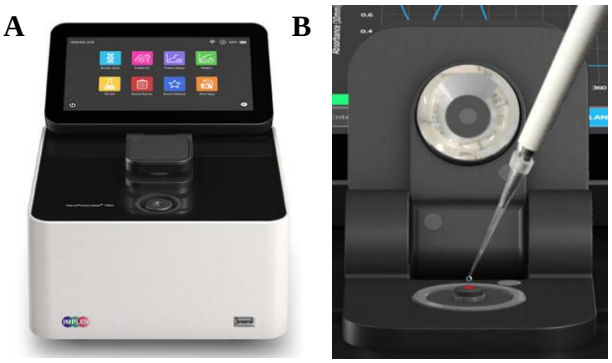


***Figure 8**. (A) NanoPhotometer® N60 and (B) its sample surface.*

For the spectrophotometry assessment, the NanoPhotometer® N60 by the German brand Implen was used (*Figure 8A*). This high-performance micro volume spectrophotometer, amongst other applications, is able to analyse genetic material, as low as 0.3 µL sample volume. For each measurement, the concentration is calculated as well as the 260/230 and the 260/280 purity ratios.

The UV/Vis spectrophotometric analysis is based on a reading at 260 nm, used afterwards to calculate the concentration via the Beer-Lambert law (*Equation 1*). This law empirically relates the absorption of the light to the properties of the sample, where the factors involved are transmittance (*T*), outgoing light (I_0), incoming light (*I*), a molecular extinction coefficient specific for each compound (ε), the concentration of the substance under analysis (*C*) and the distance the light travels (*d*).

$$A = -\log T = -\log(I/I_0) = \varepsilon \cdot C \cdot d \rightarrow C = A/(\varepsilon \cdot d)$$

***Equation 1**. Beer-Lambert law.*

By adding the 230 and 280 nm wavelength to each reading, the 260/230 and 260/280 ratios can be calculated respectively and, therefore, the impurities in the sample can be assessed. For purified samples, the 260/230 ratio should fall within the range of 1.8 and 2.2. A lower coefficient may indicate the presence of residual phenol or guanidine in the solution, while a high ratio suggests an inappropriate solution used for the blank

measurement. On the other hand, the 260/280 ratio evaluates the amount of protein in the sample, being a value close to 2.0 indicative of minimal protein presence in RNA samples. Additionally, an absorbance reading higher than mAbs at 320 nm suggests the presence of undissolved particles or salt residues in the sample.

Operating this device is very simple. To obtain our measurements we started by cleaning the quartz sample surface (*Figure 8B*) with a wipe and selecting the nucleic acids program, particularly the RNA program. Next, 1 µL of the same RNase-free water used for the RNA elution was applied to the quartz surface as the blank solution, followed by the same amount of each sample.

3.3.2. Bioanalyzer system[22,23]

The Agilent 2100 Bioanalyzer is a well-established system for automated electrophoresis that allows for sample quantification and quality control through lab-on-chip capillary gel-electrophoresis. Compared to the traditional gel methods, it requires a significantly less sample and reagent consumption and provides a faster analysis time. The resulting data is provided in a timely manner and delivers objective assessment of sizing, quantitation, integrity and purity.

Maintaining integrity and purity is crucial for reliable experimental results, as RNA samples are susceptible to degradation by RNases. The results of this testing are presented as gel-like images and as electrophoretic data (*Figure 9*), making it easier to detect the possible degradative effects. Some of the indicative features of degradation are the decreasing ratio of ribosomal bands, the additional peaks below the ribosomal bands, the decrease in overall RNA signal and the shift towards shorter fragments.

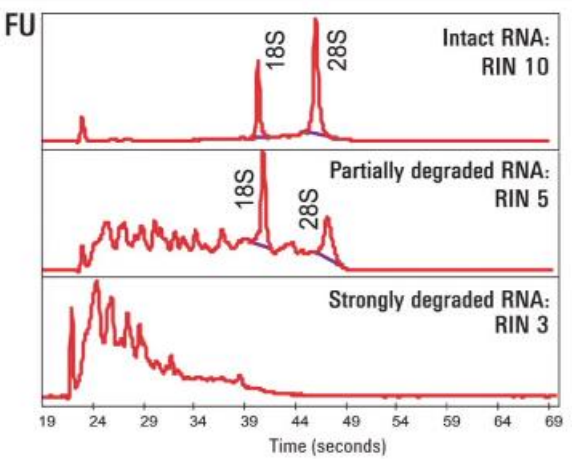


***Figure 9**. Examples of electrophoretic data showing different levels of RNA degradation. FU, fluorescence units. Extracted from: agilent.com.*

Moreover, the Bioanalyzer system provides an RNA Integrity Number or RIN value, an objective metric number determined from the shape of the curve in the electropherogram. The software then classifies the total RNA quality on a numbering

system from 1 to 10, where 1 correlates to a completely degraded RNA and 10 to a highly intact RNA. Thus, interpretation and comparison of the samples is facilitated[24].

To accurately quantify our samples, we used both the Bioanalyzer RNA 6000 Nano (*Figure 10A*) and Pico (*Figure 10B*) assays. This approach allowed us to cover a wider range of concentrations since the Nano kit can detect samples from 5 ng and the Pico kit from as low as 50 pg of total RNA.

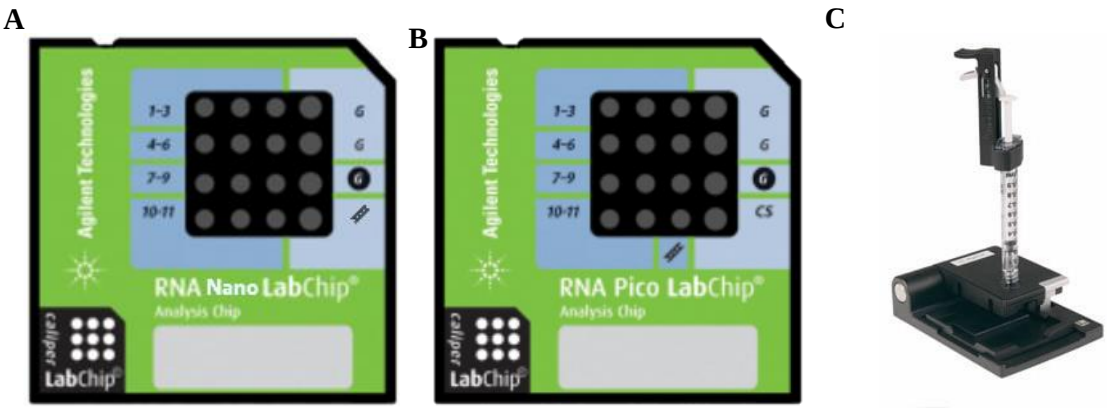


***Figure 10**. Agilent RNA 6000 (A) Nano and (B) Pico LabChip®. (C) Chip priming station.*

The protocol to be followed for both assays is very similar and started by placing a new chip on the priming station (*Figure 10C*) and adding 9 µL of the gel-dye mix previously prepared to the well marked as Ⓖ. The priming station was then closed, and the plunger of the syringe pressed down for the gel-dye mix to be distributed through the different channels of the cell. Once the pressurisation was over, 9 µL of the gel-dye mix was pipetted in each of the wells marked as ***G***.

Afterwards, and only for the Pico kit, 9 µL of the conditioning solution were added into the well market as ***CS***. Even though the composition or the function of this solution remains undisclosed, it is possibly related to stabilise and maintain the integrity of the smallest fragments for them to be detected.

Following the protocol, and again for both kits, 5 µL of the corresponding marker were dispensed into the well marked with the ladder symbol (🪜) and each of the sample wells, twelve for the Nano kit and eleven for the Pico kit. Finally, 1 µL of the ladder and the samples were loaded into the chip, both previously heated to 70ºC for 2 minutes so as to denature the double-helix and minimize the secondary structure. The chip was then vortexed for 60 seconds at 2,400 rpm for all the solutions to properly mix.

At this point, the chip was to be inserted in the Agilent 2100 Bioanalyzer instrument, whose electrodes had been already washed with a first electrode cleaner filled with an RNase decontaminating solution (RNaseZAP) and a second one filled with RNase-free water. Once the lid of the device was carefully closed so the electrodes in the cartridge fit into the wells of the chip, the 2100 Expert Software was initialised, and the appropriate assay selected from the *Assays* menu. Lastly, the chip run was started.

After the run was finished, the chip was discarded while the electrodes had to be washed again with the electrode cleaner filled with RNase-free water to ensure no residues were left from the present assay.

3.4. Testing of lncRNA expression

The reverse transcription quantitative real-time polymerase chain reaction or RT-qPCR is a rapid and sensitive technique used to measure RNA levels and detect changes in gene expression. In this project, our objective was to identify the presence of those lncRNAs that have already been proven to be aberrantly expressed in lung cancer, either by inhibition or overexpression.

To do so, we employed the iTaq™ Universal Probes One-Step Kit (Bio-Rad Laboratories Inc., CA, USA). In one-step reactions, the RNA is first reversely transcribed in a 1:1 reaction to cDNA, a much more stable molecule, and then amplified and detected, in a single run.

The kit includes two essential components for these assays: an iScript™ reverse transcriptase and a reaction mix 2x. The former also contains an RNase H^+ MMLV enzyme engineered to deliver uncompromised sensitivity and true representation of the target RNA level, as well as a potent blend of RNase inhibitors which prevents RNA degradation during the reaction. Otherwise, the reaction mix, optimised for one-step assays, contains an antibody-mediated hot-start Taq DNA polymerase, dNTPs, $MgCl_2$, enhancers, stabilisers and a blend of passive reference dyes, such as ROX.

For our analysis, we employed six probes, each consisting of unlabelled PCR primers and a dual-labelled fluorescent probe. Four of them were designed to detect the lncRNAs known to be altered in lung cancer, namely *HOTAIR*, *MALAT1*, *GAS5* and *MEG3*. The remaining two probes targeted the housekeeping genes *ACTB* and *GAPDH*, which are constitutively expressed in all the samples regardless of their condition[20,25–27].

In order to carry out this assay, we first prepared 6 mixes, one per each gene of study, according to guidelines in the *Table 2*, except for the RNA. Next, the solutions were mixed to ensure their homogeneity and dispensed in 9 µL aliquots into the wells of a PCR plate. Afterwards, 1 µL of the RNA samples, or RNase-free water for the negative controls, was also loaded to the wells. The plate was then sealed and vortexed before being inserted in the Real-Time PCR System Bio-Rad® CFX96™.

***Table 2**. Components of the RT-qPCR and their required volumes. The total volume was calculated for 16 wells: 3 technical replicates for each of the 4 samples (12), 2 negative controls and 2 for spare.*

Component	Volume per 10 µL reaction	Total volume per gene	Final concentration
iTaq Reaction Mix 2X	5 µL	80 µL	1X
iScript Reverse Transcriptase	0.25 µL	4 µL	1X
Fluorogenic probe and primers	0.5 µL	8 µL	250 nM probe 500 nM primers
RNA	1 µL	16 µL	100 ng – 100 fg
RNase-free water	3.25 µL	52 µL	-

Once the plate was set, the thermal cycling protocol was programmed on the RT-PCR instrument (*Figure 11*), and the run started.

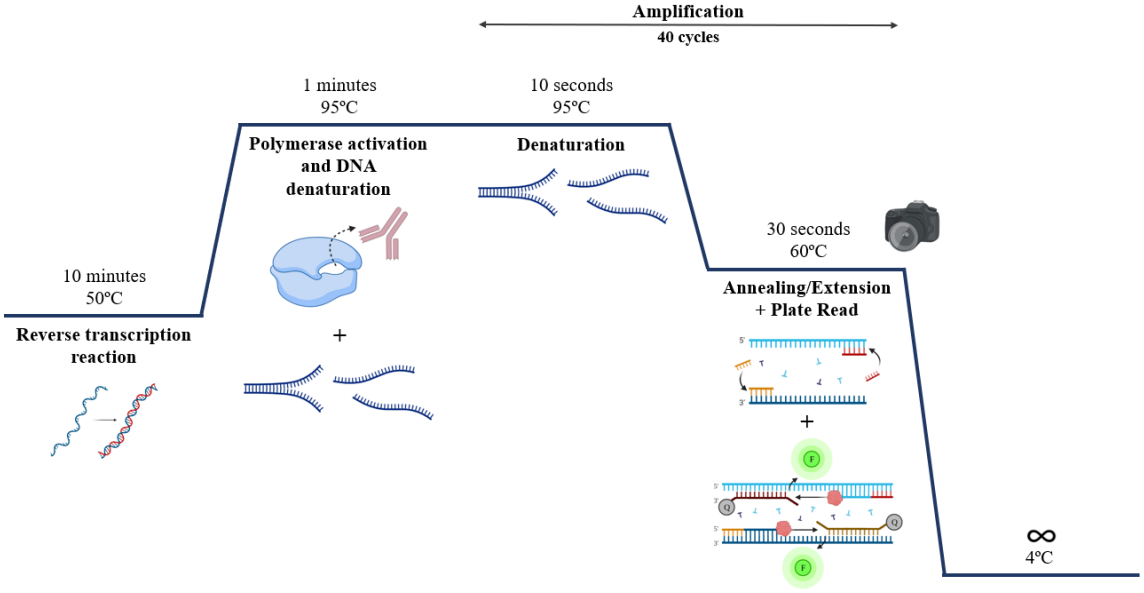


***Figure 11**. Diagram of the thermal cycling protocol on the Real-Time PCR System Bio-Rad® CFX96™.*

The resulting data displayed the C_t (threshold cycle) or C_q (quantification cycle) values, which represent the cycle number at which the fluorescence of the PCR product is detected above the background level. In simpler terms, the C_t value indicates the cycle at which the amplification plot intersects with the threshold, and it is inversely related to

the initial amount of DNA. The lower the C_t value, the more DNA is present at the beginning of the test.

As we did not include control samples in this particular assay, we were unable to utilize the $2^{-\Delta\Delta Ct}$ method to calculate the RNA expression. However, the C_t values for the four lncRNA were normalised using the C_t values of the two housekeeping genes ($2^{\Delta Ct}$). This normalisation allowed for meaningful comparisons between samples. It is important to note that certain criteria were applied when analysing the data, discarding the values when: (I) the C_t value was over 37, (II) there was no expression from the housekeeping genes as they acted as positive controls or, contrarily, (III) there was expression from the negative control wells.

3.5. Preamplification protocol[28–30]

To obtain accurate results from a qPCR, particularly for transcripts present in low concentrations, a preamplification protocol is necessary. This procedure consists of three steps: (I) reverse transcription, (II) preamplification and (III) quantitative PCR.

In order to determine the optimal conditions for this expression level study, we analysed three of the samples obtained from two slightly distinct extraction protocols (depending on the addition or not of an RNA spike-in) and starting from three different starting volumes.

For the reverse transcription step, the StaRT Reverse Transcription kit (AnyGenes®, FRA) was employed. This high-performance system enables the synthesis of a large yield of first-strand cDNA even from low-concentrated and challenging RNA sequences.

***Table 3**. Components of the reverse transcription protocol and their required volumes. The total volume was calculated for 20 reactions: 3 initial volumes (50, 100 and 200 µL) and 2 distinct extraction protocols (with or without spike-in) per each of the 3 samples (18), 1 negative control and 1 for spare.*

Component	Volume per 10 µL reaction	Total volume	Final concentration
StaRT Buffer 10X	2 µL	40 µL	2X
Random Primers 10X	2 µL	40 µL	2X
dNTP mix 10X (40 mM)	2 µL	40 µL	2X (8 mM)
StaRT Reverse Transcriptase 50 U/µL	1 µL	20 µL	5 U/µL
RNase-free water	3 µL	60 µL	-

Firstly, we prepared the Reverse Transcription mix in 1.5 mL tubes, according to *Table 3*. The mix included random primers, which are short oligodeoxyribonucleotides with random base sequences of approximately 6 nucleotides. This primers are ideal for cDNA synthesis as they can bind to any transcript sequence. Afterwards, 10 µL of this mixture was dispensed into the tubes containing 10 µL of the sample RNA, which had been previously incubated for 5 minutes at 65ºC in order to minimise the presence of secondary structures.

Finally, the tubs were sealed and slightly spun before being inserted in the thermal cycler. Once the protocol was programmed on the instrument the run started with a 10-minute incubation at 25ºC to temper the mixture, followed by 120 minutes at 37ºC for the reverse transcription procedure and a final 5-minute incubation at 85ºC to inactivate the enzyme and avoid interferences in downstream reactions. The resulting cDNA product was then stored at -20ºC.

For the second step, the amplification, we used the specific cDNA preamplification SpeAmp^{n} system also from AnyGenes® which includes a Perfect Master Mix and a second reactive component, called AmpPure^{n}. The master mix contains a thermo-stable Taq DNA Polymerase and dNTPs, as well as $MgCl_2$ and buffer for optimal enzyme performance. AmpPure^{n}, on the other hand, contains a pool of enzymes for the purification of the PCR products at the end of the assay. Indeed, a set of specific primers called Prim^{nx} was also necessary.

To proceed with the cDNA amplification, we added the required components to each tube (*Table 4*). After being centrifuged at 1,000 x g for 30 seconds, the tubes were placed in the thermal cycler. The protocol consisted of a 10-minute incubation at 95ºC to activate the enzyme and denature possible dimers, followed by 10 short cycles for the amplification: 10 seconds at 95ºC for denaturation and 30 seconds at 60ºC for annealing and extension.

Table 4*. Components of the preamplification protocol and their required volumes. The total volume was calculated for 21 reactions: 19 samples from the previous step, 1 negative control and 1 for spare.*

Component	Volume per 25 µL reaction	Total volume	Final concentration
Perfect Master Mix 2X	12.5 µL	262.5 µL	1X
Prim^{nx} 0.5 µM	7.5 µL	157.5 µL	0.15 µM
cDNA Template	5 µL	-	-

Once the run was complete, 3 µL of AmpPure^{n} were added to each tube before placing them again in the thermal cycler for the purification step: 30 minutes at 37ºC and 20 minutes at 80ºC. Despite its composition or precise function not being disclosed, it is possibly related to the removal of the adapters usually attached to the primers in this type of reactions. The primers, while necessary for the amplification step, could potentially interfere with subsequent reactions. Again, the final product was stored at -20ºC.

For the third and final step, the Perfect Master Mix SYBR Green Low ROX from AnyGenes® was utilised. This kit is based on a highly efficient Hot-start Taq polymerase that allows a highly precise and robust qPCR results, while preventing non-specific binding and amplification. The detection in this assay was performed through SYBR Green, an asymmetrical cyanine dye that preferentially binds to double-stranded DNA. For this assay, we employed six probes. Three of them aimed to detect lncRNAs already known to be altered in lung cancer, namely *MALAT1*, *NEAT1* and *H19*. The remaining three probes targeted the housekeeping genes *ACTB*, *GAPDH* and *HPRT1*[31–34], which are commonly used as reference genes.

To prepare the qPCR reaction, we created six mixes according to the details provided in *Table 5*, excluding the RNA. The solutions were then thoroughly mixed and dispensed into the wells of a PCR plate, 9 µL at a time. Next, we loaded into the wells 1 µL of RNase-free water for the negative control or 1 µL of the RNA samples previously diluted with 82 µL of the same RNase-free water (volume depending on the PCR plate format). After being sealed, the plate was vortexed 60 seconds at 1,000 x g and inserted in the QuantStudio™ 7 Flex Real-Time PCR System (Thermo Fisher Scientific Inc, MA, USA). Afterwards, the thermal cycling protocol was programmed (*Figure 12*), and the run was started.

***Table 5**. Components of the qPCR and their required volumes. The total volume per gene was calculated for 22 reactions: 20 samples from the previous step, 1 negative control and 1 for spare.*

Component	Volume per 10 µL reaction	Total volume per gene	Final concentration
Perfect Master Mix SYBR Green 2X	5 µL	110 µL	1X
Primer Set 20X (10 µM)	0.5 µL	11 µL	1X (0.5 µM)
RNA	1 µL	22 µL	-
RNase-free water	3.5 µL	77 µL	-

Again, none control samples were included in this assay, therefore the C_t values for the three lncRNA were normalised using the C_t values of the three housekeeping genes ($2^{\Delta Ct}$). As well, the same exclusion criteria as in *section 3.4* were applied when analysing the data.

Otherwise, the specific thermal cycling protocol used in this qPCR allowed for a melting curve analysis to be performed after the main quantification run. By plotting the second derivative of fluorescence against temperature, a single sharp peak will be obtained if only the target product is present in the post-amplified sample. Contrarily, multiple peaks will appear in the presence of side products. These side products can originate from nonspecific product detection, which is sometimes associated with the use of SYBR Green or from the RNA spike-in that was added as a quality control during the extraction procedure for half of the samples analysed in this experiment.

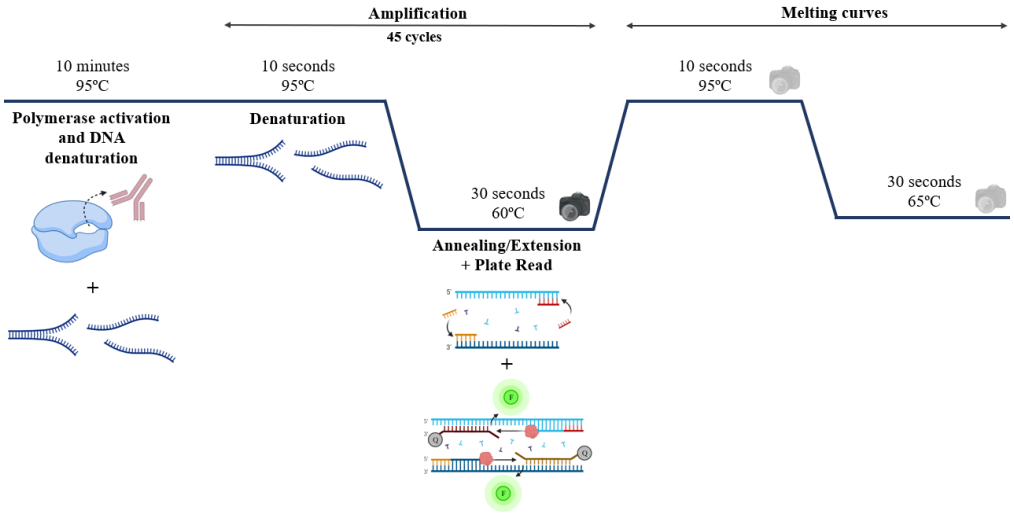


***Figure 12**. Diagram of the thermal cycling protocol on the QuantStudio™ 7 Flex Real-Time PCR System.*

3.6. Decision criteria

The aim of these assays is to determine the optimal experimental conditions for lncRNAs extraction and expression level analysis. This involves considering three key factors: the initial volume of samples, the addition or omission of an RNA spike-in and the need for a preamplification step. Regardless of the purity and integrity of the RNAthe preferred conditions should yield the maximum amount of RNA, resulting in C_q values within a reliable range (below 35 cycles) when analysed. Moreover, the influence of these parameters on the qPCR will be evaluated by assessing their impact on the melting curve analysis.

4. RESULTS AND DISCUSSION

4.1. Establishing the optimal experimental conditions

Malignant pleural effusion is present in 15% of lung cancer patients at the time of diagnosis, indicating an advanced disease and a poor survival[12]. Currently, routine markers such as lactate dehydrogenase, glucose or the carbohydrate antigen 125 (CA-125) are employed to distinguish MPE, even with their poor sensitivity and specificity[35]. Nevertheless, pleural fluid also harbours many tumour-derived products (*Figure 3*), including long non-coding RNAs [15,16].

Recently, there has been a growing interest surrounding lncRNAs, and their potential application as a diagnostic biomarkers in lung cancer. Several differentially expressed transcripts, such as *MALAT1* or *ZBED5-AS1*, have shown promising capabilities in this regard[20,35].

However, thus far, no specific lncRNA signature has been defined to accurately distinguish benign pleural effusion from malignant pleural effusion. This holds true for both individually lncRNA and their potential complementarity to existing diagnostic methods. In order to conduct these types of analysis on large cohorts, it is essential to establish the optimal conditions of the assay. This particular project aims to address this issue.

The first step involved determining the most appropriate initial sample volume to maximize the extraction of genetic material, considering the protocol and the available resources (see *section 3.3*). To achieve this, RNA was isolated from pleural fluid, using three different volumes, and subsequently quantified using two different methods. The resulting findings are shown in *Table 6*.

The concentration results obtained through the NanoPhotometer®, exhibit some irregularities, but it is plausible to perceive a drop in the precision of the instrument as the values are considerably low. To overcome this limitation and achieve higher sensitivity, the Agilent 2100 Bioanalyzer System was later used in order to achieve higher sensitivity.

Considering the results from the Nano assay, it is evident that using an initial volume of 100 µL yields to higher concentrations in two of the samples (523 and 425) compared to

using 200 µL. However, in the case of sample 521, the concentration obtained from 100 µL is slightly lower than that from the 200 µL. It must be noted that the Nano assay is configured to detect samples starting from a concentration of 5 ng/µL, in consequence all the concentrations below this threshold were discarded and the experiment repeated with the Pico assay. This assay was particularly relevant for sample 465, as it demonstrated higher concentration when starting with 200 µL. Regarding the remaining values, the observed discrepancy between the two assays maybe attributable to chip saturation.

***Table 6**. RNA concentrations and purity ratios for the four samples.*

Sample		**NanoPhotometer®**			**Bioanalyzer**	**Bioanalyzer**
ID	Initial volume	Concentration	A260/280	A260/230	**Pico assay**	**Nano assay**
523	50 µL	5.7 ng/µL	1.220	0.224	0.067 ng/µL	2 ng/µL
	100 µL	15.16 ng/µL	1.441	0.409	0.127 ng/µL	226 ng/µL
	200 µL	5.20 ng/µL	1.171	0.156	0.224 ng/µL	25 ng/µL
521	50 µL	15.92 ng/µL	1.502	0.589	0.148 ng/µL	1 ng/µL
	100 µL	8.64 ng/µL	1.479	0.312	0.164 ng/µL	33 ng/µL
	200 µL	8.92 ng/µL	1.343	0.091	0.192 ng/µL	38 ng/µL
465	50 µL	6.92 ng/µL	1.006	0.336	-	2 ng/µL
	100 µL	10.08 ng/µL	1.306	0.221	0.150 ng/µL	2 ng/µL
	200 µL	4.2 ng/µL	1.205	0.150	0.359 ng/µL	2 ng/µL
425	50 µL	16.88 ng/µL	1.486	0.062	0.070 ng/µL	1 ng/µL
	100 µL	7.56 ng/µL	1.454	0.062	0.154 ng/µL	20 ng/µL
	200 µL	4.56 ng/µL	1.000	0.039	0.142 ng/µL	1 ng/µL

On one hand, using the lowest initial volume of 50 µL yielded the expected lowest concentration. On the other hand, starting with 200 µL may lead to a saturation of the RNA extraction column resulting in heterogenous values. Taking all this data into account, it appears that an initial volume of 100 µL is the optimal choice across the three methods and four samples. However, the confirmation of this finding required subsequent PCR tests.

The observed heterogeneity in these results may be attributable to the diverse composition of the pleural fluids, even when originating from the same pathology.

While the initial sample processing protocol, including centrifugation steps, aims to minimise these differences, certain components such as lipids can persist and remain present in the supernatant[10].

After the isolation and quantification of RNA, an accurate examination of the RNA quality is necessary to ascertain the suitability of the samples since the presence of impurities can hinder downstream analysis processes.

For properly purified samples, the expected purity ratios 260/230 and 260/280 are tipically 1.8-2.2 and 2 respectively. Hence, our samples seem to exhibit high protein contamination and the presence of phenol and guanidine (used in the extraction protocol). It should be noted that the minimal concentration of the samples may have affected, the precision of the NanoPhotometer® results.

Alternatively, the Bioanalyzer System Nano kit provides a parameter to assess the integrity and purity of the samples. On one hand, based on the electrophoretic data (*Supplementary Image S1*), all samples present the marker band between 20 and 25 seconds. However, only five of the samples were concentrated enough for another band to be detected, bands which generally appeared displaced regarding the expected point (ribosomal bands). On the other hand, the RIN value was calculated for four of the samples: those with a concentration over 5 ng/µL and a properly loaded marker. The best RIN value obtained was 5.8, indicating that the samples range from partially to strongly degraded. However, this values may also be attributed to the origin of the lncRNA since they have not been isolated from the cell nucleus but have released into the PF freely or encapsulated in extracellular vesicles. We also obtained electrophoretic data from the Pico kit (*Supplementary Image S2*) with very similar results that lead to the same conclusion.

Even though the quality of the samples was proven not to be optimal and despite their poor concentration, we then proceeded to detect the presence of four lncRNA previously proven to present an altered expression in lung cancer: *HOTAIR*, *MALAT1*, *MEG3* and *GAS5*.

The method employed for the detection was the reverse transcription quantitative real-time polymerase chain reaction or RT-qPCR, which resulted in the graphical representation of the fluorescence increase for the different samples (*Figure 13*). For

this experiment, only the samples obtained from the 100 µL of initial volume were analysed.

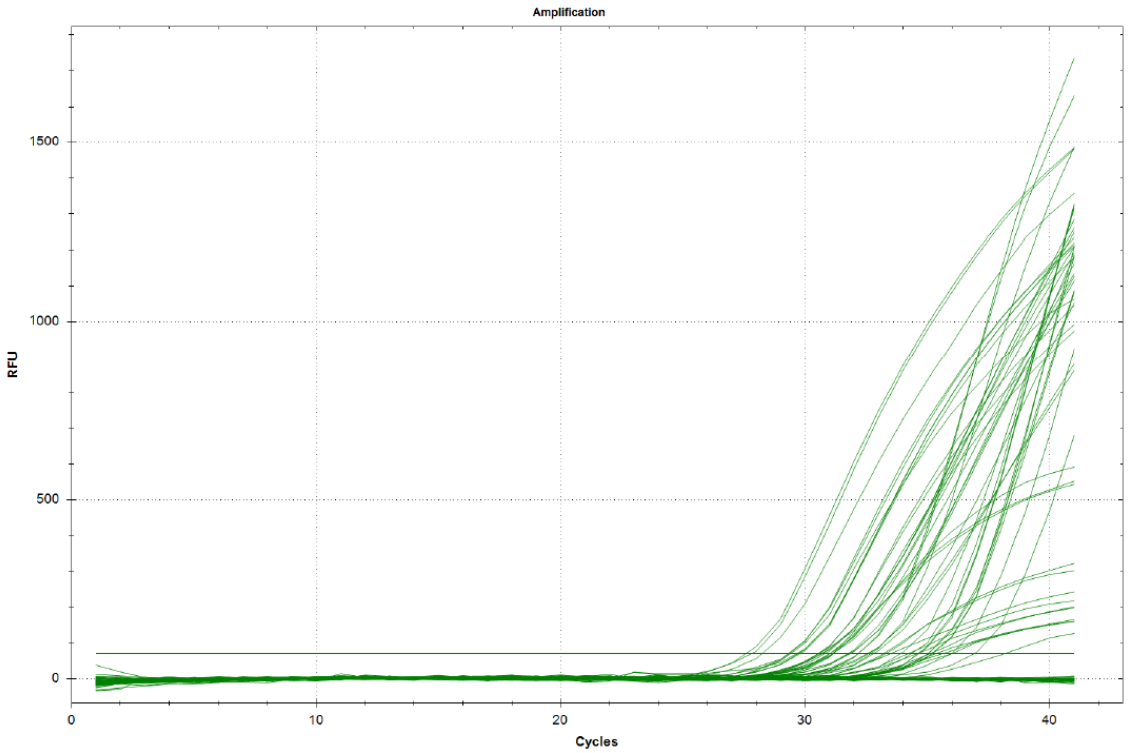


***Figure 13**. Results obtained from the RT-qPCR. RFU, relative fluorescence units.*

The results obtained, including both graphical and numerical (*Supplementary Table S1*) representation highlight the challenge posed by the low concentration of our samples. The lowest C_q value obtained was approximately 27, indicating the limited abundance of the target transcripts. As a result, two of the transcripts, *HOTAIR* and *MEG3*, were not even detected despite *HOTAIR* typically being overexpressed in lung cancer[36]. Moreover, the expression values ($2^{\Delta Ct}$) of *MALAT1* and *GAS5* exhibited notable variation among the four samples, all of which were from the same subtype of lung cancer.

To address these challenges, both the minimal RNA concentration and the great variability between the samples, two different strategies were developed. Otherwise, this range of RNA quantity would remain under the detection limit of the most commonly used techniques for measuring RNA, making it rather complex to normalise the starting quantity of the transcripts.

Firstly, to ensure quality control of the RNA isolation and assess the efficiency of cDNA synthesis and amplification a synthetic RNA, cel-miR-39 (Qiagen, NLD), was

employed as a spike-in during the extraction procedure. The cel-miR-39 was added immediately after the lysis reagent (QIAzol), and the amount recovered after the extraction directly correlates with the total RNA recovered. The expression level of this spike-in transcript can be later determined by subjecting the cDNA to qPCR. By employing this approach, the level of expression of any target transcripts can be normalised to the cel-miR-39 through the $2^{-\Delta\Delta Ct}$ method.

The second strategy implemented to enhance the qPCR results involved a preamplification step. Thus, enabling the detection of those transcripts present in the samples but in extremely low concentrations. The results obtained from this assay (*Supplementary Table S2*) facilitate the comparison of how the different experimental conditions impact the analysis of expression levels.

Regarding the initial volume, C_q values evidence that starting with 200 µL yields the highest RNA quantity, despite the initial quantification values suggesting otherwise. This observation is exemplified by the C_q values obtained for the *GAPDH* gene from the sample 523 (*Figure 14*).

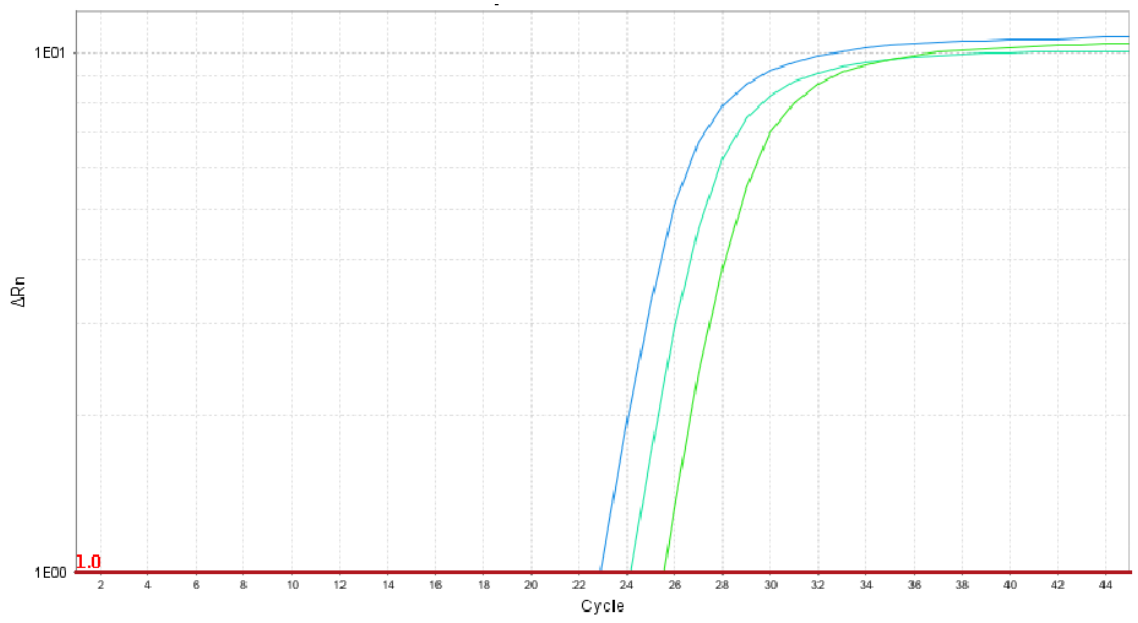


***Figure 14**. Graphical results displaying the C_q values obtained for the* GAPDH *gene from the 523 sample when adding cel-miR-39 during the extraction process. The initial volumes represented are 50 µL (green), 100 µL (cyan) and 200 µL (blue). Rn, normalized reporter value.*

Otherwise, the addition of cel-miR-39 has shown a significant a positive effect on the detection of both lncRNAs and housekeeping genes. When observing *Figure 15A*, two well-defined blocks exist, being the samples without the RNA spike-in in the group

with the higher C_q values. Similarly, in *Figure 15B*, three distinct blocks can be distinguished. Once again, the group from the far-right contains samples without the spike-in while the signal from the far-left corresponds to a sample obtained from 200 µL where cel-miR-39 was added during the extraction process.

Furthermore, the effects of cel-miR-39 are particularly noteworthy when considering the values resulting from the *MALAT1* analysis. Out of the nine samples without the spike-in, six were not detected whereas a signal was obtained for all the samples that included the synthetic miRNA.

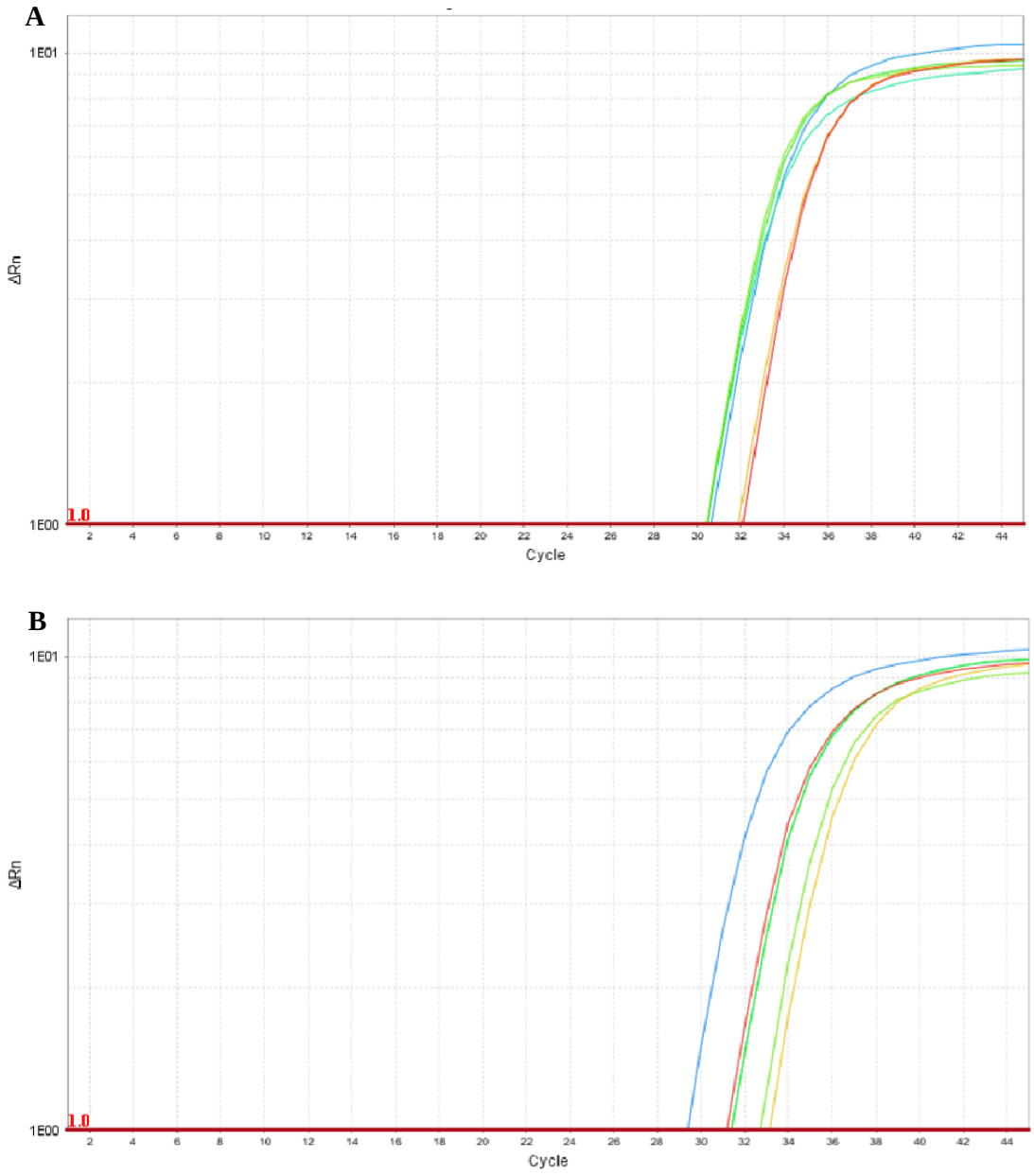


***Figure 15**. Graphical results displaying the C_q values obtained for the (A)* NEAT1 *and (B)* HPRT1 *genes from the 523 sample when adding or not cel-miR-39 during the extraction process. Rn, normalized reporter value.*

When evaluating the impact of the preamplification process, it can only be assessed by comparing those genes analysed in both qPCR experiments: *MALAT1*, *ACTB* and *GAPDH*. However, there may have been an issue with the *MALAT1* primers since, as stated before, only a portion of the samples were detected, thus the C_q values obtained for this gene will not be considered in this discussion. Regarding the C_q values for the two housekeeping genes, they have become significantly lower in both cases, displaying the positive effect of the preamplification: the results from all studied genes fall below 30 cycles (except for *MALAT1*).

Observing the melting curves obtained after the qPCR, the vast majority of them exhibit the same pattern as shown in *Figure 16A*. This single peak demonstrates the high specificity of the employed probes, proving that only the target product was amplified. However, for *NEAT1* all the plots displayed a second peak between 75-80ºC (*Figure 16B*), meaning that the probes employed to detect this transcript also bind non-specifically to side products, albeit with a lower affinity.

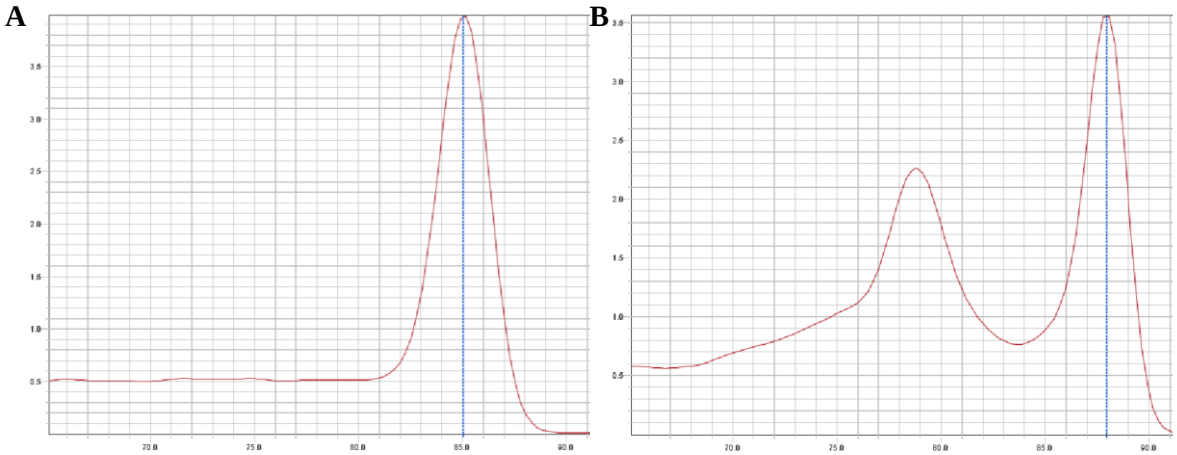


***Figure 16**. Graphical results displaying the Melt Curve Plot obtained for the (A)* GAPDH *and (B)* NEAT1 *genes from the 523 sample when omitting the cel-miR-39. The X axis represents the temperature (ºC) and the Y axis the second derivative of the normalized reporter value. The two main peaks correspond to (A) 85.04ºC and (B) 87.93ºC.*

The final aspect to evaluate in these assays is the suitability of the three housekeeping genes for the data normalization. In this regard, both *ACTB* and *GAPDH* exhibited a consistent expression among all samples (*Supplementary Table S1*). On the other hand, this second qPCR where the preamplification protocol was tested did not include technical replicates, thus the reliability of *HPRT1* could not be assessed, even though it is confirmed to be the most stable gene for expression studies on non-small cell lung cancer by Gresner *et al.* (2009)[34].

To sum up, considering all the data obtained from the various assays, it can be concluded that using an initial volume of 200 μL, incorporating cel-miR-39 as a spike-in, and implementing the preamplification step yield the most desirable results. Moreover, *ACTB* and *GAPDH* have been demonstrated to be suitable for data normalization in this expression study.

4.2. Future perspectives

Once these optimal conditions have been established, the next step to be taken in the project where my study has been placed would be to select the method through which the levels of expression will be tested. A practical and cost-effective option is the use of gene expression assays in array format. These qPCR arrays are highly suitable for expression profiling allowing for the analysis of tens to thousands of targets.

Recent bioinformatic analysis have identified more than 1,500 differentially expressed lncRNAs specific to lung adenocarcinoma[37], which is the most common subtype of lung cancer. Therefore, understanding the role of these transcripts in carcinogenesis is of utmost importance in order to accurately select lncRNA species with potential application as specific biomarkers of malignancy in patients with malignant pleural effusion secondary to lung cancer. Hereunder, the physio-pathological characteristics of those lncRNAs analysed in this project will be discussed.

GAS5 is a non-protein coding gene located in chromosome 1 and composed of 12 exons. Several are the biological processes regulated by this transcript through three different mechanisms of action. Firstly, *GAS5* can act as a decoy for glucocorticoid receptors (GR) thus suppressing the expression of GR-regulated genes[18,38]. It can also regulate gene expression via histone methylation and demethylation, primarily through its small RNA derivatives such as piRNAs[38]. Lastly, *GAS5* functions as a miRNA sponge, regulating signalling pathways and biological functions. For example, by inhibiting miR-103, it enhances the expression of *PTEN* leading to growth arrest and apoptosis[18,38].

In the context of cancer, *GAS5* serves as a tumour suppressor. In lung cancer, this transcript not only enhances apoptosis, but has also been shown to significantly inhibit cell migration and invasion through the EMT (Epithelial-Mesenchymal Transition) process[39].

HOTAIR is an approximately 2,200 nucleotide-long transcript that undergoes splicing and polyadenylation, resulting in a mature form with 6 exons[17]. In the context of cancer, *HOTAIR* can promote different oncogenic processes, from tumour growth to EMT via distinct pathways depending on the type of cancer. Particularly, in lung cancer *HOTAIR* is capable to promote dedifferentiation and proliferation by down-regulating mainly two genes: *HOXA5* which is pivotal to morphogenesis of the embryonic respiratory tract and $p21^{WAF1/CIP1}$, a mediator of growth arrest and apoptosis induced by p53 in response to DNA damage[40]. Additionally, *HOTAIR* promotes the EMT, hence mediating cell invasiveness, by: (I) repressing the expression genes related to cell-adhesion, (II) repressing EMT inhibitors such as *PTEN* or *WIF-1* (inhibitor of the Wnt/β-catenin pathway) or (III) promoting the expression of EMT effectors, including metalloproteases involved in extracellular matrix degradation[40].

The mechanisms underlying *HOTAIR*'s ability to regulate gene silencing involve its interactions with PRC2 and LSD1, acting as a bridging scaffold for these histone-modifying complexes[17,40]. The former catalyses the silencing histone mark H3K27 via tri-methylation due to the enzymatic activity of EZH2[17,41]; while the latter, a histone H3K4 demethylase also interacts with the repressor complex CoREST/REST to demethylate this active histone mark[17,40]. Moreover, while the involvement of *HOTAIR* in cross-talk with miRNAs has been demonstrated in various cancer, further investigation is still needed to understand its impact in the context of lung cancer[40].

H19 is a maternally imprinted gene located in chromosome 11. In the context of lung cancer, this transcript is able to regulate cell proliferation by, again, recruiting PRC2 and repressing the expression of *PTEN* through methylation[42]. Studies in human NSCLC cell lines also suggest that *H19* plays a role in the process of metastasis by modulating cell adhesion proteins such as E-cadherin or N-cadherin, which are crucial for tumour cell migration and invasion[42].

Moreover, *H19* has been identified as a relevant miRNA regulator via sequestration. Among the different examples, its interaction with miR-200a strands out since it results in the de-repression of *ZEB1* and *ZEB2*, two transcription factors related to enhancing lung cancer proliferation and metastasis[42,43]. Alternatively, *H19* can also induce miR-675-5p expression which, in turn, upregulates *BCL2*, leading to the decreased

expression of *TP53* and *BAX*. This molecular cascade contributes to evasion of apoptosis, a hallmark of cancer cells[42].

MALAT1 is gene also located in chromosome 11 which gives rise to 17 transcripts, all identified as lncRNAs. This versatile gene has been shown to regulate several cancer hallmarks. Firstly, it can modulate growth and invasion by sponging miR-26a/b and promoting fucosylation through FUT4, consequently activating PI3K/AKT/mTOR signalling pathway[42,44]. Similarly, *MALAT1* induces the expression of *NEDD9* which promotes EMT and metastasis and activates angiogenesis factors like VEGF, TWIST and SLUG by sequestering miR-145-5p and miR-126-5p respectively[42,44].

Despite *MALAT1* is generally considered an oncogene and its expression being correlated with poor overall survival of patients with malignancies such as NSCLC, recent studies have suggested its possible role as a tumour suppressor. In breast cancer, for example, *MALAT1* sequestered and inhibited TEAD, a transcription factor involved in promoting metastasis. For certain lung cancer patients, a reduced expression of this transcript was detected in blood samples, maybe resulting from the influence of mRNA profiles of the immune cells present in blood[44].

MEG3 is located in the chromosome 14 and, again, maternally imprinted. Physiologically is expressed in many tissues and a major tumour suppressor. This transcript exerts its tumour-suppressive functions through various mechanisms. On the one hand, this transcript promotes apoptosis by (I) decreasing the levels of MDM2, thus activating TP53, and (II) enhancing *BAX* expression and, consequently, decreasing that of *BCL2*. In studies using NSCLC cell lines, *MEG3* has been proven to contribute to the recruitment of PRC2's EZH2 which, in turn, will lead to the repression of several genes such as *ZEB1* and *ZEB2*[42].

In addition to its role as a tumour suppressor, *MEG3* is reported to be an important miRNA regulator. For example, it positively regulates the post-transcriptional expression of *SLC34A2*, pH-sensitive sodium-dependent phosphate transporter, by sponging miR-650. Through this *MEG3/miR-650/SLC34A2* axis *MEG3* may play a role in regulating the stem cell-like state, as well as cell migration and invasion[42,45].

NEAT1 is a single exon, intergenic lncRNA located on the chromosome 11, usually expressed in similar levels in most healthy tissues. However, in many solid tumours

including lung cancer, *NEAT1* is upregulated and correlated with aggressive disease and poor patient outcomes[46]. In lung cancer, *NEAT1* is overexpressed in around 90% of the cases and, studies in cell lines, have determined its role in promoting growth, migration and invasion[18,46].

The pathological mechanisms of *NEAT1* in lung cancer are multifaceted. On the one hand, one of the pathological mechanisms of *NEAT1* in lung cancer is its sequestration of miR-98-5p which directly targets MAPK6, a protein related to proliferation and invasion processes[46]. On the other hand, it has recently been established that *NEAT1* can act as a guide for DNMT1, hence epigenetically inhibiting the expression of *TP53*, *STING* and *cGAS*. *TP53* is a well-known tumour suppressor gene involved in proliferation and migration, and its suppression facilitates the survival and invasion of lung cancer cells. Otherwise, the inhibition of the cGAS/STING signalling pathway mediates immune evasion of tumour cells from T cells[47].

Collectively, the available evidence strongly supports a correlation between expression on lncRNA and the promotion and maintenance of various cancer hallmarks. Promising strides have already been made in using lncRNAs as reliable markers for cancer diagnosis and prognosis. Ongoing studies, as the present project, will continue to explore and exploit their potential in this regard.

Moreover, the emergence of liquid biopsy as a highly effective diagnostic method compared to conventional approaches[48], opened up new possibilities. By combining the detection and analysis of lncRNAs and liquid biopsy techniques, such as the analysis of malignant pleural effusions, an optimistic and non-invasive diagnostic toll can be developed. This approach holds great promise for improved diagnosis and monitoring of cancer in a minimally invasive manner [20,35,48].

5. CONCLUSION

In this project, we have conducted a comprehensive evaluation of various experimental factors to determine the optimal condition for extracting and analysing long noncoding RNAs (lncRNAs) from pleural fluid samples. Specifically, we assessed the impact of three different initial sample volumes, the inclusion of an RNA spike-in (cel-miR-39) during the extraction process and the implementation of a preamplification step.

According to the results obtained in this project, we have identified the following optimal experimental conditions:

1. The recommended initial volume for pleural fluid samples is 200 μL.
2. It is advisable to include an RNA spike-in (cel-miR-39) during the extraction procedure.
3. A preamplification step is required before performing the qPCR analysis.

Furthermore, we have successfully validated the suitability of both *ACTB* and *GAPDH* as reference genes for data normalisation in pleural fluid samples secondary to lung cancer.

Once the appropriate methodology has been established and considering the active role of lncRNAs in promoting and suppressing oncogenic functions, the analysis of those lncRNA species present in the pleural fluid holds significant potential. It is conceivable that through further research and development, the identification and characterisation of specific lncRNAs in pleural fluid could pave the way for the creation of a robust algorithm for the diagnosis and prognosis of patients with lung cancer and associated malignant pleural effusion.

BIBLIOGRAPHY

1. Cancer Today. https://gco.iarc.fr/today/online-analysis-pie?v=2020&mode=cancer&mode_population=continents&population=900&populations=900&key=total&sex=0&cancer=39&type=0&statistic=5&prevalence=0&population_group=0&ages_group%5B%5D=0&ages_group%5B%5D=17&nb_items=7&group_cancer=1&include_nmsc=1&include_nmsc_other=1&half_pie=0&donut=0.
2. Amorín Kajatt, E. & Peru Med Exp Salud Publica Revisión, R. *CÁNCER DE PULMÓN, UNA REVISIÓN SOBRE EL CONOCIMIENTO ACTUAL, MÉTODOS DIAGNÓSTICOS Y PERSPECTIVAS TERAPÉUTICAS LUNG CANCER: A REVIEW OF CURRENT KNOWLEDGE, DIAGNOSTIC METHODS AND THERAPEUTIC PERSPECTIVES*.
3. Dela Cruz, C. S., Tanoue, L. T. & Matthay, R. A. Lung Cancer: Epidemiology, Etiology, and Prevention. *Clinics in Chest Medicine* vol. 32 605–644 Preprint at https://doi.org/10.1016/j.ccm.2011.09.001 (2011).
4. Oliver, A. L. Lung Cancer: Epidemiology and Screening. *Surgical Clinics of North America* vol. 102 335–344 Preprint at https://doi.org/10.1016/j.suc.2021.12.001 (2022).
5. Inamura, K. Clinicopathological characteristics and mutations driving development of early lung adenocarcinoma: Tumor initiation and progression. *International Journal of Molecular Sciences* vol. 19 Preprint at https://doi.org/10.3390/ijms19041259 (2018).
6. Rodriguez-Canales, J., Parra-Cuentas, E. & Wistuba, I. I. Diagnosis and molecular classification of lung cancer. in *Cancer Treatment and Research* vol. 170 25–46 (Kluwer Academic Publishers, 2016).
7. Succony, L., Rassl, D. M., Barker, A. P., McCaughan, F. M. & Rintoul, R. C. Adenocarcinoma spectrum lesions of the lung: Detection, pathology and treatment strategies. *Cancer Treatment Reviews* vol. 99 Preprint at https://doi.org/10.1016/j.ctrv.2021.102237 (2021).
8. Cáncer de Pulmón: Todo lo que Necesitas Saber | Asociación Española Contra el Cáncer. https://www.contraelcancer.es/es/todo-sobre-cancer/tipos-cancer/cancer-pulmon.
9. Gayen, S. Malignant Pleural Effusion: Presentation, Diagnosis, and Management. *American Journal of Medicine* **135**, 1188–1192 (2022).

10. Porcel Pérez, J. M. ABC del líquido pleural. *Seminarios de la Fundacion Espanola de Reumatologia* vol. 11 77–82 Preprint at https://doi.org/10.1016/j.semreu.2010.02.003 (2010).
11. Porcel, J. M., Esquerda, A., Vives, M. & Bielsa, S. Etiología del derrame pleural: Análisis de más de 3.000 toracocentesis consecutivas. *Arch Bronconeumol* **50**, 161–165 (2014).
12. Aydin, Y., Turkyilmaz, A., Intepe, Y. S. & Eroglu, A. Malignant Pleural Effusions: Appropriate Treatment Approaches. *Eurasian J Med* **41**, 186 (2009).
13. Lung and Bronchus Cancer — Cancer Stat Facts. https://seer.cancer.gov/statfacts/html/lungb.html.
14. Nooreldeen, R. & Bach, H. Current and future development in lung cancer diagnosis. *International Journal of Molecular Sciences* vol. 22 Preprint at https://doi.org/10.3390/ijms22168661 (2021).
15. Sorolla, M. A., Sorolla, A., Parisi, E., Salud, A. & Porcel, J. M. Diving into the pleural fluid: Liquid biopsy for metastatic malignant pleural effusions. *Cancers* vol. 13 Preprint at https://doi.org/10.3390/cancers13112798 (2021).
16. Li, W. *et al.* Liquid biopsy in lung cancer: significance in diagnostics, prediction, and treatment monitoring. *Molecular Cancer* vol. 21 Preprint at https://doi.org/10.1186/s12943-022-01505-z (2022).
17. Balas, M. M. & Johnson, A. M. Exploring the mechanisms behind long noncoding RNAs and cancer. *Non-coding RNA Research* vol. 3 108–117 Preprint at https://doi.org/10.1016/j.ncrna.2018.03.001 (2018).
18. Bhan, A., Soleimani, M. & Mandal, S. S. Long noncoding RNA and cancer: A new paradigm. *Cancer Research* vol. 77 3965–3981 Preprint at https://doi.org/10.1158/0008-5472.CAN-16-2634 (2017).
19. Jiang, L., Li, Z. & Wang, R. Long non-coding rnas in lung cancer: Regulation patterns, biologic function and diagnosis implications (review). *Int J Oncol* **55**, 585–596 (2019).
20. Wang, W. W. *et al.* Combination of long noncoding RNA MALAT 1 and carcinoembryonic antigen for the diagnosis of malignant pleural effusion caused by lung cancer. *Onco Targets Ther* **11**, 2333–2344 (2018).
21. Implen NanoPhotometer N60/N50 | UV-Vis Spectrophotometer | #1 Alternative. https://www.implen.de/product-page/implen-nanophotometer-n60-microvolume-spectrophotometer/.

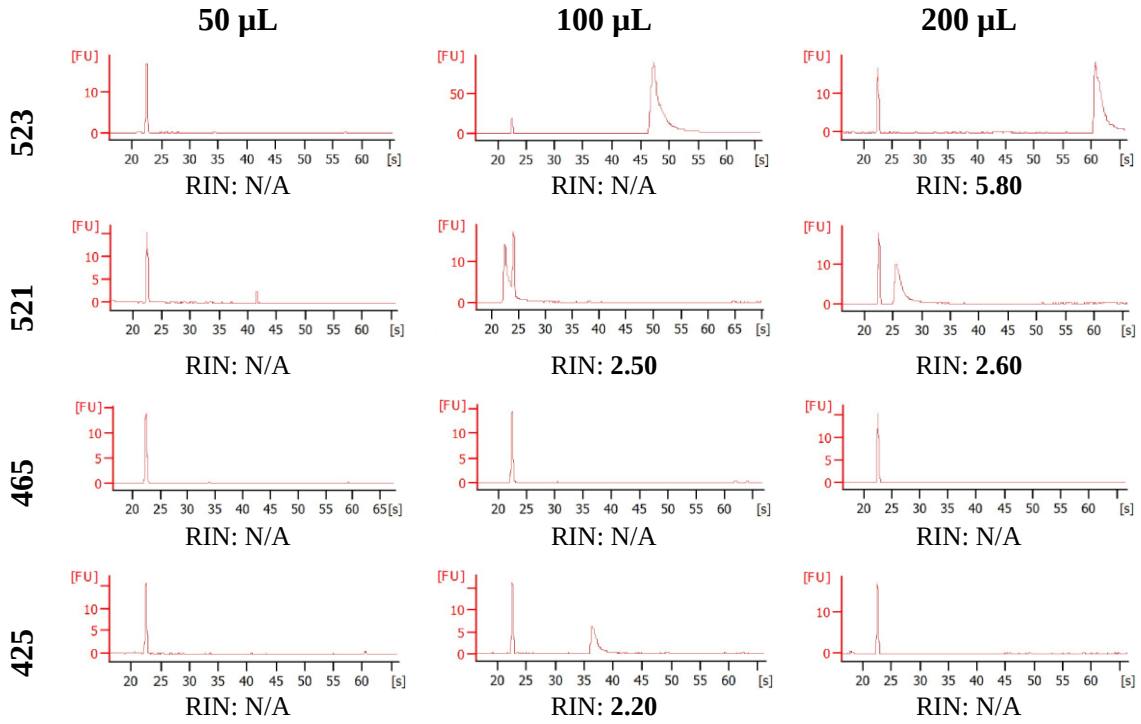
22. High Sensitivity RNA Electrophoresis, RIN, Bioanalyzer Pico | Agilent. https://www.agilent.com/en/product/automated-electrophoresis/bioanalyzer-systems/bioanalyzer-rna-kits-reagents/bioanalyzer-high-sensitivity-rna-analysis-228255.
23. RNA Electrophoresis, RNA Integrity, Bioanalyzer RIN | Agilent. https://www.agilent.com/en/product/automated-electrophoresis/bioanalyzer-systems/bioanalyzer-rna-kits-reagents/bioanalyzer-rna-analysis-228256.
24. Fleige, S. & Pfaffl, M. W. RNA integrity and the effect on the real-time qRT-PCR performance. *Molecular Aspects of Medicine* vol. 27 126–139 Preprint at https://doi.org/10.1016/j.mam.2005.12.003 (2006).
25. Iizuka, T. *et al.* The combined use of long non-coding RNA HOTAIR and polycomb group protein EZH2 as a prognostic marker of lung adenocarcinoma. *Cancer Treat Res Commun* **31**, (2022).
26. Zhu, L. *et al.* LncRNA GAS5 inhibits invasion and migration of lung cancer through influencing EMT process. *J Cancer* **12**, 3291–3298 (2021).
27. Zhao, Y., Zhu, Z., Shi, S., Wang, J. & Li, N. Long non-coding RNA MEG3 regulates migration and invasion of lung cancer stem cells via miR-650/SLC34A2 axis. *Biomedicine and Pharmacotherapy* **120**, (2019).
28. StaRT Reverse Transcription kits for high quality cDNA - AnyGenes. https://www.anygenes.com/home/products/start-reverse-transcritpion-kits/.
29. Preamplification reagents to push the qPCR limits - AnyGenes. https://www.anygenes.com/home/products/single-cell-profiling/preamplification-reagents/.
30. Perfect Master Mix SYBR Green for best PCR results ! - AnyGenes. https://www.anygenes.com/home/products/qpcr-master-mixes/qpcr-mix-sybr-green/.
31. Goyal, B. *et al.* Diagnostic, prognostic, and therapeutic significance of long non-coding RNA MALAT1 in cancer. *Biochimica et Biophysica Acta - Reviews on Cancer* vol. 1875 Preprint at https://doi.org/10.1016/j.bbcan.2021.188502 (2021).
32. Ma, F. *et al.* LncRNA NEAT1 Interacted With DNMT1 to Regulate Malignant Phenotype of Cancer Cell and Cytotoxic T Cell Infiltration via Epigenetic Inhibition of p53, cGAS, and STING in Lung Cancer. *Front Genet* **11**, (2020).

33. Zhao, Y., Feng, C., Li, Y., Ma, Y. & Cai, R. LncRNA H19 promotes lung cancer proliferation and metastasis by inhibiting miR-200a function. *Mol Cell Biochem* **460**, 1–8 (2019).
34. Gresner, P., Gromadzinska, J. & Wasowicz, W. *Reference genes for gene expression studies on non-small cell lung cancer*. www.actabp.pl (2009).
35. Huang, X. *et al.* Construction and analysis of expression profile of exosomal lncRNAs in pleural effusion in lung adenocarcinoma. *J Clin Lab Anal* **36**, (2022).
36. Loewen, G., Jayawickramarajah, J., Zhuo, Y. & Shan, B. Functions of lncRNA HOTAIR in lung cancer. *Journal of Hematology and Oncology* vol. 7 Preprint at https://doi.org/10.1186/s13045-014-0090-4 (2014).
37. Wu, X., Sui, Z., Zhang, H., Wang, Y. & Yu, Z. Integrated Analysis of lncRNA–Mediated ceRNA Network in Lung Adenocarcinoma. *Front Oncol* **10**, (2020).
38. Ji, J., Dai, X., Yeung, S. C. J. & He, X. The role of long non-coding RNA GAS5 in cancers. *Cancer Management and Research* vol. 11 2729–2737 Preprint at https://doi.org/10.2147/CMAR.S189052 (2019).
39. Zhu, L. *et al.* LncRNA GAS5 inhibits invasion and migration of lung cancer through influencing EMT process. *J Cancer* **12**, 3291–3298 (2021).
40. Loewen, G., Jayawickramarajah, J., Zhuo, Y. & Shan, B. Functions of lncRNA HOTAIR in lung cancer. *Journal of Hematology and Oncology* vol. 7 Preprint at https://doi.org/10.1186/s13045-014-0090-4 (2014).
41. Iizuka, T. *et al.* The combined use of long non-coding RNA HOTAIR and polycomb group protein EZH2 as a prognostic marker of lung adenocarcinoma. *Cancer Treat Res Commun* **31**, (2022).
42. Poulet, C. *et al.* Molecular Sciences Exosomal Long Non-Coding RNAs in Lung Diseases. doi:10.3390/ijms21100000.
43. Zhao, Y., Feng, C., Li, Y., Ma, Y. & Cai, R. LncRNA H19 promotes lung cancer proliferation and metastasis by inhibiting miR-200a function. *Mol Cell Biochem* **460**, 1–8 (2019).
44. Goyal, B. *et al.* Diagnostic, prognostic, and therapeutic significance of long non-coding RNA MALAT1 in cancer. *Biochimica et Biophysica Acta - Reviews on Cancer* vol. 1875 Preprint at https://doi.org/10.1016/j.bbcan.2021.188502 (2021).
45. Zhao, Y., Zhu, Z., Shi, S., Wang, J. & Li, N. Long non-coding RNA MEG3 regulates migration and invasion of lung cancer stem cells via miR-650/SLC34A2 axis. *Biomedicine and Pharmacotherapy* **120**, (2019).

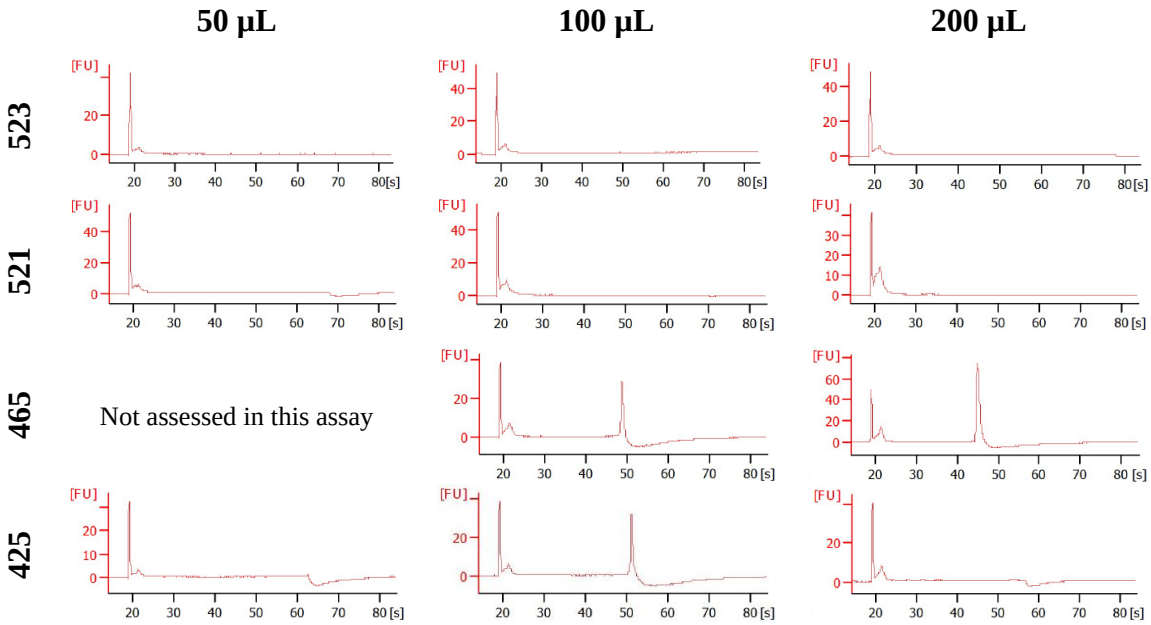
46. Smith, N. E., Spencer-Merris, P., Fox, A. H., Petersen, J. & Michael, M. Z. The Long and the Short of It: NEAT1 and Cancer Cell Metabolism. *Cancers* vol. 14 Preprint at https://doi.org/10.3390/cancers14184388 (2022).
47. Ma, F. *et al.* LncRNA NEAT1 Interacted With DNMT1 to Regulate Malignant Phenotype of Cancer Cell and Cytotoxic T Cell Infiltration via Epigenetic Inhibition of p53, cGAS, and STING in Lung Cancer. *Front Genet* **11**, (2020).
48. Shin, S., Kim, J., Kim, Y., Cho, S.-M. & Lee, K.-A. Assessment of real-time PCR method for detection of EGFR mutation using both supernatant and cell pellet of malignant pleural effusion samples from non-small-cell lung cancer patients. *Clinical Chemistry and Laboratory Medicine (CCLM)* **55**, (2017).

ANNEXES

Supplementary images



***Supplementary Figure S1**. Electrophoretic data and RIN values obtained via the Bioanalyzer 2100 Nano assay.*



***Supplementary Figure S2.** Electrophoretic data obtained via the Bioanalyzer 2100 Pico assay.*

Supplementary tables

Supplementary Table S1. *Results obtained through the RT-qPCR of the isolated RNA. Some of the values were discarded to maintain the standard deviation under 0.5.*

Gene	Sample	C_q	C_q Mean	Standard Deviation	ΔCt ACTB	ΔCt GAPDH	$2^{\Delta Ct}$ ACTB	$2^{\Delta Ct}$ GAPDH
MALAT1	523	35.01 35.17 ~~35.97~~	35.09	0.11	2.48	4.38	5.57	20.77
	521	32.66 32.16 32.14	32.32	0.29	1.63	4.47	3.09	22.16
	465	35.02 35.16 34.62	34.93	0.28	3.16	5.24	8.94	37.88
	425	34.59 34.31 33.88	34.26	0.36	3.39	4.93	10.46	30.55
HOTAIR	523	N/A N/A N/A	~	~	~	~	~	~
	521	N/A N/A N/A	~	~	~	~	~	~
	465	N/A N/A N/A	~	~	~	~	~	~
	425	N/A N/A N/A	~	~	~	~	~	~
MEG3	523	N/A N/A N/A	~	~	~	~	~	~
	521	N/A N/A N/A	~	~	~	~	~	~
	465	N/A N/A N/A	~	~	~	~	~	~
	425	N/A N/A N/A	~	~	~	~	~	~
GAS5	523	34.55 34.19 ~~35.31~~	34.37	0.25	1.76	3.66	3.38	12.61

	521	33.18 33.47 33.28	33.31	0.15	2.62	5.46	6.13	44.02
	465	35.08 35.98 35.59	35.55	0.45	3.78	5.86	13.71	58.08
	425	31.14 30.98 31.12	31.08	0.09	0.21	1.75	1.15	3.37
ACTB	523	32.51 32.56 32.77	32.61	0.14				
	521	30.63 30.78 30.67	30.69	0.08				
	465	31.72 31.82 31.78	31.77	0.05				
	425	30.52 31 31.1	30.87	0.31				
GAPDH	523	30.52 30.5 31.12	30.71	0.35				
	521	27.78 27.55 28.22	27.85	0.34				
	465	29.69 29.76 29.62	29.69	0.07				
	425	29.32 29.38 29.28	29.33	0.05				
MALAT1	NTC	N/A N/A						
HOTAIR		N/A N/A						
MEG3		N/A N/A						
GAS5		N/A N/A						
ACTB		N/A N/A						
GAPDH		N/A						

***Supplementary Table S2**. Results obtained through the qPCR of the isolated RNA after the preamplification protocol.*

Gene	Sample	Initial Volume	Cel-miR-39?	C_q	ΔCt ACTB	ΔCt GAPDH	ΔCt HPRT1	$2^{\Delta Ct}$ ACTB	$2^{\Delta Ct}$ GAPDH	$2^{\Delta Ct}$ HPRT1
MALAT1	523	50 µL	Yes	32,101	8,228	8,269	2,424	299,805	308,479	5,365
		50 µL	No	32,978	8,276	9,697	3,494	309,880	830,089	11,270
		100 µL	Yes	26,538	-1,027	4,109	-3,147	0,491	17,257	0,113
		100 µL	No	N/A	~	~	~	~	~	~
		200 µL	Yes	27,507	2,607	6,327	-0,174	6,091	80,264	0,886
		200 µL	No	N/A	~	~	~	~	~	~
	521	50 µL	Yes	29,419	4,510	6,291	0,632	22,783	78,328	1,549
		50 µL	No	N/A	~	~	~	~	~	~
		100 µL	Yes	29,681	5,324	6,944	-0,831	40,064	123,104	0,562
		100 µL	No	N/A	~	~	~	~	~	~
		200 µL	Yes	28,608	7,139	9,063	1,696	140,956	534,953	3,241
		200 µL	No	34,201	9,454	13,518	6,913	701,115	11.731,197	120,503
	465	50 µL	Yes	33,331	10,613	10,803	4,321	1.566,013	1.786,055	19,985
		50 µL	No	N/A	~	~	~	~	~	~
		100 µL	Yes	30,177	6,882	8,792	2,819	117,926	443,216	7,054
		100 µL	No	34,335	9,556	10,830	5,192	752,506	1.819,931	36,544
		200 µL	Yes	27,568	5,808	7,455	0,628	56,014	175,414	1,545
		200 µL	No	N/A	~	~	~	~	~	~
	NTC	-	-	N/A						
		-	-	N/A						
NEAT1	523	50 µL	Yes	28,723	4,849	4,890	-0,955	28,819	29,653	0,516
		50 µL	No	30,400	5,697	7,119	0,916	51,890	139,000	1,887
		100 µL	Yes	28,709	1,144	6,280	-0,976	2,210	77,717	0,508
		100 µL	No	30,189	5,264	6,081	-1,228	38,427	67,699	0,427
		200 µL	Yes	28,889	3,989	7,709	1,208	15,880	209,255	2,310
		200 µL	No	28,710	4,971	5,921	-2,253	31,371	60,569	0,210
	521	50 µL	Yes	29,484	4,575	6,356	0,697	23,833	81,936	1,621
		50 µL	No	28,702	5,546	7,137	-0,561	46,718	140,715	0,678

		100 µL	Yes	29,607	5,251	6,870	-0,904	38,078	117,001	0,534
		100 µL	No	29,171	7,470	9,242	2,184	177,238	605,482	4,545
		200 µL	Yes	27,683	6,215	8,139	0,772	74,273	281,877	1,708
		200 µL	No	29,095	4,348	8,412	1,807	20,358	340,640	3,499
	465	50 µL	Yes	30,092	7,375	7,564	1,083	165,946	189,263	2,118
		50 µL	No	29,170	2,433	3,507	-4,757	5,400	11,366	0,037
		100 µL	Yes	29,304	6,009	7,919	1,946	64,409	242,075	3,853
		100 µL	No	28,805	4,026	5,300	-0,338	16,293	39,405	0,791
		200 µL	Yes	28,139	6,379	8,026	1,199	83,215	260,598	2,296
		200 µL	No	29,897	7,233	8,498	2,609	150,483	361,612	6,102
	NTC	-	-	35,426						
		-	-	38,101						
H19	523	50 µL	Yes	24,117	0,243	0,285	-5,561	1,184	1,218	0,021
		50 µL	No	23,989	-0,713	0,708	-5,494	0,610	1,634	0,022
		100 µL	Yes	22,734	-4,831	0,305	-6,951	0,035	1,236	0,008
		100 µL	No	24,692	-0,233	0,584	-6,725	0,851	1,499	0,009
		200 µL	Yes	21,755	-3,145	0,575	-5,926	0,113	1,490	0,016
		200 µL	No	23,581	-0,158	0,791	-7,382	0,896	1,731	0,006
	521	50 µL	Yes	26,937	2,027	3,809	-1,851	4,076	14,013	0,277
		50 µL	No	24,924	1,768	3,359	-4,338	3,407	10,262	0,049
		100 µL	Yes	26,626	2,270	3,889	-3,885	4,822	14,817	0,068
		100 µL	No	23,432	1,731	3,503	-3,554	3,319	11,339	0,085
		200 µL	Yes	22,754	1,286	3,210	-4,157	2,438	9,255	0,056
		200 µL	No	24,929	0,182	4,247	-2,358	1,135	18,985	0,195
	465	50 µL	Yes	26,197	3,479	3,669	-2,813	11,152	12,719	0,142
		50 µL	No	29,218	2,482	3,555	-4,709	5,585	11,755	0,038
		100 µL	Yes	24,791	1,496	3,406	-2,567	2,821	10,602	0,169
		100 µL	No	27,333	2,554	3,829	-1,810	5,874	14,207	0,285
		200 µL	Yes	23,705	1,945	3,592	-3,235	3,849	12,055	0,106
		200 µL	No	24,720	2,056	3,321	-2,568	4,159	9,995	0,169
	NTC	-	-	N/A						
		-	-	N/A						

ACTB	523	50 µL	Yes	23,874
		50 µL	No	24,702
		100 µL	Yes	27,565
		100 µL	No	24,924
		200 µL	Yes	24,900
		200 µL	No	23,739
	521	50 µL	Yes	24,910
		50 µL	No	23,156
		100 µL	Yes	24,357
		100 µL	No	21,701
		200 µL	Yes	21,468
		200 µL	No	24,747
	465	50 µL	Yes	22,718
		50 µL	No	26,737
		100 µL	Yes	23,295
		100 µL	No	24,779
		200 µL	Yes	21,760
		200 µL	No	22,664
	NTC	-	-	34,991
		-	-	38,306
GAPDH	523	50 µL	Yes	23,832
		50 µL	No	23,281
		100 µL	Yes	22,429
		100 µL	No	24,107
		200 µL	Yes	21,180
		200 µL	No	22,789
	521	50 µL	Yes	23,128
		50 µL	No	21,565
		100 µL	Yes	22,737
		100 µL	No	19,929
		200 µL	Yes	19,544

		200 µL	No	20,683	
	465	50 µL	Yes	22,528	
		50 µL	No	25,663	
		100 µL	Yes	21,385	
		100 µL	No	23,505	
		200 µL	Yes	20,113	
		200 µL	No	21,399	
	NTC	-	-	42,312	
		-	-	N/A	
HPRT1	523	50 µL	Yes	29,678	
		50 µL	No	29,484	
		100 µL	Yes	29,685	
		100 µL	No	31,416	
		200 µL	Yes	27,681	
		200 µL	No	30,963	
	521	50 µL	Yes	28,788	
		50 µL	No	29,263	
		100 µL	Yes	30,511	
		100 µL	No	26,987	
		200 µL	Yes	26,911	
		200 µL	No	27,288	
	465	50 µL	Yes	29,010	
		50 µL	No	33,927	
		100 µL	Yes	27,358	
		100 µL	No	29,143	
		200 µL	Yes	26,940	
		200 µL	No	27,288	
	NTC	-	-	N/A	
		-	-	N/A	