UNIVERSITÀ
DEGLI STUDI
DI GENOVA

DEPARTMENT OF HEALTH SCIENCES

RESEARCH DOCTORATE

IN

HEALTH SCIENCES - CURRICULUM PREVENTION OF CANCER AND CHRONIC-DEGENERATIVE DISEASES

(CICLE XXXIV)

microRNA pattern in tumoral lung tissues from former- and non-smokers: a possible footprint for environmental and genetic risk factors evaluation.

Gabriela Fernanda Coronel Vargas

2022

Supervisor: *Prof. Alberto Izzotti*

INDEX

Abstract

For the last decade the interest of microRNA as a possible early biomarker for lung-cancer has increased. Less attention has been focused in using microRNA as a biomarker to study the contribution of environmental exposure to air pollutants in non-smokers lung cancer patients.

The aim of this thesis is the identification of environmental related microRNA pattern in collected tissues from a total of 64 formal- and non- smoker patients recruited in a three-year observational study. The identified patterns may be used as early predictors of lung cancer, as well as environmental-related footprints. Through microRNA-chip array analysis of lung tissue, it was studied the expression of 2549 microRNAs. A differential analysis between healthy and tumoral tissue showed the presence of 273 microRNA differentially regulated, 222 were down-regulated and 51 up-regulated. Differential analysis was also applied to identify environmental pollution related microRNAs and finding microRNA deregulation in Passive Smoking at home (n=8), Passive smoking at work (n=1), Vehicle traffic at home (n=53), home distance from Etna Volcano (n=21), and home Type radon risk (n=19) exposures.

A second biomarker, Benzo(a)Pyrene-DNA adducts levels in blood, was also studied to understand its correlation to the above-mentioned environmental factors. The specificity of this biomarker was minor than microRNA pattern biomarker, but it was strongly correlated to vehicle traffic pollution.

The analysis of the microRNA environmental signatures indicates the contribution of environmental factors to the analysed lung cancers in the following decreasing rank: (a) vehicle traffic, (b) passive smoke, (c) radon, and (d) volcano ashes. These results provide evidence that microRNA analysis can be used to investigate the contribution of environmental factors in human lung cancer occurring in non-smokers

Introduction

Lung cancer, also known as lung carcinoma, is a malignant lung tumour that represent worldwide an important public health issue. In Europe lung cancer have the third highest incidence after breast and colorectal cancer, and the leading cause of cancer dead in both sexes, according to the International Agency for Research on Cancer (IARC). In 2018 there were 388,000 deaths caused by lung cancer in Europe, with a higher incidence in Central and Eastern Europe where there is a relatively poor prognosis of the disease after diagnosis [1]. About 85% of cases are related to cigarette smoking. Symptoms can include cough, chest discomfort or pain, weight loss, and, less commonly, haemoptysis. Patients present metastatic disease with or without any clinical symptoms.

Lung cancer screening programs in USA and in different European countries have formulated strategies that include primary prevention (health promotion and environmental protection), secondary prevention (screening and early detection), and an integrated healthcare services as main action elements against lung cancer. Early diagnosis and prevention are fundamental to reduce lung cancer mortality in general population. Due to the asymptomatic nature in the early stage of the illness most of the cases are diagnosed at advanced stages, when a poor prognosis occurs with a 5-year relative survival estimated in 21.7% (2011 -2017 data) [2]. Some risk prediction models include biomarker information, such as germline mutations or protein-based biomarkers as independent risk predictors. One of the most used tools for lung cancer detection is Low-Dose Computed Tomography (LDCT). Most of National European lung cancer screening protocols use LDCT in high-risk population (smokers aged between 55 to 80 years old) for an early detection and the reduction of overall mortality. However, health technology assessment of LDCT for lung cancer screening in high-risk populations made by the National Institute for Health Research (NHR) have conclude that LDCT, with ≤ 9.80 years of follow-up, was associated with a non-statistically significant decrease in lung cancer mortality. Even though LDCT remains the best screening strategy because cost-effectiveness-screening programmes are predicted to be more effective than no screening, due to its result in more lung cancer diagnoses [3].

It has been suggested that screening eligibility should be based on personal risk factors including sociodemographic situation, environmental exposures (second-hand smoke, asbestos and radon), smoking habits, clinical risk factors, age, sex, ethnicity, personal and family history of cancer, history of emphysema and chronic obstructive pulmonary disease (COPD), to minimize harms associated with screening as false-positive findings and unnecessary invasive diagnostic procedures [4]. Personalized lung cancer screening strategies can improve sensitivity and specificity of diagnosis. However, different risk prediction models need further validation to assess the accuracy of the models.

Smoking habits increase lung cancer Relative Risk (RR). For example, in Japanese population smoking increased RRs (95%CIs) for lung cancer 3.59 (3.25-3.96), chronic obstructive pulmonary disease 3.57 (2.72-4.70), ischemic heart disease 2.21 (1.96-2.50) and stroke 1.40 (1.25-1.57) [5]. It is well known that tobacco smoking epidemic is a public health threat that kills 8 million of people around the word each year. WHO estimates that in 2020 more than 7 million deaths were caused by direct tobacco smoke exposure, and around 1.2 million caused by second-hand smoke. However, worldwide still exist a considerable percentage of non-smokers that develop lung cancer. Existing lung cancer risk models separate risk and mortality prediction for never, former, and current smokers.

Risk factors for lung cancer in non-smokers include age, second-hand smoke, environmental exposures, genetic factors, underlying lung diseases, oncogenic viruses, and estrogen. Increasing in age is a common predisposition factor (average age at diagnosis 66 years old), although some tumours can be found in young patients with genetic predisposition [6]. Female sex predominance in lung cancer non-smoker patients is explained with the exposure to estrogens and other female hormones, as the expression of estrogen receptors are more common in non-smoker tumours compared to smokers' [7]. When mentioning environmental exposures, is necessary to concentrate attention in indoor (as carbon-based cooking fumes), and outdoor (car traffic and industrial) air pollution, asbestos exposure (that leads a six-fold increase in RR), and Radon exposure (as in Uranium miners) [8].

Due to the necessity of a better early tumoral predictors, microRNA (miRNA) analysis has been proposed as a new biomarker for a variety of tumours for the last ten years. miRNAs are non-coding RNA molecules that have different regulatory roles in cell differentiation, proliferation, and survival. They are able to inhibit complementary mRNA targets, regulating translation through RNA degradation. miRNAs are found to be dysregulated in numerous diseases, and frequently altered owing to mutations or transcriptional changes of enzymes that regulate miRNA biogenesis [9]. Serum, blood and tissue-specific miRNAs have been previously proposed for lung cancer screening, but only few clinical studies have been published with low quality evidence to support its implementation in clinical practice. As far as we know until the writing of this thesis, any previous study focused on profiling miRNA expression in non- and former- smokers exposed to different environmental pollution.

The hereby presented thesis aims to analyse the collected data from a total of 64 recruited patients of a three-year observational study. Specialized medical personal collected questionnaires from all patients, using google sheets with close answers. Forty-six patients were chosen to study miRNA expression in lung tissue (38 tumoral and 12 healthy) and in blood (41 patients) trough miRNA microarray analysis. Fifty-four blood samples were used to analysis the presence of Benzo [a]Pyrene-DNA adducts. 52 tumoral lung tissues were analysed by Ion-Torrent to detect oncogene mutations. Questionaries from all 64 recruited individuals were utilized to segregate data for different criteria as explained in Figure 1. Questionnaires collected environmental exposure, and lifestyles variants. Thanks to this information it was possible to build a list of differential regulated tumour related miRNAs according to an environmental exposure stratification.

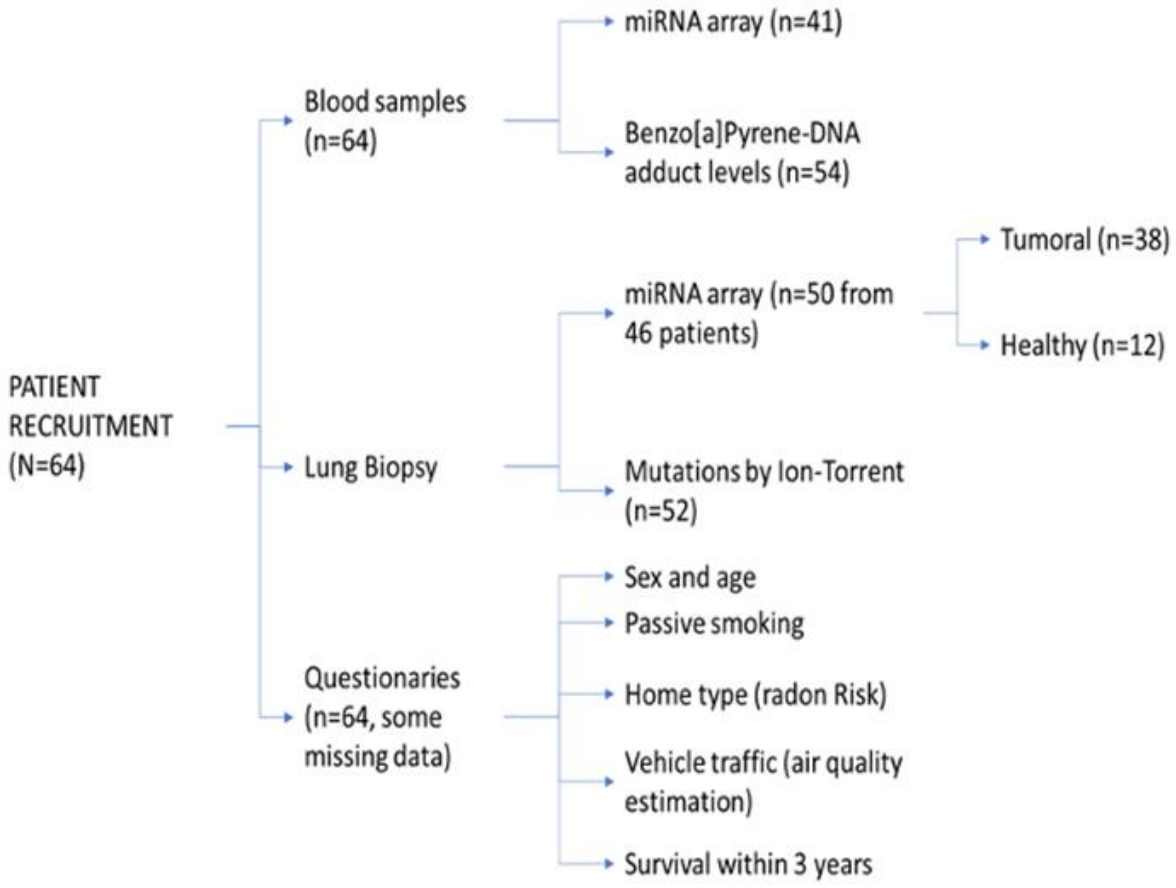


Figure 1 Patients enrolled, characteristics and sample sizes of analytical determinations carried out.

It was hypothesised that using the presented list of differential regulated tumour-related miRNAs may increase the tumour early diagnosis in non- and former-smoker patients exposed to four different environmental pollutants: car traffic air pollution, second-hand smoke, volcanic ashes, and Radon. However, even if results seem promising, it is necessary to continue the proposed research in a widely number of patients and in animal models before use them in clinical application to have more accurate prediction results. Moreover, it seems to be necessary to standardise a range of expression for each proposed miRNA in future studies.

1. miRNAs in Lung Cancer

1.1. miRNA pattern analysis as tumour predictor. Past, present, and future challenges.

MiRNAs are evolutionary conserved non transcriptional single-stranded RNA molecules of about 19 to 25 nucleotides in length, that regulate posttranscriptional silencing of target genes. In 1993 Victor Ambros and his group, from Harvard University, discovered the first miRNA (lin-4) in *Caenorhabditis elegans* which contains sequences complementary to a repeated sequence element in the 3′ untranslated regions (3′-UTR) of the lin-14 mRNA. They are present in all eukaryotes and are integrated in almost all known biological processes. In animals, mature miRNAs down-regulate protein synthesis in cytosol by pairing to complementary sequences in the 3′-UTRs of target messenger RNAs (mRNAs). In mammals, nucleotides 2 to 8 of the miRNA 5′-end constitute a "seed region" that binds imperfectly to a mRNA complementary recognition sequence at the 3′-UTR. This complementary pairing is mediated by a protein complex called miRNA-induced Silencing Complex (miRISC). The core protein of RISC is the Argonaute protein (Ago). Humans have approximately 2000 sequenced miRNAs, which accounts around 1-5% of the total genome. A single miRNA can target hundreds of mRNAs and influence the expression of many genes often involved in a functional interacting pathway [10]. As multiple miRNAs can target the same mRNA, there is no linear correlation between miRNA and mRNA expression, so an entire miRNA profile is necessary to understand the dysregulations that play an important role in disease progression, and in consequence in miRNA potential use as diagnostic and prognostic tools.

MiRNAs can be isolated from cells, tissues, and body fluids such as serum, plasma, tears, and urine. MiRNA expression can be detected in both tissue samples and cell-free biological fluids. Current methodologies used for detecting miRNAs include quantitative PCR (qPCR), in situ hybridization, microarrays and RNA sequencing. The hybridization-based arrays, which are mostly used for miRNA analysis in recruited patients mentioned in this thesis, have the advantage of allowing a large number of parallel measurements per sample at a relatively low cost. Due to limited specificity, findings from hybridization-based arrays are typically validated with a second method

such as qPCR or in situ hybridization [11], but this validation necessity increases the cost of the research.

Preanalytical and analytical variables interfere with miRNA analysis. From tissue selection and sampling, going through sample proceeding and miRNA measurement to data analysis, we can find variables that affect results. For example, miRNA expression between tissues is different, so lung tissue and circulating miRNA expression in blood may differ. Moreover, when clinical studies are proposed the individual variability (i.e. sex, age, drug assumption, genetics, ethnicity, diet and lifestyle) play an important role in expression patterns. The finding of non-tumoral and tumoral miRNA predictive patterns that take in to account these variables are still necessary.

The proposal of using miRNA profiling as a non-invasive biomarker begun in 2008, when circulating miRNAs were discovered. Circulating miRNAs have been observed in blood, encapsulated within exosomes, microvesicles, as part of lipoprotein–miRNA complexes and as free unbound miRNAs. Compared with lung biopsy, blood and bronchoalveolar lavage fluids are relatively easy and safe to obtain, and the proposal to use these tissues for miRNA expression pattern analysis have been previously suggested for lung cancer diagnosis [12]. miRNAs are stable in circulation and storage, resistant to RNAses, show disease-specific expression, and reflect microenvironmental and cellular changes prior to and throughout diseases [13]. However, lack of standardized protocols and a consensus on the most appropriate endogenous control to use to normalize relative miRNA expressions, led to difficulties in producing strong data in circulating miRNA as biomarkers. Moreover, it has being highlighted different inconsistencies across circulating miRNA studies [14], advising caution in front of enthusiastic proposals of using specific blood and other fluids circulating miRNAs analysis in diagnostics.

Despite of the indicated limits of miRNA for diagnostics, prognosis, and treatments, the first miRNA-based patent was published in Europe in 2008 (about miRNA sequencing). In 2010, prognosis and treatment use of miRNAs were patented in 2010; patent N° US 7709616 B2 of May 4, 2010, describes a list of 760,616 like polynucleotides that were associated with prostate and lung cancer. The polynucleotides are human miRNAs and miRNA-precursors. Moreover, the patent includes methods and compositions that can be used for diagnosis, prognosis, and treatment of those medical

conditions. Today there are about 4,000 miRNA registered patents in USA database and about 2,000 in the European database, of which about 1,500 are related to cancer applications. The biopharmaceutical development of miRNA-based drugs is strongly financed, industrial research is focused on the understanding of pharmacokinetics and pharmacodynamics (absorption, distribution, metabolism, and excretion) mechanisms of the proposed miRNA-based drugs, as well as efficient delivery systems [15].

A kind of miRNAs related directly with oncogenes, called OncomiRs, are of particular interest. OncomiRs were described for the first time in 2006 by Esquela-Kerscher A. and Slack F. as "microRNA genes that function as both tumour suppressors and oncogenes" because "about 50% of annotated human miRNAs are located in areas of the genome, known as fragile sites, that are associated with cancer" [16]. For example, it is widely known that let-7 family members (which are downregulated in lung cancer tissues) negatively regulate RAS oncogene which have a key function in cell proliferation, survival, and differentiation. It has been demonstrated that components of the miRNA-machinery are implicated in tumorigenesis. The expression of the endoribonuclease Dicer, the RNAse that cleaves pre-microRNA in miRNA maturation process, is downregulated in lung cancer [17]. This downregulation is correlated with shorter post-operative survival for the mature single strand miRNA. Today some examples of proposed OncomiRs are: miR-15, miR-16, miR-17, miR-19, miR-21, miR-155 and miR569 as they are largely overexpressed in different types of tumours. On the other hand, let-7 family is considered as an anti-OncomiR, as it is largely down-regulated in tumours. Some of the biological processes regulated by let-7 are metabolism (blocking glucose uptake) and the endothelial cell migration [18].

A recent metanalysis suggest that mature miRNAs miR-21 and the let-7 family members may be an important prognostic biomarker in non-small cell lung cancer (NSCLC) [19]. The oncogenic role of miR-21 differential regulation in patients with NSCLC have been reported in different studies. MiR-21 is upregulated in NSCLC, alters the apoptotic mechanism, cellular growth, and proliferation pathways. On the other hand, it has been observed downregulation in let-7 family members in NSCLS. Let-7 family transfection reduces considerably cell proliferation rates as it hence as tumour suppressor [20]. Target genes of let-7 include c-Myc, signal transducer and activator of transcription 3 (STAT3), and Janus-activated kinase 2 (JAK2), all genes involved in cell proliferation and cell cycle.

In NSCLC, let-7 down-regulation was reported as significantly correlated with patient outcome, but other studies have not established a direct link between low expression of let-7 and prognostic of NSCLC patients [21, 22].

1.2. Lung Cancer, miRNA regulation, smoking habits and environmental exposures.

Lung Cancers are carcinomas (malignancies that arise from epithelial cells) that can be classified according to histological type (microscopic examination of the tissue) by size and the appearance of the malignant cells. For therapeutic purposes two broad classes are distinguished: Small-Cell Lung Cancer (SCLC) and Non-Small-Cell Lung Cancer (NSCLC). Research of miRNA regulation in Lung Cancer is focused on how regulation variates between SCLC which represent about 15% of cases, and NSCLC with about 85% of cases.

As it is widely known, SCLC is the most aggressive and almost fatal between the two lung cancer types, despite a good initial response to chemo-radiation therapy. SCLC present metastasis at the moment of the diagnosis in 80% of the patients. SCLC cells have a high density of neurosecretory granules, which give the tumour a paraneoplastic syndrome association. SCLC is present in smokers most of the times. SCLC is the most aggressive lung cancer

By the other side, the clinical behaviour of NSCLC is more variable and depends on histologic type. In NSCLC, about 40% of patients will present a metastatic disease outside of the chest at the time of diagnosis [23]. NSCLC is present in smokers as well as in former- and non-smokers. The tree main subtypes of NSCLC are:

- adenocarcinoma, present in about 40% of NSCLC cases, associated with smokers, former-smokers and second-hand smoking, it is usually originated from peripheral lung tissue,
- squamous-cell carcinoma, present in about 30% of NSCLC cases, typically occur close to large airways,
- large-cell carcinoma, present in about 10 to 15% of NSCLC cases, their cells present big cytoplasm, large nuclei, and conspicuous nucleoli.
- pulmonary enteric adenocarcinoma, a rare subtype of adenocarcinoma [24].

It has been hypothesized that lung tumour presents genetic, transcriptional and microenvironmental differences between smoker and non-smoker patients, but until today any study has analysed miRNA transcriptional alterations in lung tumour based only in non-smoker population considering different environmental exposures. For example, a bioinformatic analysis, that used 210 non-smokers and 292 smokers data from Cancer Genome Atlas and Gene Expression Omnibus databases integrated with Department of Thoracic Surgery of Zhongshan Hospital-Fudan University's biobank, suggests that exist a differential regulation in mRNA and miRNA expression as well as significant difference in somatic mutation frequencies in lung adenocarcinoma between smokers and non-smokers groups [25]. After qRT-PCR validation of bioinformatic findings using lung tissues from Fudan University's biobank, researches found that NTS and NNAT mRNA, as well as miR-377-5p and miR-34a-3p were the most up-regulated factors in non-smoking patients with adenocarcinoma. By the other hand, TFF2 and REG4 mRNA were the most downregulated in smokers. However, even if the alteration of expression (up- or down- regulation) in the results of qRT-PCR was generally consistent with bioinformatics, fold change values in qRT-PCR were not as large as those calculated in the bioinformatics results. Strong limitations for this research were declared, as there was a lack of information regarding whether the patients were frequently exposed to second-hand smoke or kitchen fumes, as well as a lack of high-quality data about smoking habits (i.e., if they were smokers or former-smokers, smoking years, etc.).

A strong evidence between smoking habits, miRNA regulation and lung cancer is available in literature, indicating data from in-vitro, in-vivo and clinical observations. Differences between miRNA profiles based on gender have being suggested in animal models of adenoma-free and adenoma bearing mice exposed to mainstream cigarette smoking [26]. A review by Momi N. et al. (2014) [27] has proposed a complex model to explain the miRNAs alteration related to tobacco smoking and lung cancer, where already cited miRNAs like let-7 family, miR-21, miR-30, miR-34 family, and miR-143, are some of the responsible of important gene expression and protein regulation processes in lung cancer. As explained in Figure 2, by Momi N. et al. (2014), the presence of smoke and nicotine/ Nicotine-derived nitrosamine ketone (NNK) in cellular environment plays a primary role in the disruption of miRNA-directed regulation of pathways related to cell cycle regulation, survival, angiogenesis, inflammation, and metastasis.

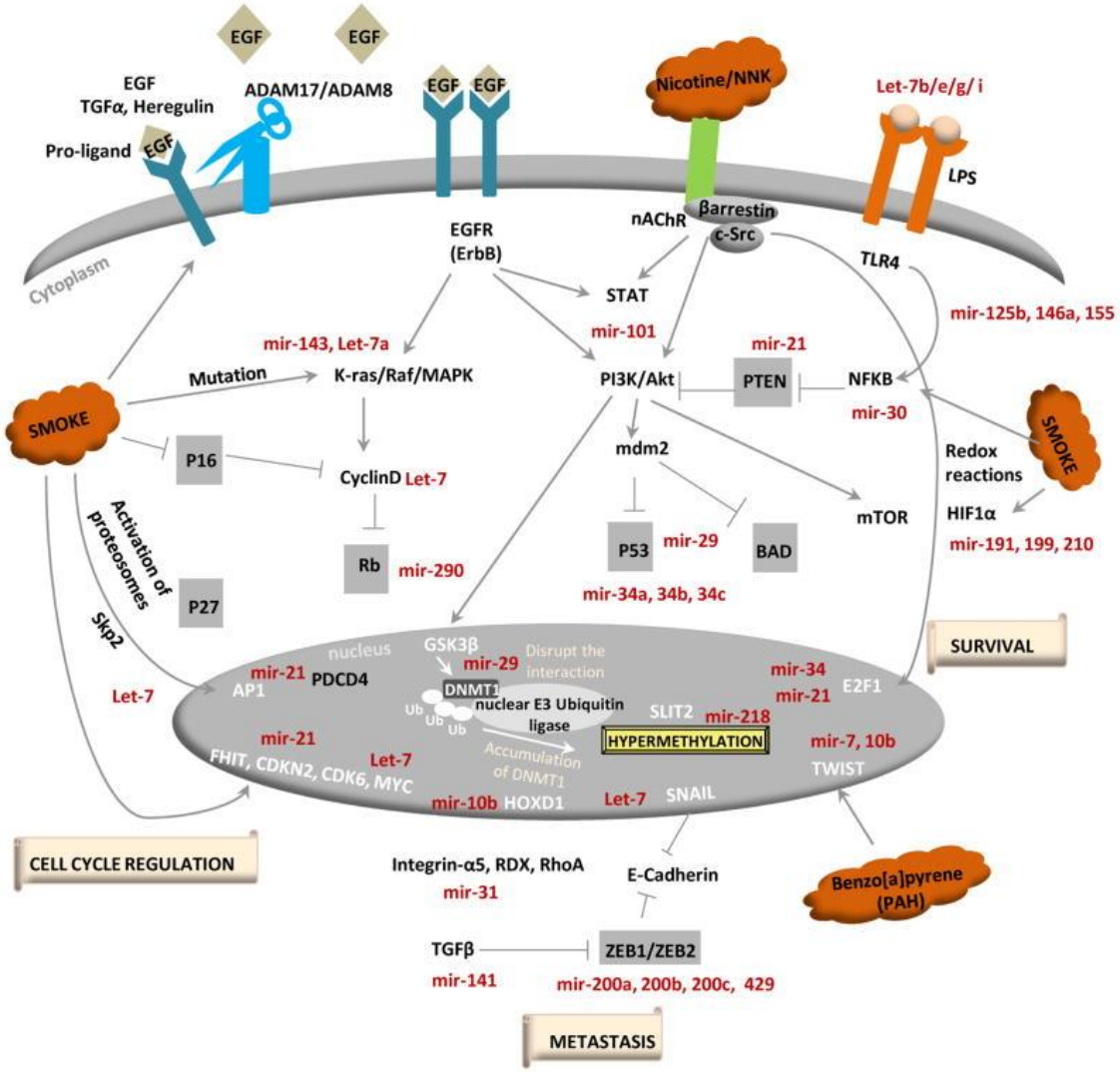


Figure 2 Disruption of miRNA-directed regulation in smokers according to Momi N. et al. (2014).

By the other side, limited data is available to analyse the role of miRNA regulation in environmental risk factors for non-smokers lung cancer patients. Some literature is available for air pollution-induced miRNA alterations. Different response to air pollutants as particulate matter, ultrafine particles, nitrogen oxides, black carbon and carbon oxides (CO and CO2) may be related to different expression of miR-92a-3p, miR-484 and miR-186-5p, linking traffic-related exposure to disease risk [28]. Lung malignancies linked to asbestos exposure has been related to miR-126, miR-205, and miR-222 [29]. Further to inducing alterations in the microRNA machinery, air pollution also induces genotoxic alterations including formation of DNA adducts, DNA damage and mutations.

Differential regulation of microRNA machinery in lung tissue after Radon (Rn) exposure has been demonstrated in-vitro studies using human bronchial epithelial cells (HBE), and in-vivo using rat tissues [30]. It has been observed a downregulation for let-7 family (let-7b-3p and let-7a-2-3p) that alters K-RAS oncogene pathway, therefore altering cell signalling, intracellular metabolism, reactive oxygen species (ROS) production, and reduced mitochondrial activity. Let-7 family and K-RAS correlation to lung cancer has been previously proposed as the inhibition of tumour growth may occur via suppression of K-RAS expression by let-7a transfection [31, 32].

The research proposed in this thesis aims to clarify only in part some questions about the correlation between miRNA regulation, environmental pollution, and individual variables. As explained before, this kind of correlation variates among studies when analysing only tumour and healthy tissues, and tumoral tissues from smokers and no-smokers. The collected data from patients (as previously explained in Figure 1) allowed us to cluster miRNA profiles according to:

1. Lifestyle: non-smokers and former smokers
2. Environmental factors: exposure to vehicle traffic, Etna Volcano, second-hand smoke, Radon.
3. Individual variables: lung tissue mutations in oncogenes, 3-years-survival.

Neither epigenetic nor genetic alteration, when used alone, are accurately predict the lung cancer risk in exposed subjects. Indeed, the adverse effects of mutations can be silenced by a functional microRNA machinery and the alteration of the miRNA machinery is devoid of remarkable consequences in absence of genotoxic damage. This research integrates genomic and postgenomic analyses to shed light on the differential contribution of environmental factors to lung carcinogenesis.

As mentioned before, this issue has been explored in a peculiar environmental situation characterized by the presence of an active volcano (Etna) near to the analysed population (Catania, Italy). In fact, volcanic dust from Etna has being related to a higher risk for pleural mesothelioma and other non-malignant respiratory diseases [33], to a possible pathogenic role in the epidemiology of amyotrophic lateral sclerosis [34], and to neurodegenerative diseases [35], but not to lung cancer. Etna's volcanic dust is also a vector of atmospheric pollutants as polycyclic aromatic hydrocarbons (PAHs) and particulate mercury [36].

1.3. BaP-DNA adducts as biomarker

BaP-DNA adducts are covalent modifications of the DNA. This is the result from exposure to specific carcinogens, so the level of adducts in cells can serve as a biomarker for a significant [37]. It has been observed that DNA adducts tend to be higher among subjects heavily exposed to air pollution [38] as vehicle traffic. BaP is a well-studied pro-carcinogen, and its adducts directly alter regulation of transcription of tumor suppressors or oncogenes. It is well known that Cytochrome P4501A1 plays a central role in the activation step of BaP, therefore the formation of BaP-DNA adducts [39]. The structure of adducts is well known [40] (**Figure 3**). Because DNA adduct levels in tumor tissue and in blood lymphocytes have been associated with lung cancer, it has been proposed as potential biomarker of risk for lung cancer [37].

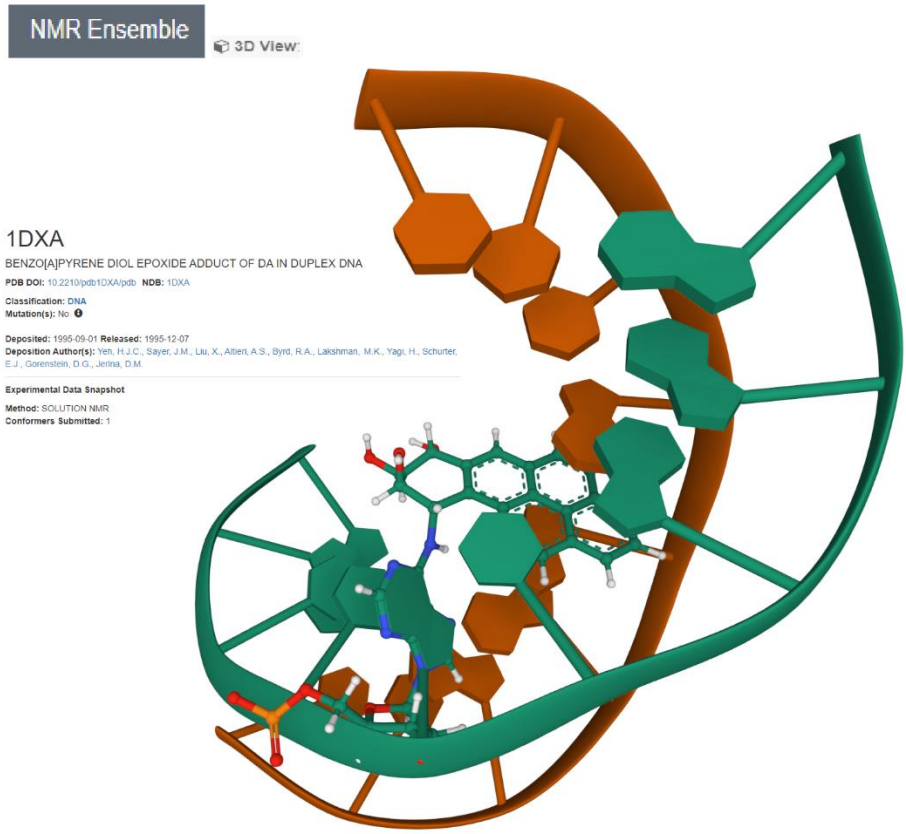


Figure 3 BENZO[A]PYRENE DIOL EPOXIDE ADDUCT OF DA IN DUPLEX DNA 3D structure, ENSEMBL entry 1DXA, available at https://www.rcsb.org/structure/1dxa.

2. Methods

Most of hereby explained methods can be found in both articles already published [41, 42] of this study. The author of this thesis was primarily involved in Bioinformatic analysis of the experimental results. A more detailed explication of bioinformatic analysis excluded from published articles is exposed below.

2.1. Patient's recruitment, questionnaires, and environmental exposure evaluation

The patient's recruitment was carried out in four hospitals of Catania (University Hospital "G. Rodolico - San Marco", "Garibaldi-Nesima" Hospital, "Cannizzaro" Hospital, "Morgagni" Clinic) and "San Vincenzo" Hospital of Taormina (Messina province). The study protocol was per-formed according to the Declaration of Helsinki and approved by the Ethic Committee Catania 1 (n. 11778 released on March 17th, 2015), and Ethic Committee Catania 2 (346/C.E. released on May 28th, 2015), respectively.

The criteria used for patient enrolment were to be over 18 years of age, have lung cancer for which surgery treatment has been indicated, have been non-smokers or former smokers for at least 5 years, have signed the written informed consent. No restriction was made regarding the sex of patients or morphology of the reported neoplastic lesions. Both the neoplastic and healthy tissue samples were taken from the same patient and, the tissue samples were obtained directly from the pathological anatomies of the hospitals involved in the project. Instead, the blood samples were collected by the thoracic surgery units of the hospitals. The interviews were carried out directly in the thoracic surgery wards by the cancer registry doctors involved in the study. A total of sixty-four patients were enrolled.

Data was collected by trained personnel using a semi-structured questionnaire to obtain information on sociodemographic, lifestyles data including smoke history, nutrition, home characteristic and its location (for Radon and urban traffic pollution exposure evaluation) (**Figure 1**). Data from the 64 questionnaires are presented codified by number in **Table S1**. Patients average age is 69.02 years old (min = 43, max = 84), females are 34.4%, and 20.3% of patients died within three years after the biopsy. A total of 15 subjects had never smoked, 20 were former smokers for more

than 20 years, 13 from 11 to 19 years, and 9 from 10 to 5 years, smoking habits data was missed for 14 patients. The presence of mutation in oncogenes AKT1, ALK, BRAF, DDR2, EGFR, ERBB2, ERBB4, FBXW7, FGFR3, KRAS, MAP2K1, MET, NOTCH1, PIK3CA, PTEN, SMAD4, STK11, TP53 for 52 out of 64 lung biopsies are also reported. As shown in **Figure 4,** all patients lived near Etna Volcano (average = 56 Km, min = 13 Km, max =152 Km).

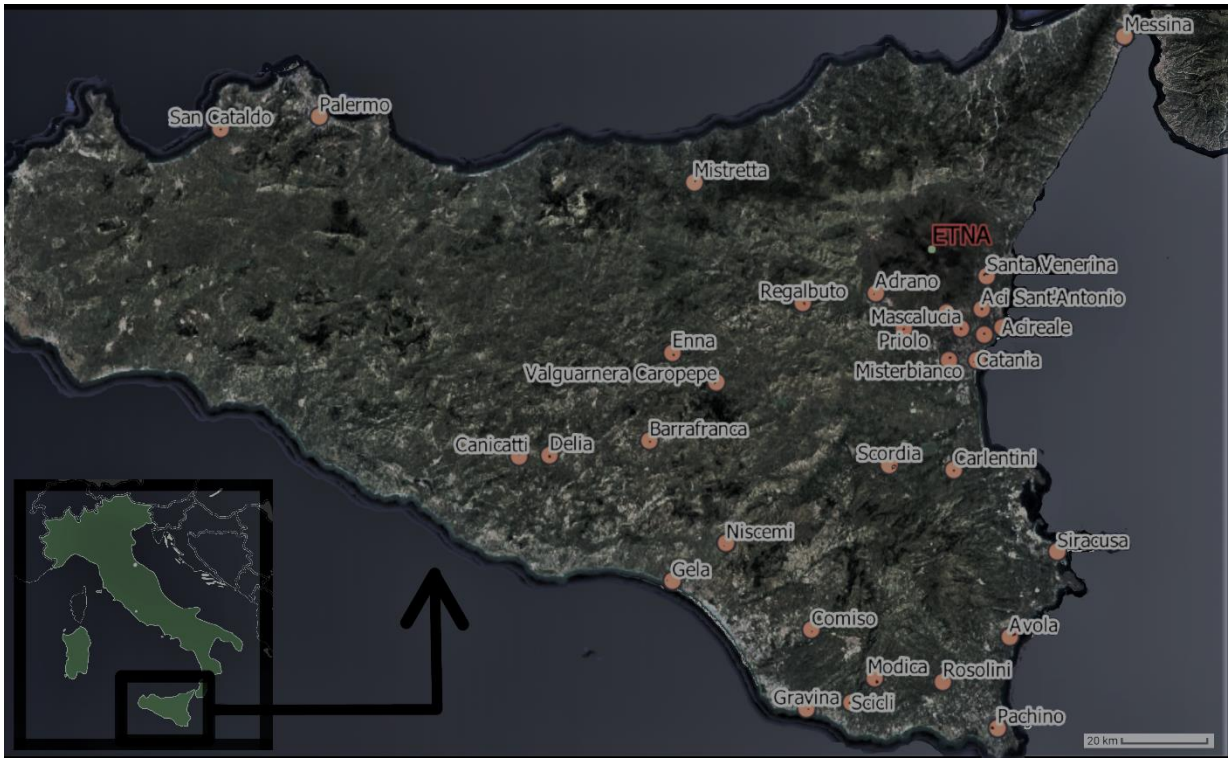


Figure 4 Sicilian Municipalities of residence of patients in this study. Detailed data for each individual can be found in questionary data.

Measurement of environmental exposure to vehicle traffic, radon, and volcano ashes were not available. However, environmental exposure was estimated and classified according to information given by patients in the questionnaires. For vehicle traffic, patients choose if their home were surrounded by intense, moderated, low and none traffic. For Radon exposure, patients declared what type of house they live in; ground and first floor departments as well as rural houses and villas were categorized as high-risk condition. For volcano exposure, it was calculated median distance of patient's houses from Etna, therefore patients were categorized by high and low exposure as under and over the median (≤60 Km and >60 Km).

2.2. Lung biopsy

Lung biopsy specimens (n=64) were collected at the onset of disease from patients who were diagnosed with lung cancer between 2015 and 2018, and referred to the Catania, Messina, Enna Cancer Registry, Italy. The study was approved by ethics committee, informed consent was obtained by "G. Rodolico -San Marco" University Hospital. All patients were classified as cases according to the 2021 ICD-10-CM Diagnosis Code C34.90. MicroRNA were comparatively evaluated in cancer and surrounding normal tissue as identified by histopathological analysis.

2.3. DNA Extraction

Genomic DNA (gDNA) was extracted from 25 mg of fresh frozen lung biopsy DNA using the DNeasy ® Blood & Tissue kit (Qiagen, Milan, Italy), as described by the manufacturer's protocol. The purification of gDNA was automated on the QIAcube ® instrument (Qiagen, Milan, Italy). The gDNA quality and quantity were assessed with a NanoDrop® 1000 spectrometer and with a Qubit® 3.0 Fluorometer using a dsDNA HS Assay Kit (Thermo Fisher Scientific, Carlsbad, CA, USA).

2.4. Somatic Mutation Identification

The mutational status of the oncogenes associated with lung cancer was analysed by sequencing using the Colon and Lung Cancer Research Panel v.2 (Thermo Fisher Scientific, Carlsbad, CA, USA), which screens 92 amplicons in hotspots and target regions of these genes. For each sample, 15 ng of gDNA was amplified using the Ion AmpliSeq™ Library Kit 2.0 (Thermo Fisher Scientific, Carlsbad, CA, USA) according to the protocol for gDNA isolated from fresh frozen samples.

The quality control of the libraries was assessed by TapeStation® 2200 using the High Sensitivity D1000 assay® (Agilent Technologies, Santa Clara, CA, USA) and with a Qubit® 2.0 Fluorometer using the dsDNA HS Assay Kit (Thermo Fisher Scientific, Carlsbad, CA, USA). Then, seven multiplexed libraries (100 pM) were amplified and enriched by OneTouch™ and the OneTouch™ ES, respectively using Ion PGM™ Hi-Q™ View OT2 Kit (Thermo Fisher Scientific, Carlsbad, CA, USA). Finally, the template was loaded onto a 316 v.2 chip and sequenced using the Ion PGM™ Hi-Q™ View Sequencing

Kit on the Ion PGM™ platform (Thermo Fisher Scientific, Carlsbad, CA, USA). The sequencing data were analysed using the Ion Torrent Software Suite with the plugin Torrent Variant Caller v.5.10.0.18 (Thermo Fisher Scientific, Carlsbad, CA, USA) applying somatic, high stringency parameters. We considered gene variants with a variant allele frequency up to 1%, if covered at least 1000×. All gene variants were annotated by Ion Reporter™ Software v. 5.10.

2.5. Total RNA Extraction

The total RNA was extracted from lung biopsies and blood plasma using standardized protocols that combined phenol/guanidine-based lysis of samples and silica-membrane-based purification.

Briefly, 3 mL of whole blood were collected in Ethylenediaminetetraacetic acid (EDTA) tubes and layered onto 3 mL Histopaque-1077® (Sigma-Aldrich Chemie Gmbh, Munich, Germany) through centrifugation at 400× g for 30 min. Plasma and lymphocytes were separately collected and stored at −20 °C at the Laboratory of Molecular Epidemiology (University of Catania) until analysis. Next, the total RNA from the plasma was extracted using the miRNeasy® Serum/Plasma Kit (Qiagen, Milan, Italy), as described by the manufacturer's protocol.

For lung biopsies, 30 mg of fresh starting material was first stabilized in 2.5 mL of RNAlater solution and stored at −20 °C at the Laboratory of Molecular Epidemiology (University of Catania) until analysis. Next, lung biopsies were disrupted using the TissueRuptor® II for 20–40 s and homogenized in 700 µL QIAzol® Lysis Reagent (Qiagen, Milan, Italy). The total RNA was purified from the homogenate using the miRNeasy® Mini Kit (Qiagen, Milan, Italy), as described by the manufacturer's protocol. The purification of RNA was automated on the QIAcube® instrument (Qiagen, Milan, Italy). The quantification of RNA was assessed with a Qubit® 3.0 Fluorometer using the HS RNA Assay kit (Thermo Fisher Scientific, Carlsbad, CA, USA).

2.6. miRNA extraction

Total RNA was extracted from lung biopsies using a standardized protocol which combines phenol/guanidine-based lysis of samples and silica-membrane–based purification. In brief, 30 mg of starting material were first disrupted and homogenized in 700 µl QIAzol® Lysis Reagent, using the

TissueRuptor® II (Qiagen, Milan, Italy) for 20-40 s. Next, total RNA was purified from the homogenate using the miRNeasy® Mini Kit (Qiagen, Milan, Italy), as described by the manufacturer's protocol. Purification of RNA was automated on the QIAcube® instrument (Qiagen, Milan, Italy).

2.7. miRNA-microarray analysis

miRNA expression profiling was carried out by Agilent platform following the miRNA Microarray protocol v.3.1.1 (Agilent Technologies, Santa Clara, CA, USA). Briefly, 50 ng of total RNA, containing miRNAs and Spike-in controls underwent dephosphorylation and labelling step with Cyanine 3-pCp. The Cy3-labeled RNA was then purified using Micro Bio-Spin® P-6 Gel Column (Bio-Rad Laboratories, Inc., Hercules, CA, USA) and hybridized on Human miRNA microarray slide 8x60K (Agilent Technologies; including 2,549 miRNAs, miRBase 21.0) at 55°C for 20 hours. After washing, the slides were scanned by G2565CA scanner (Agilent Technologies) and the images were extracted by Feature Extraction software v.10 (Agilent Technologies). Microarray raw data was deposited in Gene Expression Omnibus (http://www.ncbi.nlm.nih. gov/geo/; GEO accession number GSE169587, ID: 200169587).

2.8. Benzo[a]Pyrene-DNA Adduct Levels in Human White Blood Cells

Hydrolyzed BPDE adducts or Tetrol I-1 and Tetrol II-2 were analyzed in lymphocyte DNA through the modified high-performance liquid chromatography–fluorescence (HPLC–FL) method described by Alexandrov et al. [43], and Oliveri Conti et al. [44].

Briefly, lymphocytes were separated from whole blood samples using HISTOPAQUE® -1077 (Sigma-Aldrich Chemie Gmbh, Munich, Germany). The lymphocyte DNA was extracted using the DNeasy® Blood and Tissue kit according to customer's procedure (Qiagen, Milan, Italy).

Hence, DNA was subjected to a procedure of hydrolysis and purification, and Tetrols were quantified according to the methodology of Oliveri Conti et al. [44]. HCl, also if hypergrade certified, can contain traces of fluorescent active contaminants that could interfere with the peak detection of the studied analytes and reduce analytical sensitivity of the method.

To avoid this important bias in the sample preparative phase, all the HCl impurities visibly reactive to the FL detector were deleted by chemical purification. To improve the sensibility of detection, Thermo Scientific™ PEEK Capillary Tubing (0.005 in.) was used. The extracted and purified DNA was dissolved in 1 mL of water and analysed in a Varian Pro Star System HPLC using a TOSOH (C18 RP 25 × 0.46 cm, 5 µm) column with the following elution program: 15 min with 20% water/acetonitrile of equilibrium phase, 5 min with 20% water/acetonitrile and 60 min to acetonitrile (100%) (slop of 1) and, finally 10 min to 100% acetonitrile.

An isocratic program (0.85 mL/min) was used and the FL detector (FLD) was programmed to 344 nm (ext.) and 388 nm (em.) for the excitation and emission wavelengths, respectively. The sensitivity of the FLD was fixed to a high modality. The wavelength of UV–VIS detector (UV) was set at 238 nm, permitting the dual detection of both Tetrols (I-1 and II-2).

The chromatographic system was calibrated using external certified pure standards of Tetrol I-1 and Tetrol II-2 (purity 99.0%) (Chemical Carcinogen Reference Standard Repository, Kansas City, MO, USA).

Recoveries were 94% and 82% for Tetrol I-1 and Tetrol II-2, respectively. The processing of reagent blank disclosed no trace of Tetrol I-1 and Tetrol II-2. The linearities (R2) obtained of FLD were 0.9980 and 0.9990 for Tetrol I-1 and Tetrol II-2, respectively. For UV, the Rs2 were 0.9850 e 0.9803 for Tetrol I-1 and Tetrol II-2, respectively. MDL were 2.0 pg/mL and 3.1 pg/mL for Tetrol I-1 and Tetrol II-2, respectively. The validated method permitted detecting Tetrol I-1 and Tetrol II-2 in a minimum of 3µg of extracted DNA.

2.9. Statistics and Bioinformatic analysis

All Lung-tissue-miRNA raw data files from Agilent Technologies Microarray Scanner System G2505C were imported to GeneSpring® using miRNA analysis type, Technology 70156_v21_0, without baseline transformation. Protocol used was: Analysis type = Expression, Experiment type = Generic Single Color, Normalization algorithm = none, percentile target = 75, baseline transformation = none. Intensities of replicated spots on each array were averaged.

Microarrays' raw data were deposited in Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/; GEO accession number GSE169587, ID: 200169587).

Given the explorative nature of this study and the wide variability between patients, normalization was not performed in data import step. Volcano Plot analysis for all miRNA entities between averaged tumoral and healthy tissue was run, without multiple test correction, FC ≥ 2 and $p \leq 0.05$. A total of 273 miRNAs (including Agilent positive control hur detected because of the lack of normalization) were identified as possible cancer related markers (Lung Cancer Related miRNAs or LCRMs) and exported as a new entry list for further experiments.

Quality Control of LCRMS as tumoral markers, was performed by 3D principal component analysis (PCA) scores. Hierarchical Clustering and Principal Component Analysis using miRNA-arrays data were also run in GeneSpring®. Protocol used for Clustering Algorithm: Hierarchical, Clustered By: Normalized intensity values, Clustered On: Entities, Similarity Measure: Euclidean, Linkage Rule: Wards. Protocol used for PCA Conditions (per sample), Mean Cantered: true, Scaled: true.

Target detection based on different databases is one of GeneSpring®'s functions. After detecting the most representative LCRMs targeted genes, they were imported to run subsequently the Gene Ontology - Biological Proces (GO-BP) analysis. The simplest input to such analysis is a list of genes that is most differentially expressed or frequently mutated in a dataset. A typical analysis workflow consists of two steps: (1) a gene list is defined by filtering experimental data for genes with significant gene-level statistics, and (2) enrichment analysis is performed to determine processes and pathways over-represented in the gene list [45].

Most significative Gene Ontology (i.e. the top 10 genes with a highest KS prediction) was found using Bioconductor (https://www.bioconductor.org/) miRNAtap R Package [46, 47]. The entire code can be found at https://github.com/CoronelVargasG/RTSO-GCVThesis/blob/main/Rtso_Predictions.R (last accession 16/07/2021). An example of the code used for hsa-let7a-5p gene targets and GO-BP prediction is shown in **Figure 5.** Before using the code library

installation and activation of packages miRBaseConverter, miRNAtap, miRNAtap.db, topGO, org.Hs.eg.db, and GOplot in RStudio console, must be execute.

```
allGO2genes = annFUN.org(whichOnto='BP', feasibleGenes = NULL,
                         mapping="org.Hs.eg.db", ID = "entrez")

#miRNAs from RTSO

hsa_let_7a_5p   =       'let-7a-5p'
predictions_hsa_let_7a_5p       =       getPredictedTargets(  hsa_let_7a_5p , species = 'hsa', method = 'geom', min_src = 2)
rankedGenes_hsa_let_7a_5p       =               predictions_hsa_let_7a_5p       ,['rank_product']
selection_hsa_let_7a_5p =       function(x) TRUE
GOdata_hsa_let_7a_5p    =       new('topGOdata', ontology = 'BP', allGenes = rankedGenes_hsa_let_7a_5p  ,
                        annot = annFUN.GO2genes, GO2genes = allGO2genes,  geneSel = selection_hsa_let_7a_5p     ,
                        nodeSize=10)
results.ks_hsa_let_7a_5p        =       runTest(        GOdata_hsa_let_7a_5p    , algorithm = "classic", statistic = "ks")
allRes_hsa_let_7a_5p    =       GenTable(       GOdata_hsa_let_7a_5p    , KS = results.ks_hsa_let_7a_5p , orderBy = "KS", topNodes = 10)
allRes_data_hsa_let_7a_5p       =       allRes_hsa_let_7a_5p    [,c('GO.ID','Term','KS')]
Top_10_miRNAs_ranked_GOdata = rbind.data.frame(allRes_data_hsa_let_7a_5p        , OTHER MIRNAS)
write.csv(Top_10_miRNAs_ranked_GOdata, file = 'Top_10_miRNAs_ranked_GOdata.csv')
```

Figure 5 An example of the code used for prediction of the most affine gene targets and GO-BP prediction

The Gene Ontology is a system of classification in which genes are assigned to a set of predefined bins depending on their functional characteristics. The ontology covers three domains: (a) “molecular function”, that are all molecular-level activities performed by gene products (b) “cellular component” that classify locations relative to cellular structures in which a gene product performs a function, and (c) “Biological Process”, the larger processes with broad and nailed functional terms accomplished by multiple molecular activities. GO-BP enrichment analysis is a bioinformatic tool largely used to understand the relationship between a set of genes and their biological function, creating a computational hypothetical model of larger processes, or “biological programs”, accomplished by multiple molecular activities that may be deregulated. describes our knowledge of the biological domain with respect to three aspects

Result of the top 10 GO data (2660 GO terms) for each LCRM can be found at https://github.com/CoronelVargasG/RTSO-GCVThesis/blob/main/Top_10_miRNAs_ranked_GOdata.csv).

After identifying and classifying the most significant GO terms for each LCRMs, Revigo tool (http://revigo.irb.hr/) [48] was used to create a large-similarity-semantic-based TreeMap.

MiRNAs related to five different environmental exposures, or Environmental Exposure miRNA Signal (EESs) were determined analysing only LCRMs list. Environmental exposures variables were inserted in GeneSpring for each patient using the questionnaires information according to (a) Passive smoking at home, (b) Passive smoking at work, (c) vehicle traffic at home, (d) distance in kilometres from etna volcano, and (e) Radon risk according to home type. Volcano Plot analysis (FC ≥ 2, p-value ≤ 0.05, without multiple testing correction) on averaged interpretations were run for each exposure, obtaining the down and up regulated significative miRNAs used as EESs.

The Total gene signal in Agilent Arrays can be normalized between arrays, and the Agilent recommendation is either not to normalize or to normalize to the 75th percentile signal intensity [49]. There is also evidence that it might be preferable to use non-corrected signals for the processing of microRNA data, rather than background-corrected signals. However, it was chosen to use an Agilent Technologies miRNA profiling assay that is based on a highly efficient labelling method that has little sequence bias. Furthermore, the probe design strategy used with Agilent arrays provides both sequence and size discrimination for mature miRNAs, increasing confidence in LCRMs and EESs results. In fact, in quality control step using Principal Component Analysis and Hierarchical Clustering it is clear the clustering of healthy tissues despite of a large heterogeneity of tissues.

To understand the relationship between environmental exposure signatures and their biological significance in lung cancer tissues, a target detection for each environmental exposure signature was done using TargetScan prediction database. This database was chosen as it is the most updated database, and the number of target genes are reported for different cut-offs. The most interesting genes targeted by EESs that are potentially related with each environmental exposure are below discussed.

After target detection, a prediction model was build using Neural Network class prediction algorithm (GeneSpring 14.9) [50] for each Environmental Exposure signature to test the overall accuracy prediction (range from 0 to 1) for chosen miRNAs. EESs prediction overall accuracy were compared with the whole chip array miRNAs'(n=2,570) accuracy.

Bivariate Correlations and regression variable plots between Benzo[a]Pyrene-DNA adducts and the different environmental exposures were calculated with IBM SPSS statistics (Version 22). Correlations were used according to the nature of the data: Pearson for parametric, or Spearman's Rho for non-parametric.

3. Results

3.1. Lung tissues analysed with miRNA-array. Patients' information

Available tumoral tissues were higher (from 38 patients) than healthy (from 12 patients). Detailed data from patients which biopsies were miRNA-array analysed is presented in **Table 1**. Biopsies of both tumoral and healthy tissues were available from only 4 patients (05, 06, 07, and 08). For the other 42 biopsies there were used only tumoral or only healthy tissues for each patient. A total of 46 patients were analysed, of which 34.8% are females, with an average of 64.31 years old, and 65.2 were males with an average of 70.17 years old. Average age of patients for all analysed miRNA-arrays is 68.13. When analysing exposure, we can see that 76.1 % were exposed to passive smoking at home or at work, 67.4% had a high-risk exposure to Radon, and 34.8% have a high exposure to vehicle traffic at home, the median distance from Etna volcano is 54.26 Km. The follow up reported that 23.9% of patients died within 3 years after biopsy.

Table 1 Detailed data of the patients from which tissues were extracted

Patient code	Tissue type	Sex	Age	Passive smoking at home or at work	Radon Risk	Vehicle traffic exposure	3 years survival	Distance from Etna (Km)
05	healthy, tumoral	Male	69	YES	other	low	yes	65
06	healthy, tumoral	Male	67	NO	high	low	no	114
07	healthy, tumoral	Female	71	YES	high	low	yes	93
08	healthy, tumoral	Female	66	YES	high	low	yes	73
09	healthy	Male	78	YES	high	low	no	20
11	healthy	Male	70	NO	high	high	NA	NA
12	healthy	Female	78	YES	high	low	yes	60
13	healthy	Male	65	YES	high	low	no	93
14	healthy	Male	71	YES	high	low	yes	100
15	healthy	Female	72	YES	high	high	yes	99
16	healthy	Male	75	YES	high	high	yes	18
18	healthy	Male	61	YES	other	low	yes	60

26	tumoral	Male	69	YES	other	high	yes	93
27	tumoral	Male	58	YES	high	high	yes	20
30	tumoral	Male	82	YES	high	low	yes	13
32	tumoral	Female	67	YES	other	high	yes	98
33	tumoral	Male	50	NO	high	low	yes	16
34	tumoral	Female	68	YES	high	low	no	26
35	tumoral	Male	61	YES	high	high	yes	20
36	tumoral	Male	54	YES	NA	high	NA	108
37	tumoral	Male	78	YES	other	high	yes	93
38	tumoral	Female	78	YES	high	high	no	93
39	tumoral	Female	47	NO	other	high	yes	66
40	tumoral	Male	75	YES	high	low	yes	22
41	tumoral	Male	69	YES	high	low	yes	NA
42	tumoral	Male	77	YES	high	high	no	19
43	tumoral	Male	70	YES	high	low	yes	79
44	tumoral	Male	76	YES	high	low	yes	51
45	tumoral	Male	64	YES	high	low	yes	16
46	tumoral	Female	80	NO	NA	low	no	21
47	tumoral	Female	43	NO	high	low	no	26
48	tumoral	Female	60	YES	other	low	yes	77
49	tumoral	Male	67	YES	high	low	no	15
50	tumoral	Male	78	NO	high	low	yes	24
52	tumoral	Female	54	NO	high	low	yes	NA
53	tumoral	Female	54	YES	other	high	yes	26
54	tumoral	Female	65	YES	other	low	NA	NA
55	tumoral	Male	78	NO	other	low	no	26
56	tumoral	Male	67	YES	high	low	yes	98
57	tumoral	Male	77	YES	high	low	no	92
58	tumoral	Male	78	YES	other	low	yes	26
59	tumoral	Male	62	YES	high	low	yes	13
60	tumoral	Female	53	NO	other	high	yes	100
61	tumoral	Female	73	NO	high	high	yes	26
63	tumoral	Male	80	YES	other	low	yes	60
64	tumoral	Male	79	YES	high	high	yes	21

3.2. Lung Cancer Related miRNAs (LCRMs) identification

The scatter plot analysis of miRNAs between average healthy and lung cancer tissues presents general trend toward down regulation in cancer tissue we can see a trend to downregulation (m =

0.86, R^2= 0.9) as shown in **Figure 6a**. Volcano plot analysis between healthy and tumoral tissue highlighted a list of 273 miRNAs that were altered more than 2-fold and above the statistical significantly threshold (p-value < 0.05) as clearly demonstrated in **Figure 6b**. In this study the 273 miRNAs mentioned are considered as a cancer footprint called **Lung Cancer Related miRNAs (LCRMs).**

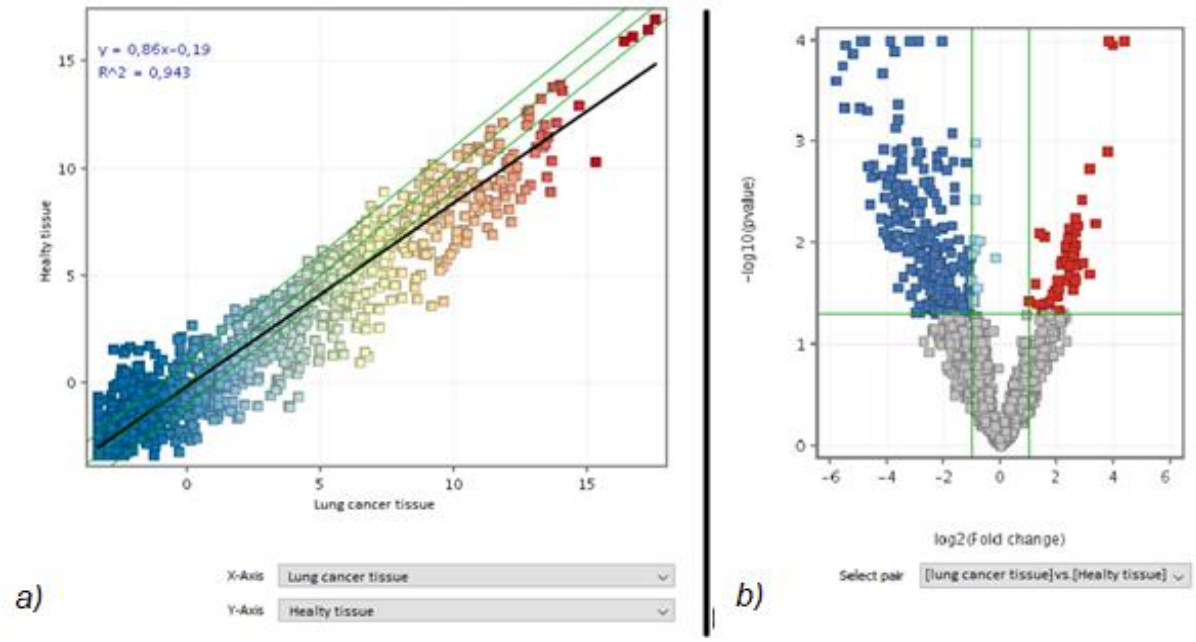


Figure 6 (a) Scatter plot analysis comparing miRNA expression (dots) according to their level of expression in healthy (vertical axis) vs cancer (horizontal axis) tissues of the examined patients. miRNA colour reflects the signal intensity (red high, yellow intermediate, blue low). The diagonal green lines indicate the 2-fold variation interval. The best-fit regression line is reported in black. Its slope towards the horizontal axis reflects the overall downregulation of miRNA expression in cancer as compared to healthy lung tissue. (b) Volcano plot analysis identifying miRNAs whose expression was altered more than 2-fold (horizontal axis) and above the statistical significance threshold (P<0.05) (vertical axis) in cancer vs healthy lung tissue. either downregulated (blue) or upregulated (red).

All LCRMs can be found in **Table 2**. Of these miRNAs, 222 were down-regulated and 51 up-regulated. This list includes well established oncogenic miRNAs such as an extensive downregulation of the highly conservate let-7 miRNA family (let-7a-5p, let-7b-5p, let-7c-5p, let-7d-5p, let-7e-5p, let-7f-5p, let-7g-5p, and let-7i-5p), established tumour-suppressors in lung and other cancers [51].

Table 2 Lung Cancer Related miRNAs (LCRMs) The list of 270 miRNAs (3 positive control "hur-" values were excluded) that resulted as statistically significative after Volcano T-Test between averaged signal of tumoral and healthy tissues.

Systematic name	p-value	Regulation	Fold Change	mirbase accession
hsa-let-7a-5p	0.003457483	down	-9.93267	MIMAT0000062
hsa-let-7b-5p	0.0071048727	down	-4.1737056	MIMAT0000063
hsa-let-7c-5p	0.0027343032	down	-5.8710303	MIMAT0000064
hsa-let-7d-5p	0.0026750325	down	-16.202303	MIMAT0000065
hsa-let-7e-5p	0.009982946	down	-14.271941	MIMAT0000066
hsa-let-7f-5p	0.004075876	down	-14.1158905	MIMAT0000067
hsa-let-7g-5p	0.0064250096	down	-15.127661	MIMAT0000414
hsa-let-7i-5p	0.021721233	down	-5.3294306	MIMAT0000415
hsa-miR-1-3p	1.7767643E-4	down	-46.82256	MIMAT0000416
hsa-miR-100-5p	0.0065184673	down	-9.21915	MIMAT0000098
hsa-miR-101-3p	0.0018394814	down	-6.491789	MIMAT0000099
hsa-miR-101-5p	0.03154675	down	-3.348052	MIMAT0004513
hsa-miR-103a-3p	0.038089182	down	-4.3529544	MIMAT0000101
hsa-miR-106b-3p	0.011158621	up	4.7460146	MIMAT0004672
hsa-miR-107	0.022401135	down	-5.6110463	MIMAT0000104
hsa-miR-10a-5p	0.009016526	down	-15.704841	MIMAT0000253
hsa-miR-10b-5p	0.008520666	down	-16.504122	MIMAT0000254
hsa-miR-1227-5p	0.03029482	down	-2.2108335	MIMAT0022941
hsa-miR-1238-5p	0.023602227	up	4.8538966	MIMAT0022947
hsa-miR-1247-5p	0.0067282193	down	-3.2875102	MIMAT0005899
hsa-miR-125a-5p	9.974076E-4	down	-7.804573	MIMAT0000443
hsa-miR-125b-5p	0.0015694721	down	-7.678459	MIMAT0000423
hsa-miR-126-3p	4.944838E-4	down	-26.00859	MIMAT0000445
hsa-miR-126-5p	2.4788603E-4	down	-54.862705	MIMAT0000444
hsa-miR-128-3p	0.04017438	down	-6.1065207	MIMAT0000424
hsa-miR-129-2-3p	0.038250394	down	-5.874468	MIMAT0004605
hsa-miR-1296-5p	0.032698315	up	3.9346223	MIMAT0005794
hsa-miR-1306-3p	1.1124488E-4	up	15.246751	MIMAT0005950
hsa-miR-130a-3p	0.0039922795	down	-8.522441	MIMAT0000425
hsa-miR-133a-3p	1.1501503E-5	down	-28.701208	MIMAT0000427
hsa-miR-133a-5p	0.0030193352	down	-5.6037283	MIMAT0026478
hsa-miR-133b	1.3488866E-4	down	-36.514313	MIMAT0000770
hsa-miR-135a-5p	9.1589776E-5	down	-27.485888	MIMAT0000428
hsa-miR-138-5p	0.004896176	down	-9.4946995	MIMAT0000430
hsa-miR-139-3p	0.0018985552	down	-3.78872	MIMAT0004552
hsa-miR-139-5p	2.1379872E-4	down	-17.9196	MIMAT0000250
hsa-miR-140-3p	0.0012392761	down	-4.2493033	MIMAT0004597

hsa-miR-140-5p	0.010334321	down	-15.18544	MIMAT0000431
hsa-miR-142-3p	0.021352166	down	-6.275377	MIMAT0000434
hsa-miR-142-5p	0.018636087	down	-3.2859912	MIMAT0000433
hsa-miR-143-3p	0.0029421495	down	-13.065256	MIMAT0000435
hsa-miR-143-5p	0.04291127	down	-3.356232	MIMAT0004599
hsa-miR-144-3p	5.3506505E-5	down	-14.967624	MIMAT0000436
hsa-miR-144-5p	1.0900437E-4	down	-44.189434	MIMAT0004600
hsa-miR-145-3p	0.0032639632	down	-12.659062	MIMAT0004601
hsa-miR-145-5p	2.9446861E-5	down	-8.330996	MIMAT0000437
hsa-miR-146a-5p	0.04517611	down	-4.9618096	MIMAT0000449
hsa-miR-146b-5p	0.0060929195	down	-10.381786	MIMAT0002809
hsa-miR-147b	0.03965416	up	2.4136007	MIMAT0004928
hsa-miR-150-5p	0.0018827947	down	-14.296803	MIMAT0000451
hsa-miR-151a-5p	0.026424088	down	-5.6179104	MIMAT0004697
hsa-miR-151b	0.01971659	down	-5.525759	MIMAT0010214
hsa-miR-152-3p	0.010101993	down	-6.5972843	MIMAT0000438
hsa-miR-1537-3p	0.010019664	down	-5.1548657	MIMAT0007399
hsa-miR-15a-5p	0.019115578	down	-8.025987	MIMAT0000068
hsa-miR-15b-3p	0.037224907	down	-4.462699	MIMAT0004586
hsa-miR-15b-5p	0.008560004	down	-7.314562	MIMAT0000417
hsa-miR-16-2-3p	0.034894716	down	-4.267132	MIMAT0004518
hsa-miR-16-5p	0.010695414	down	-12.1582985	MIMAT0000069
hsa-miR-181a-2-3p	0.017184366	down	-5.96992	MIMAT0004558
hsa-miR-181a-3p	0.04945864	down	-4.950125	MIMAT0000270
hsa-miR-182-3p	0.010721036	up	4.75117	MIMAT0000260
hsa-miR-183-3p	0.007660201	up	6.2313943	MIMAT0004560
hsa-miR-184	7.194409E-4	down	-13.022917	MIMAT0000454
hsa-miR-185-5p	0.023636332	down	-4.4132204	MIMAT0000455
hsa-miR-186-5p	0.043665413	down	-3.685546	MIMAT0000456
hsa-miR-187-5p	6.02259E-4	down	-12.302708	MIMAT0004561
hsa-miR-18a-5p	0.044035003	down	-6.4913006	MIMAT0000072
hsa-miR-18b-5p	0.04321834	down	-4.911199	MIMAT0001412
hsa-miR-190a-5p	0.0012564617	down	-12.553396	MIMAT0000458
hsa-miR-191-5p	0.012516888	down	-4.595004	MIMAT0000440
hsa-miR-1913	0.014741061	up	4.4493365	MIMAT0007888
hsa-miR-193a-5p	0.039410394	down	-2.3893714	MIMAT0004614
hsa-miR-195-5p	0.0022041805	down	-22.629496	MIMAT0000461
hsa-miR-199a-3p	0.008664297	down	-9.2620735	MIMAT0000232
hsa-miR-199a-5p	0.013618913	down	-4.5943766	MIMAT0000231
hsa-miR-199b-5p	0.007679599	down	-17.657621	MIMAT0000263

hsa-miR-203a-3p	0.010997548	down	-13.447799	MIMAT0000264
hsa-miR-204-5p	0.0083169	down	-12.60669	MIMAT0000265
hsa-miR-205-3p	0.020174691	up	8.715979	MIMAT0009197
hsa-miR-20a-5p	0.04280471	down	-6.1101165	MIMAT0000075
hsa-miR-20b-5p	0.027864084	down	-7.881743	MIMAT0001413
hsa-miR-21-3p	0.008667924	up	2.9119663	MIMAT0004494
hsa-miR-214-3p	0.023492103	down	-3.504781	MIMAT0000271
hsa-miR-214-5p	0.021508496	down	-6.83216	MIMAT0004564
hsa-miR-218-5p	4.6615294E-4	down	-45.84953	MIMAT0000275
hsa-miR-22-5p	0.004706939	down	-12.123566	MIMAT0004495
hsa-miR-221-3p	0.030677848	down	-4.129919	MIMAT0000278
hsa-miR-221-5p	0.005295244	down	-12.992223	MIMAT0004568
hsa-miR-222-3p	0.049154088	down	-2.1237488	MIMAT0000279
hsa-miR-223-3p	0.00603705	down	-12.803483	MIMAT0000280
hsa-miR-223-5p	0.013339674	down	-6.1945033	MIMAT0004570
hsa-miR-224-3p	0.04533735	down	-5.148501	MIMAT0009198
hsa-miR-2277-3p	0.010260569	up	6.1839	MIMAT0011777
hsa-miR-23a-3p	0.0102553135	down	-5.1024346	MIMAT0000078
hsa-miR-23a-5p	0.033158194	down	-4.635053	MIMAT0004496
hsa-miR-23b-3p	0.015326893	down	-5.947073	MIMAT0000418
hsa-miR-24-3p	0.037936654	down	-2.909435	MIMAT0000080
hsa-miR-26a-1-3p	0.01689716	down	-3.010686	MIMAT0004499
hsa-miR-26a-5p	0.011874804	down	-4.865339	MIMAT0000082
hsa-miR-26b-3p	0.0139598325	down	-5.665662	MIMAT0004500
hsa-miR-26b-5p	0.00420298	down	-24.594625	MIMAT0000083
hsa-miR-27a-3p	0.04915082	down	-3.2384896	MIMAT0000084
hsa-miR-27a-5p	0.019851089	down	-3.8878102	MIMAT0004501
hsa-miR-27b-3p	0.030439751	down	-4.8090467	MIMAT0000419
hsa-miR-28-5p	0.025002686	down	-6.595582	MIMAT0000085
hsa-miR-2861	0.022837397	down	-2.1317737	MIMAT0013802
hsa-miR-29a-3p	0.03019676	down	-3.2799473	MIMAT0000086
hsa-miR-29a-5p	0.030590033	down	-5.209106	MIMAT0004503
hsa-miR-29b-1-5p	0.011967432	down	-8.264298	MIMAT0004514
hsa-miR-29b-2-5p	0.0150986435	down	-5.095968	MIMAT0004515
hsa-miR-29b-3p	0.027163703	down	-4.8850327	MIMAT0000100
hsa-miR-29c-3p	0.010836463	down	-4.0718822	MIMAT0000681
hsa-miR-29c-5p	0.021635728	down	-5.5772324	MIMAT0004673
hsa-miR-3065-3p	0.007735019	down	-6.404753	MIMAT0015378
hsa-miR-3065-5p	0.034041088	down	-5.067239	MIMAT0015066
hsa-miR-30a-3p	0.0025913727	down	-13.062014	MIMAT0000088

hsa-miR-30a-5p	0.00171489	down	-4.7035885	MIMAT0000087
hsa-miR-30b-5p	0.0026990736	down	-12.653434	MIMAT0000420
hsa-miR-30c-2-3p	0.0075573986	down	-5.009342	MIMAT0004550
hsa-miR-30c-5p	0.008662366	down	-9.249711	MIMAT0000244
hsa-miR-30d-5p	0.003681377	down	-3.0694556	MIMAT0000245
hsa-miR-30e-3p	0.014569481	down	-7.417907	MIMAT0000693
hsa-miR-3149	0.01669215	up	6.3766117	MIMAT0015022
hsa-miR-3188	0.03555318	down	-2.7369933	MIMAT0015070
hsa-miR-32-5p	0.00875309	down	-9.987482	MIMAT0000090
hsa-miR-324-3p	0.011736013	down	-2.2862568	MIMAT0000762
hsa-miR-326	0.003936705	down	-7.7617455	MIMAT0000756
hsa-miR-328-3p	0.047037996	down	-3.3454077	MIMAT0000752
hsa-miR-331-3p	0.018208742	down	-3.8520381	MIMAT0000760
hsa-miR-335-5p	0.003513045	down	-16.721733	MIMAT0000765
hsa-miR-338-3p	0.004935093	down	-11.228805	MIMAT0000763
hsa-miR-339-5p	0.029144458	down	-3.6802807	MIMAT0000764
hsa-miR-340-5p	0.01973299	down	-6.4123063	MIMAT0004692
hsa-miR-342-3p	0.0152998995	down	-4.362609	MIMAT0000753
hsa-miR-342-5p	0.0022582049	down	-14.524604	MIMAT0004694
hsa-miR-34a-3p	0.0061846743	down	-10.137312	MIMAT0004557
hsa-miR-34a-5p	0.033176474	down	-3.9531028	MIMAT0000255
hsa-miR-34b-5p	0.012189897	down	-8.055832	MIMAT0000685
hsa-miR-34c-5p	0.027579289	down	-12.216356	MIMAT0000686
hsa-miR-3607-3p	0.0064467057	down	-6.376298	MIMAT0017985
hsa-miR-361-3p	0.009842584	down	-5.972705	MIMAT0004682
hsa-miR-362-3p	0.0055012233	down	-13.304245	MIMAT0004683
hsa-miR-362-5p	0.046346925	down	-3.9805145	MIMAT0000705
hsa-miR-3620-5p	0.049587507	up	2.7487247	MIMAT0022967
hsa-miR-363-3p	0.002120064	down	-17.233507	MIMAT0000707
hsa-miR-3659	8.8144276E-5	up	14.118807	MIMAT0018080
hsa-miR-365a-3p	0.012557839	down	-7.036109	MIMAT0000710
hsa-miR-3660	0.021218026	up	5.925379	MIMAT0018081
hsa-miR-3663-5p	0.0037759703	up	7.122243	MIMAT0018084
hsa-miR-371b-5p	0.0028327415	down	-3.1190054	MIMAT0019892
hsa-miR-374a-5p	0.005773257	down	-18.297781	MIMAT0000727
hsa-miR-374b-5p	0.008157671	down	-12.894991	MIMAT0004955
hsa-miR-374c-5p	0.0031787313	down	-11.641886	MIMAT0018443
hsa-miR-424-5p	0.041307494	down	-7.369182	MIMAT0001341
hsa-miR-4252	0.006666235	up	6.59057	MIMAT0016886
hsa-miR-4254	0.04602684	up	4.1047406	MIMAT0016884

hsa-miR-4284	0.019753091	down	-3.9356518	MIMAT0016915
hsa-miR-4290	0.0057441983	up	6.205229	MIMAT0016921
hsa-miR-4306	0.019319808	down	-2.4259553	MIMAT0016858
hsa-miR-4317	0.0077170506	down	-7.3803926	MIMAT0016872
hsa-miR-4318	0.042405307	down	-3.7857358	MIMAT0016869
hsa-miR-4324	1.2660767E-4	down	-13.353904	MIMAT0016876
hsa-miR-4328	1.7068113E-6	down	-30.341303	MIMAT0016926
hsa-miR-4440	0.028792787	up	3.7428448	MIMAT0018958
hsa-miR-4443	0.008061844	up	2.591256	MIMAT0018961
hsa-miR-4481	0.009133603	up	5.6318855	MIMAT0019015
hsa-miR-449a	0.03263379	down	-8.910406	MIMAT0001541
hsa-miR-450a-5p	0.049076617	down	-7.218559	MIMAT0001545
hsa-miR-4516	0.013548022	down	-2.567815	MIMAT0019053
hsa-miR-451a	4.6308115E-4	down	-31.666576	MIMAT0001631
hsa-miR-452-5p	0.033777237	down	-4.1409574	MIMAT0001635
hsa-miR-4521	3.1114705E-6	down	-7.476263	MIMAT0019058
hsa-miR-4532	8.101723E-4	down	-3.3207798	MIMAT0019071
hsa-miR-454-3p	0.0060575623	down	-15.596857	MIMAT0003885
hsa-miR-455-3p	0.01230269	down	-9.817139	MIMAT0004784
hsa-miR-455-5p	0.01936017	down	-6.943761	MIMAT0003150
hsa-miR-4634	0.017753901	down	-2.1694999	MIMAT0019691
hsa-miR-4655-3p	0.016215164	down	-4.4984784	MIMAT0019722
hsa-miR-4695-3p	0.029992351	up	3.6855245	MIMAT0019789
hsa-miR-4716-5p	0.013365842	up	4.945816	MIMAT0019826
hsa-miR-4730	0.0064160675	up	10.25193	MIMAT0019852
hsa-miR-4731-3p	0.038894046	up	3.0328867	MIMAT0019854
hsa-miR-4763-5p	0.034002375	up	3.8233552	MIMAT0019912
hsa-miR-4770	0.00342172	down	-8.7577715	MIMAT0019924
hsa-miR-4793-3p	9.542979E-5	up	20.642054	MIMAT0019966
hsa-miR-483-3p	0.049434304	up	4.230489	MIMAT0002173
hsa-miR-486-3p	0.0014036096	down	-4.7451825	MIMAT0004762
hsa-miR-486-5p	9.567448E-5	down	-9.390969	MIMAT0002177
hsa-miR-489-3p	0.0090776775	down	-8.160073	MIMAT0002805
hsa-miR-490-3p	0.0039292807	down	-4.759613	MIMAT0002806
hsa-miR-491-3p	0.014214844	up	5.6008396	MIMAT0004765
hsa-miR-497-3p	0.0015522552	down	-3.215224	MIMAT0004768
hsa-miR-497-5p	0.002430224	down	-5.8345428	MIMAT0002820
hsa-miR-5001-5p	0.0139284525	down	-2.0481567	MIMAT0021021
hsa-miR-500a-3p	0.04790341	down	-3.5717416	MIMAT0002871
hsa-miR-500b-5p	0.011883656	down	-4.8304405	MIMAT0016925

hsa-miR-501-5p	0.03635382	down	-4.249691	MIMAT0002872
hsa-miR-502-3p	0.02129167	down	-5.070547	MIMAT0004775
hsa-miR-504-3p	0.045502603	up	3.8429651	MIMAT0026612
hsa-miR-505-5p	0.0050095418	down	-7.1938596	MIMAT0004776
hsa-miR-511-3p	0.0011597008	down	-17.620407	MIMAT0026606
hsa-miR-516b-5p	0.03268391	down	-4.76173	MIMAT0002859
hsa-miR-517-5p	0.0349356	down	-2.7619033	MIMAT0002851
hsa-miR-517a-3p	0.0016572509	down	-11.549932	MIMAT0002852
hsa-miR-517c-3p	0.0018382822	down	-11.30683	MIMAT0002866
hsa-miR-521	0.0012756404	down	-7.0095363	MIMAT0002854
hsa-miR-522-3p	0.0012633094	down	-7.532452	MIMAT0002868
hsa-miR-532-3p	0.002999619	down	-8.658704	MIMAT0004780
hsa-miR-532-5p	0.030358983	down	-5.574835	MIMAT0002888
hsa-miR-548aa	0.039962247	up	3.0047872	MIMAT0018447
hsa-miR-548aw	0.004633741	down	-4.355515	MIMAT0022471
hsa-miR-548b-3p	0.0032465165	down	-5.1606574	MIMAT0003254
hsa-miR-548c-3p	0.009904102	down	-3.576939	MIMAT0003285
hsa-miR-548f-3p	9.7490294E-5	down	-4.2154775	MIMAT0005895
hsa-miR-548q	0.0029641336	down	-9.744756	MIMAT0011163
hsa-miR-548x-3p	0.030913204	down	-3.1084898	MIMAT0015081
hsa-miR-551b-3p	0.0022457927	down	-23.905922	MIMAT0003233
hsa-miR-5701	0.001176325	down	-11.373918	MIMAT0022494
hsa-miR-582-5p	0.0016798502	down	-22.03724	MIMAT0003247
hsa-miR-585-3p	0.031742055	down	-3.1072986	MIMAT0003250
hsa-miR-590-5p	0.021334562	down	-10.855298	MIMAT0003258
hsa-miR-595	0.021010172	up	5.3233123	MIMAT0003263
hsa-miR-598-3p	0.011142222	down	-9.39129	MIMAT0003266
hsa-miR-6068	0.027576268	down	-2.2664123	MIMAT0023693
hsa-miR-6073	4.2795646E-4	down	-12.323209	MIMAT0023698
hsa-miR-6075	0.03862527	down	-3.5490692	MIMAT0023700
hsa-miR-610	0.008825424	up	5.1574306	MIMAT0003278
hsa-miR-624-5p	0.02706973	down	-4.596437	MIMAT0003293
hsa-miR-628-5p	0.047291912	down	-4.5065737	MIMAT0004809
hsa-miR-642b-5p	0.0017398077	down	-4.8533616	MIMAT0022736
hsa-miR-6500-5p	0.016187562	up	4.347074	MIMAT0025454
hsa-miR-6516-3p	0.0014973093	down	-11.282737	MIMAT0030418
hsa-miR-652-3p	0.011874442	down	-5.034457	MIMAT0003322
hsa-miR-653-3p	0.032237496	down	-2.1033967	MIMAT0026625
hsa-miR-660-5p	0.017879974	down	-8.998719	MIMAT0003338
hsa-miR-664a-3p	0.04644798	down	-2.4924498	MIMAT0005949

hsa-miR-664b-3p	0.040972516	down	-2.5689476	MIMAT0022272
hsa-miR-6716-3p	0.0012428381	up	13.550869	MIMAT0025845
hsa-miR-6722-5p	0.01525621	up	4.7678285	MIMAT0025853
hsa-miR-6730-3p	0.02344632	up	4.014452	MIMAT0027362
hsa-miR-6743-3p	0.029110048	up	5.8719335	MIMAT0027388
hsa-miR-6771-5p	0.003915364	down	-6.763831	MIMAT0027442
hsa-miR-6779-3p	0.012219111	up	6.0111957	MIMAT0027459
hsa-miR-6794-3p	0.037364304	up	3.268569	MIMAT0027489
hsa-miR-6804-5p	0.013194768	up	5.809477	MIMAT0027508
hsa-miR-6806-5p	0.0077665085	down	-6.1864467	MIMAT0027512
hsa-miR-6817-5p	0.023182526	up	6.154842	MIMAT0027534
hsa-miR-6826-5p	0.03771252	up	2.0217252	MIMAT0027552
hsa-miR-6865-3p	0.008403192	down	-5.610171	MIMAT0027631
hsa-miR-6872-3p	0.02524211	up	2.3172631	MIMAT0027645
hsa-miR-6886-3p	0.015608697	up	7.481634	MIMAT0027673
hsa-miR-6891-3p	0.0070896316	up	5.3324924	MIMAT0027683
hsa-miR-6895-5p	0.012817935	down	-5.096586	MIMAT0027690
hsa-miR-7108-3p	0.031871215	up	3.4731681	MIMAT0028114
hsa-miR-7111-3p	0.029639475	down	-2.9728024	MIMAT0028120
hsa-miR-7159-5p	0.016500017	up	5.289905	MIMAT0028228
hsa-miR-744-5p	0.0026703363	down	-10.007609	MIMAT0004945
hsa-miR-770-5p	0.0018573835	up	8.820649	MIMAT0003948
hsa-miR-7704	0.025055826	down	-2.0992706	MIMAT0030019
hsa-miR-8063	0.015703373	down	-2.599198	MIMAT0030990
hsa-miR-8077	0.028724372	up	4.096299	MIMAT0031004
hsa-miR-874-5p	0.023569982	down	-4.3161435	MIMAT0026718
hsa-miR-887-3p	0.028764278	down	-3.3909423	MIMAT0004951
hsa-miR-940	0.0016014805	down	-2.3139522	MIMAT0004983
hsa-miR-95-3p	0.0482305	down	-7.970016	MIMAT0000094
hsa-miR-98-5p	0.0017507627	down	-25.522917	MIMAT0000096
hsa-miR-99a-3p	0.0025130045	down	-6.4765115	MIMAT0004511
hsa-miR-99a-5p	0.002272892	down	-14.822669	MIMAT0000097
hsa-miR-99b-5p	0.006868263	down	-4.111178	MIMAT0000689

3.3. Quality Control Analysis of LCRMs as indicators in healthy and tumoral tissues

LCRMs down regulation trend was well distinguishable between cancer and healthy tissue as indicated by Hierarchical Cluster Analysis, where healthy tissue profiles (yellow bar) were clustered in the upper part of the hierarchical tree separately from cancer tissue profiles (blue bar), colour range indicates LCRMs intensity signal (**Figure 7a**). In Principal Component Analysis of Variance, healthy tissue samples (yellow dots) clustered in the lower left part of the 3D space. Instead, cancer tissue samples (blue dots) are located along the 3D space, demonstrating how different each patient's tumoral tissue was (**Figure 7b).**

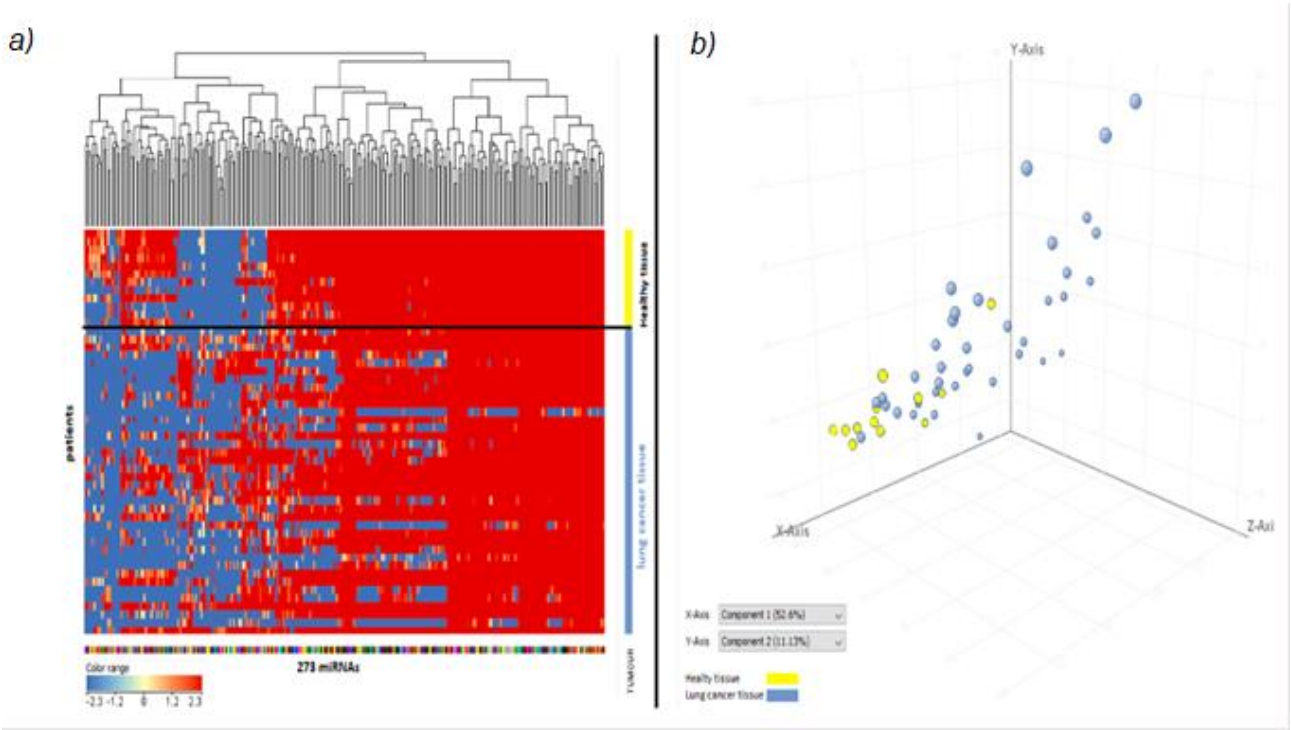


Figure 7 (a) Clustering Hierarchical Analysis reporting the Expression of the 273 LCRMs (x axes) in Healthy tissue (vertical axis, yellow bar) and lung cancer tissue (vertical axis, blue bar) in the 50 samples tested (y axis). (b) Principal Component Analysis of Variance reporting identifying samples from healthy tissues (yellow dots) and samples from cancer tissues (blue dots) according to the variance of the expression of the 273 LCRMs, dots size = Principal Component Analysis score.

3.4. Gene Ontology and Biological Processes (GO-BP) enrichment analysis

After identifying the top 10 most significative predicted target genes for each one of the LCRMs, the GO-BP enrichment analysis was run using TOP-Go Bioconductor R package and REVIGO online tool as explained in Chapter 2.9. In **Figure 8** is presented the treemap of the most representative Biological Processes for LCRMs (around 2000 GO terms), where the most representative BPs are: regulation of RNA splicing, tissue migration, monosaccharide transmembrane transport, protein modification by small protein conjugation or removal and cellular protein-containing complex assembly. Other less representative BPs are: small molecule metabolic process, lipid metabolic process, carbohydrate derivative biosynthetic process, myeloid leukocyte activation, cellular ketone metabolic process, cell cycle process, regulation of transcription DNA-templated, cellular catabolic process, sulfuric compound metabolic process, gene expression, biological process involved in interspecies interaction between organisms, cellular process, biological process involved in symbiotic interaction, cellular component organization or biogenesis, protein localization to cell periphery, biological regulation, glycosylation, chromatin organization, macromolecule modification, ion transport, chromosome organization, multicellular organismal process, and cellular protein metabolic process.

The most representative GO term is regulation of RNA splicing (GO:0008380) that include other processes as regulation of cell cycle, cellular response to organic cyclic compound, positive regulation of cell population proliferation, and other 55 target functions. GO:0008380 definition is "*The process of removing sections of the primary RNA transcript to remove sequences not present in the mature form of the RNA and joining the remaining sections to form the mature form of the RNA*", and it includes 404 genes in GO library, including negative regulation of RNA splicing (**Figure 9a**). Negative regulation of mRNA is the final biological function of miRNAs, they act as imperfect sequence guides to recruit a ribonucleoprotein (RNP) complex to the complementary RNA. The microRNA-RNP complex is called the RNA-induced silencing complex (or RISC), that uses a small RNA (microRNA) to direct sequence-specific recruitment of the RISC to its target RNA modulating its expression. This is conceptually similar to mRNA splicing, where small nuclear RNAs (U1, U2 and U4-

6) act through complementarity to sequences at the splice and branch sites within the intron that determine mRNA splicing [52].

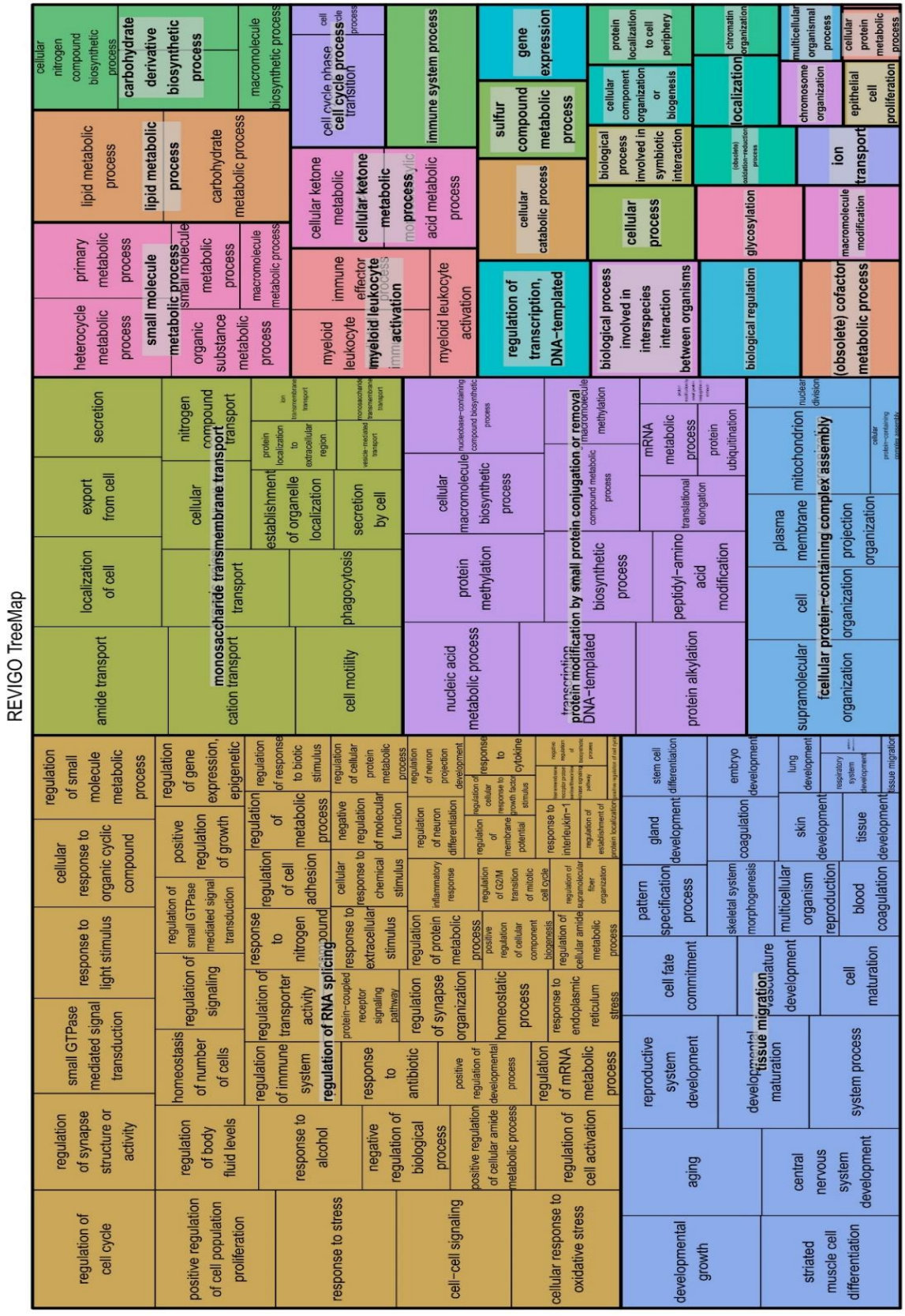


Figure 8 treemap of the most representative Biological Processes for LCRMs

It is plausible that a large set of genes targeted by LCRMs, are also responsible of RNA splicing regulation as a feedback mechanism of gene regulation in tumoral mechanisms. The second most significative GO is 0090130, tissue migration, a well-known biological process in embryogenesis as well as in tumoral processes [53].Tissue migration GO includes 208 genes and 444 GO annotations (Figure 9b). Almost all predicted BPs connected to tissue migration belong to developmental growth as shown in Figure 7, but in the GO treemap there are also present GO terms connected more specifically to tumoral processes as aging, cell fate, tissue development, and stem cell differentiation. However, the hypothesis of LCRMs altering RNA splicing as well as tissue migration should be study in deep in future.

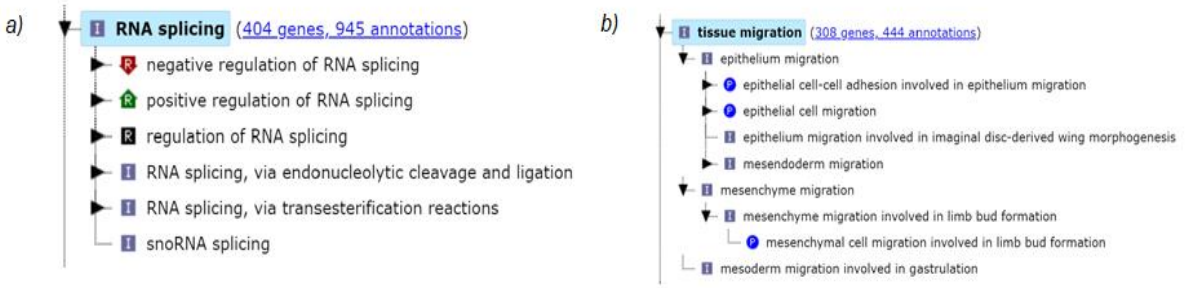


Figure 9 GO tree view of a) GO: 0008380 and b) GO: 0090130. Information available at: http://www.informatics.jax.org/vocab/gene_ontology/

3.5. LCRMs related to cancer isotype

From 64 patients, only 22 had an histotype record: 10 with NSCLC and 33 with SCLC. Using only the list of LCRMs described in **Table 2,** it was analysed how miRNA patterns changed between NSCLC and the most aggressive SCLC. In the scatter plot (**Figure 10a**) we can see a slightly (m = 1.02) but significative (R^2= 0.9) trend to upregulation in NSCLC compared to SCLC. The volcano plot t-test (FC≥2 and p-value ≤ 0.05) identified that 25 LCRMs were upregulated and only 1 (miR-326) was downregulated in NSCLC (**Figure 10b**), which means that the same miRNAs are down-regulated in SCLC.

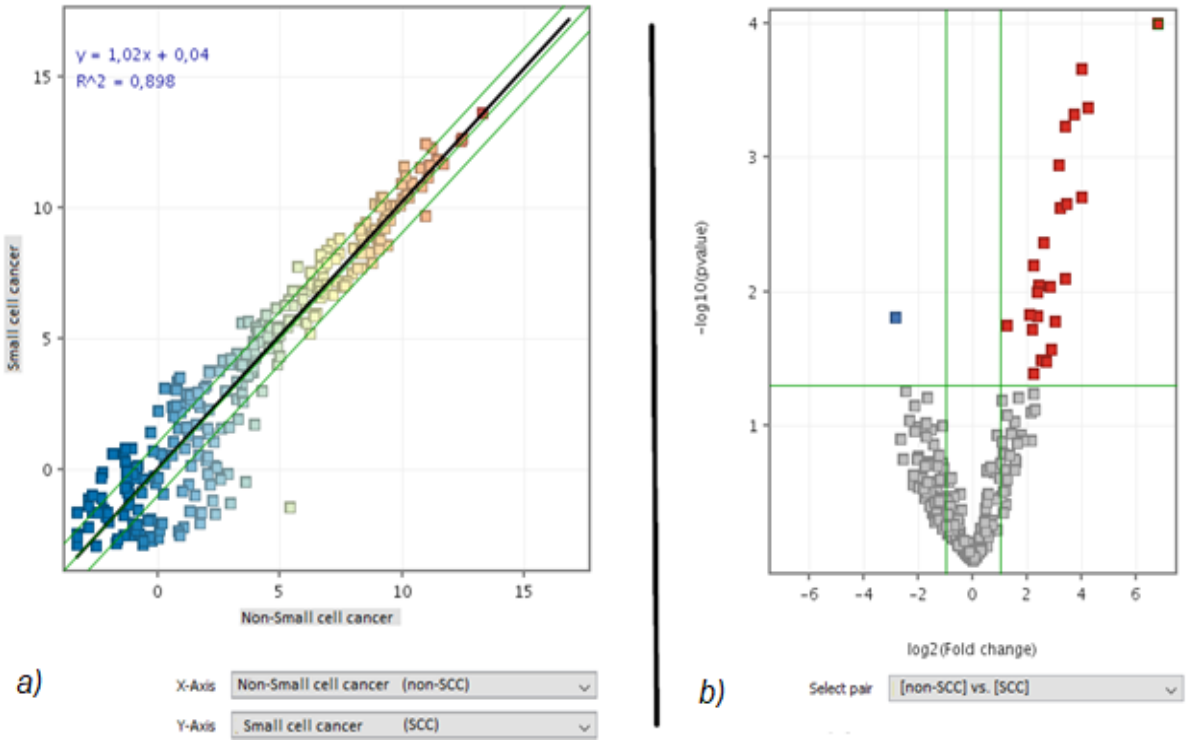


Figure 10 a) scatter plot and b) volcano plot t-test between non-small-cells lung cancer (NSCLC) and small-cells lung cancer (SCLC) using only Lung Cancer Related miRNAs.

The miRNAs resulting from the volcano plot t-test can be found in **Table 3**. Volcano plot analysis indicated that 26 out of the 273 cancer related miRNAs were differentially expressed between NSCLC and SCLC, which were used to run a GO-BP analysis to clarify what biological processes may be involved in the difference between histotypes.

Table 3 miRNAs result from the volcano plot analysis of only Lung Cancer Related miRNAs (n = 273) between non-small-cells lung cancer (NSCLC) and small-cells lung cancer (SCLC). An up-regulation of 25 miRNAs can be found for NSCLC.

miRNA	p-value	Regulation	FC
hsa-miR-1238-5p	0.011302715	up	9.62
hsa-miR-1296-5p	0.007524842	up	8.62
hsa-miR-1306-3p	0.021063296	up	8.58
hsa-miR-205-3p	4.19297E-06	up	220.64
hsa-miR-2277-3p	0.00932873	up	10.59
hsa-miR-3149	0.000521124	up	25.38
hsa-miR-326	0.019960763	down	-8.85
hsa-miR-4290	0.000451764	up	16.39
hsa-miR-4440	0.002965666	up	10.05
hsa-miR-4443	0.03872927	up	2.41
hsa-miR-4481	0.015899722	up	7.33
hsa-miR-4716-5p	0.000232035	up	20.00
hsa-miR-4763-5p	0.008582705	up	8.26
hsa-miR-4793-3p	0.002927522	up	16.70
hsa-miR-483-3p	0.002941658	up	13.22
hsa-miR-504-3p	0.002079689	up	14.95
hsa-miR-595	0.000426941	up	21.36
hsa-miR-6730-3p	0.016724579	up	6.91
hsa-miR-6743-3p	0.043577574	up	7.71
hsa-miR-6779-3p	0.016434822	up	8.81
hsa-miR-6794-3p	0.005262722	up	7.19
hsa-miR-6817-5p	0.044506542	up	7.97
hsa-miR-6826-5p	0.02230857	up	2.63
hsa-miR-6886-3p	0.014682693	up	12.86
hsa-miR-6891-3p	0.01124693	up	7.93
hsa-miR-7108-3p	0.012961295	up	6.21

The treemap of the most important biological processes altered by miRNAs in **Table 3** identified that the most deregulated process is GO: 0050878, or regulation of body fluid levels, that refers to

"Any process that modulates the levels of body fluids". GO: 0050878 counts with 368 genes and 624 annotations that includes body fluid secretion, homeostasis, multicellular organismal water homeostasis, regulation of mucus secretion, and secretion in different organs (**Figure 11a**). It is not clear how this predicted process is different between NSCLC and SCLC, but we can hypothesise that it may be linked to one of the principal difference between this isotypes, as SCLC cells exhibit neuroendocrine features evidenced by the expression of a variety of markers like thyroid transcription factor 1 (TTF1) [54], synaptophysin (SYP), chromogranin A (CHGA), and gastrin releasing peptide (GRP) [55].

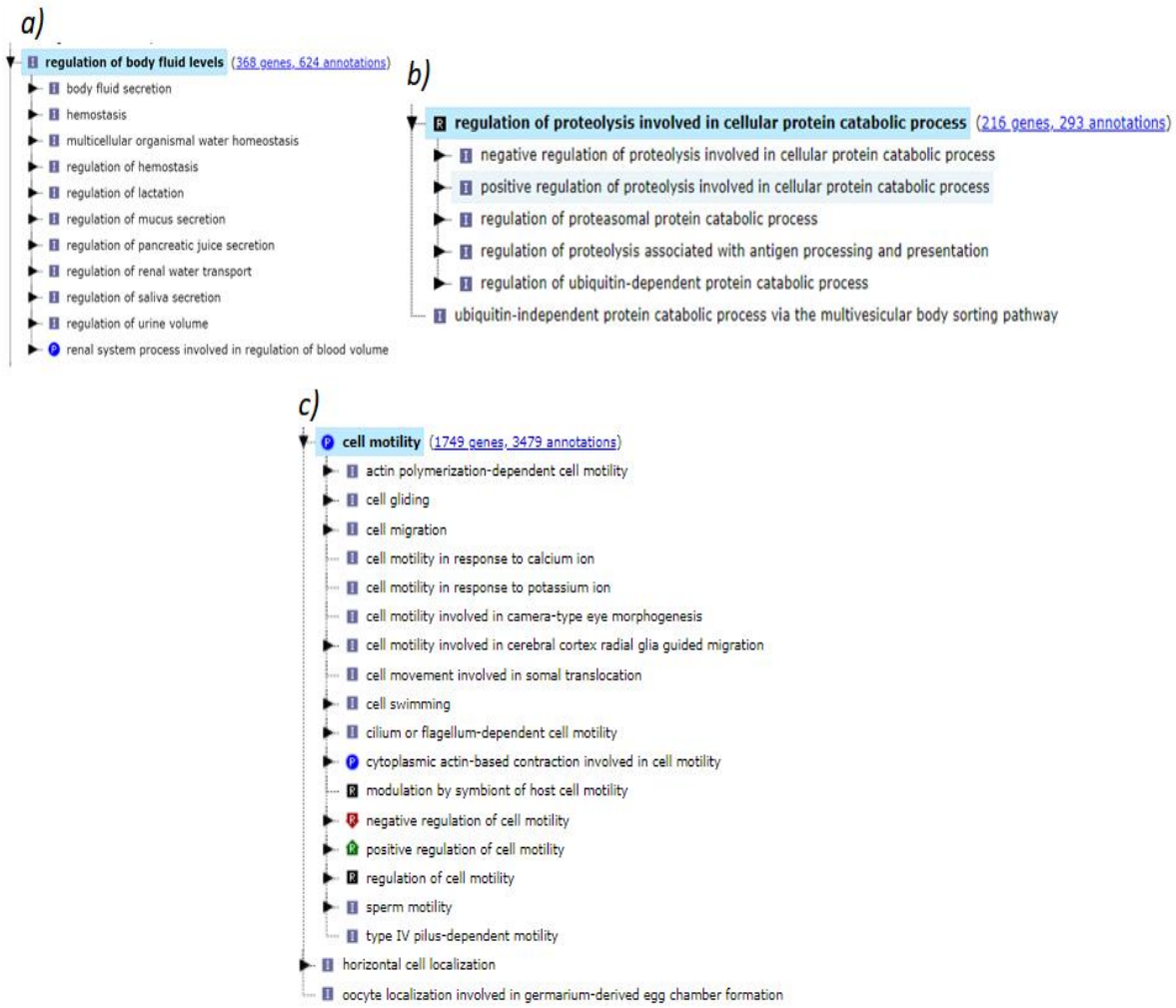


Figure 11 Most of representative predicted biological processes impacted by significative t-test dysregulated miRNAS between non-small cell cancer vs. small cell lung cancer tissues. a)GO: 0050878, b)GO:1903050, and c)GO:0048870. Information available at: http://www.informatics.jax.org/vocab/gene_ontology/

After GO: 0050878, the second most significative biological process is GO:1903050, or regulation of proteolysis involved in cellular protein catabolic process (**Figure 11b**). The third, and most

interesting predicted biological process is GO:0048870, or cell motility, that is other of the characteristic difference between small and non-small histotype. Cell motility GO includes 1749 genes and 3479 annotations (Figure 11c), that makes a vast biological process across the organism. We also should notice that in the case of the Table 4 miRNAs, 25 out of 26 are upregulated in NSCLC, what means that mRNA from genes connected with cell motility process are inhibited as expected for NSCLC.

The mentioned predicted biological processes, that match with the well-known biological behaviour and phenotype of the respective types of lung cancer, would be attributable to the differential expression of miRNAs in Table 3, making them a plausible histotype specific marker. The treemap of the most important biological processes altered between NSCLC vs. SCLC is in Figure 12.

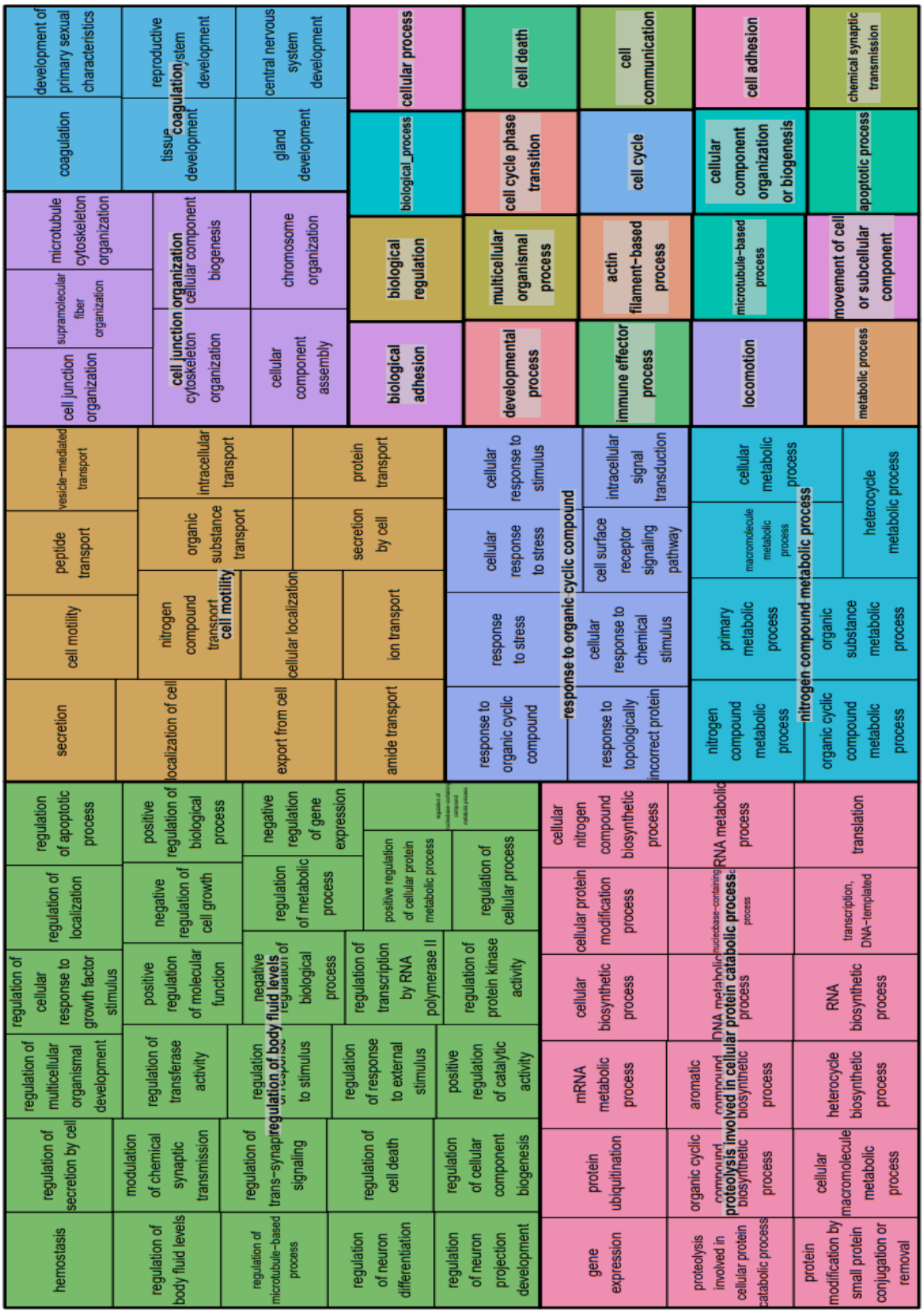


Figure 12 Treemap of the predicted biological processes impacted by significative t-test dysregulated miRNAS between non-small cell cancer vs. small cell lung cancer.

3.6. Oncogenes mutations and miRNA profile variation

Lung cancer in non-smokers have a genetic predisposition. Many studies have shown that patients with a family history of lung cancer are at increased risk [56]. Certain acquired mutations in oncogenes and tumour-suppressors are found more frequently in never-smokers compared with smokers.

Mutations in the Epidermal Growth Factor Receptor gene (EGFR) are found in approximately 40 percent of never smokers' biopsies. EGFR, a tyrosine kinase receptor, plays a critical role in cell differentiation, development, proliferation and homeostasis. Another common mutation in non-smokers are those in oncogenic Kirsten Rat Sarcoma Virus gene (KRAS). This kind of mutations involve point mutations in codons 12 or 13 in exon. KRAS encodes GTPase activity in proteins that regulate cell growth, differentiation, and apoptosis and serves as a downstream mediator of EGFR-induced cell signalling. KRAS mutation may affect chromosomal translocations and rearrangements in a hotspot gene called Anaplastic Lymphoma Receptor Tyrosine kinase (ALK), which in consequence is also more frequently mutated in non-smokers [57].

Tumoral Protein 53 (TP53) is a tumour-suppressor gene integrates numerous signals that control cell life and death. It is the most frequently mutated gene in different types of cancer. Mutation in TP53 most of the times results in the expression of a protein that has lost its function, and therefore is unable to coordinate transcription process that ultimately contribute to tumour suppression. In other cases, mutant TP53 proteins also acquire oncogenic properties that enable invasion promoting, metastasis, proliferation and cell survival. However, there's no evidence that TP53 mutation have a higher frequency in non-smokers rather than in smokers [58].

For this study it was verify the existence of mutations in a group of oncogenes and tumour-suppressors from patients' tumoral tissue biopsies. Other than the already explained EGFR, KRAS, ALK and TP53, there were analysed Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha oncogene (PIK3CA), Erb-B2 Receptor Tyrosine Kinase2 oncogene (ERBB2), Serine/threonine kinase 11 tumour-suppressor (STK11), B-Raf kinase proto-oncogene (BRAF), Phosphatase and Tensin

Homolog Oncogene (PTEN), Mitogen-Activated Protein 2 Kinase 1 proto-oncogene (MAP2K1), Fibroblast Growth Factor Receptor oncogene (FGFR), and other 4 genes which mutations were not found.

Despite the presence of lung cancer, no mutation was observed in 10 out of the 52 examined patients (19,2%). Mutations were observed in 42 patients with the following frequency: TP53 (36.54%), KRAS (30.77%), EGFR (25%), PIK3CA (13.46%), PTEN (9.62%), STK11 (5.77%), BRAF (3.85%), MAP2K1 (1.92%), FGFR (1.92%), and ERBB2 (1.92%) (Figure 13).

	Frequency in patients	Percentage
AKT1	1	1.92
ALK	1	1.92
BRAF	2	3.85
DDR2	0	0.00
EGFR	13	25.00
ERBB2	1	1.92
ERBB2	1	1.92
FBXW7	1	1.92
FGFR3	1	1.92
KRAS	16	30.77
MAP2K1	1	1.92
MET	0	0.00
NOTCH1	6	11.54
PIK3CA	7	13.46
PTEN	5	9.62
SMAD4	0	0.00
STK11	3	5.77
TP53	19	36.54
No Mutations	10	19.23
Total Patients	52	

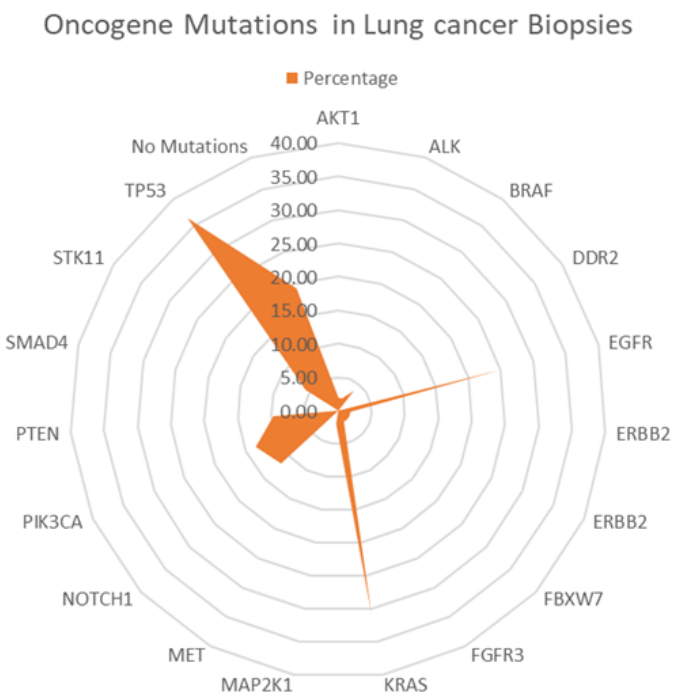


Figure 13. Radar chart showing the frequency of mutations in lung byopsies of analyzed patients. TP53 (36.54%), KRAS (30.77%), EGFR (25%), PIK3CA (13.46%), NOTCH1(11.54%) are the most frequent mutations on analyzed patients

Only 14 Patients (26%) carried mutations targetable by available precision medicine therapies (EGFR 25 mutations, BRAF 2 mutations, AKT 2 mutations), but it was impossible to understand if patients had access to these therapies. Moreover, 25 out of 52 patients presented more than one mutation as explained in Table 4.

Table 4 Frequency of single, double, triple, quadruple and quintuple mutations in lung biopsies from analysed patients.

		Frequency	*Percentage*
Number of mutations	*0*	10	19.2
	1	17	32.7
	2	19	36.5
	3	2	3.8
	4	3	5.8
	5	1	1.9
	Total	52	100.0

From the 38 patients in which both miRNA and mutations have been analysed, 33 were patients carrying at least 1 mutation. Mutational status affects miRNA expression but as tumoral tissue samples were biased with the presence of more than one mutation, not enough data was available to analyse a specific miRNA footprint for each mutation. However, a Volcano Plot T-test was run where possible, using the list of the 273 LCRMs. There were found altered miRNAs only for BRAF, EGFR, KRAS, NOTCH, PIK3CA, PTEN, STK11, and TP53. The number of miRNAs altered in tumoral tissue as associated with mutation is reported in Table 5. Moreover, it was verified if any altered miRNA in the list directly targeted the tested gene using Targetscan database (http://www.targetscan.org/). The most interesting result were the alterations of miR-15b-3p and hsa-miR-21-3p, both of which targeted directly KRAS gene. KRAS and TP53 were the most frequently mutated genes across tumoral tissues, but the miRNA signature in KRAS is stronger as it present 13 entities. Other results are limited, probably because of the lack of single-mutated tissue. For example, the signature of STK1 is the strongest as it present 31 entities, but not reliable as it was obtained using only 2 tumoral tissues.

Table 5 Volcano plot moderated T-Test of miRNA-chip-arrays from tumoral lung biopsies using only Cancer Associated miRNAs (from Table3). P-value <= 0.05, FC>=2.0, no correction, [mutated] vs. [no mutated]. miRNAs entities are listed in Annex3, Table S2. Only 4 miRNAs targeted directly considered genes.

Gene	Mutated/Total patients	Number of entities (altered miRNAs)	up	down	N° miRNAs targeting oncogene
BRAF	2/33	7	1	6	0
EGFR	9/33	1	0	1	0
ERBB2	1/33	-	-	-	-
ERBB4	1/33	-	-	-	-
FGFR3	1/33	-	-	-	-
KRAS	10/33	13	13	0	2 (hsa-miR-15b-3p, hsa-miR-21-3p)
NOTCH1	3/33	1	1	0	0

PIK3CA	3/33	0	0	0	0
PTEN	2/33	0	0	0	0
STK11	2/33	31	30	1	1 (hsa-miR-548aa)
TP53	10/33	4	4	0	1 (hsa-miR-205-3p)

Notwithstanding the low reliability of results, a list of all deregulated miRNAs fount after Volcano Plot test are reported in **Table 6**. As we can see, mutation determines an up regulation in most of miRNAs in KRAS, STK11, NOTCH and TP53. The only mutation that determines miRNAs down-regulation are BRAF and EGFR. These results are not clear and should be studied in-depth.

Table 6 Significant altered miRNAs (Volcano plot moderated t-test, [mutated tissues] vs. [no mutated tissues], FC>=2, p<=0,05, no correction) between tumoral tissues of a total of 33 patients by Gene mutation, using only Cancer Associated miRNAs from Table 3. TargetScan predicted if each miRNA targeted directly considered genes.

Mutation	systematic_name	p-value	Regulation	FC	mirbase accession No	Direct Target
BRAF	hsa-miR-1306-3p	0.024846	down	-32.20	MIMAT0005950	NO
BRAF	hsa-miR-139-3p	0.01998	down	-11.09	MIMAT0004552	NO
BRAF	hsa-miR-193a-5p	0.049849	down	-7.84	MIMAT0004614	NO
BRAF	hsa-miR-3620-5p	0.009772	down	-16.80	MIMAT0022967	NO
BRAF	hsa-miR-3659	0.015827	down	-35.46	MIMAT0018080	NO
BRAF	hsa-miR-521	0.00369	up	13.25	MIMAT0002854	NO
BRAF	hsa-miR-610	0.042202	down	-15.90	MIMAT0003278	NO
EGFR	hsa-miR-744-5p	0.048781	down	-6.35	MIMAT0004945	NO
KRAS	hsa-miR-106b-3p	0.009659	up	5.78	MIMAT0004672	NO
KRAS	hsa-miR-1247-5p	0.026188	up	2.56	MIMAT0005899	NO
KRAS	hsa-miR-1306-3p	0.045089	up	5.07	MIMAT0005950	NO
KRAS	hsa-miR-1537-3p	0.035001	up	4.21	MIMAT0007399	NO
KRAS	hsa-miR-15b-3p	0.048594	up	4.27	MIMAT0004586	YES
KRAS	hsa-miR-191-5p	0.041432	up	3.98	MIMAT0000440	NO
KRAS	hsa-miR-1913	0.005619	up	7.55	MIMAT0007888	NO
KRAS	hsa-miR-21-3p	0.016535	up	3.35	MIMAT0004494	YES
KRAS	hsa-miR-4440	0.027506	up	5.09	MIMAT0018958	NO
KRAS	hsa-miR-4793-3p	0.04182	up	6.36	MIMAT0019966	NO
KRAS	hsa-miR-6516-3p	0.046655	up	6.33	MIMAT0030418	NO
KRAS	hsa-miR-6804-5p	0.001817	up	12.40	MIMAT0027508	NO
KRAS	hur_5	0.033032	up	2.67	-----	------

NOTCH1	hsa-miR-3065-5p	0.02177	down	-28.11	MIMAT0015066	NO
STK11	hsa-miR-100-5p	0.044152	up	50.39	MIMAT0000098	NO
STK11	hsa-miR-106b-3p	0.027296	up	18.62	MIMAT0004672	NO
STK11	hsa-miR-144-5p	0.049938	up	73.34	MIMAT0004600	NO
STK11	hsa-miR-15b-3p	0.00585	up	44.38	MIMAT0004586	NO
STK11	hsa-miR-182-3p	0.018396	up	29.20	MIMAT0000260	NO
STK11	hsa-miR-190a-5p	0.017345	up	40.56	MIMAT0000458	NO
STK11	hsa-miR-191-5p	0.005873	up	33.11	MIMAT0000440	NO
STK11	hsa-miR-1913	0.010622	up	37.54	MIMAT0007888	NO
STK11	hsa-miR-26b-3p	0.010553	up	43.63	MIMAT0004500	NO
STK11	hsa-miR-29a-5p	0.049188	up	38.69	MIMAT0004503	NO
STK11	hsa-miR-29b-1-5p	0.011177	up	118.38	MIMAT0004514	NO
STK11	hsa-miR-339-5p	0.037138	up	16.63	MIMAT0000764	NO
STK11	hsa-miR-34a-3p	0.047655	up	50.77	MIMAT0004557	NO
STK11	hsa-miR-4252	0.014764	up	55.13	MIMAT0016886	NO
STK11	hsa-miR-4318	0.001537	up	54.74	MIMAT0016869	NO
STK11	hsa-miR-4328	0.035986	up	22.42	MIMAT0016926	NO
STK11	hsa-miR-4730	0.03915	down	-45.98	MIMAT0019852	NO
STK11	hsa-miR-4770	0.010971	up	39.62	MIMAT0019924	NO
STK11	hsa-miR-489-3p	0.004512	up	43.13	MIMAT0002805	NO
STK11	hsa-miR-517a-3p	0.019567	up	22.36	MIMAT0002852	NO
STK11	hsa-miR-517c-3p	0.021522	up	21.18	MIMAT0002866	NO
STK11	hsa-miR-522-3p	0.026955	up	6.25	MIMAT0002868	NO
STK11	hsa-miR-548aa	0.048199	up	13.60	MIMAT0018447	YES
STK11	hsa-miR-5701	0.035075	up	35.87	MIMAT0022494	NO
STK11	hsa-miR-585-3p	0.024521	up	4.84	MIMAT0003250	NO
STK11	hsa-miR-6073	0.026801	up	36.71	MIMAT0023698	NO
STK11	hsa-miR-624-5p	0.009898	up	32.12	MIMAT0003293	NO
STK11	hsa-miR-628-5p	0.032169	up	27.20	MIMAT0004809	NO
STK11	hsa-miR-6516-3p	0.047177	up	34.70	MIMAT0030418	NO
STK11	hsa-miR-6872-3p	0.016967	up	8.28	MIMAT0027645	NO
STK11	hsa-miR-8077	8.18E-04	up	115.11	MIMAT0031004	NO
TP53	hsa-miR-147b	0.023046	up	3.07	MIMAT0004928	NO
TP53	hsa-miR-205-3p	0.018824	up	11.09	MIMAT0009197	YES
TP53	hsa-miR-4290	0.017774	up	6.99	MIMAT0016921	NO
TP53	hsa-miR-6891-3p	0.043351	up	4.85	MIMAT0027683	NO

3.7. Survival prediction using miRNA footprint.

In this study 9 out of 35 monitored patients died within 3 years since biopsy. We explored what miRNAs expression profile may be predictive of clinical outcome in the following years after surgery. It was found that miRNA expression profile in cancer tissue was different between survivors and non-survivors, as shown by scatter plot (**Figure 14a**) and volcano plot analyses (**Figure 14b**).

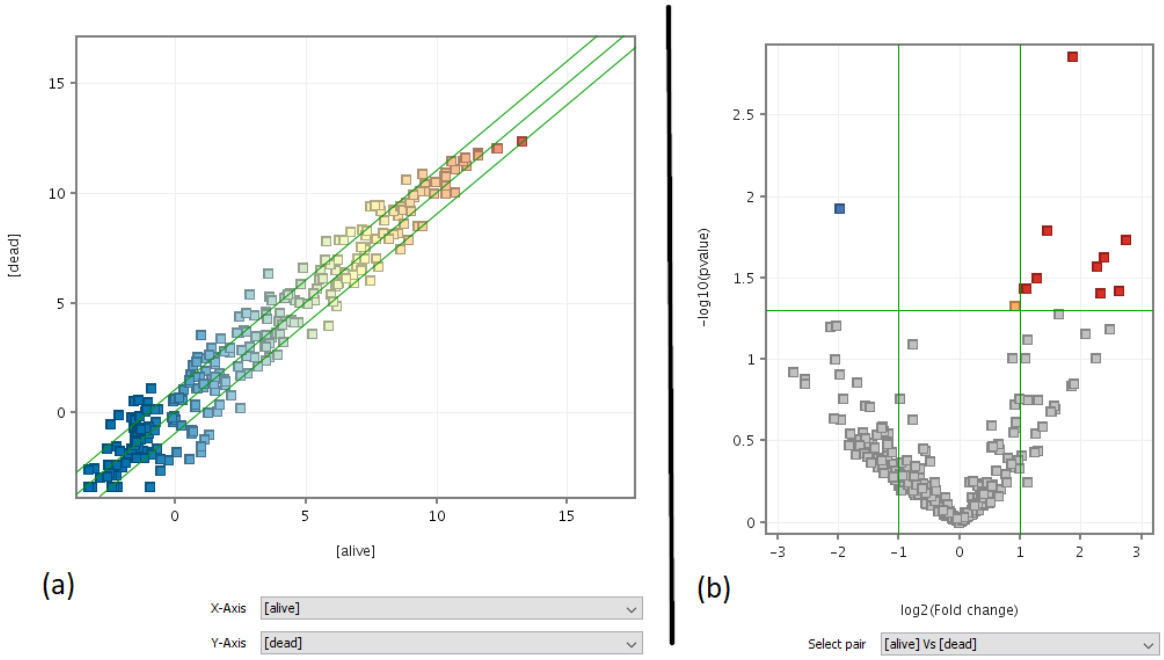


Figure 14 (a) Scatter Plot analysis = Entity list: Cancer Related miRNAs (273), interpretation: averaged [Alive] vs. [dead], FC>=2.0. (b) Moderated T-Test Volcano Plot analysis = Entity list: Cancer Related miRNAs (273), interpretation: averaged [Alive] vs. [dead], without Multiple Testing Correction, p-value cut-off = 0.05, Fold change cut-off = 2.0.

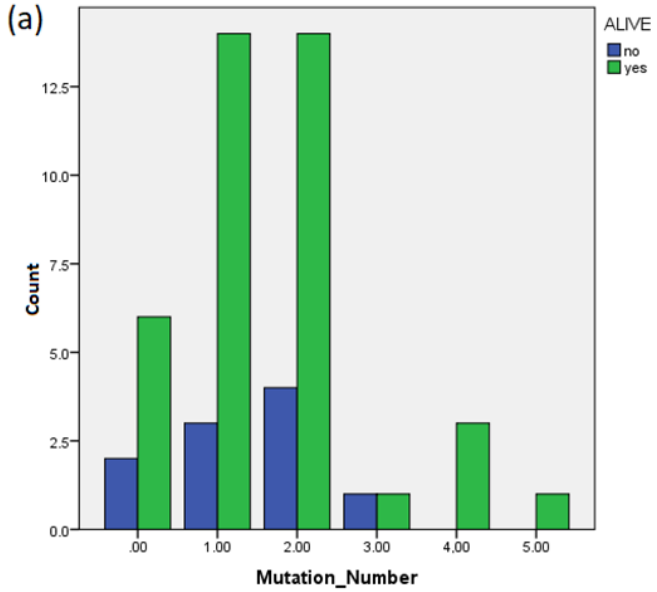
The list of the 11 miRNAs which may be predictive for patient survival (10 upregulated (red dots) in survivors as compared to non survivors and 1 downregulated (blue dot) are reported in **Table 7**. A prediction model using the list of these 11 miRNAs related to survival and GeneSpring Neural Network prediction algorithm was run, obtaining an overall accuracy of 0.92 (+0.11), a higher result than those obtained for all miRNA entities (accuracy 0.81). This high accuracy shows the potential of these 11 cancer related miRNAs from lung biopsies to be used as survival predictors.

Table 7 Cancer Related miRNAs altered (FC>= 2, p<=0.05) in Volcano Plot Analysis between average signal in samples of patients alive vs. dead within 3 years since biopsy.

systematic_name	p-value	Regulation	FC
hsa-miR-1227-5p	0.03180667	up	2.42
hsa-miR-147b	0.011820709	down	-3.96
hsa-miR-187-5p	0.03807269	up	6.18
hsa-miR-23a-5p	0.01835003	up	6.71
hsa-miR-2861	0.036854673	up	2.08
hsa-miR-3663-5p	0.03956902	up	5.03
hsa-miR-371b-5p	0.001374278	up	3.66
hsa-miR-6068	0.016317874	up	2.72
hsa-miR-6075	0.02674605	up	4.83
hsa-miR-6771-5p	0.023678219	up	5.23
hsa-miR-7704	0.03648895	up	2.15

Conversely, the mutation status poorly predicted the survival. Indeed, the rate of survivors (47 out of 60) and non-survivors (13 out of 60) was not different between mutation free (10 out of 52) and mutation carrier (42 out of 52) patients. The number of mutations carried by the same patients was not different between survivors and non-survivors as demonstrated by the Chi-squared test ($p = 0.803$) as shown in Figure 15.

(a)



(b) **Contingency table Mutation Number * Alive**

			ALIVE		Total
			no	yes	
Mutation_Number	.00	Count	2	6	8
		% in Mutation_Number	25.0%	75.0%	100.0%
	1.00	Count	3	14	17
		% in Mutation_Number	17.6%	82.4%	100.0%
	2.00	Count	4	14	18
		% in Mutation_Number	22.2%	77.8%	100.0%
	3.00	Count	1	1	2
		% in Mutation_Number	50.0%	50.0%	100.0%
	4.00	Count	0	3	3
		% in Mutation_Number	0,0%	100,0%	100.0%
	5.00	Count	0	1	1
		% in Mutation_Number	0.0%	100.0%	100.0%
Total		Count	10	39	49
		% in Mutation_Number	20.4%	79.6%	100.0%

(c) **Chi-squared test**

	Value	Degrees of freedom	Asymp. Sig. (2-sided)
Pearson's chi-squared test	2.324[a]	5	.803[a]
Likelihood-ratio test	2.906	5	.715
Valid cases	49		

a. Null hypothesis mantains. Variables are independent

Figure 15 The number of mutations did not predict patient survival. (a) Bar plot of the number of patients with zero to five mutations detected separated by survival. (b) Same data of (a) summarized in a contingency table. (c) The Chi-squared test maintained the null hypothesis: the survival and mutation number were independent variables.

3.8. Contribution of environmental exposures to lung carcinogenesis as inferred from miRNA profiling.

Environmental exposures considered were (a) passive smoke at home, (b) passive smoke at work; (c) airborne car traffic pollution; (d) volcano ashes; and (e) radon exposure risk. These exposures were evaluated for each patient only by semi-structured questionnaires as shown in **Table 2** data, but not by pollution quantification by measuring each pollution source. To understand how cancer related miRNAs were altered by each exposure, there were created 5 miRNA signatures, that hereby are called Environmental Exposure Signals (EESs).

EESs were obtained comparing LCRMs intensity signal (Volcano Plot analysis FC>=2, p<=0,05) in tumoral lung tissue between patients undergoing low or high exposure according to questionnaires data. The number of miRNAs composing each EESs is the following: (a) n = 8; (b) n = 1; (c) n = 53; (d) n = 21; (e) n = 19. Volcano Plot after comparing miRNA expression in lung between unexposed vs. exposed subjects for each environmental signature are reported in **Figure 16**.

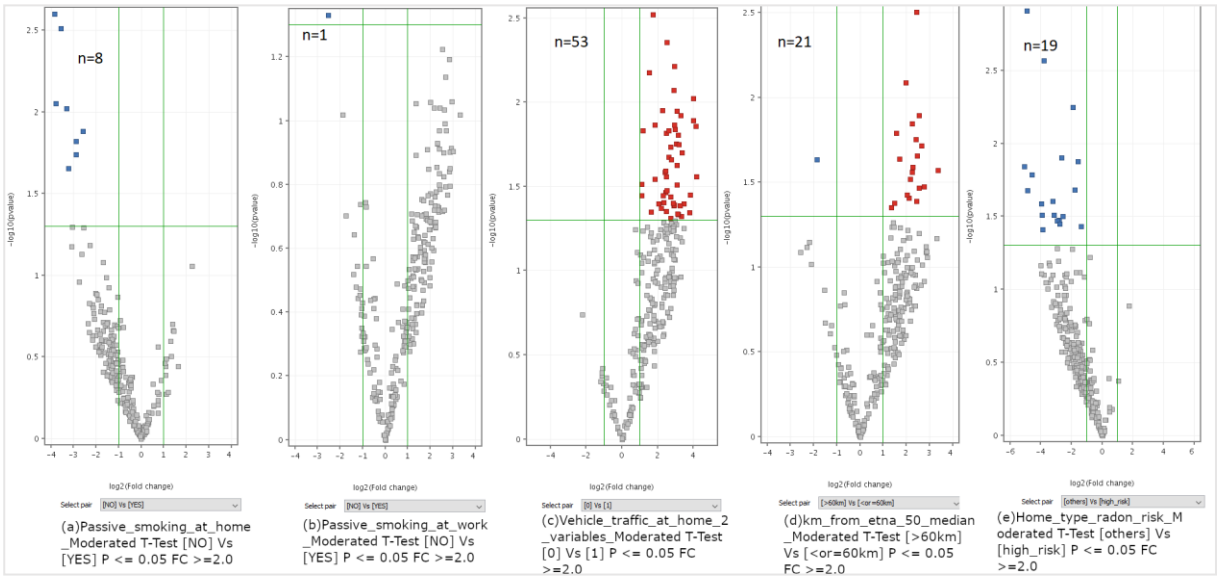


Figure 16 Moderated T-Test Volcano Plot analyses identifying miRNA environmental signatures (EESs) among the 273 LCRMs. Interpretation: [no-exposure and low exposure] vs. [medium and high-exposure], p-value cut-off = 0.05, Fold change cut-off = 2.0.; Y-axis: -log10(p-value); X-axis: log2(Fold change).

miRNA pattern changed for each exposure: for passive smoking and Radon Risk all of miRNAs are downregulated, but for Vehicle traffic and volcano exposure, they are upregulated. This result makes us hypothesis that organism response to different exposures brings also different miRNA regulation, but as we don't have patients exposed exclusively to one of each environmental pollution, this result may also be explained as a bias based on patients' heterogeneity. To understand the value of each EESs it was search in bibliography if miRNA entities were present in other studies. Because passive smoke at work differed form passive smoke at home for only one miRNA., these two signatures were aggregated in a single signature referred to passive smoke for bibliographic search. The entire list of deregulated miRNAs, and their presence in bibliography, is present in **Table 8**. It was found that any miRNAs in our list is mentioned in other studies for passive smoking, Radon exposure, and volcano ashes exposure. For air quality, but not specifically for vehicle traffic, it was found citation for let-7a-5p, miR-125a-5p, miR-142-3p, miR-150-5p, miR-223-3p, miR-23a-3p, let-7c-5p, miR-185-

5p, miR-99b-5p, miR-146a-5p, miR-199a-5p, miR-214-3p, miR-4516, [59 - 65] . It was unexpected that none of our radon EESs were present in bibliography, as there exist at least 11 different papers that study miRNA expression after Radon exposure.

Table 8 Cancer Related miRNAs run on a Moderated T-Test Volcano Plot analysis (FC>= 2, p<=0.05) for each environmental exposure: (a) passive smoke at home (No vs. Yes), (b) passive smoke at work (No vs. Yes); (c) airborne car traffic pollution (low vs. high); (d) volcano ashes (>60Km vs. <=60Km); (e) radon risk (according to house type low vs. high)

systematic name	p-value	Regulation ([low] Vs [high])	FC ([low] Vs [high])	environmental exposure
hsa-miR-2277-3p	0.01829	down	-7.38783	passive smoking at home
hsa-miR-328-3p	0.013113	down	-5.98362	passive smoking at home
hsa-miR-4254	0.002496	down	-14.3598	passive smoking at home
hsa-miR-483-3p	0.015154	down	-7.43976	passive smoking at home
hsa-miR-491-3p	0.003077	down	-11.924	passive smoking at home
hsa-miR-6743-3p	0.022234	down	-9.42842	passive smoking at home
hsa-miR-6779-3p	0.009505	down	-9.89658	passive smoking at home
hsa-miR-6886-3p	0.008845	down	-13.8967	passive smoking at home
hsa-miR-4695-3p	0.04655064	down	-5.80325	Passive smoking at work
hsa-let-7a-5p	0.041112	up	7.4232	Vehicle traffic at home
hsa-let-7b-5p	0.011125	up	4.773517	Vehicle traffic at home
hsa-let-7c-5p	0.015205	up	5.49175	Vehicle traffic at home
hsa-miR-100-5p	0.040598	up	7.583029	Vehicle traffic at home
hsa-miR-103a-3p	0.039793	up	7.255231	Vehicle traffic at home
hsa-miR-125a-5p	0.034176	up	5.446658	Vehicle traffic at home
hsa-miR-125b-5p	0.021168	up	6.086423	Vehicle traffic at home
hsa-miR-130a-3p	0.048873	up	6.662414	Vehicle traffic at home
hsa-miR-133b	0.047599	up	9.969255	Vehicle traffic at home
hsa-miR-142-3p	0.046327	up	8.855112	Vehicle traffic at home
hsa-miR-146a-5p	0.023557	up	8.37029	Vehicle traffic at home
hsa-miR-150-5p	0.039885	up	10.98931	Vehicle traffic at home
hsa-miR-151b	0.015652	up	8.787968	Vehicle traffic at home
hsa-miR-15b-5p	0.03082	up	8.508455	Vehicle traffic at home
hsa-miR-16-2-3p	0.044513	up	5.532554	Vehicle traffic at home
hsa-miR-185-5p	0.039332	up	5.07439	Vehicle traffic at home
hsa-miR-199a-5p	0.04242	up	4.570111	Vehicle traffic at home
hsa-miR-203a-3p	0.02749	up	18.03255	Vehicle traffic at home
hsa-miR-204-5p	0.041068	up	9.511877	Vehicle traffic at home
hsa-miR-20a-5p	0.045853	up	8.663452	Vehicle traffic at home

hsa-miR-214-3p	0.026044	up	5.304047	Vehicle traffic at home
hsa-miR-221-3p	0.017622	up	8.297637	Vehicle traffic at home
hsa-miR-221-5p	0.013803	up	17.72528	Vehicle traffic at home
hsa-miR-222-3p	0.013626	up	3.554951	Vehicle traffic at home
hsa-miR-223-3p	0.012787	up	15.80766	Vehicle traffic at home
hsa-miR-224-3p	0.017806	up	9.024755	Vehicle traffic at home
hsa-miR-23a-3p	0.035768	up	4.962604	Vehicle traffic at home
hsa-miR-23a-5p	0.011235	up	8.435381	Vehicle traffic at home
hsa-miR-30c-2-3p	0.004407	up	5.743132	Vehicle traffic at home
hsa-miR-3188	0.044807	up	3.079721	Vehicle traffic at home
hsa-miR-324-3p	0.035711	up	2.143047	Vehicle traffic at home
hsa-miR-339-5p	0.008492	up	7.478019	Vehicle traffic at home
hsa-miR-342-3p	0.018368	up	6.661599	Vehicle traffic at home
hsa-miR-365a-3p	0.036355	up	6.536557	Vehicle traffic at home
hsa-miR-4306	0.002998	up	3.347996	Vehicle traffic at home
hsa-miR-4324	0.033328	up	5.706607	Vehicle traffic at home
hsa-miR-4516	0.006668	up	2.859822	Vehicle traffic at home
hsa-miR-452-5p	0.014409	up	7.888119	Vehicle traffic at home
hsa-miR-4532	0.014752	up	2.258895	Vehicle traffic at home
hsa-miR-455-3p	0.009444	up	15.94311	Vehicle traffic at home
hsa-miR-4634	0.030527	up	2.128231	Vehicle traffic at home
hsa-miR-4655-3p	0.021817	up	6.810059	Vehicle traffic at home
hsa-miR-4731-3p	0.01469	up	6.153936	Vehicle traffic at home
hsa-miR-501-5p	0.012	up	9.779916	Vehicle traffic at home
hsa-miR-505-5p	0.013655	up	7.634665	Vehicle traffic at home
hsa-miR-517a-3p	0.027523	up	5.597306	Vehicle traffic at home
hsa-miR-517c-3p	0.025637	up	5.422509	Vehicle traffic at home
hsa-miR-582-5p	0.044966	up	13.7203	Vehicle traffic at home
hsa-miR-6895-5p	0.006121	up	7.657433	Vehicle traffic at home
hsa-miR-744-5p	0.019918	up	10.30978	Vehicle traffic at home
hsa-miR-98-5p	0.035529	up	13.94902	Vehicle traffic at home
hsa-miR-99b-5p	0.040104	up	4.172732	Vehicle traffic at home
hur_6	-	-	-	Vehicle traffic at home
hsa-miR-133a-3p	0.017678	up	5.314951	km from etna 50 median
hsa-miR-135a-5p	0.034151	up	5.869018	km from etna 50 median
hsa-miR-193a-5p	0.044354	up	2.526421	km from etna 50 median
hsa-miR-27a-5p	0.023292	down	-3.61896	km from etna 50 median
hsa-miR-29b-2-5p	0.014229	up	4.74146	km from etna 50 median
hsa-miR-340-5p	0.033516	up	6.845372	km from etna 50 median
hsa-miR-3607-3p	0.012756	up	5.837188	km from etna 50 median

hsa-miR-361-3p	0.025633	up	4.868399	km from etna 50 median
hsa-miR-449a	0.026818	up	10.2289	km from etna 50 median
hsa-miR-4770	0.03923	up	4.277719	km from etna 50 median
hsa-miR-500a-3p	0.037476	up	4.086222	km from etna 50 median
hsa-miR-500b-5p	0.008137	up	3.92188	km from etna 50 median
hsa-miR-505-5p	0.030313	up	4.50934	km from etna 50 median
hsa-miR-5701	0.019156	up	6.29237	km from etna 50 median
hsa-miR-628-5p	0.022147	up	5.505154	km from etna 50 median
hsa-miR-642b-5p	0.016176	up	2.964167	km from etna 50 median
hsa-miR-6516-3p	0.041004	up	5.372472	km from etna 50 median
hsa-miR-652-3p	0.027532	up	4.759585	km from etna 50 median
hsa-miR-664a-3p	0.042052	up	2.812089	km from etna 50 median
hsa-miR-664b-3p	0.023029	up	3.220316	km from etna 50 median
hsa-miR-874-5p	0.003118	up	5.434077	km from etna 50 median
hsa-miR-16-2-3p	0.024773	down	-9.45544	Radon risk Home type
hsa-miR-182-3p	0.033553	down	-7.05349	Radon risk Home type
hsa-miR-22-5p	0.030786	down	-15.112	Radon risk Home type
hsa-miR-221-5p	0.016387	down	-23.7398	Radon risk Home type
hsa-miR-30c-2-3p	0.012436	down	-6.19418	Radon risk Home type
hsa-miR-3660	0.001222	down	-29.6997	Radon risk Home type
hsa-miR-4306	0.005655	down	-3.77123	Radon risk Home type
hsa-miR-4440	0.033708	down	-7.56958	Radon risk Home type
hsa-miR-4443	0.0132	down	-3.02148	Radon risk Home type
hsa-miR-452-5p	0.030769	down	-8.7803	Radon risk Home type
hsa-miR-454-3p	0.01435	down	-33.4518	Radon risk Home type
hsa-miR-455-3p	0.038657	down	-14.6433	Radon risk Home type
hsa-miR-4793-3p	0.002683	down	-13.7501	Radon risk Home type
hsa-miR-598-3p	0.025903	down	-15.7451	Radon risk Home type
hsa-miR-6500-5p	0.035551	down	-6.86215	Radon risk Home type
hsa-miR-6826-5p	0.037038	down	-2.57302	Radon risk Home type
hsa-miR-6872-3p	0.020769	down	-3.46987	Radon risk Home type
hsa-miR-7159-5p	0.031503	down	-5.91743	Radon risk Home type
hsa-miR-98-5p	0.020851	down	-28.8801	Radon risk Home type

EESs were compared by Venn diagram analysis with the miRNAs composing the individual cancer-related signature LCRMs of each patient (Figure 17). This approach was used to identify the relative contribution of environmental risk factors to cancer development in each patient. In a first moment it was hypotheses that a major number of deregulated EESs per patient may predict their survival, as a minor deregulation may be a signal of adaptative mechanisms against tumoral

development, but after a statistical test we understood that this correlation was not significative. As shown in **Table 1**, among patients analysed with miRNA chip-array only 11 died within 3 years after surgery, and we have EESs available data from 9 of them: patient 6 presents 43 deregulated EESs, patient 34 has 78, patient 38 has 59, patient 42 has 50, patient 46 has 92, patient 47 has 61, patient 49 has 95, patient 55 has 92, and patient 57 has 45.

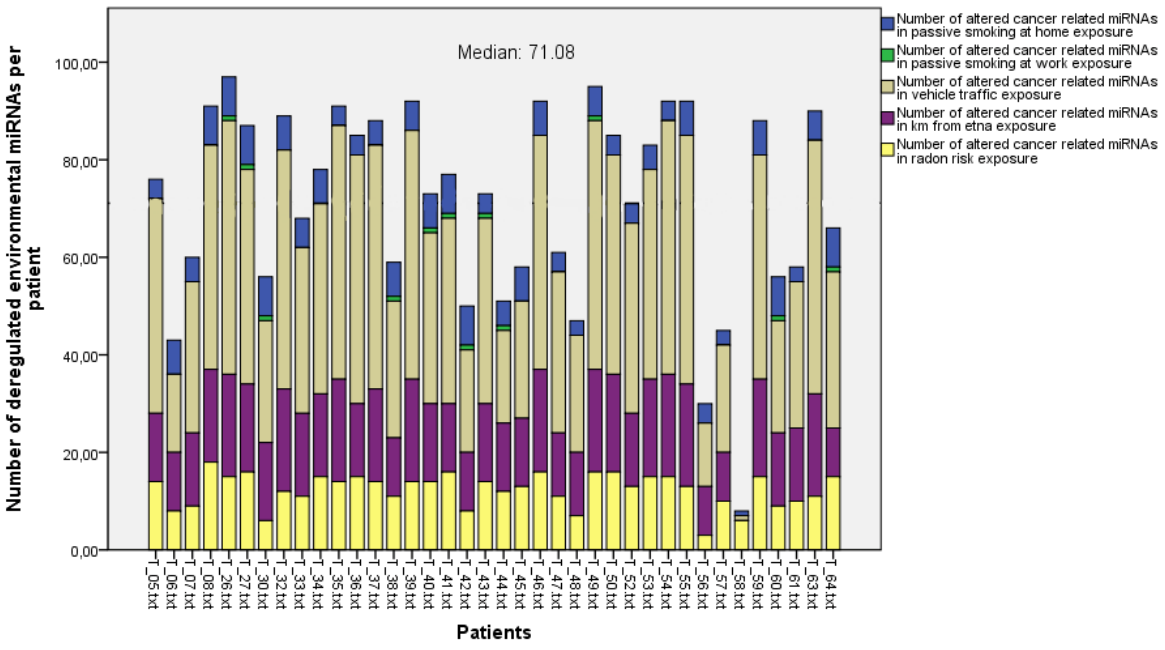
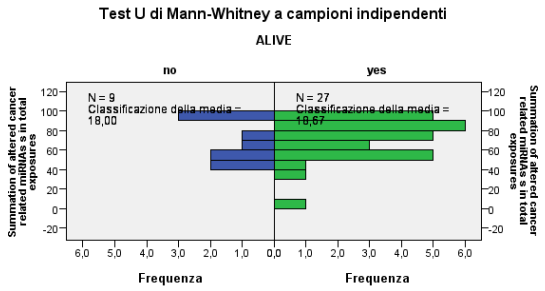


Figure 17 Number of altered miRNAs in EESs (x-axes) per patient (y-axes, T codes). Median is shown as a black vertical line

As explained before, not significative data between number of deregulated miRNA and patient's tumoral types, or survival, was found. In fact, U Mann-Witney non-parametric test was run to understand if the distribution of the sum of EESs per patient was equal between survival ($p = 0.886$) as shown in **Figure 18a**, and histotype ($p = 0.167$) as shown in **Figure 18b**. The null hypothesis was maintained.

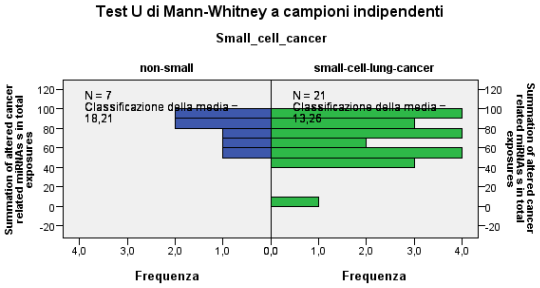
As conclusion we can only say that not surprisingly the strongest EES signal is vehicle traffic. This is supported by bibliography for let-7a-5p, let-7c-5p, miR-125a-5p, miR-142-3p, miR-146a-5p, miR-150-5p, miR-185-5p, miR-199a-5p, miR-214-3p, miR-223-3p, miR-23a-3p, miR-4516, miR-99b-

5p, which have been previously identified as altered after exposure to air pollution. It remains uncertain if Radon and volcano ashes EESs may be reliable markers.



Numero di casi totale	36
U di Mann-Whitney	126,000
W di Wilcoxon	504,000
Statistica test	126,000
Errore standard	27,357
Statistica test standardizzata	,164
Sign. asintotica (test a 2 vie)	,869
Sign. esatta (test a 2 vie)	,886

a)



Numero di casi totale	28
U di Mann-Whitney	47,500
W di Wilcoxon	278,500
Statistica test	47,500
Errore standard	18,835
Statistica test standardizzata	-1,380
Sign. asintotica (test a 2 vie)	,167
Sign. esatta (test a 2 vie)	,172

b)

Figure 18 U Mann-Witney non-parametric test between the sum of EESs per patient and survival status

The biological effects of these miRNA alterations was evaluated by analysing the genes targeted by each environmental exposure signature. A gene target detection for each environmental exposure signature was run on GeneSpring using TargetScan as database. The number of estimated target genes for each signature and different p-values cut-off are summarized in Table 9.

Table 9 Number of predicted target genes according to Targetscan database by each environmental exposure with different p-value cut off.

Environmental Exposure miRNA Signature	*Number of predicted target genes by p-value cut off*				
	0.05	0.01	0.005	0.001	0.0001
Passive Smoking at home (n=8)	8,726	8,726	8,726	6	1
Passive smoking at work (n=1)	1,796	3	0	0	0
Vehicle traffic at home (n=53)	16,132	16,132	16,132	16,132	7
Home distance from Etna Volcano (n=21)	14,662	14,662	14,662	20	4
Home Type radon risk (n=19)	13,762	13,762	15	0	0

The most significant targeted genes for each signature are summarized in Table 10, and we found that these genes are also expressed in lung tissue. The RPKM (Reads Per Kilobase of transcript, per Million mapped reads) value in lung tissue for each targeted gene was found using NCBI Gene database (https://www.ncbi.nlm.nih.gov/gene). Is very complex to correlate EESs to an individual health outcome as each tissue sample have a unique miRNA profile due to patient's heterogeneity. However, gene target prediction supports the hypothesis that the deregulation of miRNAs in EESs may have a strong influence in lung tissue, even if it was not understood how they are correlated to environmental exposures.

Table 10 Most significative gene target detection per signature by p-value cut-off. Each targeted gene had different RPKM(Reads Per Kilobase of transcript, per Million mapped reads) values in lung.

Environmental Exposure signature	*p-value cut-off*	*Genes*	*Gene name*	*RPKM in lung*
Passive Smoking at home (n=8)	0.001	*PTX4*	pentraxin 4	0.006
		NAXD	NAD(P)HX dehydratase	11.04
		MAPK3	mitogen-activated protein kinase 3	30.87
		VPS16	core subunit of CORVET and HOPS complexes	8.88
		CACNA1S	calcium voltage-gated channel subunit alpha1 S	0.11
		SHARPIN	SHANK associated RH domain interactor	10.01
Passive smoking at work (n=1)	0.01	*HBG2*	hemoglobin subunit gamma 2	ND
		RNASE12	ribonuclease A family member 12	ND
		IFT88	intraflagellar transport 88	2.84
Vehicle traffic at home (n=53)	0.0001	*LEFTY1*	left-right determination factor 1	0.00
		RTTN	rotatin	0.76
		THYN1	thymocyte nuclear protein 1	9.95
		CASKIN1	CASK interacting protein 1	0.87
		SERPING1	serpin family G member 1	219.55
		OGFOD2	2-oxoglutarate and iron dependent oxygenase domain containing 2	2.74
		PKDCC	protein kinase domain containing, cytoplasmic	11.39
Home distance from Etna Volcano (n=21)	0.001	*ARHGEF33*	Rho guanine nucleotide exchange factor 33	0.12
		COX17	COX17	13.49

		RAI2	retinoic acid induced 2	10.76
		KIF12	kinesin family member 12	0.87
		COL26A1	collagen type XXVI alpha 1 chain	0.20
		RMDN2	regulator of microtubule dynamics 2	1.47
		GADD45A	growth arrest and DNA damage inducible alpha	9.03
		DTNA	dystrobrevin alpha	1.44
		HTRA4	HtrA serine peptidase 4	0.32
		TAS2R30	taste 2 receptor member 30	ND
		STRN3	striatin 3	4.72
		BRINP3	BMP/retinoic acid inducible neural specific 3	0.13
		EYS	eyes shut homolog	0.01
		JAG2	jagged canonical Notch ligand 2	3.18
		HSD17B12	hydroxysteroid 17-beta dehydrogenase 12	25.64
		NIN	ninein	4.25
		NAA35	N-alpha-acetyltransferase 35, NatC auxiliary subunit	2.93
		ZNF37A	zinc finger protein 37A	1.72
		GLT8D2	glycosyltransferase 8 domain containing 2	3.69
		DDX59	DEAD-box helicase 59	1.99
Home Type radon risk (n=19)	0.005	*CNPY3*	canopy FGF signaling regulator 3	9.45
		SUB1	SUB1 regulator of transcription	15.25
		CLTC	clathrin heavy chain	28.19
		ZNF280A	zinc finger protein 280A	ND
		ALS2CR12	(or FLACC1) flagellum associated containing coiled-coil domains 1	1.09
		TMEM139	transmembrane protein 139	5.56
		BTBD3	BTB domain containing 3	7.59
		WDR7	WD repeat domain 7	2.54
		RAB15	member RAS oncogene family	4.32
		KRT84	keratin 84	ND
		MAZ	MYC associated zinc finger protein	15.90
		ATAD2B	ATPase family AAA domain containing 2B	0.93
		PSPH	phosphoserine phosphatase	1.00
		PGBD1	piggyBac transposable element derived 1	1.16
		BEX2	brain expressed X-linked 2	4.10

GeneSpring 14.9 was used to build a prediction model using Environmental Exposure signatures. Neural Network class prediction algorithm was used to test the overall accuracy prediction for each environmental exposure of all miRNAs in the array as compared to the overall accuracy prediction for each environmental miRNA signature entities. The network was tested in tumoral tissue. For passive smoking at home, and Home Type radon risk the accuracy of the environmental signature remarkably increase as compared to the overall miRNAs. Conversely, for vehicle traffic at home and home distance from Etna volcano the overall accuracy was only slightly higher (Table 11). For Passive smoking at work signature the accuracy remains equal, demonstrating that one miRNA-signature is not enough to predict this exposure.

Table 11 Results of the Neural Network class prediction for each environmental exposure using both All miRNAS (2570 entities) and only environmental exposure signal (EES). Prediction overall accuracy variation is higher in 4 out of 5 signatures if using only EESs. Prediction Overall Accuracy have a range between 0 and 1.

Environmental Exposure (EES's number of miRNA entities)	*n. of endpoints [endpoint value]*	*ALL miRNAs (n=2570)*	*EESs (variation)*
Passive Smoking at home (n=8)	YES vs. NO	0.73	0.96 (+0.23)
Passive smoking at work (n=1)	YES vs. NO	0.68	0.68 (+0)
Vehicle traffic at home (n=53)	High risk vs. other	0.77	0.81 (+0.04)
Home distance from Etna Volcano (n=21)	<60km vs. ≥60km	0.60	0.82 (+0.22)
Home Type radon risk (n=19)	other vs. High risk	0.83	0.96 (+0.13)

Vehicle traffic at home signature have the highest number of EESs entities (n=53) and the use of this signature in the Neural Network prediction have the accuracy of 0.81. Distance from Etna volcano had the second highest number of EESs entities (n=21), with an overall accuracy of 0.82. Radon risk had the third weight in miRNA signature (n=19) with an overall accuracy of 0.96. Passive smoke had the fourth weight in miRNA signature (9) with an overall accuracy of 0.82.

3.9. B(a)P-DNA adducts and environmental exposures

Pearson's Bivariate Correlation analysis shows that there's not a statistically significant association between B(a)P-DNA adducts and environmental exposures. As with miRNA analysis, the lack of statistical significance may occur due to the limited number of patients analysed and their

heterogeneity. Detailed data of BaP-DNA adducts in lymphocytes and different environmental exposures from all recruited patients as reported by questionnaires are detailed in **Table 12**. This table differs from **Table 1** as BaP-DNA adducts were measured in 58 out of 64 patients, but complete exposure data was available for 54 patients.

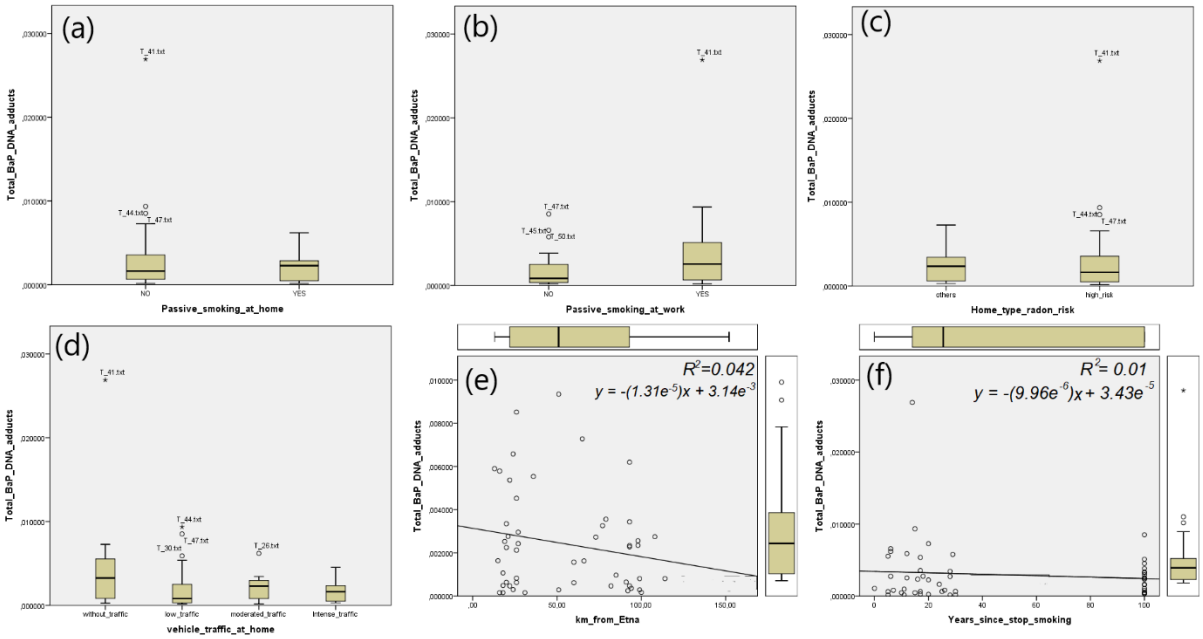
Table 12 BaP-DNA adducts and exposure data from all 64 recruited patients. Difference with Table 2 is that BaP-DNA adducts were measured in 58 patients. However, some exposure data are missed.

Patient code	Total BaP-DNA adducts	Passive smoking at home	Passive smoking at work	Home type radon risk	Vehicle traffic at home	Distance from Etna in Km
01	0.000490	YES	NO	high_risk	Intense_traffic	18
02	0.000470	YES	YES	others	low_traffic	22
03	0.000360	YES	NO	others	moderated_traffic	94
04	0.000650	NO		high_risk	low_traffic	91
05	0.007280	NO	YES	others	without_traffic	65
06	0.000810	NO	NO	high_risk	without_traffic	114
07	0.000470	YES	NO	high_risk	without_traffic	93
08	0.002730	YES	YES	high_risk	without_traffic	73
09	0.000640	NO	YES	high_risk	low_traffic	20
10	0.000290	NO	YES	high_risk	low_traffic	51
11	0.001080	NO	NO	high_risk	moderated_traffic	
12	0.001570	YES	NO	high_risk	low_traffic	60
13	0.002280	YES	YES	high_risk	low_traffic	93
14	0.000170	NO	YES	high_risk	low_traffic	100
15	0.000810	YES	NO	high_risk	moderated_traffic	99
16	0.001070	NO	YES	high_risk	moderated_traffic	18
17	0.000970	NO	NO	high_risk	low_traffic	85
18	0.000620	YES	YES	others	low_traffic	60
19	0.000290	NO	NO	high_risk	low_traffic	99
20	0.002430	YES	YES	others	low_traffic	27
21	0.000470	YES	YES	high_risk	moderated_traffic	82

22	0.000160	NO	NO	high_risk	moderated_traffic	31
23	0.000160	YES	NO	high_risk	low_traffic	18
24	0.000300	NO	NO	others	Intense_traffic	24
25	0.000320	YES	NO	others	low_traffic	152
26	0.006200	YES	YES	others	moderated_traffic	93
27	0.002250	YES	NO	high_risk	moderated_traffic	20
28		YES	YES	others	Intense_traffic	50
29		NO	YES	others	Intense_traffic	107
30	0.005900	YES	YES	high_risk	low_traffic	13
31	0.002960	YES	NO	others	moderated_traffic	27
32	0.002350	YES	YES	others	Intense_traffic	98
33	0.000160	NO	NO	high_risk	low_traffic	16
34	0.002130	YES	NO	high_risk	low_traffic	26
35	0.003350	NO	YES	high_risk	moderated_traffic	20
36	0.002750	YES	YES		moderated_traffic	108
37	0.003440	NO	YES	others	moderated_traffic	93
38	0.002340	YES	YES	high_risk	moderated_traffic	93
39	0.001630	NO		others	Intense_traffic	66
40	0.005370	YES	YES	high_risk	low_traffic	22
41	0.026900	NO	YES	high_risk	without_traffic	
42	0.002530	YES	YES	high_risk	moderated_traffic	19
43	0.003560	NO	YES	high_risk	without_traffic	79
44	0.009350	NO	YES	high_risk	low_traffic	51
45	0.005790	YES	NO	high_risk	without_traffic	16
46	0.002760	NO	NO		without_traffic	21
47	0.008520	NO	NO	high_risk	low_traffic	26
48	0.003260	NO	YES	others	without_traffic	77
49	0.001640	NO	YES	high_risk	without_traffic	15
50	0.006580	NO	NO	high_risk	without_traffic	24
51	0.005540	YES	YES	high_risk	without_traffic	36
52	0.003830	NO	NO	high_risk	without_traffic	

53	0.004530	YES		others	Intense_traffic	26
54	0.005120	NO	YES	others	without_traffic	
55	0.000830	NO	NO	others	without_traffic	26
56	0.002560	NO	YES	high_risk	low_traffic	98
57	0.000250	YES	YES	high_risk	without_traffic	92
58	0.000630	NO	YES	others	without_traffic	26
59		NO	YES	high_risk	low_traffic	13
60		NO	NO	others	moderated_traffic	100
61		NO	NO	high_risk	Intense_traffic	26
62		YES	YES	others	moderated_traffic	67
63	0.000490	NO	YES	others	low_traffic	60
64	0.000470	YES	NO	high_risk	Intense_traffic	21

Pearson correlation test results were: (a) passive smoking at home from 56 patients, correlation coefficient = - 0.126, p= 0.354 as shown in **Figure 19a**; (b) passive smoking at work from 53 patients, correlation coefficient = 0.223, p= 0.109 as shown in **Figure 19b**; (c) radon risk related to home type from 54 patients, correlation coefficient= - 0.064, p=0.644, as shown in **Figure 19c**; (d) vehicle traffic at home from 56 patients, correlation coefficient = - 0.273, p= 0.079, as shown in **Figure 19d**; (e) distance from Etna available from 52 patients, correlation coefficient = - 0.204, p-value= 0.147, R^2 = 0.042, as shown in **Figure 19e**; and (f) years since smoking cessation available from 54 patients, correlation coefficient = - 0.099, p-value= 0.477, R^2 = 0.01, as shown in **Figure 19f**. Linear regression shows that the level of B(a)P-DNA adducts in lymphocytes was inversely related with the distance from Etna volcano, as well as years since smoking cessation, even if both correlations were not significative.



			Passive smoking at home	Passive smoking at work	Home type radon risk	vehicle traffic at home	km from Etna	Years since stop smoking
Total_BaP_DNA_adducts	Pearson	Correlation Coefficient	-0.126	0.223	0.064	-0.237	-0.204	-0.099
		Sign. (two tails)	0.354	0.109	0.644	0.079	0.147	0.477
		N	56	53	54	56	52	54

Figure 19 Box plot analysis for Total BaP-DNA adducts in: (a) passive smoking at home, (b) passive smoking at work, (c) radon risk home type, (d) vehicle traffic at home, and linear regression for (e) distance from Etna, and (f) years since smoking cessation. Bivariate (Pearson) Correlation shows that in our study B(a)P-DNA adducts are not significantly correlate to environmental exposures mentioned.

When using linear correlation test instead of Pearson correlation test, we can see that the correlation between pollution and BaP remains not significant (Table 13).

Table 13 Linear correlation results using BaP-DNA adducts as dependent variant and pollution as independent variant.

Independent variant	R	R^2	ANOVA Regression sum of squares	ANOVA p-value
Passive Smoking at home	0.126	0.016		
Passive Smoking at work	0.223	0.05		
Home Type Radon Risk	0.064	0.004		
Vehicle Traffic at home	0.238	0.56		
Km from Etna	0.204	0.042	0.000	0.147
Years since stop smoking	0.78	0.01	0.000	0.672

4. Discussion and conclusions

Our results provide evidence that miRNAs are massively dysregulated in lung cancer as compared to surrounding normal tissue, as previously demonstrated by several studies [66]. A major problem in using miRNA analysis for lung cancer prediction and early diagnosis is the reproducibility of the results and the invasiveness of the biopsy approach. Our Lung Cancer Related miRNAs, or LCRMs, signature at least in part overlap with the most common lung cancer miRNA-related signatures found in literature. Indeed 6 miRNAs of our signature were also present in other papers that analysed miRNA expression in lung cancer [67, 68]. These are let-7a, let-7b, miR-10a, miR-15a, miR-21, miR-23a, miR-29a, miR-29b, miR-29c, miR-30a, miR-30b, miR-30c, miR-34a, miR-126, miR-128, miR-145, miR-191, miR-205, miR-221, miR-222, and miR-326.

An important limit in this study is the lack of normalization of chip-array imported data. The goal of microarray bioinformatic data analysis is to remove systematic differences between samples that do not represent true biological variation. This is usually done at the initial data normalization stage of the analysis process. Different normalization methods have been used on miRNA microarray expression profiling data sets, but there is currently no clear consensus about their relative performances. Some have even chosen to omit normalization [69]. Due to the wide variability between samples, and the explorative nature of this research, normalization was avoided. Therefore, LCRMs and EESs may be biased, and miRNAs proposed as signal should be confirmed by PCR analysis in a new subset of patients.

According to the most recent metanalysis about miRNAs in Lung Cancer patients when this thesis was written [70], about 92 studies with high quality standards have been published about miRNA regulation in lung cancer tissue compared to noncancerous surrounded tissues in humans. Metanalysis' authors declared that high quality studies have shed light on how miRNAs are expressed in different lung cancers stages, and how expression changes as response to chemotherapy in patients with advanced stages of lung cancer. In the metanalysis authors found a total of 176 up-regulated and 114 down-regulated miRNAs related to lung cancer in serum, sputum, plasma, peripheral blood mononuclear cell, and lung tissue. A total of 132 miRNAs were confirmed from different studies to be differentially regulated in cancer tissues compared to surrounded healthy lung tissue.

In Table 13 are reported lung tissue-related miRNAs using the data of [70], with 2 or more citations. In Figure 20 a Venn diagram shows all mentioned miRNAs up- and down-regulated in tumoral lung tissue, and miRNAs that overlapped in both regulations according to the metanalysis. The number of cited downregulated miRNAs (72) is major than the upregulated (70). miR-21, miR-182, miR-205, miR-210 and miR-9 are the most citated as up-regulated, while miR-30a, miR-486, miR-126, and miR-451 are the most citated as downregulated. Inconsistent results (i.e. down- and up-regulation were different between studies) were found for 10 miRNAs: miR-205, miR-9, miR-31, miR-150, miR-224, miR-34a, miR-133a, miR-203, miR-218, miR-375.

Table 14 Data from Zhong S, et al. (2021) about deregulated miRNAs in different studies about lung cancer tissues.Most citated miRNA is miR-21 with 7 citations, followed by miR-182 with 6 citations.

Regulation	microRNA	Citation Frequency
up-	miR-21	7
up-	miR-182	6
up-	miR-205	5
up-	miR-210	5
up-	miR-9	5
up-	miR-183	4
up-	miR-31	3
up-	miR-93	3
up-	miR-96	3
up-	miR-106b	2
up-	miR-1290	2
up-	miR-135b	2
up-	miR-150	2
up-	miR-155	2
up-	miR-196a	2
up-	miR-196b	2
up-	miR-20a	2
up-	miR-210-3p	2
up-	miR-221	2
up-	miR-222	2
up-	miR-224	2
up-	miR-34a	2

Regulation	microRNA	Citation Frequency
down-	miR-30a	7
down-	miR-486	7
down-	miR-126	4
down-	miR-451	4
down-	miR-139	3
down-	miR-144	3
down-	miR-144-3p	3
down-	miR-218	3
down-	miR-30d	3
down-	miR-34a	3
down-	miR-34b	3
down-	let-7e	2
down-	miR-100	2
down-	miR-125a	2
down-	miR-126-3p	2
down-	miR-133a	2
down-	miR-140-3p	2
down-	miR-145	2
down-	miR-205	2
down-	miR-27a	2
down-	miR-30a-3p	2

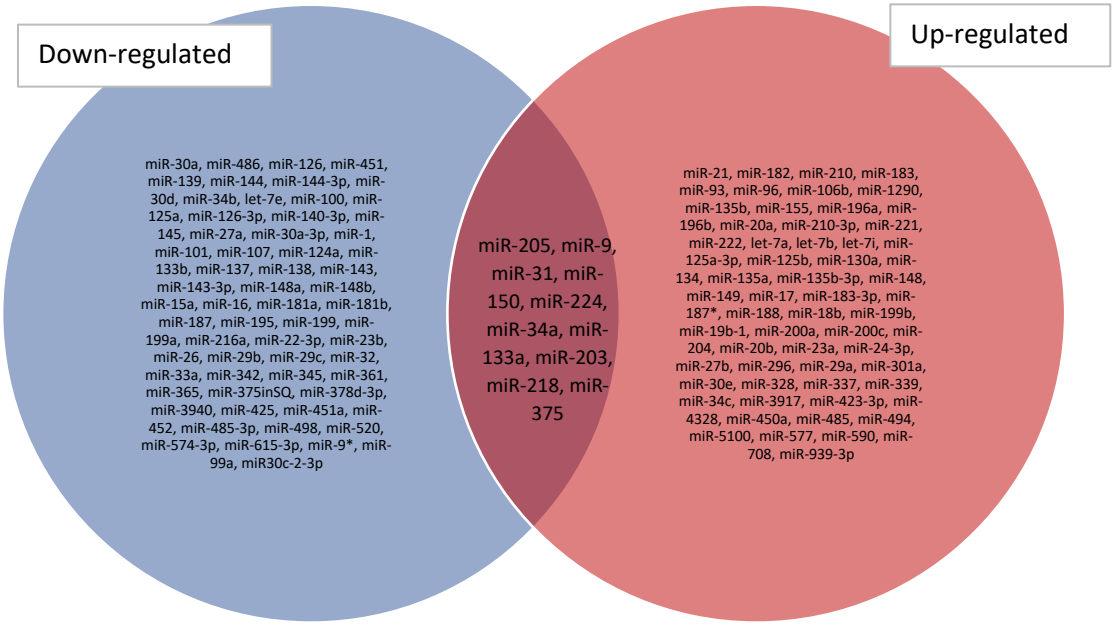


Figure 20 Downregulated and upregulated miRNAs in tumoral lung tissue as cited by different studies according to data from Zhong S, et al. (2021). Some miRNAs have inconsistent results, and the massevelly downregulation tipically found in chip array analysis is not evident.

As interesting as this result can be, lung cancer-related miRNA metanalysis continue to report studies with different miRNA-profiling platforms (microchip array of different generations, PCR amplification and different technologies used for sequencing) and different data-analysis methods. This still cause high inconsistency as discussed by Zhong S, et al. Moreover, other metanalysis about let-7 family members have discuss how it is largely downregulated in lung cancer tissue [71, 72], but this is not reflected according to Zhong S, et al. results.

If compared with the data of miRNAs altered in cancer tissue versus healthy tissue obtained in Izzotti et al. (2021) (that is the data used for this thesis available in Results chapter), it is evident that only a small part of results overlaps with Zhong S, et al. (2021) As we can see in **Figure 21**, 4 out of 117 miRNAs (miR-21, miR-182, miR-183, miR-106b) overlap as up-regulated, and 38 out of 224 (miR-30a, miR-486, miR-126, miR-139, miR-144, miR-30d, miR-34b, let-7e, miR-100, miR-125a, miR-145, miR-27a, miR-1, miR-101, miR-107, miR-133b, miR-138, miR-143, miR-15a, miR-16, miR-181a, miR-187, miR-195, miR-199a, miR-23b, miR-29b,

miR-29c, miR-32, miR-342, miR-361, miR-451a, miR-452, miR-99a, miR-150, miR-224, miR-34a, miR-133a, miR-218) overlap as downregulated, while 34 miRNAs present inconsistent results.

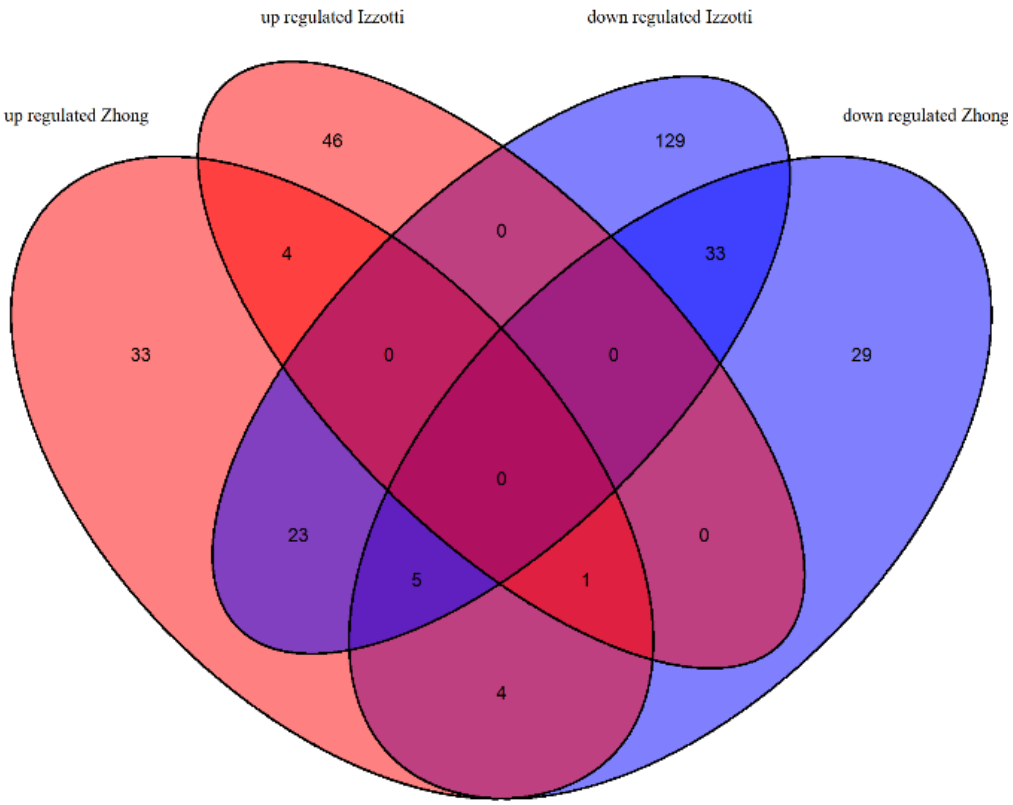


Figure 21 Comparisonbetween miRNAs mentioned in Zhong S, et al. (2021) and in our results from Izzotti et al. (2021) using Venn Diagram

All the mentioned limitations, especially when analysing miRNA regulation data, need to be overcome. The first question remains what are the biological processes and the individual variability factors that define the inconsistence of miRNA expression in lung tissues across different studies.

The identification of novel biomarkers based on miRNA profiles from accessible biological samples, like blood, would help in a near future a better understand of patient's health state. Outcomes like a better malignant tumour tissue early detection, over time therapy

effectiveness prediction, and patient's survival prediction rates, may become a reality. The identification of useful circulating miRNAs for predictive outcomes may require models based on standardized tissue-specific and blood-based profiles in oncologic patients.

As seeing in our results, the presence of different mutations variates the Scatter Plot in each case. From Cancer Related miRNAs significantly altered in each mutation, only 4 were predicted by TargetScan to target the considered genes: hsa-miR-15b-3p and hsa-miR-21-3p for KRAS, hsa-miR-548aa for STK11, and hsa-miR-205-3p for TP53. This suggest that the dysregulation in these miRNAs may worse mutation condition.

The possibility of using miRNAs as survival predictors is one of the most interesting proposals. Accordingly, 11 miRNAs were found as statistically significant dysregulate in dead patients within 3 years since the biopsy: hsa-miR-1227-5p, hsa-miR-147b, hsa-miR-187-5p, hsa-miR-23a-5p, hsa-miR-2861, hsa-miR-3663-5p, hsa-miR-371b-5p, hsa-miR-6068, hsa-miR-6075, hsa-miR-6771-5p and hsa-miR-7704. From this list, four miRNAs seem to be the most promising survival markers. It has been observed that miR-187-5p suppresses cancer cell progression in non-small cell lung cancer (NSCLC) through down-regulation of CYP1B1 [73], miR-147b promotes lung adenocarcinoma cell aggressiveness through glycoprotein 4 (MFAP4) regulation [74], miR-2861 has being proposed as a biomarker of Lung Cancer Stem Cells [75], miR-6075 has being used as a biomarker for lung cancer high-accuracy diagnosis prediction models [76].

The proposal of using miRNAs as survival predictors instead of, our analysis confirms that survival is not correlated to the number of oncogene mutations.

5. Bibliography

1. Ferlay J, Colombet M, Soerjomataram I, Dyba T, Randi G, Bettio M, et al. Cancer incidence and mortality patterns in Europe: Estimates for 40 countries and 25 major cancers in 2018. European Journal of Cancer [Internet]. Elsevier BV; 2018 Nov;103:356–87. Available from: http://dx.doi.org/10.1016/j.ejca.2018.07.005
2. National Institute of Health, National Cancer Institution, N.I.H. Lung and Bronchus Cancer; Cancer Stat Facts. [Online]. Available from: https://seer.cancer.gov/statfacts/html/lungb.html [Accessed 23 June 2021
3. Snowsill T, Yang H, Griffin E, Long L, Varley-Campbell J, Coelho H, et al. Low-dose computed tomography for lung cancer screening in high-risk populations: a systematic review and economic evaluation. Health Technology Assessment [Internet]. National Institute for Health Research; 2018 Nov;22(69):1–276. Available from: http://dx.doi.org/10.3310/hta22690
4. Toumazis I, Bastani M, Han SS, Plevritis SK. Risk-Based lung cancer screening: A systematic review. Lung Cancer [Internet]. Elsevier BV; 2020 Sep;147:154–86. Available from: http://dx.doi.org/10.1016/j.lungcan.2020.07.007
5. Lee PN, Forey BA, Thornton AJ, Coombs KJ. The relationship of cigarette smoking in Japan to lung cancer, COPD, ischemic heart disease and stroke: A systematic review. F1000Research [Internet]. F1000 Research Ltd; 2018 Feb 19;7:204. Available from: http://dx.doi.org/10.12688/f1000research.14002.1
6. Dias M, Linhas R, Campainha S, Conde S, Barroso A. Lung cancer in never-smokers – what are the differences? Acta Oncologica [Internet]. Informa UK Limited; 2017 Feb 17;56(7):931–5. Available from: http://dx.doi.org/10.1080/0284186x.2017.1287944
7. Kawaguchi T, Koh Y, Ando M, Ito N, Takeo S, Adachi H, et al. Prospective Analysis of Oncogenic Driver Mutations and Environmental Factors: Japan Molecular Epidemiology for Lung Cancer Study. Journal of Clinical Oncology [Internet]. American Society of Clinical Oncology (ASCO); 2016 Jul 1;34(19):2247–57. Available from: http://dx.doi.org/10.1200/jco.2015.64.2322
8. Dubin S, Griffin D. Lung Cancer in Non-Smokers. Mo Med. 2020 Jul-Aug;117(4):375-379. PMID: 32848276
9. Rupaimoole R, Slack FJ. MicroRNA therapeutics: towards a new era for the management of cancer and other diseases. Nature Reviews Drug Discovery [Internet]. Springer Science and Business Media LLC; 2017 Feb 17;16(3):203–22. Available from: http://dx.doi.org/10.1038/nrd.2016.246
10. Fabian MR, Sonenberg N. The mechanics of miRNA-mediated gene silencing: a look under the hood of miRISC. Nature Structural & Molecular Biology [Internet]. Springer Science and Business Media LLC; 2012 Jun;19(6):586–93. Available from: http://dx.doi.org/10.1038/nsmb.2296
11. Lu TX, Rothenberg ME. MicroRNA. Journal of Allergy and Clinical Immunology [Internet]. Elsevier BV; 2018 Apr;141(4):1202–7. Available from: http://dx.doi.org/10.1016/j.jaci.2017.08.034
12. Wang Y, Guan J, Wang Y. Could microRNA be used as a diagnostic tool for lung cancer? Journal of Cellular Biochemistry [Internet]. Wiley; 2019 Jun 25;120(11):18937–45. Available from: http://dx.doi.org/10.1002/jcb.29214

13. O'Brien K. TaqMan Low Density Array: MicroRNA Profiling for Biomarker and Oncosuppressor Discovery. MicroRNA Profiling [Internet]. Springer New York; 2016 Nov 9;71–84. Available from: http://dx.doi.org/10.1007/978-1-4939-6524-3_8
14. Leidner R. S., Li L., Thompson C. L. Dampening enthusiasm for circulating microRNA in breast cancer. PLoS ONE. 2013;8(3) doi: 10.1371/journal.pone.0057841.e57841
15. Chakraborty C, Sharma AR, Sharma G, Doss CGP, Lee S-S. Therapeutic miRNA and siRNA: Moving from Bench to Clinic as Next Generation Medicine. Molecular Therapy - Nucleic Acids [Internet]. Elsevier BV; 2017 Sep;8:132–43. Available from: http://dx.doi.org/10.1016/j.omtn.2017.06.005
16. Esquela-Kerscher A, Slack FJ. Oncomirs — microRNAs with a role in cancer. Nature Reviews Cancer [Internet]. Springer Science and Business Media LLC; 2006 Apr;6(4):259–69. Available from: http://dx.doi.org/10.1038/nrc1840
17. Karube Y, Tanaka H, Osada H, Tomida S, Tatematsu Y, Yanagisawa K, et al. Reduced expression of Dicer associated with poor prognosis in lung cancer patients. Cancer Science [Internet]. Wiley; 2005 Feb;96(2):111–5. Available from: http://dx.doi.org/10.1111/j.1349-7006.2005.00015.x
18. Jiang S. A Regulator of Metabolic Reprogramming: MicroRNA Let-7. Translational Oncology [Internet]. Elsevier BV; 2019 Jul;12(7):1005–13. Available from: http://dx.doi.org/10.1016/j.tranon.2019.04.013
19. Pop-Bica C, Pintea S, Magdo L, Cojocneanu R, Gulei D, Ferracin M, et al. The Clinical Utility of miR-21 and let-7 in Non-small Cell Lung Cancer (NSCLC). A Systematic Review and Meta-Analysis. Frontiers in Oncology [Internet]. Frontiers Media SA; 2020 Oct 19;10. Available from: http://dx.doi.org/10.3389/fonc.2020.516850
20. Esquela-Kerscher A, Trang P, Wiggins JF, Patrawala L, Cheng A, Ford L, et al. . The let-7 microRNA reduces tumor growth in mouse models of lung cancer. Cell Cycle. (2008) 7:759–64. 10.4161/cc.7.6.5834
21. Adjuvant chemotherapy after complete resection in non-small-cell lung cancer. Lung Cancer [Internet]. Elsevier BV; 1996 Nov;15(3):397. Available from: http://dx.doi.org/10.1016/s0169-5002(96)80072-9
22. Silva J, Garcia V, Zaballos A, Provencio M, Lombardia L, Almonacid L, et al. Vesicle-related microRNAs in plasma of nonsmall cell lung cancer patients and correlation with survival. European Respiratory Journal [Internet]. European Respiratory Society (ERS); 2010 Jul 1;37(3):617–23. Available from: http://dx.doi.org/10.1183/09031936.00029610
23. Msd manual, M.S.D. Lung Carcinoma. [Online]. Available from: https://www.msdmanuals.com/professional/pulmonary-disorders/tumors-of-the-lungs/lung-carcinoma [Accessed 5 July 2021]
24. Lu C, Onn A, Vaporciyan AA, et al. (2010). "Chapter 78: Cancer of the Lung". Holland-Frei Cancer Medicine (8th ed.). People's Medical Publishing House. ISBN 978-1-60795-014-1
25. Sui Q, Liang J, Hu Z, Chen Z, Bi G, Huang Y, et al. Genetic and microenvironmental differences in non-smoking lung adenocarcinoma patients compared with smoking patients. Translational Lung Cancer Research [Internet]. AME Publishing Company; 2020 Aug;9(4):1407–21. Available from: http://dx.doi.org/10.21037/tlcr-20-276

26. Izzotti A, Balansky R, Ganchev G, Iltcheva M, Longobardi M, Pulliero A, et al. Blood and lung microRNAs as biomarkers of pulmonary tumor-igenesis in cigarette smoke-exposed mice. Oncotarget [Internet]. Impact Journals, LLC; 2016 Oct 5;7(51)
27. Momi N, Kaur S, Rachagani S, Ganti AK, Batra SK. Smoking and microRNA dysregulation: a cancerous combination. Trends in Molecular Medicine [Internet]. Elsevier BV; 2014 Jan;20(1):36–47. Available from: http://dx.doi.org/10.1016/j.molmed.2013.10.005
28. Espín-Pérez A, Krauskopf J, Chadeau-Hyam M, van Veldhoven K, Chung F, Cullinan P, et al. Short-term transcriptome and microRNAs responses to exposure to different air pollutants in two population studies. Environmental Pollution [Internet]. Elsevier BV; 2018 Nov;242:182–90.
29. Santarelli L, Gaetani S, Monaco F, Bracci M, Valentino M, Amati M, et al. Four-miRNA Signature to Identify Asbestos-Related Lung Malignancies. Cancer Epidemiology Biomarkers & Prevention [Internet]. American Association for Cancer Research (AACR); 2018 Sep 26;28(1):119–26
30. Chen Z, Wang D, Gu C, Liu X, Pei W, Li J, et al. Down-regulation of let-7 microRNA increased K-ras expression in lung damage induced by radon. Environmental Toxicology and Pharmacology [Internet]. Elsevier BV; 2015 Sep;40(2):541–8. Available from: http://dx.doi.org/10.1016/j.etap.2015.08.009
31. He X, Chen J, Zhang Z, Li C, Peng Q, Peng H. The let-7a microRNA protects from growth of lung carcinoma by suppression of k-Ras and c-Myc in nude mice. Journal of Cancer Research and Clinical Oncology [Internet]. Springer Science and Business Media LLC; 2009 Dec 22;136(7):1023–8. Available from: http://dx.doi.org/10.1007/s00432-009-0747-5
32. Johnson SM, Grosshans H, Shingara J, Byrom M, Jarvis R, Cheng A, et al. RAS Is Regulated by the let-7 MicroRNA Family. Cell [Internet]. Elsevier BV; 2005 Mar;120(5):635–47. Available from: http://dx.doi.org/10.1016/j.cell.2005.01.014
33. BRUNO C, COMBA P, ZONA A. Adverse Health Effects of Fluoro-Edenitic Fibers: Epidemiological Evidence and Public Health Priorities. Annals of the New York Academy of Sciences [Internet]. Wiley; 2006 Sep 1;1076(1):778–83. Available from: http://dx.doi.org/10.1196/annals.1371.020
34. Nicoletti A, Vasta R, Venti V, Mostile G, Lo Fermo S, Patti F, et al. The epidemiology of amyotrophic lateral sclerosis in the Mount Etna region: a possible pathogenic role of volcanogenic metals. European Journal of Neurology [Internet]. Wiley; 2016 Feb 29;23(5):964–72. Available from: http://dx.doi.org/10.1111/ene.12973
35. Giacoppo S, Galuppo M, Calabrò RS, D'Aleo G, Marra A, Sessa E, et al. Heavy Metals and Neurodegenerative Diseases: An Observational Study. Biological Trace Element Research [Internet]. Springer Science and Business Media LLC; 2014 Aug 9;161(2):151–60. Available from: http://dx.doi.org/10.1007/s12011-014-0094-5
36. Stracquadanio M, Dinelli E, Trombini C. Role of volcanic dust in the atmospheric transport and deposition of polycyclic aromatic hydrocarbons and mercury. Journal of Environmental Monitoring [Internet]. Royal Society of Chemistry (RSC); 2003;5(6):984. Available from: http://dx.doi.org/10.1039/b308587b
37. Pierre P. Massion, Lecia V. Sequist, William Pao. Biology of Lung Cancer, Murray and Nadel's Textbook of Respiratory Medicine (Sixth Edition), W.B. Saunders, 2016. Pages

912-926. ISBN 9781455733835. Available from https://doi.org/10.1016/B978-1-4557-3383-5.00051-8.
38. Peluso M, Merlo F, Munnia A, Bolognesi C, Puntoni R, Parodi S. (32)P-postlabeling detection of DNA adducts in peripheral white blood cells of greenhouse floriculturists from western Liguria, Italy. Cancer Epidemiol Biomarkers Prev. 1996 May;5(5):361-9. PMID: 9162302
39. Shiizaki K, Kawanishi M, Yagi T. Modulation of benzo[a]pyrene–DNA adduct formation by CYP1 inducer and inhibitor. Genes and Environment [Internet]. Springer Science and Business Media LLC; 2017 Apr 10;39(1). Available from: http://dx.doi.org/10.1186/s41021-017-0076-x
40. Yeh HJC, Sayer JM, Liu X, Altieri AS, Byrd RA, Lakshman MK, et al. NMR Solution Structure of a Nonanucleotide Duplex with a dG Mismatch Opposite a 10S Adduct Derived from Trans Addition of a Deoxyadenosine N6-Amino Group to (+)-(7R,8S,9S,10R)-7,8-Dihydroxy-9,10-epoxy- 7,8,9,10-tetrahydrobenzo[a]pyrene: An Unusual syn Glycosidic Torsion Angle at the Modified dA. Biochemistry [Internet]. American Chemical Society (ACS); 1995 Oct 17;34(41):13570–81. Available from: http://dx.doi.org/10.1021/bi00041a037
41. Izzotti A, Coronel Vargas G, Pulliero A, Coco S, Vanni I, Colarossi C, et al. Relationship between the miRNA Profiles and Oncogene Mutations in Non-Smoker Lung Cancer. Relevance for Lung Cancer Personalized Screenings and Treatments. Journal of Personalized Medicine [Internet]. MDPI AG; 2021 Mar 5;11(3):182. Available from: http://dx.doi.org/10.3390/jpm11030182
42. Izzotti A, Coronel Vargas G, Pulliero A, Coco S, Colarossi C, Blanco G, et al. Identification by MicroRNA Analysis of Environmental Risk Factors Bearing Pathogenic Relevance in Non-Smoker Lung Cancer. Journal of Personalized Medicine [Internet]. MDPI AG; 2021 Jul 15;11(7):666. Available from: http://dx.doi.org/10.3390/jpm11070666
43. Alexandrov, K.; Rojas ,M.; Geneste, O., Castegnaro, M.; Camus, A.M.; Petruzzelli, S.; Giuntini, C.; Bartsch H. An improved fluorometric assay for dosimetry of benzo(a)pyrene diol-epoxide-DNA adducts in smokers' lung: comparisons with total bulky adducts and aryl hydrocarbon hydroxylase activity. Cancer Res. 1992,52(22),6248–6253
44. Oliveri Conti, G.; Calogero, A.E.; Giacone, F.; Fiore, M.; Barchitta, M.; Agodi, A.; Ferrante, M. B(a)P adduct levels and fertility: A cross sectional study in a Sicilian population. Mol Med Rep. 2017,15,3398–3404. DOI:10.3892/mmr.2017.6396
45. Creixell P, Reimand J, Haider S, Wu G, Shibata T, Vazquez M, Mustonen V, Gonzalez-Perez A, Pearson J, Sander C, Raphael BJ, Marks DS, Ouellette BFF, Valencia A, Bader GD, Boutros PC, Stuart JM, Linding R, Lopez-Bigas N, Stein LD; Mutation Consequences and Pathway Analysis Working Group of the International Cancer Genome Consortium. Pathway and network analysis of cancer genomes. Nat Methods. 2015 Jul;12(7):615-621. doi: 10.1038/nmeth.3440. PMID: 26125594; PMCID: PMC4717906
46. Pajak M, Simpson TI (2020). miRNAtap: miRNAtap: microRNA Targets - Aggregated Predictions. R package version 1.22.0. DOI: 10.18129/B9.bioc.miRNAtap), and topGO R Package

47. Alexa A, Rahnenfuhrer J (2020). topGO: Enrichment Analysis for Gene Ontology. R package version 2.40.0. Doi: 10.18129/B9.bioc.topGO
48. Supek F, Bošnjak M, Škunca N, Šmuc T. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. Gibas C, editor. PLoS ONE [Internet]. Public Library of Science (PLoS); 2011 Jul 18;6(7):e21800. Available from: http://dx.doi.org/10.1371/journal.pone.0021800
49. López-Romero P, González MA, Callejas S, Dopazo A, Irizarry RA. Processing of Agilent microRNA array data. BMC Re-search Notes [Internet]. Springer Science and Business Media LLC; 2010;3(1):18. Available from: http://dx.doi.org/10.1186/1756-0500-3-18
50. Agilent technologies, inc. 278 Neural Network. In: Agilent technologies, inc (ed.) GeneSpring User Manual. Santa Clara, CA 95051 USA: ; 2016. p. 823-825. Available at https://www.agilent.com/cs/library/usermanuals/public/GeneSpring-manual.pdf
51. Lee H, Han S, Kwon CS, Lee D. Biogenesis and regulation of the let-7 miRNAs and their functional implications. Protein & Cell [Internet]. Springer Science and Business Media LLC; 2015 Sep 23;7(2):100–13. Available from: http://dx.doi.org/10.1007/s13238-015-0212-y.
52. Mohr A, Mott J. Overview of MicroRNA Biology. Seminars in Liver Disease [Internet]. Georg Thieme Verlag KG; 2015 Jan 29;35(01):003–11. Available from: http://dx.doi.org/10.1055/s-0034-1397344
53. Brabletz T, Jung A, Spaderna S, Hlubek F, Kirchner T. Migrating cancer stem cells — an integrated concept of malignant tumour progression. Nature Reviews Cancer [Internet]. Springer Science and Business Media LLC; 2005 Sep;5(9):744–9. Available from: http://dx.doi.org/10.1038/nrc1694
54. Kitamura H, Yazawa T, Okudela K, Shimoyamada H, Sato H. Molecular and Genetic Pathogenesis of Lung Cancer: Differences Between Small-Cell and Non-Small-Cell Carcinomas. The Open Pathology Journal [Internet]. Bentham Science Publishers Ltd.; 2008 Oct 30;2(1):106–14. Available from: http://dx.doi.org/10.2174/1874375700802010106
55. George J, Lim JS, Jang SJ, Cun Y, Ozretić L, Kong G, et al. Comprehensive genomic profiles of small cell lung cancer. Nature [Internet]. Springer Science and Business Media LLC; 2015 Jul 13;524(7563):47–53. Available from: http://dx.doi.org/10.1038/nature14664
56. Kanwal M, Ding X-J, Cao Y. Familial risk for lung cancer. Oncology Letters [Internet]. Spandidos Publications; 2016 Dec 20;13(2):535–42. Available from: http://dx.doi.org/10.3892/ol.2016.5518
57. Chapman AM, Sun KY, Ruestow P, Cowan DM, Madl AK. Lung cancer mutation profile of EGFR, ALK, and KRAS: Meta-analysis and comparison of never and ever smokers. Lung Cancer [Internet]. Elsevier BV; 2016 Dec;102:122–34. Available from: http://dx.doi.org/10.1016/j.lungcan.2016.10.010
58. Zhang R, Tian P, Chen B, Wang T, Li W. The prognostic impact of TP53 comutation in EGFR mutant lung cancer patients: a systematic review and meta-analysis. Postgraduate Medicine [Internet]. Informa UK Limited; 2019 Mar 15;131(3):199–206. Available from: http://dx.doi.org/10.1080/00325481.2019.1585690
59. Rodosthenous RS, Coull BA, Lu Q, Vokonas PS, Schwartz JD, Baccarelli AA. Ambient particulate matter and microRNAs in extracellular vesicles: a pilot study of older

individuals. Particle and Fibre Toxicology [Internet]. Springer Science and Business Media LLC; 2015 Dec;13(1). Available from: http://dx.doi.org/10.1186/s12989-016-0121-0
60. Pergoli L, Cantone L, Favero C, Angelici L, Iodice S, Pinatel E, et al. Extracellular vesicle-packaged miRNA release after short-term exposure to particulate matter is associated with increased coagulation. Particle and Fibre Toxicology [Internet]. Springer Science and Business Media LLC; 2017 Aug 24;14(1). Available from: http://dx.doi.org/10.1186/s12989-017-0214-4
61. Li J, Wang T, Wang Y, Xu M, Zhang L, Li X, et al. Particulate matter air pollution and the expression of microRNAs and pro-inflammatory genes: Association and mediation among children in Jinan, China. Journal of Hazardous Materials [Internet]. Elsevier BV; 2020 May;389:121843. Available from: http://dx.doi.org/10.1016/j.jhazmat.2019.121843
62. Chen R, Li H, Cai J, Wang C, Lin Z, Liu C, et al. Fine Particulate Air Pollution and the Expression of microRNAs and Circulating Cytokines Relevant to Inflammation, Coagulation, and Vasoconstriction. Environmental Health Perspectives [Internet]. Environmental Health Perspectives; 2018 Jan 16;126(1):017007. Available from: http://dx.doi.org/10.1289/ehp1447
63. Rodosthenous RS, Kloog I, Colicino E, Zhong J, Herrera LA, Vokonas P, et al. Extracellular vesicle-enriched microRNAs interact in the association between long-term particulate matter and blood pressure in elderly men. Environmental Research [Internet]. Elsevier BV; 2018 Nov;167:640–9. Available from: http://dx.doi.org/10.1016/j.envres.2018.09.002
64. Li Z, Ma J, Bi J, Guo H, Chan MTV, Wu WKK, et al. MicroRNA signature of air pollution exposure-induced congenital defects. Journal of Cellular Physiology [Internet]. Wiley; 2019 Mar 18;234(10):17896–904. Available from: http://dx.doi.org/10.1002/jcp.28422
65. Li X, Lv Y, Hao J, Sun H, Gao N, Zhang C, et al. Role of microRNA-4516 involved autophagy associated with exposure to fine particulate matter. Oncotarget [Internet]. Impact Journals, LLC; 2016 Jun 13;7(29):45385–97. Available from: http://dx.doi.org/10.18632/oncotarget.9978
66. Joshi P. MicroRNA in lung cancer. World Journal of Methodology [Internet]. Baishideng Publishing Group Inc.; 2014;4(2):59. Available from: http://dx.doi.org/10.5662/wjm.v4.i2.59
67.] Tutar Y, Özgür A, Tutar E, Tutar L, Pulliero A, Izzotti A. Regulation of oncogenic genes by MicroRNAs and pseudogenes in human lung cancer. Biomedicine & Pharmacotherapy [Internet]. Elsevier BV; 2016 Oct;83:1182–90. Available from: http://dx.doi.org/10.1016/j.biopha.2016.08.043
68. Bianchi F, Nicassio F, Marzi M, Belloni E, Dall'Olio V, Bernard L, et al. A serum circulating miRNA diagnostic test to identify asymptomatic high-risk individuals with early stage lung cancer. EMBO Molecular Medicine [Internet]. EMBO; 2011 Jul 11;3(8):495–503. Available from: http://dx.doi.org/10.1002/emmm.201100154
69. Pradervand S, Weber J, Thomas J, Bueno M, Wirapati P, Lefort K, et al. Impact of normalization on miRNA microarray expression profiling. RNA [Internet]. Cold Spring

Harbor Laboratory; 2009 Jan 20;15(3):493–501. Available from: http://dx.doi.org/10.1261/rna.1295509

70. Zhong S, Golpon H, Zardo P, Borlak J. miRNAs in lung cancer. A systematic review identifies predictive and prognostic miRNA candidates for precision medicine in lung cancer. Translational Research [Internet]. Elsevier BV; 2021 Apr;230:164–96. Available from: http://dx.doi.org/10.1016/j.trsl.2020.11.012
71. Zhang X, Kang H, Chen G, Li X. Prognostic Role of MicroRNA Let-7 in Non-Small Cell Lung Cancer: A Systematic Review and Meta-Analysis. Research Square; 2020 Jul 28; Available from: http://dx.doi.org/10.21203/rs.3.rs-46072/v1
72. Roth JA, Carlson JJ. Prognostic Role of ERCC1 in Advanced Non–Small-Cell Lung Cancer: A Systematic Review and Meta-Analysis. Clinical Lung Cancer [Internet]. Elsevier BV; 2011 Nov;12(6):393–401. Available from: http://dx.doi.org/10.1016/j.cllc.2011.04.005
73. Mao M, Wu Z, Chen J. MicroRNA-187-5p suppresses cancer cell progression in non-small cell lung cancer (NSCLC) through down-regulation of CYP1B1. Biochemical and Biophysical Research Communications [Internet]. Elsevier BV; 2016 Sep;478(2):649–55. Available from: http://dx.doi.org/10.1016/j.bbrc.2016.08.001
74. Feng Y-Y, Liu C-H, Xue Y, Chen Y-Y, Wang Y-L, Wu X-Z. MicroRNA-147b promotes lung adenocarcinoma cell aggressiveness through negatively regulating microfibril-associated glycoprotein 4 (MFAP4) and affects prognosis of lung adenocarcinoma patients. Gene [Internet]. Elsevier BV; 2020 Mar;730:144316. Available from: http://dx.doi.org/10.1016/j.gene.2019.144316
75. Zhao M, Li L, Zhou J, Cui X, Tian Q, Jin Y, et al. MiR-2861 Behaves as a Biomarker of Lung Cancer Stem Cells and Regulates the HDAC5-ERK System Genes. Cellular Reprogramming [Internet]. Mary Ann Liebert Inc; 2018 Apr;20(2):99–106. Available from: http://dx.doi.org/10.1089/cell.2017.0045
76. Asakura K, Kadota T, Matsuzaki J, Yoshida Y, Yamamoto Y, Nakagawa K, et al. A miRNA-based diagnostic model predicts resectable lung cancer in humans with high accuracy. Communications Biology [Internet]. Springer Science and Business Media LLC; 2020 Mar 19;3(1). Available from: http://dx.doi.org/10.1038/s42003-020-0863-y