UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
FACULDADE DE MEDICINA

Ana Laura Seneda

Identificação de novos microRNAs associados ao metabolismo maligno do câncer de pulmão

Tese apresentada à Faculdade de Medicina, Universidade Estadual Paulista "Júlio de Mesquita Filho", Câmpus de Botucatu, para obtenção do título de Doutora em Cirurgia e Medicina Translacional.

Orientadora: Profa. Associada Patricia Pintor dos Reis
Co-Orientador: Prof. Dr. Luis Alejandro Jose Mur

Botucatu
2023

Ana Laura Seneda

Identificação de novos microRNAs associados ao metabolismo maligno de câncer de pulmão

Tese apresentada à Faculdade de Medicina, Universidade Estadual Paulista "Júlio de Mesquita Filho", Câmpus de Botucatu, para obtenção do título de Doutora em Cirurgia e Medicina Translacional.

Orientadora: Profa. Associada Patricia Pintor dos Reis
Co-Orientador: Prof. Dr. Luis Alejandro Jose Mur

Botucatu
2023

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP
BIBLIOTECÁRIA RESPONSÁVEL: MARIA CAROLINA A. CRUZ E SANTOS-CRB 8/10188

Seneda, Ana Laura.
Identificação de novos microRNAs associados ao metabolismo maligno do câncer de pulmão / Ana Laura Seneda. - Botucatu, 2023

Tese (doutorado) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Faculdade de Medicina de Botucatu
Orientador: Patricia Pintor dos Reis
Coorientador: Luis Alejandro Jose Mur
Capes: 40102050

1. Pulmões - Câncer. 2. Genes. 3. Tumores malignos. 4. MicroRNAs.

Palavras-chave: Câncer de pulmão; Genes alvos; Metabolismo maligno; Novos microRNAs.

unesp

UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Botucatu

ATA DA DEFESA PÚBLICA DA TESE DE DOUTORADO DE ANA LAURA SENEDA, DISCENTE DO PROGRAMA DE PÓS-GRADUAÇÃO EM CIRURGIA E MEDICINA TRANSLACIONAL, DA FACULDADE DE MEDICINA.

Aos 11 dias do mês de agosto do ano de 2023, às 14:00 horas, por meio de Videoconferência, realizou-se a defesa de TESE DE DOUTORADO de ANA LAURA SENEDA, intitulada **Identificação de novos microRNAs associados ao metabolismo maligno de câncer de pulmão**. A Comissão Examinadora foi constituida pelos seguintes membros: Profa. Dra. PATRICIA PINTOR DOS REIS (Orientador(a) - Participação Presencial) do(a) Depto. de Cirurgia e Ortopedia / FM/Botucatu - Unesp, Prof. Dr. CRISTIANO DE PÁDUA SOUZA (Participação Virtual) do(a) Hospital de Câncer de Barretos - Fundação Pio XII, Prof. Dr. ROBSON FRANCISCO CARVALHO (Participação Presencial) do(a) Depto. de Biologia Estrutural e Funcional / IB/Botucatu - Unesp. Após a exposição pela doutoranda e arguição pelos membros da Comissão Examinadora que participaram do ato, de forma presencial e/ou virtual, a discente recebeu o conceito final: APROVADA. Nada mais havendo, foi lavrada a presente ata, que após lida e aprovada, foi assinada pelo(a) Presidente(a) da Comissão Examinadora.

Profa. Dra. PATRICIA PINTOR DOS REIS

Faculdade de Medicina - Câmpus de Botucatu -
Av.: Prof. Mário Rubens Guimarães Montenegro, s/nº, 18618687

Dedicatória

A Deus por me permitir chegar até aqui e concluir este trabalho.

À minha família, meus pais, Mariza da Silva Seneda e Carlos Alberto Seneda, e minhas irmãs, Letícia Fernanda Seneda e Maria Júlia Seneda, por todo carinho e apoio incondicional, compreendendo minhas ausências, e sempre me incentivando a buscar por meus objetivos, e por não medirem esforços para que eu chegasse até aqui.
A vocês, minha imensa e eterna gratidão.

Dedico, ainda, a todos os pacientes e seus familiares que são afetados por esta doença. Espero fortemente que este trabalho possa contribuir significativamente para melhorar sua qualidade de vida, bem como contribuir para futuras diretrizes terapêuticas.

Agradecimentos

Agradeço imensamente à minha orientadora, Dra. Patrícia Pintor dos Reis, pelas orientações, pela paciência, pela atenção, pelo tempo disposto e incentivo.
Você foi fundamental para a conclusão deste trabalho. Admiro demais a pesquisadora que é, um grande exemplo a ser seguido, a qual me inspiro muito. Sou muito grata por nossa parceria ao longo desta jornada. Você me ensinou muito. Obrigada por tudo.

Ao meu co-orientador, Prof. Dr. Luis Alejandro Jose Mur, pelo tempo disposto, atenção, ensinamentos e incentivo.

Aos professores membros da banca, Dr. Cristiano Pádua de Souza e Dr. Robson Francisco Carvalho pelas valorosas contribuições.

A todos os amigos e amigas, que me acompanharam e torceram por mim nesta jornada, especialmente Carolina Fazio Campos, Iael Weissberg Minutentag e Tainara Francini Felix, pelo carinho, ajuda, suporte aos dias difíceis e mais

ainda por torná-los mais leves com palavras de apoio. Vocês são incríveis.

À equipe do Laboratório de Oncologia Molecular e Translacional, especialmente ao Dr. Márcio de Carvalho por toda ajuda técnica e à Dra. Sandra Aparecida Drigo Linde pelo apoio.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela bolsa concedida.

Aos colaboradores do projeto, Dr. Mateus Barros-Filho e Dr. Wan Lam.

À Faculdade de Medicina de Botucatu (FMB-UNESP) e à Unidade de Pesquisa Experimental (UNIPEX) pela estrutura disponibilizada e a todos os seus funcionários.

A todos os funcionários da Biblioteca e da Seção Técnica de Pós-Graduação, em especial à secretária do Programa de Cirurgia e Medicina Translacional, Márcia Fonseca Piagentini Cruz, por toda ajuda e apoio.

Ao NCBI dbGaP e ao TCGA por autorização a obtenção dos dados utilizados para a análise computacional.

A todos que, direta ou indiretamente, contribuíram para a realização desse trabalho.

"Em algum lugar, algo incrível está esperando para ser descoberto"

Carl Sagan

RESUMO

Introdução: O câncer de pulmão é a principal causa de morte por câncer no mundo e compreende dois principais subtipos histológicos: adenocarcinoma (LUAD) e carcinoma de células escamosas (LUSC). MicroRNAs (miRNAs) são pequenos RNAs não codificantes que regulam a expressão gênica e são conhecidos por modular processos celulares, incluindo o metabolismo, que está alterado nas células cancerígenas. De acordo com o miRBase, um banco de dados de sequências de miRNAs, atualmente são conhecidos 2.654 miRNAs humanos maduros. No entanto, estudos recentes sugerem a existência de novos miRNAs com um potencial papel no desenvolvimento e progressão da doença. **Objetivos:** Identificar novas sequências de miRNAs no câncer de pulmão. **Métodos:** Os dados de sequenciamento de miRNA publicamente disponíveis (miRNA-Seq) do LUAD e LUSC foram obtidos do *The Cancer Genome Atlas* (TCGA). O miRMaster foi utilizado para predição de novos miRNAs, com critério de filtragem de FC> 2 e p <0,001. Os genes alvo dos miRNAs foram investigados usando mirDB. Em seguida, a expressão dos genes alvo foi validada (*in silico*) usando conjuntos de dados de RNA-seq (LUAD e LUSC) obtidos do *Xena Browser*. Análise de correlação de Spearman entre miRNAs e genes alvo foi realizada. PathDB foi usado para identificar vias metabólicas incluindo genes alvo de miRNA. **Resultados:** 126 sequências de novos miRNAs previstos foram identificadas no LUAD e 225 no LUSC. Destas, 18 novas sequências foram correlacionadas negativamente (r> 50%; p <0,0001) com 47 genes alvo no LUAD, e 66 foram correlacionadas negativamente com 473 genes alvo no LUSC. Destes, 13 novos miRNAs foram comumente superexpressos em LUAD e LUSC. Em seguida, identificamos (*in silico*) o papel dos novos miRNAs no metabolismo do câncer. Encontramos duas vias principais: metabolismo de carboidratos (136 vias incluindo 30 genes alvo de miRNA) e metabolismo energético (64 vias e 26 genes alvo de miRNA). **Conclusões:** Um subconjunto de novos miRNAs está desregulado no câncer de pulmão versus tecidos pulmonares normais, com papéis potenciais no metabolismo maligno associado à

tumorigênese pulmonar. Os miRNAs identificados modulam a expressão de genes alvo com funções biológicas nas vias do metabolismo de carboidratos e energia.

Palavras-chaves: novos microRNAs, câncer de pulmão, genes alvos, metabolismo maligno.

ABSTRACT

Introduction: Lung cancer is the main cause of cancer death worldwide and comprises two main histological subtypes: lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC). MicroRNAs (miRNAs) are small non-coding RNAs that regulate gene expression, and are known to modulate cellular processes, including metabolism, which is altered in cancer cells. According to miRBase, a database of miRNA sequences, 2,654 mature human miRNAs are currently known. However, recent studies suggested the existence of novel miRNAs with a potential role in disease development and progression. **Objectives:** To identify novel miRNA sequences in lung cancer. **Study design and methods:** Publicly available miRNA sequencing (miRNA-Seq) data from LUAD and LUSC were obtained from The Cancer Genome Atlas (TCGA). miRMaster was used for novel miRNA prediction, with a filtering criterion of FC>2 and $p<0.001$. miRNA target genes were investigated using mirDB. Next, target gene expression was validated (*in silico*) using LUAD and LUSC RNA-seq datasets obtained from Xena Browser. Spearman correlation analysis between miRNAs and target genes were performed. PathDB was used to identify metabolic pathways including miRNA target genes. **Results:** 126 sequences of predicted novel miRNAs were identified in LUAD, and 225 in LUSC. Of these, 18 novel sequences were negatively correlated (r>50%; $p<0.0001$) with 47 target genes in LUAD, and 66 were negatively correlated with 473 target genes in LUSC. Of these, 13 novel miRNAs were commonly overexpressed in LUAD and LUSC. Next, we identified (*in silico*) the role of novel miRNAs in cancer metabolism. We found two main pathways: carbohydrate metabolism (136 pathways including 30 miRNA target genes) and energy metabolism (64 pathways and 26 miRNA target genes). **Conclusions:** A subset of novel miRNAs are deregulated in lung cancer *vs*. normal lung tissues, with potential roles in malignant metabolism associated with lung tumorigenesis. The identified miRNAs modulate the expression of target genes with biological roles in pathways of carbohydrate and energy metabolism.

Key-words: novel microRNAs, lung cancer, target genes, malignant metabolism.

SUMMARY

1. INTRODUCTION

1.1 Lung Cancer: Epidemiology and Risk Factors

The World Health Organization (WHO) (2020) reported that lung cancer is the fourth most prevalent and the second most incident cancer, and the leading cause of cancer death worldwide (**Figures 1A, 1B**, **1C**) (1). In Brazil, according to the National Institute of Cancer (INCA, 2023), lung cancer is the third most incident cancer among men (7,5% of all new cases), and the fourth most incident (6%) in women. Regarding mortality, lung cancer is the major cause of death in men (13,6% of total deaths related to cancer) and the second (11,6%) in women (2). The high mortality index is mainly related to diagnosis of advanced stage disease (3).

Although lung cancer is more commonly observed in older patients, with a median age of diagnosis of 71 years, in the United States (4), the number of new cases is increasing in individuals from 45 years of age, showing that diagnosis of lung cancer is now becoming more common in younger individuals (3).

Smoking is still the main risk factor for lung cancer development. According to INCA (2023), cigarette smoking is associated with approximately 90% of deaths by lung cancer. Although, about 10% of patients diagnosed with lung cancer have never smoked (3), showing that there are other risk factors that play a role in disease development, such as air pollution, ionizing radiation, secondhand smoke, and occupational exposure to carcinogens such as asbestos, radon, and metals (mainly arsenic, chromium and nickel). A history of previous lung diseases, and human immunodeficiency virus (HIV) infection (4). Diseases that lead to pulmonary chronic tissue inflammation, damage and fibrosis, as well as gene expression changes, including pulmonary tuberculosis (5,6) and chronic obstructive pulmonary disease (COPD) (7), also elevate the risk of lung cancer development. These chronic diseases may be associated with lung cancer cases in younger individuals (8).

It is noteworthy that smoking is more strongly related to squamous cell carcinoma and small cell lung cancer subtypes (9).

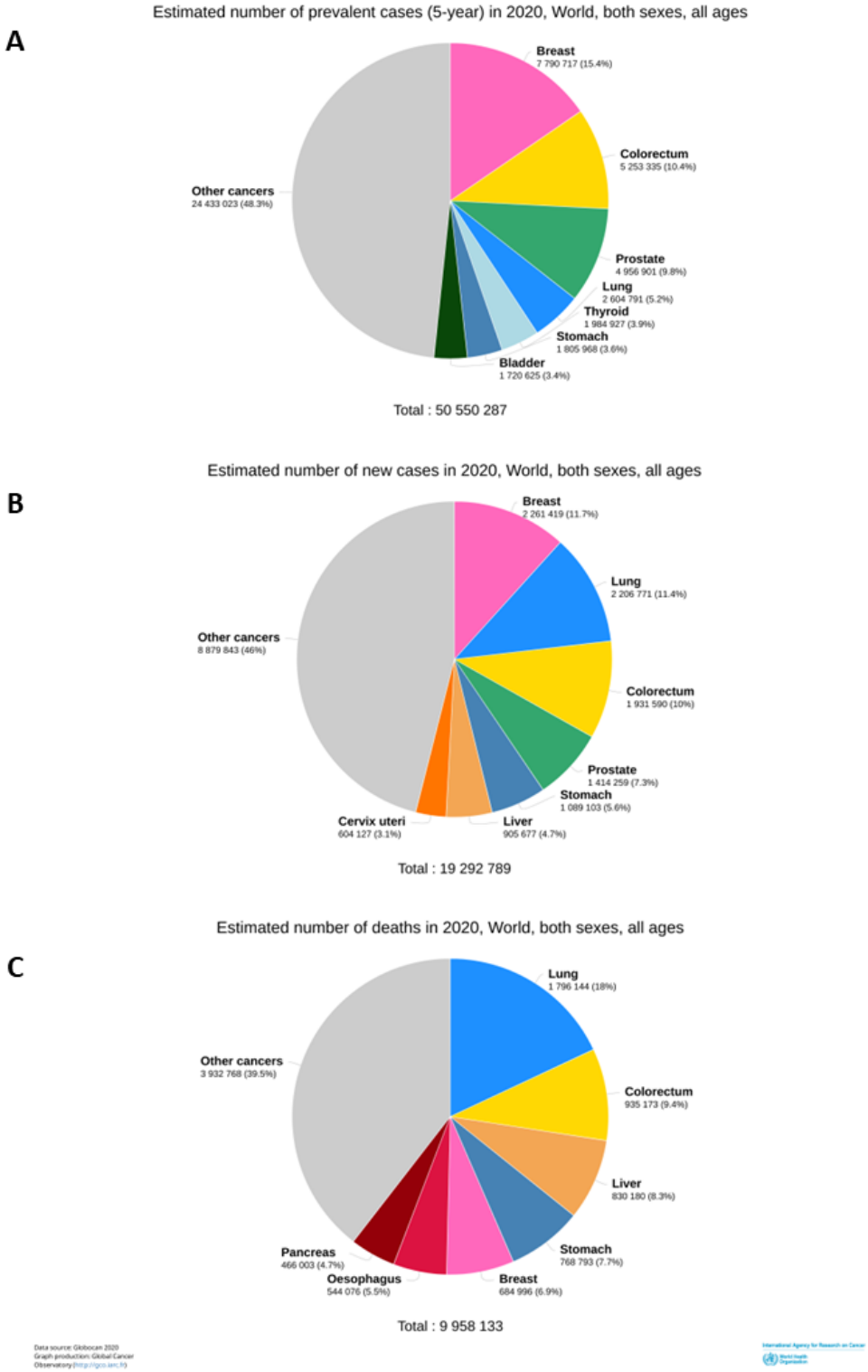

Figure 1. Prevalence, Incidence and Mortality by cancer. Lung cancer is shown in light blue. (**A**) Estimated number of cases for the most prevalent cancers; (**B**) Number of new cases (incidence) of cancer; (**C**) Estimated number of cancer-related deaths. Charts show worldwide data from both genders and for all ages for the year of 2020 (1). Accessed on: Dec., 7th 2022**.**

1.2 Histological Classification

According to the WHO, there are two main types of lung cancer: non-small cell lung cancer (NSCLC), which accounts for approximately 85% of cases, and small cell lung cancer (SCLC), which accounts for about 15% of cases. Among the NSCLC types, the most common are lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) (3,4).

1.2.1 Lung Adenocarcinoma

Lung Adenocarcinoma (LUAD) is the most common subtype of lung cancer, comprising 40% of NSCLC cases. In the lungs, adenocarcinoma arises from alveolar cells, which are located in the epithelium of the smaller airways, and tumor cells express the thyroid transcription factor-1 (TTF1) or napsin A (3,4) immunohistochemical markers. The presence of mucin granules (increased mucus production) in the cell cytoplasm (more than 5 cells in the field of view) also serves to classify lesions as adenocarcinoma (3).

The study by Noguchi *et al*. (10) described six subtypes of LUAD in 1995. This study became the basis for the most recent histological tumor classifications defined by The International Association for the Study of Lung Cancer (IASLC), The American Thoracic Society (ATS), The European Respiratory Society (ERS), and WHO (11).

Currently, it is accepted that LUAD starts and gradually develops from atypical adenomatous hyperplasia (AAH) to adenocarcinoma *in situ* (AIS), to minimally invasive adenocarcinoma (MIA) and, finally, to invasive cancer (12). AIS is defined as a lesion with a lepidic pattern and less than 3 cm in diameter. MIA measures over 3 cm in diameter (3,4); and invasive adenocarcinoma is classified according to its predominant histological pattern and morphology, with subtypes of adenocarcinoma being lepidic, acinar, papillary, micropapillary, and solid. About 90% of LUAD have mixed histology, and in this case are classified according to the predominant histological subtype (13,14). There are other variations to invasive LUAD that have received their own categories, namely: mucinous, mixed non-

mucinous and mucinous, colloid, fetal, and enteric invasive (14).

It is important to determine the histological pattern of disease, since it has implications in patient prognosis. Upon complete resection, non-invasive lesions such as adenocarcinoma *in situ* or minimally invasive adenocarcinoma, are not expected to recur; the same is observed for patients diagnosed with the lepidic pattern, which have better prognosis. The acinar and papillary subtypes have an intermediate risk of recurrence, while the micropapillary and solid subtypes have a higher risk of recurrence (14).

1.2.2. Lung Squamous Cell Carcinoma

Lung Squamous Cell Carcinoma (LUSC) corresponds to 25-30% of cases and presents a rapid progression and aggressive growth, explaining the diagnosis of advanced stage disease in most patients with squamous cell cancer of the lung (8). The presence of the immunohistochemical markers p40, p63, desmoglein-3 and cytokeratins 5 and 6, together with the absence of keratinization, are useful to confirm the histological subtype of squamous cell carcinoma, as they allow the tumor to be distinguished from adenocarcinomas that have similar morphology (3,4). LUSC can be distinguished in main four subtypes: keratinizing, non-keratinizing, basal cell carcinoma, and pre-invasive lesions (*in situ* squamous cell carcinoma) (3).

Squamous carcinoma of the lungs originates in the epithelial airway cells, which suffer morphological alterations usually associated with chronic inflammation, progressing from basal cell hyperplasia (BCH), squamous metaplasia (SM), grade I–III dysplasia, and carcinoma *in situ* (CIS) (3,4,8).

Basal Cell Hyperplasia occurs when there are three or more layers of basal cells and ciliated cells, with absence of goblet cells. The substitution of the cylindrical ciliated epithelium by the squamous epithelium is defined as the SM. Both subtypes, BCH and SM appeared commonly in smokers and in patients with COPD. Interestingly, these early, pre-neoplastic lesions may be reversible by cessation of exposure, as they result from physiological responses to damages

caused by air pollution, smoking, and inflammation. In fact, patients with pulmonary inflammatory diseases affecting the respiratory epithelium and/or exposed to high rates of air pollution frequently present SM (8).

Regarding dysplastic lesions, there are distinct grades of dysplasia: grade I (mild), which shows minimal abnormalities, slightly enlarged cells, vertically oriented nuclei, and rare mitoses; II (moderate), which presents more evident alterations, and mitoses limited to the lower third of epithelial layer; and III (severe), which exhibit cellular pleomorphism, basilar zone expanded with cellular crowding into the upper third, and mitoses confined to the lower two thirds of the airway epithelial layer (8).

1.3. Diagnosis, Staging, and Treatment

The differential diagnosis of lung cancer subtypes is defined by histopathological analysis, based on cell morphology and detection of immunohistochemical markers in tissue samples (biopsies). The clinical outcome of patients with NSCLC is strongly related to the stage of the disease at the time of diagnosis, evidencing the importance of an early diagnosis. Besides, the histological diagnosis is associated with several molecular alterations, evidencing the importance to differentiate the histological subtypes (3).

Mostly, the symptoms of lung cancer are nonspecific, the most common being cough, followed by hemoptysis, chest pain and dyspnea; other less common symptoms are laboratory alterations or paraneoplastic syndromes (4).

Reliable biomarkers for early disease detection have not been identified yet. However, several studies have demonstrated that miRNAs have a potential value as diagnostic biomarkers. The detection of miRNAs in serum was associated with image exams, such as tomography, and showed increased diagnostic performance (15,16).

Disease staging follows the TNM staging system (tumor, nodule, metastasis) and it is performed at the time of diagnosis, to guide treatment decisions (4).

Regarding therapeutic options for patients with lung cancer, surgery is an

effective treatment for early disease stages and depending on patient status, which determine eligibility to surgery. However, disease recurrence may be observed, with a 5-year overall survival varying from 83% to 36% for stage IA to stage IIIA disease, respectively. For patients with unresectable tumors, the standard treatment is the association of thoracic radiation and cytotoxic therapy (9). In patients with advanced disease stage, treatment is mainly based on combinations of platinum, such as cisplatin, with other cytotoxic agents (9).

In recent decades, treatment for lung cancer has evolved significantly, from the use of nonspecific cytotoxic therapies to more precise targeted treatments. Novel treatment strategies are directed towards histological and molecular subtypes of disease, often determined by a panel of mutations specific to LUAD, with therapies still under development for LUSC. The use of target therapies for lung cancer treatment started by the end of the 1990s, with the development of gefitinib, a tyrosine kinase inhibitor (TKI) targeting *EGFR* driver mutations. Another TKI targeting *EGFR* mutations, erlotinib, improved survival of patients with advanced NSCLC. Furthermore, the identification of new driver mutations led to development of new target therapies, targeting *ROS1* fusions, *BRAF* mutations, and *ALK* rearrangements (9).

Another class of therapies that are showing promising results in NSCLC are the immunotherapies, which consist of using immune checkpoints blockades, such as monoclonal antibodies targeting PD1/PD-L1 and CTLA-4. These treatments stimulate immune response against the tumor. Examples of FDA-approved immunotherapies for the NSCLC treatment are nivolumab and pembrolizumab, both anti-PD1. Immunotherapy has demonstrated an improvement in patient overall survival (9). However, there is still a need for the identification of biomarkers to improve disease treatment, since not all patients respond to targeted therapies and immunotherapies.

1.4 Molecular Classification

Similarly, to other solid malignancies, lung cancer is composed of cellular

subpopulations that may contain distinct molecular characteristics, defined as tumor heterogeneity (9). Such features influence molecular diagnosis and treatment decisions, as well as and the arising of drug resistance. Furthermore, a percentage of mutations in neoplastic cell subclones may be associated with disease recurrence in patients with localized lung adenocarcinoma (17). The understanding gained from previous scientific studies led to advancements in clinical practice for treating lung cancer. Previously reliant on broad cellular cytotoxic agents, therapeutic approaches have included personalized medicine tailored based on genetic alterations. In LUAD, molecular diagnostics have resulted in more effective treatments (9).

The epidermal growth factor receptor (*EGFR*) and the Kirsten rat sarcoma (*KRAS*) are among the most frequent gene mutations identified in LUAD. Interestingly, these alterations are generally found in clone-cells denominated "founders", since they play an important role in tumorigenesis and are useful for targeted therapeutics. Usually, such gene mutations are mutually exclusive, that is, if *EGFR* is mutated, *KRAS* is wild type and vice-versa; however, when they coexist, they may confer resistance to treatment with tyrosine kinase inhibitors. Other molecular changes such as *TP53* mutations, are more frequently observed in advanced-stage tumors, suggesting a role in tumor progression (9).

Regarding tumor suppressor genes, the most commonly mutated in LUAD are *TP53*, *KEAP1*, *STK11*, and *NF1*. It has been reported that *KEAP1* inactivation accompanied by *KRAS* mutations may contribute to sensitivity to glutaminase inhibition in preclinical models, showing its potential used for targeted treatment (9).

Commons alterations present in LUSC are *TP53* (>90% of cases) and *CDKN2A* inactivation. *EGFR* amplification has been also reported; however, *EGFR* mutations are not common in the squamous subtype (9).

Additionally, molecular profiles of lung cancer in smokers is different from that of non-smokers. It is common to find significantly higher frequency of mutations in patients with a history of smoking including *KRAS* and *TP53*

mutations. In non-smokers, it is more common to find changes in actionable driver genes, such as activating mutations in *EGFR* and translocations in *ROS1* and *ALK* (9). In fact, *EGFR* mutations were significantly associated with tumors arising in non-smokers while *KRAS* mutations were significantly associated in lung adenocarcinoma in smoker patients (12).

1.5 microRNAs (miRNAs)

The first miRNA, lin-4, was identified in 1993 by Ambros and colleagues in *Caenorhabditis elegans* (*C. elegans*) (18), and was described as a "small, non-protein-coding RNA, regulating the expression of a protein-coding gene" (18,19). The second, let-7, also discovered in *C. elegans*, was identified seven years later by Ruvkun and colleagues (20). To date, there are 2,654 mature miRNAs annotated in the human genome (21).

miRNAs are small (about 18-22 nucleotides in length), highly conserved, endogenous, non-coding RNAs that act post-transcriptionally to regulate gene expression. Their mechanism of action consists basically to prevent the target messenger RNA (mRNA) translation and stability, blocking protein synthesis. Thereby, miRNAs act in several cellular processes, controlling genes related to cell cycle regulation, proliferation and apoptosis, differentiation, migration, inflammation and immunity, stress response, and metabolism, among other processes. miRNAs are deregulated in several diseases, including cancer, playing an important role in tumorigenesis, including disease development and progression. Moreover, miRNAs have tissue-specific expression patterns (19,22,23).

miRNA biogenesis (canonical pathway) starts in the cell nucleus with the transcription of the miRNA gene, by an RNA polymerase II (Pol II), in a long hairpin structure, named primary miRNA (pri-miRNA). The pri-miRNA is then cleaved by a microprocessor protein complex in a precursor miRNA molecule (pre-miRNA), which contains approximately 70 nucleotides in length, and shapes a stem-loop structure, being capped at 5' and polyadenylated at 3'. This microprocessor complex comprises the DGCR8 (DiGeorge syndrome critical

region gene 8), a double-stranded RNA protein that binds to the RNase III endonuclease Drosha (19,22–24).

The pre-miRNA is then exported from the nucleus to the cytoplasm through Ran/GTP/exportin-5 (EXP-5). Following, the next cleavage by Dicer (a type of RNase III protein) generates a duplex (miRNA-3p/miRNA-5p) with approximately 18-22 nucleotides in length in each strand, and both may be functional. The mature miRNA strand is finally bound to the Argonaute (Ago) protein to form the RNA-induced silencing complex (RISC), which will bind to the 3' untranslated region (UTR) of a target mRNA, promoting its decaying and suppressing its translation (19,22,23).

The majority of miRNA genes are mapped in the human genome in intronic clusters in regions with protein-coding pre-mRNAs, some others may be transcribed as autonomous gene units, besides other miRNAs that can be encoded in long non-coding RNAs (lncRNA) (22).

The binding between miRNAs and target mRNAs is through base complementarity. Base pairing between miRNA and mRNA occurs in the seed region, which usually comprises the nucleotides 2 to 7. It has been demonstrated that only the miRNA seed sequence inside the AGO protein is accessible to bind to the target mRNA, being the seed region important to stabilize the pairing (25). Commonly, partial complementarity between miRNA and target mRNA leads to mRNA destabilization and translational repression (19,25).

1.6 miRNAs in lung cancer

All steps of miRNA biogenesis can be altered in cancer. Alterations include genomic changes such as amplifications, deletions, or translocations, point mutations, and changes in the expression of proteins and enzymes involved in miRNA biogenesis. Such alterations can lead to deregulated miRNA expression, and as a consequence miRNA-target gene expression in cancer (19,23,26).

miRNAs can have roles as oncogenes (oncomiRs) or as tumor suppressors (oncosuppressor miR), having as target genes, tumor suppressors or oncogenes,

respectively. An example of an oncomiR is miR-21, which is upregulated in the vast majority of tumors, with roles in cell proliferation and migration, and inhibition of apoptosis, helping in tumor maintenance, growth, and survival. On the other hand, a classical example of an oncosuppressor miRNA is let-7, which acts by targeting *RAS* or *MYC* genes, with roles in cancer development (19,23).

Several studies have demonstrated that miRNAs are deregulated in lung cancer cells, tissues, and body fluids. Examples of deregulated miRNAs in lung cancer include miR-205-3p, miR-205-5p, and miR-21-3p. The first two were found upregulated in NSCLC tissues and serum, with expression levels significantly higher in LUSC than LUAD; while all three miRNAs were upregulated in patient serum. Also, a significant correlation was found between relative expression levels of miR-21-3p and miR-205-3p in tissues and serum. In this same study, a potential use of miR-205-5p as a biomarker for NSCLC was suggested (27). Another study analyzed the role of miR-205-3p in cell lines, demonstrating its importance to promote cellular viability and apoptosis inhibition, by targeting the amyloid β precursor protein-binding family B member 2 (APBB2) (28).

Among miRNAs correlated with advanced disease stage in NSCLC, miR-10a-5p and miR-196a-5p were upregulated in tissues and serum, and both showed a positive correlation with lymph node metastasis (29).

Upregulation of miR-4652-5p was negatively correlated with expression of *RND1* (Rho family GTPase 1) gene in LUSC. This study suggested that RND1 under-expression, due to high miR-4652-5p levels, contributes to cell proliferation, migration, and invasion, also affecting cell adhesion, and promoting disease progression (30).

It is known that cancer development and progression occur due to a series of complex events, known as the hallmarks of cancer: (I) self-sufficiency in growth signals, (II) insensitivity to anti-growth signals, (III) tissue invasion, and metastasis, (IV) limitless replicative potential, (V) sustained angiogenesis, and (VI) evading apoptosis (31). In 2011, Hanahan & Weinberg reported additional properties of tumor cells, including more complexity to the cancer hallmarks. These were four

additional characteristics: (VII) avoiding immune destruction, (VII) tumor-promoting inflammation, (IX) genome instability and mutation, and (X) deregulating cellular energetics (32). In 2022, Hanahan reported other four important hallmarks of cancer: (XI) non-mutational epigenetic reprogramming, (XII) polymorphic microbiomes, (XIII) senescent cells, and (XIV) unlocking phenotypic plasticity (33). Of note, miRNAs were associated with each of the cancer hallmarks, highlighting the importance of these molecules in tumor development, tumor maintenance, and progression (34).

1.7 novel microRNAs

Several studies (quoted below) have indicated the existence of novel miRNAs in the human genome. In 2019, Alles *et al*. published a study suggesting a real estimate of the number of human miRNAs. They analyzed 28,866 small RNA sequencing data sets and, after northern blotting validation, they extrapolated 2,300 true human mature miRNAs, of which 1,115 were annotated on miRBase (35). Another large study using 13 different human cell lines analyzed 1,323 short RNA sequencing samples (RNA-seq) and identified 3,707 novel miRNA mature sequences. They also demonstrated that these novel sequences showed a specific tissue expression profile (36).

Furthermore, the presence of novel miRNAs has been demonstrated in several cancer types. In malignant B cells, 286 novel miRNAs were identified in a study performed with 31 samples of tumor and normal B cells (37). Novel miRNAs were also identified in breast, bladder, colon, and lung tumors (n=23 samples from each tumor type), and 12 novel miRNAs were validated (38). In addition, 146 novel miRNAs were reported in head and neck squamous cell carcinoma (HNSCC) (n=43 paired samples of tumor and normal) (39). In papillary thyroid carcinoma (PTC) (n=504 tumor samples; n=59 normal tissues), 234 novel miRNAs were identified (40). In gastric adenocarcinoma, 143 novel miRNA sequences were significantly deregulated (n=434 tumor samples; n=41 normal samples) (41). Finally, in breast cancer cell lines, 189 novel miRNA candidates were described (42). In normal liver

tissues, 103 novel miRNAs were identified, and suggested to have a potential role in hepatocellular carcinoma (n=47) (43).

The existence of novel miRNAs has also been explored in other diseases. In brain tissues of patients with Huntington's disease (n=28) and Parkinson's disease (n=29), 99 novel miRNAs were identified, and suggested to play a role in disease pathogenesis (44).

Some evidence showed that these novel miRNAs have a low expression pattern in tissues, which can explain why they were not identified up to now. Experiments with MCF-7 breast cancer cells lines demonstrated that the expression of these transcripts was reduced when Dicer was knocked-down, supporting that these sequences are participating of the canonical miRNA's biogenesis. Moreover, it was also possible to verify that these novel sequences have the same evolutive origin of the known miRNAs, however, the novel sequences presented an evolutive origin more recent than the known ones (45).

Mostly, known miRNAs are commonly located in intronic or intergenic regions, on the other hand, the novel ones were more present in exons and repeats regions. The target genes prediction also uncovered that the novel miRNAs are capable to bind their target mRNAs with higher affinity and with more strength than the known miRNAs (45).

All of the above studies support the hypothesis of the existence of novel miRNAs, but also highlight the need for validation and additional investigation about the function of novel miRNAs in disease, considering the different cellular contexts and different tissues. It is important to highlight that most studies did not include a large number of samples, neither showed validation of their results in clinical samples.

1.8 Cancer metabolism and microRNAs

Pavlova and Thompson proposed six cancer metabolic hallmarks: (I) deregulated uptake of glucose and amino acids, (II) use of opportunistic modes of nutrient acquisition, (III) use of glycolysis/TCA cycle intermediates for

biosynthesis and NADPH production, (IV) increased demand for nitrogen, (V) alterations in metabolite-driven gene regulation, and (VI) metabolic interactions with the microenvironment. The cancer cells can show some or all of these hallmarks (46).

Deregulating cellular energetics is a cancer hallmark, once the cancer cells uncontrollably proliferate with a need for energetic compensation to support the high rate of cell division. This energetic disbalance occurs mainly because the cancer cells, even in the presence of oxygen, metabolize glucose in the glycolysis pathway, without using mitochondria, even if this means a lower ATP production (32,46). This reprogramming was first observed by Otto Warburg and since then it has been known as the "Warburg effect" (47) or "aerobic glycolysis" (46).

Aerobic glycolysis produces a small amount of ATP (not good to support the high proliferation rate and energetic requirement by cancer cells); however, it generates other metabolic intermediates that will support the energetic demand, by alternating between aerobic glycolysis and oxidative phosphorylation (48).

It has been demonstrated that cancer cells upregulate the glucose transporters, mainly GLUT1, increasing the influx of glucose to the cytoplasm. The glucose supply is correlated with oncogenes, such as *RAS* and *MYC*, and mutated tumor suppressor genes. Such alterations are beneficial to cancer cells once they give them proliferation competence and ability to evade apoptosis. Also, within tumors there are hypoxic conditions, which contribute to increasing the dependence on glycolysis. The real reasons behind this glycolytic switch in cancer cells are poorly understood; however, some hypotheses were raised: first, with the increased glycolysis, there is the deviation of several glycolytic intermediates that act in many biosynthetic pathways, and this can contribute to synthesizing novel macromolecules and organelles for nascent cells. Second, in some normal conditions when it is necessary a fast cell division such in embryonic tissues, it is possible to observe the Warburg-like metabolism (32).

Glutamine metabolism has also been highlighted as a metabolic pathway associated with oncogenesis. Glutamine is important once it provides nitrogen for

biosynthesis of non-essential amino acids, nucleotides, and glucosamine-6-phosphate. Thus, cancer cells present a high consumption of glutamine, again, to support the high proliferation rates (46). Overall, metabolic alterations were reported in LUAD and LUSC, with differences according to the histological subtype. LUSC showed differential glutamine and glucose catabolism when compared to LUAD and normal tissues (49).

Additionally, lung cancers showed upregulation of glycine/serine/threonine and pyrimidine pathways and, consequently, a low level of threonine and histidine both utilized in those pathways. Lipid metabolism is also altered in patients with lung cancer. Choline is a precursor to membrane phospholipids largely used by proliferation cells, and it showed low levels in the serum of patients with lung carcinoma (50).(51)

Several miRNAs have been previously associated with metabolism reprogramming in lung cancer. Examples of altered miRNAs include miR-144 and miR-124, which play roles in glucose metabolism by targeting GLUT1 (51), HK2, p-Akt1/2 (52), and also enhancing the Warburg effect and ATP generation, respectively. Furthermore, miR-33b was associated with lipid metabolism, targeting LDHA, and enhancing aerobic glycolysis and tumor invasion (53). These and other miRNAs associated with lung cancer metabolism were reviewed (48) and are summarized in **Table 1.**

Table 1. microRNAs deregulated in lung cancer and associated with metabolic pathways*.

Down-regulated microRNAs				
microRNA	**Target Gene(s)**	**Metabolic Pathway**	**Function(s)**	**Reference**
miR-29a	*ENO1*, *PGAM1*, *PGK1*, *CMPK1*	Glucose metabolism	Facilitates glycolysis	(54)
miR-33b	*LDHA*	Lipid metabolism	Enhances aerobic glycolysis and tumor invasion	(53)
miR-124	*GLUT1*, *HK2*, *p-Akt1/2*	Glucose metabolism	Enhances Warburg effect and ATP generation	(52)
miR-126	*SLC7A5*	Amino acid metabolism	Facilitates amino acid exchange and activate mTOR	(55)
miR-133b	*PKM2*	Glucose metabolism	Enhances Warburg effect and radiation resistance	(56)
miR-143	*HK2*	Glucose metabolism	Enhances Warburg effect	(57)
miR-144-3p	*GLUT1*	Glucose metabolism	Enhances Warburg effect	(51)
miR-206	*HK2*	Glucose metabolism	Enhances Warburg effect	(58)
miR-451	*c-Myc*, *GSK3*	c-Myc	Enhanced metabolic dysregulation and migration	(59)
Up-regulated microRNAs				
microRNA	**Target Gene**	**Metabolic Pathway**	**Function(s)**	**Reference**
miR-21	*HBP1*, *CD36*, *SOCS1*, *SOCS6*, *PTEN*	Lipid metabolism and PI3K/Akt/PTEN	Promotes proliferation and increase intracellular lipid influx; PTEN inhibition	(60,61)
miR-31-5p	*FIH*	Hypoxia	Promotes hypoxia and enhances aerobic glycolysis	(62)
miR-34	*SIRT1*	Hypoxia	Facilitate mutant p53 expression via a feedback loop	(63,64)
miR-125b	*p-Akt*, *GSK3*	PI3K/Akt/PTEN	PI3K/Akt and GSK3/Wnt/-catenin activation	(65)
miR-155	*SOCS1*, *SOCS6*, *PTEN*	PI3K/Akt/PTEN	PTEN inhibition	(60)
miR-182	*PDK4*/*PDH* axis	Lipid metabolism	Enhances de novo lipogenesis and JNK activation	(66)
miR-199-5p		Hypoxia	Activates HIF-1	(67)
miR-210	*NDUFA4*, *SDHD*	Hypoxia	Mitochondria disfiguration and cristae formation	(68)

*Reviewed in (48).

2. STUDY RATIONALE

To the best of our knowledge, there are no published studies on the discovery and validation of novel miRNAs in lung cancer, particularly when comparing the main histological disease subtypes. Given our group's extensive expertise in lung cancer research, this knowledge gap justifies the choice of our study model.

In addition, we highlight that the discovery of novel miRNAs is a promising field that will likely contribute to identifying novel regulatory molecules implicated in tumorigenesis mechanisms. Such data may aid in biomarker development for diagnosis and treatment, ultimately benefiting patients.

3. HYPOTHESIS

Novel miRNAs have different expression profiles in lung adenocarcinoma and squamous cell carcinoma, and modulate malignant metabolism pathways.

4. OBJECTIVES

(1) To identify novel microRNAs (*in silico*) in a large lung cancer miRNA-Seq dataset, determining their expression levels in tumors compared to normal lung tissues;

(2) To predict and validate (*in silico*), novel miRNAs and target genes with biological roles in metabolism.

5. MATERIAL AND METHODS

Ethical considerations

The bioinformatics part of our project is included under the Faculty of Medicine (FMB) institutional Declaration # 351-2020, which exempts research projects exclusively based on publicly available data analysis, from ethics approval. The experimental procedures (validation) to be performed in our study have been subjected to our local Ethics Committee (FMB) under the protocol # CAAE: 65306922.3.0000.5411.

Source of data

First, microRNA sequencing (miR-Seq) data were obtained from The Cancer Genome Atlas (TCGA) (https://cancergenome.nih.gov/), an online database that provides original datasets for different cancer types. In order to obtain the raw, original miR-Seq data, an authorization was requested and issued on Feb 17th 2022, by NCBI dbGaP (the Database of Genotypes and Phenotypes); Project #24725: Identification of novel non-coding RNAs in colorectal and thoracic cancers. Since our study is focused on thoracic cancers, we selected 1,050 cases from the two most common subtypes of thoracic cancers with available TCGA data: LUAD and LUSC. The number and characteristics of samples included for each tumor subtype is outlined below.

Lung Adenocarcinoma: 528 samples

- 436 primary tumors
- 90 paired samples (45 normal plus 45 primary tumors)
- 2 recurrent tumors

Lung Squamous Cell Carcinoma: 522 samples

- 432 tumors
- 90 paired samples (45 normal plus 45 primary tumors)

Our study results are focused on LUAD and LUSC, which are the two major subtypes of NSCLC.

In order to obtain the TCGA datasets for these samples, we applied specific filtering criteria in the following order:

1. In the TCGA website, under the menu "repository", the .bam file format was selected;
2. miRNA-seq data was selected;
3. In the menu "exploration", bronchus and lung option was selected;
4. The options: "adenomas and adenocarcinomas", and "squamous cell neoplasms" were selected.
5. The samples were then added to the chart and were available for download.

The ".bam" files of miR-Seq were downloaded from TCGA, generated as single-end runs of 1,050 samples, were converted into .fastq files using SAMtools.

Next, the ".fastq" files were uploaded on miRMaster (https://ccb-compute.cs.uni-saarland.de/mirmaster/) (69), a tool useful to predict and quantify novel and known microRNAs. The analysis was run using miRMaster default settings. We excluded one sample due to file error and two samples due to low quality of reads.

TCGA data was generated using the Illumina HiSeq 2000 miRNA Sequencing. Data was accessed and downloaded on August 18th, 2020. miRMaster analysis was performed on November 2nd, 2020. The sample IDs used in this study are described in **Table 2** (attached file).

Our dataset had two patients with three biological replicates each. In order to handle these biological replicates from the same patients, we averaged the expression values obtained from miRMaster analysis, as previously reported (39,70). Therefore, each patient had only one expression value for downstream analyses, and 1,044 samples remained in our study.

Finally, miRNAs that had ≥1 read per million (RPM) in at least 10% of samples (104 of 1,044), in a given sample group (LUAD or LUSC), were considered

to be expressed in that group and were thus included for analyses. This data analysis strategy has been used in previous studies (40,69). **Figure 2** summarizes these steps.

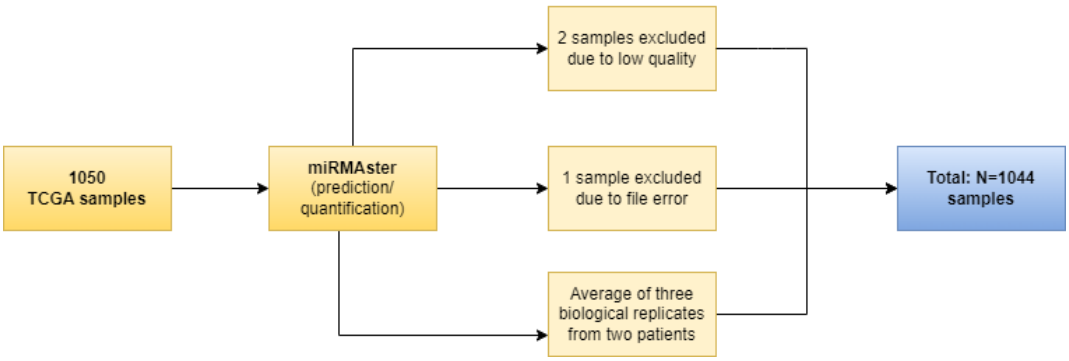

Figure 2. Flowchart showing the steps of sample selection for analyses.

Samples were divided into Discovery and Validation sets using a random selection tool in Excel. Sample groups considered the histological subtypes of lung adenocarcinoma (LUAD) or squamous cell carcinoma (LUSC), and whether samples were paired (tumor and corresponding normal, from the same patient) or unpaired (tumor only). Predicted novel miRNAs with similar sequences with known miRNAs were excluded from downstream analyses (**Figure 3A**).

Novel miRNA prediction was performed using paired samples from LUAD and LUSC in the discovery group and validated in paired samples in the validation group. This first analysis step was performed using paired samples, in order to reduce bias due to the low number of normal samples in comparison with the number of tumor samples in the TCGA dataset. Next, we identified which novel miRNAs are differentially expressed in tumors compared with paired normal.

We clarify that all samples will be used to correlate novel miRNAs with clinical data including survival (this analysis will be performed at the end of the study).

6. RESULTS

6.1 Identification of differentially expressed miRNAs in Tumor *versus* Normal

To identify differentially expressed miRNAs, the average fold change (FC) and corresponding Benjamini-Hochberg (71) adjusted *p-value* were calculated between tumor *vs.* normal, within each histological subtype (LUAD or LUSC). If a miRNA had a FC ≥ 2 and adjusted *p-value* ≤ 0.05, as determined by paired t-test, this predicted novel miRNA was considered differentially expressed (**Figure 3B**). Following these steps, we further applied a *p-value* ≤ 0.001, in order to obtain the most significant predicted novel miRNAs; this analysis resulted in 292 knowns (**Table 3** – attached file) and 126 novel miRNAs, validated as differentially expressed in LUAD, and 330 known (**Table 4** – attached file) and 225 novel miRNAs validated as differentially expressed in LUSC. These results are shown in **Figure 3C**.

Novel miRNAs and corresponding target genes are shown in **Table 5** (attached file) (LUAD) and **Table 6** (attached file) (LUSC).

The identification of known miRNAs previously described in lung cancer shows the robustness of our analysis strategy to identify novel miRNAs. For example, we identified previously known miRNAs, such as the oncomiR miR-21, which is overexpressed in several cancer types including lung cancer (72), and the miR-30 family, a tumor suppressor miRNA family under expressed in lung cancer and associated with epithelial mesenchymal transition (73).

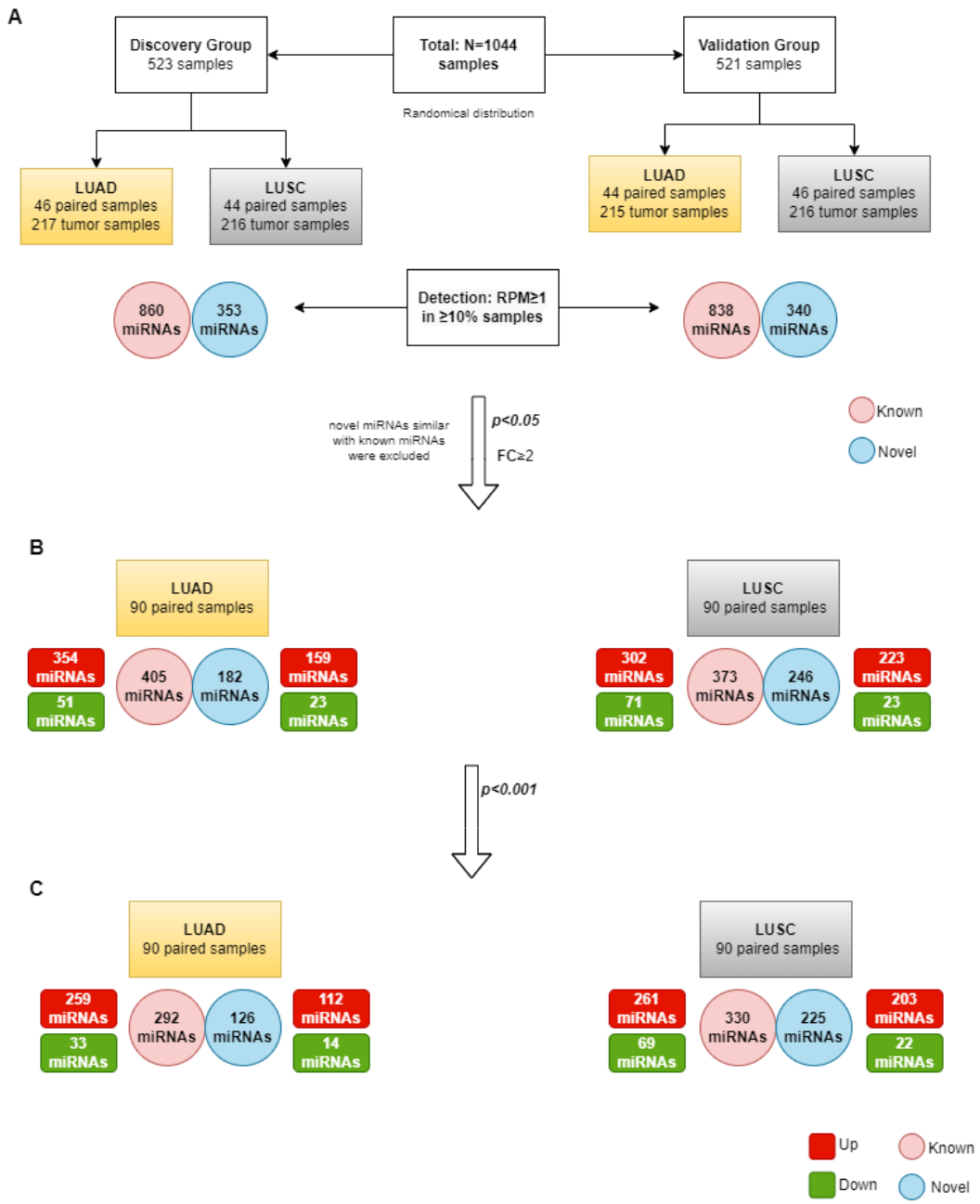

Figure 3. Analysis steps and results of known and novel miRNAs identified and validated (*in silico*) as differentially expressed in lung adenocarcinoma (LUAD) and squamous cell carcinoma (LUSC) *vs*. normal lung samples. All samples will be used in the future analysis of the study to identify potential novel miRNAs with prognostic values.

6.2 GC content and length of sequence

The quantity of guanine and cytosine (GC content) was measured to compare the characteristics of the putative novel miRNA sequences with known miRNAs; thus, we determined their GC content and the length of the mature 5p and 3p sequences. We observed that both GC content and length of sequences were very similar between known and novel miRNAs (**Figure 4**).

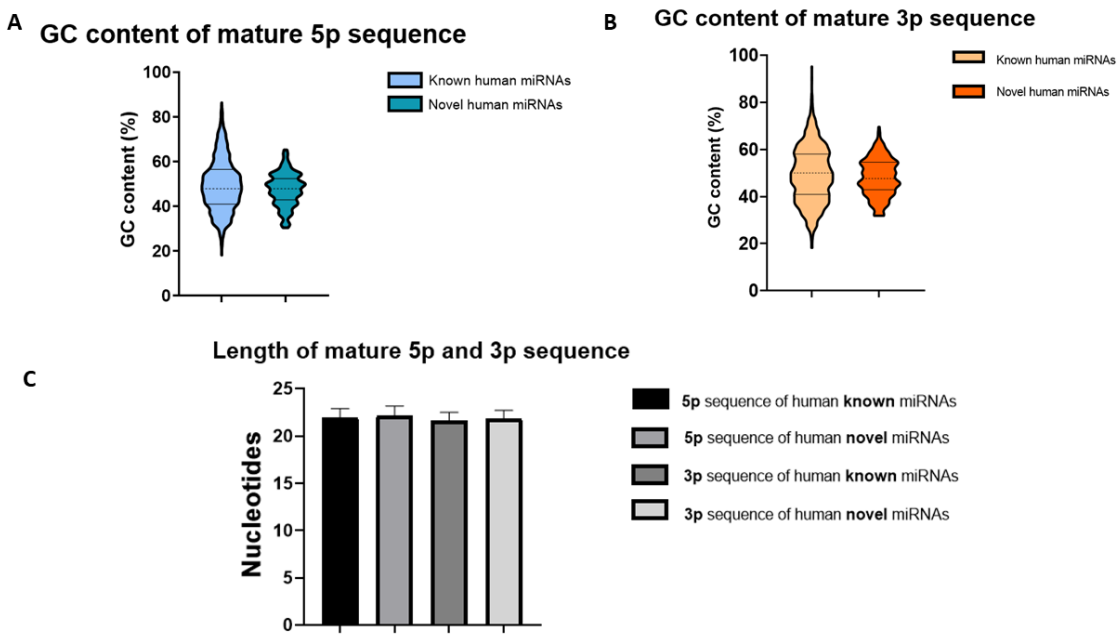

Figure 4. (A) GC content of mature 5p sequence of known and candidates to a novel miRNA. **(B)** GC content of mature 3p sequence of known and candidates to a novel miRNA. **(C)** Length of mature 5p and 3p sequence of known and candidates to a novel miRNA. These graphics was constructed on GraphPad Prism 8 software.

6.3 miRNAs are able to differentiate Tumor from Normal samples

In order to verify whether known and novel miRNAs were able to differentiate tumor from normal samples, in both LUAD and LUSC subtypes, we performed Principal Component Analysis (PCA) using the ClustVis tool (https://biit.cs.ut.ee/clustvis/) (74) (**Figures 5** and **6**).

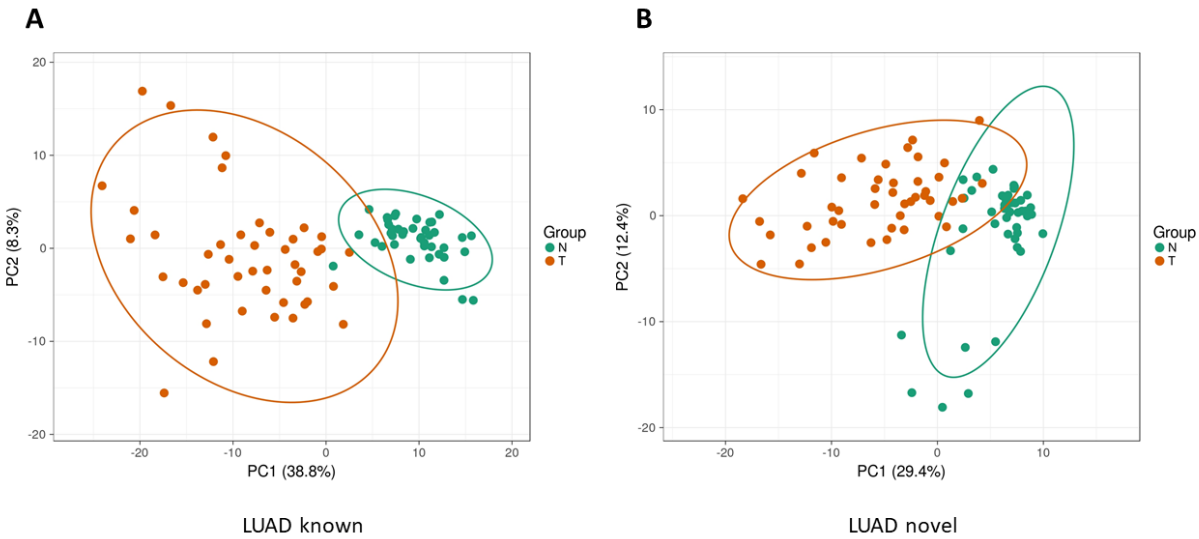

Figure 5: Principal Component Analysis (PCA) of LUAD samples. **A:** known miRNAs. **B:** novel microRNAs.

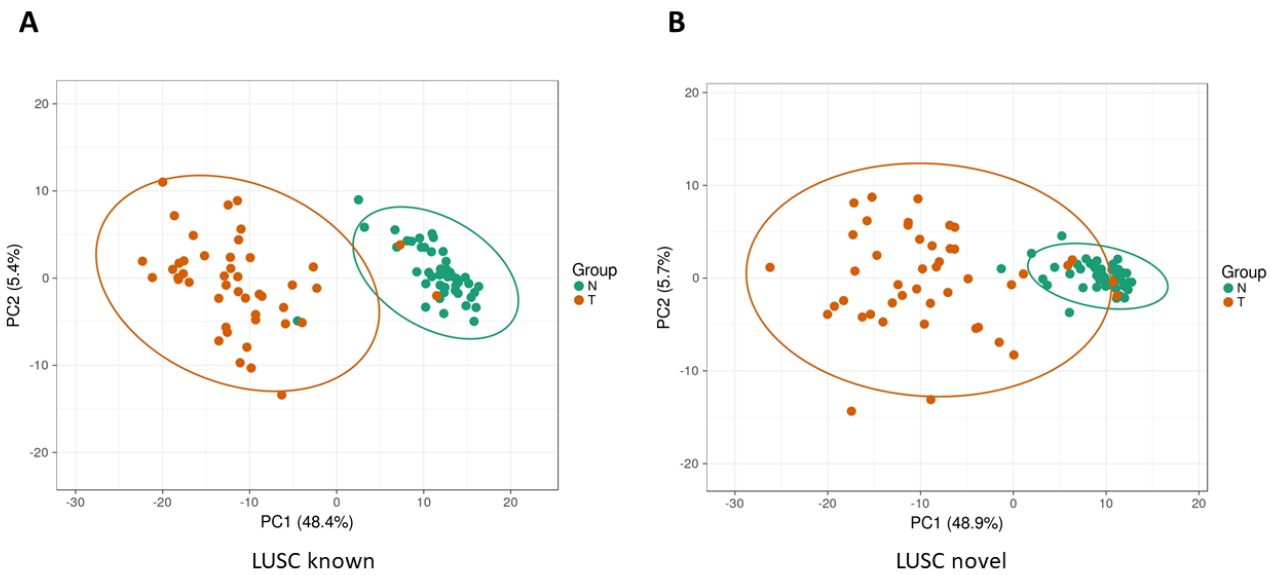

Figure 6: Principal Component Analysis (PCA) of LUSC samples. **A:** known miRNAs. **B:** novel microRNAs.

6.4 Computational identification of genes potentially targeted by novel miRNAs

Next, we sought to identify genes targeted by the predicted novel miRNAs. In this analysis, we searched novel miRNA sequences that were differentially expressed (FC ≥ 2 and *p-value* ≤ 0.001) in LUAD (126 novel miRNAs) and LUSC (225 novel miRNAs) using miRDB (http://mirdb.org/custom.html) (75). According to miRDB, a predicted target gene is likely real if the Target Score is ≥ 80 (**Tables 5* and 6***). Therefore, we applied a target score ≥ 80 to filter genes predicted to be targeted by novel miRNAs.

Table 5. Top 20 predicted target genes of novel miRNAs in LUAD.

miRNA	Target Score	Gene Symbol	Gene Description
pred-nov-miR-10651-5p	100	ZNF99	zinc finger protein 99
pred-nov-miR-10651-5p	100	ZNF493	zinc finger protein 493
pred-nov-miR-10975-5p	100	ZNF99	zinc finger protein 99
pred-nov-miR-10975-5p	100	ZNF493	zinc finger protein 493
pred-nov-miR-12186-5p	100	FGFRL1	fibroblast growth factor receptor like 1
pred-nov-miR-13707-5p	100	POU3F3	POU class 3 homeobox 3
pred-nov-miR-14480-3p	100	ONECUT2	one cut homeobox 2
pred-nov-miR-14480-3p	100	PRTG	protogenin
pred-nov-miR-14480-3p	100	POU2F1	POU class 2 homeobox 1
pred-nov-miR-14480-3p	100	ARFGEF2	ADP ribosylation factor guanine nucleotide exchange factor 2
pred-nov-miR-14480-3p	100	SOCS5	suppressor of cytokine signaling 5
pred-nov-miR-14526-5p	100	SOX7	SRY-box 7
pred-nov-miR-150-3p	100	PPP2R2D	protein phosphatase 2 regulatory subunit Bdelta
pred-nov-miR-15401-3p	100	CREB3L3	cAMP responsive element binding protein 3 like 3
pred-nov-miR-15434-5p	100	GNB1	G protein subunit beta 1

pred-nov-miR-16144-5p	100	RIMS2	regulating synaptic membrane exocytosis 2
pred-nov-miR-16806-5p	100	ATL3	atlastin GTPase 3
pred-nov-miR-18647-5p	100	GK5	glycerol kinase 5
pred-nov-miR-18647-5p	100	LIN54	lin-54 DREAM MuvB core complex component
pred-nov-miR-20988-5p	100	YOD1	YOD1 deubiquitinase

*The complete table is shown in Table 4 (attached file).

Table 6. Top 20 predicted target genes of novel miRNAs in LUSC.

miRNA	Target Score	Gene Symbol	Gene Description
pred-nov-miR-10580-3p	100	CBR4	carbonyl reductase 4
pred-nov-miR-12186-5p	100	FGFRL1	fibroblast growth factor receptor like 1
pred-nov-miR-13694-5p	100	ITGA9	integrin subunit alpha 9
pred-nov-miR-1419-5p	100	GSG1L	GSG1 like
pred-nov-miR-150-3p	100	PPP2R2D	protein phosphatase 2 regulatory subunit Bdelta
pred-nov-miR-15401-3p	100	CREB3L3	cAMP responsive element binding protein 3 like 3
pred-nov-miR-15434-5p	100	GNB1	G protein subunit beta 1
pred-nov-miR-16144-5p	100	RIMS2	regulating synaptic membrane exocytosis 2
pred-nov-miR-16511-3p	100	KMT5B	lysine methyltransferase 5B
pred-nov-miR-16511-3p	100	SUZ12	SUZ12, polycomb repressive complex 2 subunit
pred-nov-miR-16806-5p	100	ATL3	atlastin GTPase 3
pred-nov-miR-1688-3p	100	CELF4	CUGBP Elav-like family member 4
pred-nov-miR-1688-3p	100	SERPINI1	serpin family I member 1
pred-nov-miR-18647-5p	100	GK5	glycerol kinase 5
pred-nov-miR-18647-5p	100	LIN54	lin-54 DREAM MuvB core complex component
pred-nov-miR-18768-3p	100	IRGQ	immunity related GTPase Q
pred-nov-miR-19112-3p	100	LILRA1	leukocyte immunoglobulin like receptor A1
pred-nov-miR-19112-3p	100	FSHB	follicle stimulating hormone subunit beta
pred-nov-miR-19112-3p	100	SLC18A2	solute carrier family 18 member A2
pred-nov-miR-19704-5p	100	CREB3L3	cAMP responsive element binding protein 3 like 3

*The complete table is shown in Table 5 (attached file).

6.5 Computational verification of genes targeted by novel miRNAs in LUAD and LUSC

Next, we sought to verify whether target genes of novel miRNAs were deregulated in LUAD and LUSC. For this, RNA-Seq data from the *GDC TCGA Lung Adenocarcinoma* (normal n=273; LUAD n=596) (http://analysis.xenahubs.net/8023e708233ef5078fdbaad499504d5e3bea02dc/) and *GDC TCGA Lung Squamous Cell Carcinoma* (normal n=248; LUSC n=498) (http://analysis.xenahubs.net/2a127a84fcccc54637374173d2369e471eb5cd2a/), were retrieved and analyzed in the University of California Santa Cruz (UCSC) *Xena Browser tool* (https://xenabrowser.net/) (76). The links above have the detailed data analysis results.

The analysis steps used the following parameters:

- Phenotypic data type
- *primary_diagnosis.diagnoses*
- *primary_site*
- *prior_treatment.diagnoses*
- *sample_type.samples*
- Patients with prior treatment and FFPE samples were excluded
- Differential expression of primary tumor (and) recurrent tumor *vs.* solid tissue normal
- Low expression threshold: 2
- No normalization type was selected (the GDC data are already normalized)
- Differential expression analysis method: *limma_voom*
- logFC threshold: 2

In LUAD, 945 target genes were validated. Of these, 909 were under expressed and 36 were overexpressed in tumors compared with normal lung tissues (**Supplementary Table 1**).

In LUSC, 2,444 target genes were validated. Of these, the vast majority, 2,315 genes were under expressed, and 129 were overexpressed in tumors compared with normal lung tissues (**Supplementary Table 2**).

6.6 Correlation Analysis

We also performed Spearman correlation analysis, in order to strengthen the probability of target genes being real targets of predicted novel miRNAs. Considering r > -50% and p<0.0001, we found that 18 novel miRNAs were negatively correlated with 47 target genes in LUAD, and 66 novel miRNAs were negatively correlated with 473 target genes in LUSC.

Among the novel miRNAs negatively correlated in both subtypes, 14 were shared: 13 were upregulated, and only one was down-regulated (**Figure 7**).

The target genes predicted and correlated for the 13 novel sequences in LUAD and in LUSC are presented in **Table 7** and **Table 8**, respectively.

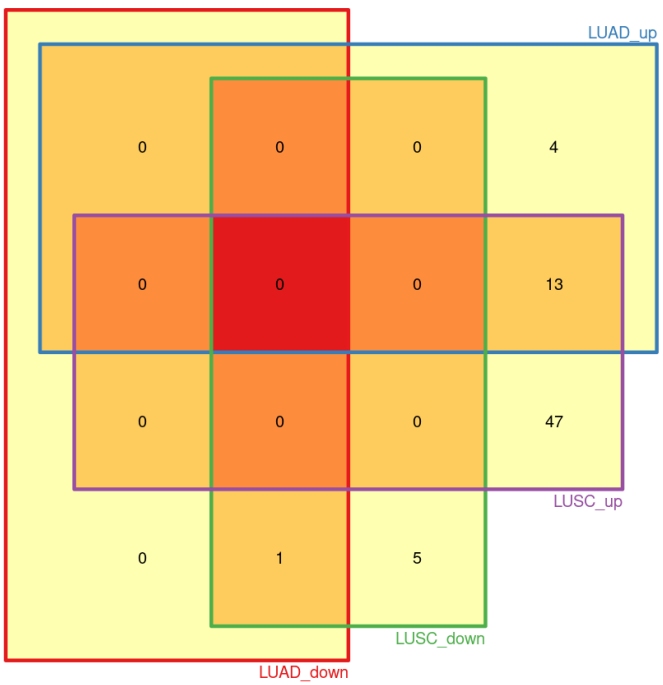

Figure 7: Venn Diagram showing the predicted novel miRNAs that were negatively correlated with target genes, in LUAD and LUSC subtypes.

6.7 Genomic distribution of the novel miRNAs and their predicted target genes

In order to visualize the genomic location of predicted novel miRNAs and their respective target genes, a Circos Plot analysis was performed on ClicO FS tool (77). Genomic localization of known miRNAs is shown in the outer light-green circle, while the inner light-orange circle shows the genomic location of the 13 predicted novel miRNAs represented by the black bars. The color lines in the middle represent the binding between the novel miRNAs to their respective target genes widespread on the genome. Each color represents a different novel miRNA as exhibited by the label (**Figure 8**).

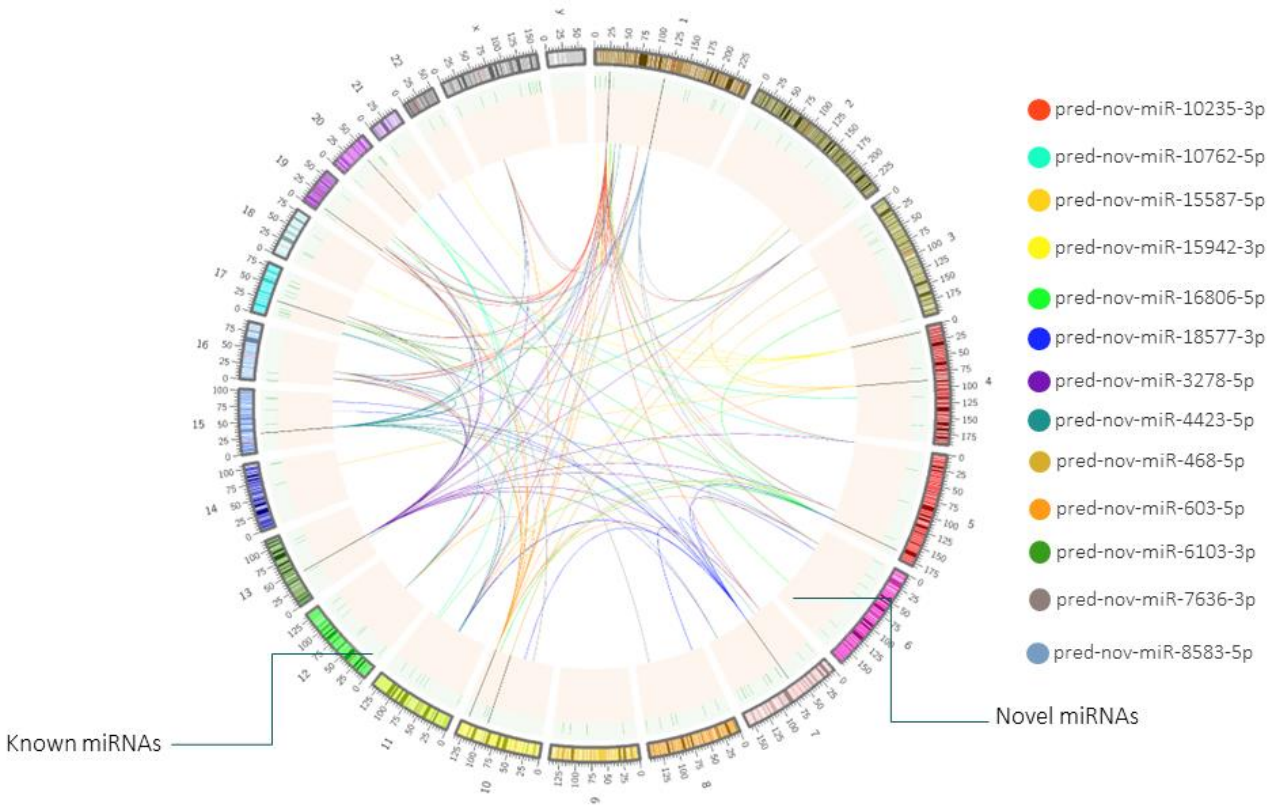

Figure 8: Circos Plot showing the genomic location of known and novel miRNAs and the genomic location of the novel miRNAs target genes. Created on ClicO FS tool (77).

6.8 Statistical Analyses

6.8.1 Tissue specificity analysis

t-Distributed Stochastic Neighbor Embedding (t-SNE) Analysis

The t-Distributed Stochastic Neighbor Embedding (t-SNE) was used to verify if the 13 novel miRNAs have tissue-specific patterns of expression. To perform this analysis, we used the expression values of the 13 novel miRNAs in a dataset of normal tissues obtained from TCGA. miRNA expression in lung normal tissues (n=90) was compared with other tissue types: brain (n=5), breast (n=90), bladder (n=19), colorectal (n=11), and liver (n=59) (**Table 9**). These tissues were chosen mainly because they share embryonic origins with the lung. This analysis was performed using the Orange Data Mining software (78) with its default settings.

Table 9. Presence (+) or absence (-) of the 13 novel miRNAs in other tissues.

microRNA	Lung	Bladder	Liver	Colorectal	Brain	Breast
pred-nov-miR-10235-3p	+	-	-	+	-	-
pred-nov-miR-10762-5p	+	-	-	+	-	-
pred-nov-miR-15587-5p	+	-	-	-	-	-
pred-nov-miR-15942-3p	+	-	-	-	-	-
pred-nov-miR-16806-5p	+	-	-	-	-	-
pred-nov-miR-18577-3p	+	-	-	-	-	-
pred-nov-miR-3278-5p	+	-	-	+	-	-
pred-nov-miR-4423-5p	+	-	-	-	-	-
pred-nov-miR-468-5p	+	-	+	+	-	+
pred-nov-miR-603-5p	+	-	-	+	-	-
pred-nov-miR-6103-3p	+	-	-	+	-	-
pred-nov-miR-7636-3p	+	-	-	-	-	-
pred-nov-miR-8583-5p	+	-	-	-	-	-

As we can see in **Figure 9**, the predicted novel miRNAs have expression patterns specific to lung tissue, represented by the grouped green circles, isolated from the other tissues. T-SNE analysis grouped together the samples with similar expression profiles, supporting the tissue specificity expression in lung.

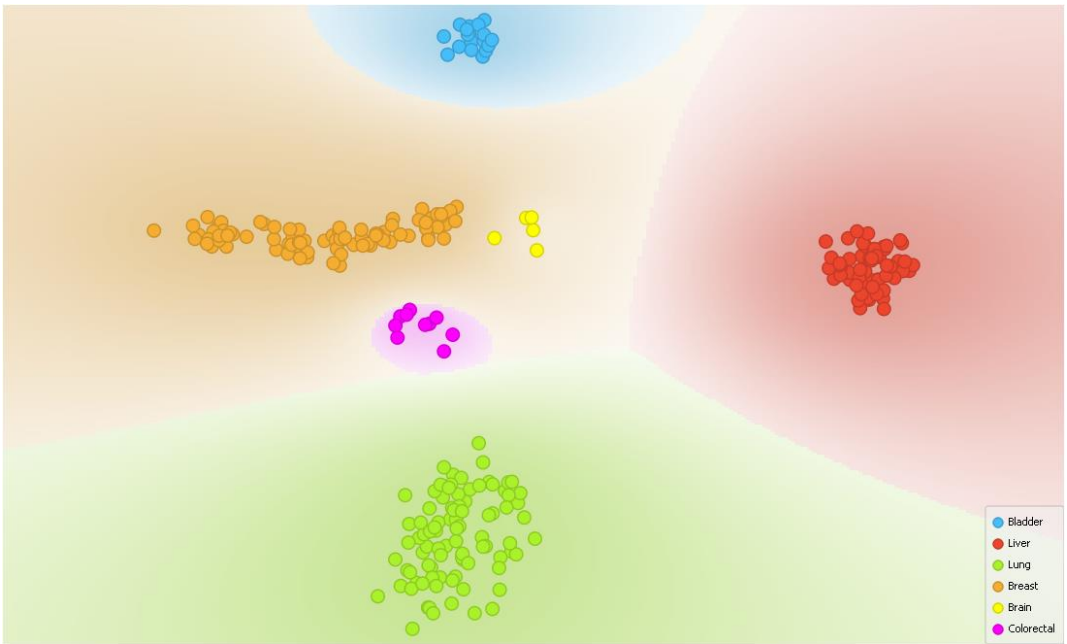

Figure 9: Tissue-specific expression patterns of the 13 predicted novel miRNAs. t-Distributed Stochastic Neighbor Embedding (t-SNE) analysis (78) of normal tissues from TCGA. This analysis was performed using normalized expression values of the 13 novel miRNA sequences in the lung.

6.8.2 Mann-Whitney and Kruskal-Wallis Tests

In addition, we applied Kruskal-Wallis and Mann-Whitney Tests in order to verify whether there was a differential expression between the 13 miRNAs in both histological subtypes and their respective normal tissues and whether this difference was significant. Eight of 13 novel miRNAs showed a higher expression value in LUSC than LUAD, and this difference was significant ($p < 0.05$) in eight novel miRNAs (pred-nov-miR-10235-3p, pred-nov-miR-15587-5p, pred-nov-miR-15942-3p, pred-nov-miR-16806-5p, pred-nov-miR-18577-3p, pred-nov-miR-3878-5p, pred-nov-miR-603-5p, and pred-nov-miR-8583-5p); all of these having higher expression levels in LUSC compared with LUAD (**Figure 10**).

Interestingly, 3 novel miRNAs (pred-nov-miR-10762-5p, pred-nov-miR-15942-3p, and pred-nov-miR-4423-5p) only showed expression in tumor tissues and they were not expressed in normal tissues.

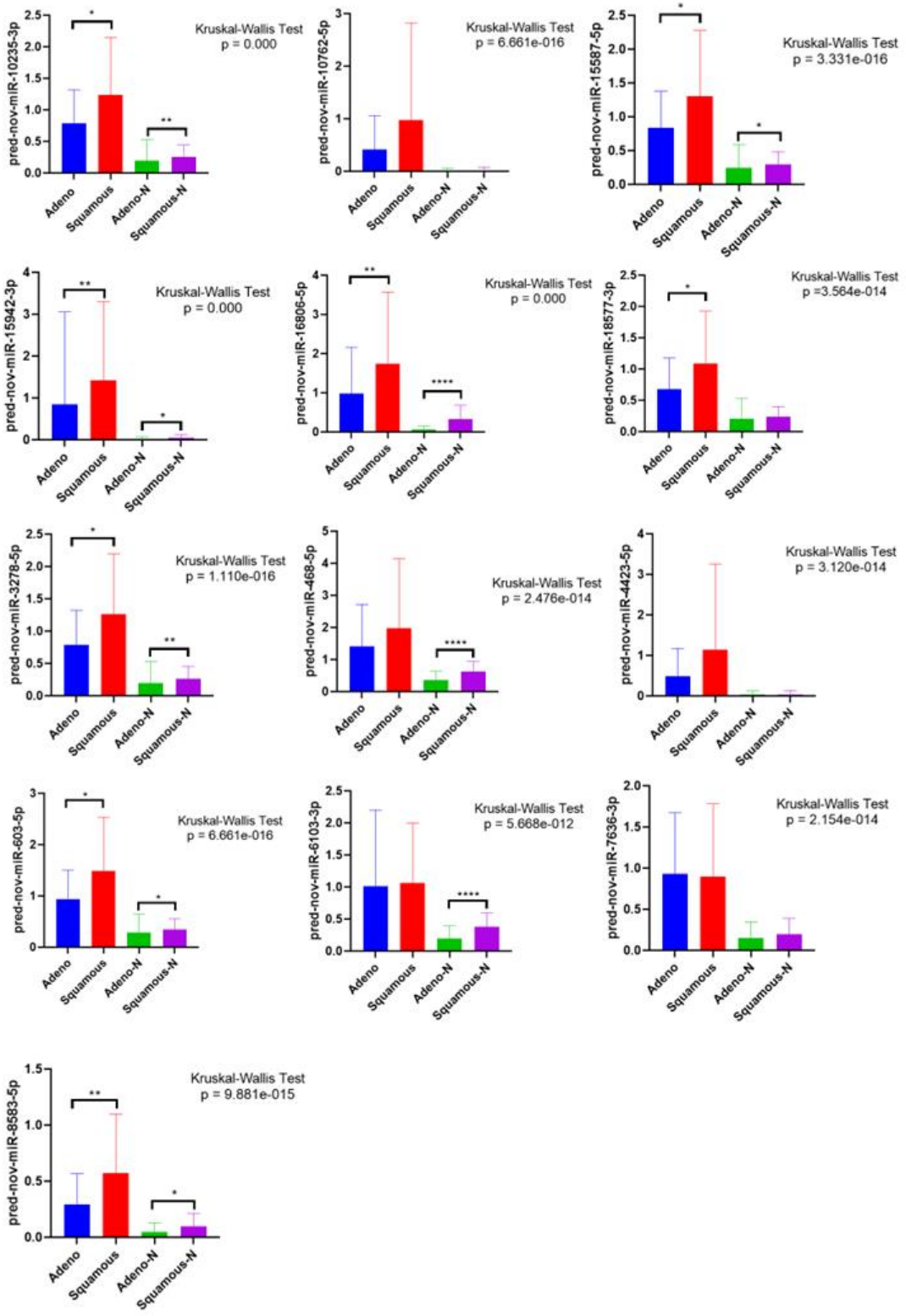

Figure 10: Kruskal-Wallis Test of the 13 predicted novel miRNAs. (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$). These analyses were performed using the GraphPad Prism 8 software.

6.9 miRNA-mRNA networks

Next, in order to investigate the relationship between the novel miRNAs and their target genes, we created an interaction network on Cytoscape (79). **Figure 11** illustrates the networks between the 13 upregulated novel miRNAs and their target genes in LUAD and LUSC. Genes represented in orange are exclusively regulated by miRNAs in LUAD. Genes in light-pink are exclusively regulated by miRNAs in LUSC. Genes represented in green are targeted by novel miRNAs in both LUAD and LUSC (**Figure 11**).

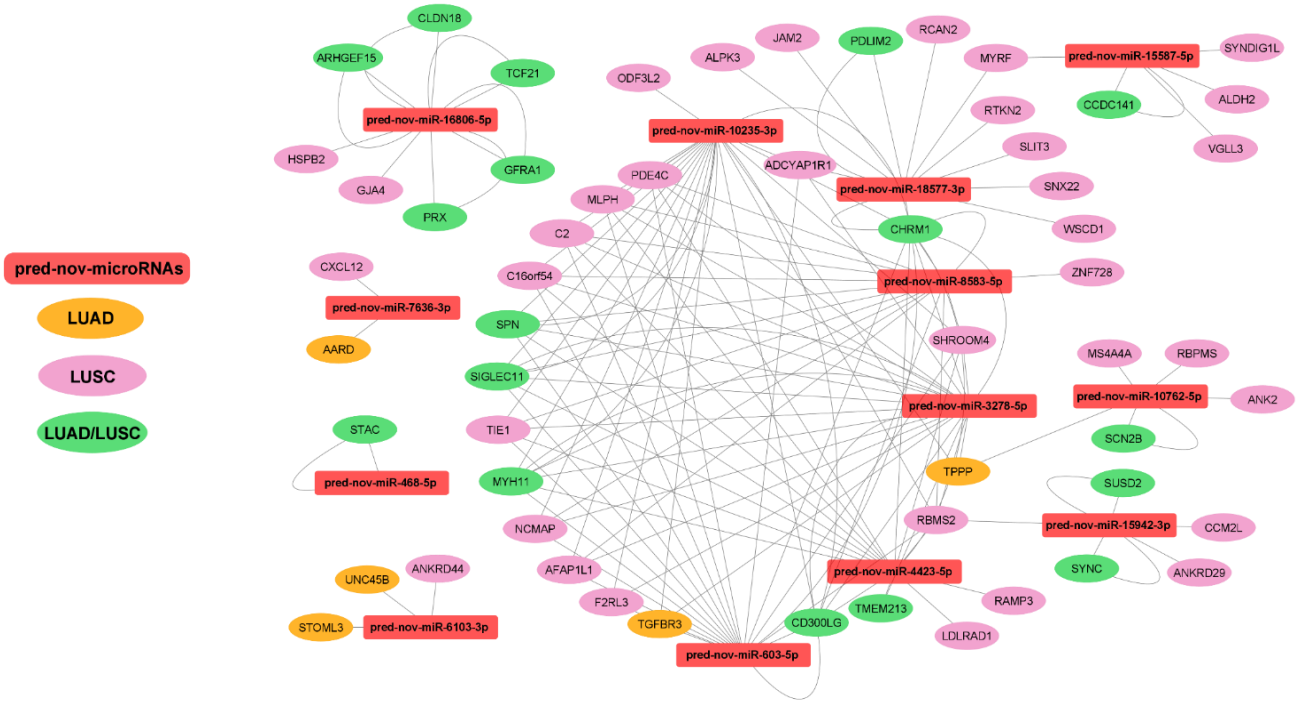

Figure 11: miRNA-mRNA networks. The 13 novel miRNAs (upregulated, shown in red) in LUAD and LUSC. Target genes of novel miRNAs are shown: in orange, exclusive to LUAD; in light-pink, exclusive to LUSC, and in green, common to LUAD and LUSC. Graph generated in Cytoscape (79).

6.10 Metabolic pathways

To identify the involvement of metabolic pathways, we searched target genes (predicted as regulated by novel miRNAs) using pathDIP (80). Several statistically significant ($p < 0.05$) pathways were identified and are detailed at **Supplementary Table 3**. Main pathways are shown in **Figure 12**, as follows: Carbohydrate metabolism, Energy metabolism, Lipidic metabolism, Calcium metabolism, Amino acid metabolism, Metabolism of vitamins and minerals, and Nucleotide metabolism. Carbohydrate and energy metabolism were more represented than other pathways, suggesting a significant energetic imbalance in cancer metabolism. Notably, a large number of target genes of novel miRNA were closely associated with these pathways.

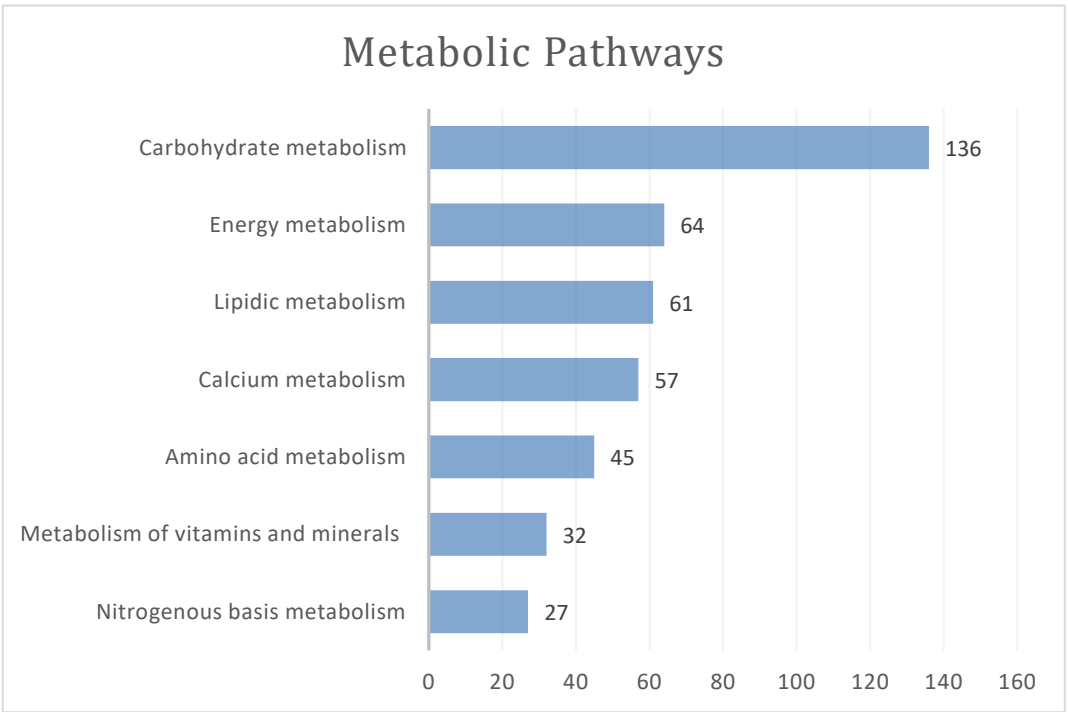

Figure 12: Metabolic pathways which the predicted novel miRNA target genes in lung cancer were related.

In the carbohydrate metabolism group, we identified 30 related genes, and the genes that most appeared were *TIE1* and *HSPB2*, each one found in 15 pathways, followed by *ANK2*, present in 13 pathways, and *GFRA1* in 12 pathways each. As the same way, in the energy metabolism group, we identified 26 genes, and the genes that most appeared were *ALDH2* and *HSPB2*, each one found in 8 pathways, followed by *CHRM1* present in 7 pathways.

The majority of these genes were found exclusively deregulated in LUSC (*TIE1*, *HSPB2*, *ANK2*, and *ALDH2*), and two (*GFRA1* and *CHRM1*) were common between subtypes.

In LUSC, the novel potential miRNAs pred-nov-miR-10235-3p, pred-nov-miR-3278-5p, pred-nov-miR-603-5p, pred-nov-miR-8583-5p, and pred-nov-miR-4423-5p were predicted to target *TIE1*; pred-nov-miR-16806-5p was predicted to target *HSPB2*; pred-nov-miR-10762-5p was predicted to target *ANK2*; and the pred-nov-miR-15587-5p was predicted to target *ALDH2*.

In both subtypes, the gene *GFRA1* was predicted as a target for the pred-nov-miR-16806-5p; and *CHRM1* was predicted as a target for the pred-nov-miR-3278-5p, pred-nov-miR-10235-3p, pred-nov-miR-603-5p pred-nov-miR-18577-3p, pred-nov-miR-8583-5p, and pred-nov-miR-4423-5p.

Supplementary Table 3. Metabolism-related pathways including target genes of novel miRNAs.

Group	Source	Gene	Pathway	Interactions	p-value	FDR
Carbohydrate metabolism						
	UniProt_Pathways	HSPB2	Carbohydrate degradation; glycolysis	155	2,58E-05	0,000251
	KEGG	ANK2	Carbohydrate digestion and absorption	162	0,0011	0,006473
	REACTOME	TGFBR3	Chondroitin sulfate/dermatan sulfate metabolism	78	0,00012	0,000965
	BioCarta	CHRM1	chrebp regulation by carbohydrates and camp	121	1,51E-06	1,97E-05
	BioCarta	CXCL12	chrebp regulation by carbohydrates and camp	273	0,000138	0,001096
	BioCarta	ADCYAP1R1	chrebp regulation by carbohydrates and camp	113	0,000367	0,002547
	BioCarta	F2RL3	chrebp regulation by carbohydrates and camp	12	0,000623	0,004
	WikiPathways	TCF21	Factors and pathways affecting insulin-like growth factor (IGF1)-Akt signaling	128	4,00E-06	4,75E-05
	WikiPathways	TIE1	Factors and pathways affecting insulin-like growth factor (IGF1)-Akt signaling	225	0,000127	0,001016
	WikiPathways	FSCN1	Factors and pathways affecting insulin-like growth factor (IGF1)-Akt signaling	184	0,000621	0,00399
	SMPDB	HSPB2	Fructose-1,6-diphosphatase deficiency	155	5,97E-05	0,000526
	KEGG	ANKRD29	Glucagon signaling	26	0,001027	0,006106
	REACTOME	PDE4C	Glucagon signaling in metabolic regulation	49	1,19E-15	6,37E-14
	REACTOME	PDE4C	Glucagon-like Peptide-1 (GLP1) regulates insulin secretion	49	0,000264	0,001919
	REACTOME	ADCYAP1R1	Glucagon-type ligand receptors	113	2,24E-07	3,48E-06
	HumanCyc	HSPB2	gluconeogenesis	155	4,10E-06	4,86E-05

REACTOME	HSPB2	Gluconeogenesis	155	1,72E-05	0,000175
SMPDB	HSPB2	Gluconeogenesis	155	5,97E-05	0,000526
REACTOME	HSPB2	Glucose metabolism	155	0,001009	0,006018
ACSN2	HSPB2	GLUCOSE_METABOLISM	155	0,00014	0,00111
ACSN2	ALDH2	GLUCOSE_METABOLISM	99	0,00016	0,001241
SMPDB	ALDH2	Glucose-6-phosphate dehydrogenase deficiency	99	0,00012	0,000971
SMPDB	ALDH2	Glucose-Alanine Cycle	99	1,91E-05	0,000192
UniProt_Pathways	ANK2	Glycan biosynthesis; glycogen biosynthesis	162	0,001102	0,006481
REACTOME	ANK2	Glycogen storage disease type 0 (muscle GYS1)	162	0,000113	0,000916
SMPDB	HSPB2	Glycogen Storage Disease Type 1A (GSD1A) or Von Gierke Disease	155	5,97E-05	0,000526
REACTOME	ANK2	Glycogen storage disease type XV (GYG1)	162	0,000113	0,000916
SMPDB	HSPB2	Glycogenosis, Type IB	155	5,97E-05	0,000526
SMPDB	HSPB2	Glycogenosis, Type IC	155	5,97E-05	0,000526
HumanCyc	HSPB2	glycolysis	155	8,54E-05	0,00072
SMPDB	HSPB2	Glycolysis	155	0,000402	0,002754
Panther Pathway	HSPB2	Glycolysis	155	0,000592	0,003836
KEGG	ALDH2	Glycolysis/Gluconeogenesis	99	2,78E-24	3,59E-22
KEGG	HSPB2	Glycolysis/Gluconeogenesis	155	5,27E-05	0,000471
WikiPathways	HSPB2	Glycolysis and Gluconeogenesis	155	5,14E-06	5,95E-05
EHMN	ALDH2	Glycolysis and Gluconeogenesis	99	0,000104	0,000857
WikiPathways	ALDH2	Glycolysis and Gluconeogenesis	99	0,000287	0,002065
INOH	ALDH2	Glycolysis Gluconeogenesis	99	1,61E-05	0,000165
INOH	GFRA1	insulin	43	2,88E-06	3,53E-05
INOH	RTKN2	insulin	44	6,05E-06	6,89E-05

	INOH	SPRED3	insulin	55	1,39E-05	0,000145
	INOH	MYH11	insulin	159	1,55E-05	0,00016
	INOH	FSCN1	insulin	184	6,09E-05	0,000534
	INOH	STAC	insulin	133	0,000141	0,001111
	INOH	ANK2	insulin	162	0,000179	0,001371
	INOH	TIE1	insulin	225	0,000261	0,001898
	INOH	SHCBP1	insulin	161	0,000396	0,002717
	INOH	GFRA1	insulin	43	2,88E-06	3,53E-05
	INOH	RTKN2	insulin	44	6,05E-06	6,89E-05
	INOH	SPRED3	insulin	55	1,39E-05	0,000145
	INOH	MYH11	insulin	159	1,55E-05	0,00016
	INOH	FSCN1	insulin	184	6,09E-05	0,000534
	INOH	STAC	insulin	133	0,000141	0,001111
	INOH	ANK2	insulin	162	0,000179	0,001371
	INOH	TIE1	insulin	225	0,000261	0,001898
	INOH	SHCBP1	insulin	161	0,000396	0,002717
	INOH	GFRA1	insulin Mam	43	3,69E-06	4,41E-05
	INOH	RTKN2	insulin Mam	44	7,74E-06	8,59E-05
	INOH	SPRED3	insulin Mam	55	1,78E-05	0,000181
	INOH	MYH11	insulin Mam	159	2,16E-05	0,000214
	INOH	FSCN1	insulin Mam	184	8,39E-05	0,000708
	INOH	STAC	insulin Mam	133	0,000185	0,001413
	INOH	ANK2	insulin Mam	162	0,000236	0,001739
	INOH	TIE1	insulin Mam	225	0,000355	0,002477
	INOH	SHCBP1	insulin Mam	161	0,000518	0,003421

PID	GFRA1	Insulin Pathway	43	5,50E-15	2,72E-13
PID	STAC	Insulin Pathway	133	2,55E-11	7,72E-10
PID	TIE1	Insulin Pathway	225	2,41E-09	5,36E-08
PID	FSCN1	Insulin Pathway	184	2,45E-05	0,00024
PID	AFAP1L1	Insulin Pathway	71	0,00011	0,000899
PID	SCN2B	Insulin Pathway	210	0,000281	0,002027
PID	RTKN2	Insulin Pathway	44	0,00049	0,003262
PID	SPRED3	Insulin Pathway	55	0,000802	0,004957
PID	ANK2	Insulin Pathway	162	0,00142	0,008012
REACTOME	SCN2B	Insulin processing	210	1,07E-08	2,12E-07
REACTOME	GFRA1	Insulin receptor signalling cascade	43	3,16E-13	1,25E-11
REACTOME	TIE1	Insulin receptor signalling cascade	225	2,22E-08	4,18E-07
REACTOME	RTKN2	Insulin receptor signalling cascade	44	1,26E-05	0,000132
REACTOME	SPRED3	Insulin receptor signalling cascade	55	2,46E-05	0,000241
REACTOME	ALPK3	Insulin receptor signalling cascade	62	4,64E-05	0,000421
REACTOME	AFAP1L1	Insulin receptor signalling cascade	71	8,44E-05	0,000712
KEGG	GFRA1	Insulin resistance	43	0,000254	0,001856
KEGG	TIE1	Insulin resistance	225	0,001736	0,009471
KEGG	PDE4C	Insulin secretion	49	5,73E-15	2,83E-13
KEGG	ANKRD29	Insulin secretion	26	2,14E-05	0,000212
KEGG	ANK2	Insulin secretion	162	0,000269	0,001949
KEGG	ADCYAP1R1	Insulin secretion	113	0,000676	0,004292
KEGG	CHRM1	Insulin secretion	121	0,000897	0,005448
BioCarta	GFRA1	insulin signaling	43	1,84E-13	7,49E-12
WikiPathways	GFRA1	Insulin Signaling	43	5,99E-13	2,28E-11

WikiPathways	TIE1	Insulin Signaling	225	1,12E-11	3,57E-10
WikiPathways	SPRED3	Insulin Signaling	55	1,17E-08	2,31E-07
KEGG	GFRA1	Insulin signaling	43	1,16E-07	1,90E-06
BioCarta	TIE1	insulin signaling	225	4,49E-07	6,56E-06
WikiPathways	STAC	Insulin Signaling	133	6,92E-07	9,72E-06
KEGG	TIE1	Insulin signaling	225	2,20E-06	2,77E-05
WikiPathways	ANK2	Insulin Signaling	162	7,92E-06	8,76E-05
KEGG	STAC	Insulin signaling	133	7,80E-05	0,000665
KEGG	RTKN2	Insulin signaling	44	9,24E-05	0,000771
WikiPathways	RTKN2	Insulin Signaling	44	0,000192	0,001455
KEGG	SPRED3	Insulin signaling	55	0,000207	0,001554
WikiPathways	FSCN1	Insulin Signaling	184	0,000214	0,001599
BioCarta	SPRED3	insulin signaling	55	0,000836	0,005138
WikiPathways	JAM2	Insulin Signaling	106	0,000977	0,005855
WikiPathways	AFAP1L1	Insulin Signaling	71	0,001763	0,00959
WikiPathways	RCAN2	Insulin Signaling	88	0,001763	0,00959
stke	RTKN2	Insulin Signaling Pathway	44	4,46E-05	0,000407
stke	SPN	Insulin Signaling Pathway	58	8,66E-05	0,000728
stke	SPRED3	Insulin Signaling Pathway	55	8,66E-05	0,000728
stke	GFRA1	Insulin Signaling Pathway	43	0,000661	0,004208
SMPDB	GFRA1	Insulin Signalling	43	1,41E-07	2,28E-06
SMPDB	SPRED3	Insulin Signalling	55	5,12E-07	7,40E-06
SMPDB	RTKN2	Insulin Signalling	44	2,35E-05	0,000231
SMPDB	TIE1	Insulin Signalling	225	0,000134	0,001064
Panther_Pathway	TCF21	Insulin_IGF_pathway_MAP_kinase_cascade	128	7,46E-05	0,000639

	Panther_Pathway	TIE1	Insulin_IGF_pathway_MAP_kinase_cascade	225	0,001147	0,006702
	Panther_Pathway	TCF21	Insulin_IGF_pathway-protein_kinase_B_signaling_cascade	128	2,88E-05	0,000277
	Panther_Pathway	SPRED3	Insulin_IGF_pathway-protein_kinase_B_signaling_cascade	55	9,98E-05	0,000824
	Panther_Pathway	TIE1	Insulin_IGF_pathway-protein_kinase_B_signaling_cascade	225	0,000461	0,003097
	Panther_Pathway	CLSPN	Insulin_IGF_pathway-protein_kinase_B_signaling_cascade	109	0,000819	0,005044
	REACTOME	RBPMS	Insulin-like Growth Factor-2 mRNA Binding Proteins (IGF2BPs/IMPs/VICKZs) bind RNA	446	0,001565	0,008689
	PID	FSCN1	Insulin-mediated glucose transport	184	1,49E-06	1,95E-05
	PID	STAC	Insulin-mediated glucose transport	133	9,93E-06	0,000107
	PID	MYH11	Insulin-mediated glucose transport	159	1,07E-05	0,000114
	PID	ALDH2	Insulin-mediated glucose transport	99	5,00E-05	0,00045
	PID	ANK2	Insulin-mediated glucose transport	162	0,000238	0,001752
	PID	SUSD2	Insulin-mediated glucose transport	11	0,000259	0,001886
	PID	SHCBP1	Insulin-mediated glucose transport	161	0,000416	0,002836
	PID	TIE1	Insulin-mediated glucose transport	225	0,001147	0,006702
	PID	TPPP	Insulin-mediated glucose transport	124	0,001263	0,007263
	HumanCyc	ALDH2	lactate fermentation (reoxidation of cytosolic NADH)	99	3,46E-06	4,17E-05
	SMPDB	ANK2	Lactose Degradation	162	0,000183	0,001398
	SMPDB	ANK2	Lactose Intolerance	162	0,000183	0,001398
	REACTOME	CHRM1	Regulation of insulin secretion	121	3,99E-09	8,55E-08
	REACTOME	PDE4C	Regulation of insulin secretion	49	7,52E-05	0,000643

	REACTOME	C2	Regulation of Insulin-like Growth Factor (IGF) transport and uptake by Insulin-like Growth Factor Binding Proteins (IGFBPs)	275	4,10E-06	4,86E-05
	REACTOME	ALDH2	Regulation of Insulin-like Growth Factor (IGF) transport and uptake by Insulin-like Growth Factor Binding Proteins (IGFBPs)	99	2,83E-05	0,000272
	REACTOME	CXCL12	Regulation of Insulin-like Growth Factor (IGF) transport and uptake by Insulin-like Growth Factor Binding Proteins (IGFBPs)	273	0,000167	0,001292
	REACTOME	GFRA1	Signaling by Insulin receptor	43	1,35E-11	4,26E-10
	REACTOME	TIE1	Signaling by Insulin receptor	225	1,20E-06	1,61E-05
	REACTOME	RTKN2	Signaling by Insulin receptor	44	7,37E-05	0,000632
	REACTOME	SPRED3	Signaling by Insulin receptor	55	0,000142	0,001124
	REACTOME	ALPK3	Signaling by Insulin receptor	62	0,000265	0,001927
	REACTOME	AFAP1L1	Signaling by Insulin receptor	71	0,000475	0,003178
Energy metabolism						
	BioCarta	CHRM1	activation of camp-dependent protein kinase pka	121	2,15E-07	3,36E-06
	BioCarta	CXCL12	activation of camp-dependent protein kinase pka	273	2,19E-05	0,000217
	BioCarta	ADCYAP1R1	activation of camp-dependent protein kinase pka	113	0,000107	0,000873
	BioCarta	F2RL3	activation of camp-dependent protein kinase pka	12	0,000332	0,002338
	REACTOME	CLSPN	Activation of PUMA and translocation to mitochondria	109	2,83E-05	0,000272
	REACTOME	TCF21	Activation of PUMA and translocation to mitochondria	128	0,001673	0,009186
	SMPDB	PDE4C	Adenosine Deaminase Deficiency	49	5,38E-13	2,06E-11
	REACTOME	PDE4C	Adenylate cyclase activating	49	4,49E-22	4,72E-20
	REACTOME	PDE4C	Adenylate cyclase inhibitory	49	8,89E-20	7,48E-18

REACTOME	CHRM1	Adenylate cyclase inhibitory	121	0,000232	0,001713
REACTOME	CHRM1	ADP signalling through P2Y purinoceptor 1	121	5,10E-08	8,98E-07
REACTOME	CHRM1	ADP signalling through P2Y purinoceptor 12	121	3,82E-05	0,000355
BioCarta	CXCL12	adp-ribosylation factor	273	0,000193	0,001462
PID	ANLN	Arf1	335	3,08E-06	3,75E-05
PID	ANLN	Arf6 downstream	335	0,000251	0,001836
PID	CHRM1	Arf6 signaling events	121	5,68E-07	8,13E-06
PID	TIE1	Arf6 signaling events	225	0,00023	0,001703
PID	F2RL3	Arf6 signaling events	12	0,000453	0,003052
PID	AFAP1L1	Arf6 signaling events	71	0,00089	0,005415
PID	SCN2B	Arf6 trafficking events	210	1,94E-06	2,47E-05
PID	ALPK3	Arf6 trafficking events	62	0,001359	0,007728
REACTOME	ANK2	ATP sensitive Potassium channels	162	0,000666	0,004235
KEGG	PDE4C	cAMP signaling	49	4,23E-18	2,99E-16
KEGG	ADCYAP1R1	cAMP signaling	113	4,13E-14	1,83E-12
KEGG	CHRM1	cAMP signaling	121	1,63E-12	5,86E-11
KEGG	ANKRD29	cAMP signaling	26	3,87E-05	0,000359
KEGG	MYH11	cAMP signaling	159	0,001041	0,006179
KEGG	SPRED3	cAMP signaling	55	0,001112	0,00653
INOH	HSPB2	Citrate cycle	155	1,47E-05	0,000152
WikiPathways	RCAN2	Energy Metabolism	88	5,30E-06	6,11E-05
UniProt_Pathways	ALDH2	Fermentation; pyruvate fermentation to lactate	99	3,46E-06	4,17E-05
REACTOME	PDE4C	Integration of energy metabolism	49	5,43E-13	2,08E-11
REACTOME	CHRM1	Integration of energy metabolism	121	5,84E-09	1,22E-07
INOH	ALDH2	Pyruvate metabolism	99	7,61E-11	2,14E-09

	KEGG	ALDH2	Pyruvate metabolism	99	9,45E-09	1,90E-07
	REACTOME	ALDH2	Pyruvate metabolism	99	0,001007	0,00601
	REACTOME	ALDH2	Pyruvate metabolism and Citric Acid (TCA) cycle	99	0,000503	0,003336
	REACTOME	HSPB2	The citric acid (TCA) cycle and respiratory electron transport	155	4,98E-05	0,000448
	SMPDB	HSPB2	Warburg Effect	155	8,04E-05	0,000683
	ACSN2	FSCN1	MITOCHONDRIA_OXIDATIVE_STRESS	184	4,46E-07	6,53E-06
	ACSN2	MYH11	MITOCHONDRIA_OXIDATIVE_STRESS	159	2,30E-06	2,89E-05
	ACSN2	PDLIM2	MITOCHONDRIA_OXIDATIVE_STRESS	64	3,42E-05	0,000322
	ACSN2	STAC	MITOCHONDRIA_OXIDATIVE_STRESS	133	3,90E-05	0,000361
	ACSN2	SHCBP1	MITOCHONDRIA_OXIDATIVE_STRESS	161	9,68E-05	0,000803
	ACSN2	TIE1	MITOCHONDRIA_OXIDATIVE_STRESS	225	0,000342	0,002396
	ACSN2	SUSD2	MITOCHONDRIA_OXIDATIVE_STRESS	11	0,000446	0,003013
	ACSN2	ANK2	MITOCHONDRIA_OXIDATIVE_STRESS	162	0,000686	0,004343
	ACSN2	HMGA2	MITOCHONDRIA_OXIDATIVE_STRESS	499	0,001202	0,006973
	REACTOME	ALDH2	Mitochondrial biogenesis	99	0,001313	0,007502
	SMPDB	PDE4C	Mitochondrial DNA depletion syndrome	49	5,38E-13	2,06E-11
	SMPDB	HSPB2	Mitochondrial Electron Transport Chain	155	0,000971	0,005828
	stke	LDLRAD1	Mitochondrial Pathway of Apoptosis: Antiapoptotic Bcl-2 Family	10	0,000328	0,002313
	stke	LDLRAD1	Mitochondrial Pathway of Apoptosis: BH3-only Bcl-2 Family	10	0,001171	0,006819
	stke	CLSPN	Mitochondrial Pathway of Apoptosis: Caspases	109	0,000273	0,001975
	stke	ANLN	Mitochondrial Pathway of Apoptosis: Caspases	335	0,000885	0,005387
	stke	RCAN2	Mitochondrial Pathway of Apoptosis: Multidomain Bcl-2 Family	88	2,10E-05	0,000209

	stke	LDLRAD1	Mitochondrial Pathway of Apoptosis: Multidomain Bcl-2 Family	10	0,000974	0,005843
	stke	CLSPN	Mitochondrial Pathway of Apoptosis: Multidomain Bcl-2 Family	109	0,001595	0,008828
	ACSN2	HSPB2	MITOCHONDRIAL_GENES	155	9,42E-05	0,000784
	ACSN2	HSPB2	MITOCHONDRIAL_METABOLISM	155	9,06E-11	2,52E-09
	ACSN2	ALDH2	MITOCHONDRIAL_METABOLISM	99	3,82E-06	4,56E-05
	KEGG	HSPB2	Oxidative phosphorylation	155	5,81E-06	6,64E-05
	ACSN2	HSPB2	OXIDATIVE_PHOSPHOYLATION_AND_TCA_CYCLE	155	7,48E-07	1,04E-05
	ACSN2	ALDH2	OXIDATIVE_PHOSPHOYLATION_AND_TCA_CYCLE	99	2,25E-05	0,000222
Lipidic metabolism						
	systems-biology.org	FSCN1	adipocyte_v2.1	184	3,92E-09	8,40E-08
	systems-biology.org	HMGA2	adipocyte_v2.1	499	8,31E-08	1,41E-06
	systems-biology.org	PDE4C	adipocyte_v2.1	49	8,65E-07	1,19E-05
	systems-biology.org	GFRA1	adipocyte_v2.1	43	1,17E-06	1,57E-05
	systems-biology.org	HSPB2	adipocyte_v2.1	155	6,51E-06	7,36E-05
	systems-biology.org	MYH11	adipocyte_v2.1	159	1,04E-05	0,000111
	systems-biology.org	CHRM1	adipocyte_v2.1	121	3,67E-05	0,000342
	systems-biology.org	TIE1	adipocyte_v2.1	225	5,36E-05	0,000478
	systems-biology.org	PDLIM2	adipocyte_v2.1	64	0,000101	0,000831
	systems-biology.org	ALDH2	adipocyte_v2.1	99	0,001663	0,009142
	systems-biology.org	RTKN2	adipocyte_v2.1	44	0,00173	0,009443
	KEGG	HMGA2	Adipocytokine signaling	499	0,000335	0,002356
	KEGG	GFRA1	Adipocytokine signaling	43	0,001044	0,006193
	KEGG	RTKN2	Adipocytokine signaling	44	0,001602	0,00886
	WikiPathways	ONECUT1	Adipogenesis	234	5,66E-08	9,88E-07

WikiPathways	DLX5	Adipogenesis	241	6,17E-06	7,01E-05
WikiPathways	TCF21	Adipogenesis	128	6,71E-05	0,000582
WikiPathways	DLX2	Adipogenesis	275	0,00058	0,00377
WikiPathways	SPRED3	Adipogenesis	55	0,001659	0,009119
EHMN	ALDH2	Butanoate metabolism	99	0,011522	0,042842
KEGG	ALDH2	Fatty acid degradation	99	4,16E-13	1,62E-11
WikiPathways	ALDH2	Fatty Acid Omega Oxidation	99	4,54E-12	1,53E-10
REACTOME	CHRM1	Fatty Acids bound to GPR40 (FFAR1) regulate insulin secretion	121	3,70E-05	0,000345
ACSN2	TIE1	FATTY_ACID_BIOSYNTHESIS	225	0,001023	0,00609
REACTOME	CHRM1	Free fatty acids regulate insulin secretion	121	0,000107	0,000876
KEGG	ALDH2	Glycerolipid metabolism	99	0,000714	0,004494
KEGG	CHRM1	Glycerolipid metabolism	121	0,001626	0,008968
EHMN	ALDH2	Glycerophospholipid metabolism	99	0,000383	0,002644
REACTOME	ANLN	LDL clearance	335	5,68E-08	9,92E-07
NetPath	TIE1	Leptin	225	2,69E-18	1,95E-16
NetPath	GFRA1	Leptin	43	4,32E-10	1,08E-08
NetPath	STAC	Leptin	133	2,00E-07	3,14E-06
NetPath	SPRED3	Leptin	55	4,49E-06	5,27E-05
NetPath	FSCN1	Leptin	184	0,000136	0,00108
NetPath	HSPB2	Leptin	155	0,000372	0,002578
NetPath	HMGA2	Leptin	499	0,000767	0,004777
NetPath	RTKN2	Leptin	44	0,001176	0,006842
NetPath	DLX5	Leptin	241	0,001222	0,007068
WikiPathways	RTKN2	Leptin Insulin Overlap	44	2,35E-05	0,000231

	WikiPathways	SPRED3	Leptin Insulin Overlap	55	3,90E-05	0,000361
	WikiPathways	GFRA1	Leptin Insulin Overlap	43	0,001091	0,006427
	WikiPathways	TIE1	Leptin signaling	225	9,70E-13	3,58E-11
	WikiPathways	GFRA1	Leptin signaling	43	3,84E-11	1,13E-09
	WikiPathways	HMGA2	Leptin signaling	499	1,41E-08	2,75E-07
	WikiPathways	STAC	Leptin signaling	133	1,18E-05	0,000125
	WikiPathways	FSCN1	Leptin signaling	184	0,000418	0,002846
	WikiPathways	TIE1	Lipid Metabolism	225	0,001147	0,006702
	PID	ANK2	LKB1 signaling events	162	3,12E-07	4,71E-06
	PID	FSCN1	LKB1 signaling events	184	1,08E-06	1,46E-05
	PID	MYH11	LKB1 signaling events	159	7,70E-05	0,000657
	PID	SHCBP1	LKB1 signaling events	161	0,000177	0,001355
	PID	CLSPN	LKB1 signaling events	109	0,000397	0,002726
	PID	SUSD2	LKB1 signaling events	11	0,000572	0,003722
	PID	TIE1	LKB1 signaling events	225	0,000614	0,003954
	KEGG	PDE4C	Regulation of lipolysis in adipocytes	49	2,55E-17	1,65E-15
	KEGG	TIE1	Regulation of lipolysis in adipocytes	225	2,14E-05	0,000212
	KEGG	ADCYAP1R1	Regulation of lipolysis in adipocytes	113	7,72E-05	0,000658
	KEGG	RTKN2	Regulation of lipolysis in adipocytes	44	0,000744	0,004652
	KEGG	SPRED3	Regulation of lipolysis in adipocytes	55	0,001213	0,007025
	KEGG	CHRM1	Regulation of lipolysis in adipocytes	121	0,001239	0,007146
	REACTOME	ANLN	VLDLR internalisation and degradation	335	8,50E-08	1,43E-06
Calcium metabolism						
	REACTOME	DMP1	Ca2+	24	9,21E-05	0,000769
	REACTOME	RCAN2	Ca2+	88	0,000192	0,001459

BioCarta	ANKRD29	ca-calmodulin-dependent protein kinase activation	26	8,60E-10	2,06E-08
BioCarta	CCM2L	ca-calmodulin-dependent protein kinase activation	3	0,001837	0,009924
REACTOME	PDE4C	Ca-dependent events	49	5,10E-23	5,87E-21
REACTOME	FSCN1	Ca-dependent events	184	0,000478	0,003193
REACTOME	JAM2	Ca-dependent events	106	0,001118	0,006559
REACTOME	RCAN2	Calcineurin activates NFAT	88	5,31E-06	6,13E-05
PID	DMP1	Calcineurin-regulated NFAT-dependent transcription in lymphocytes	24	7,28E-05	0,000625
REACTOME	RAMP3	Calcitonin-like ligand receptors	16	5,64E-18	3,94E-16
WikiPathways	CHRM1	Calcium Regulation in the Cardiac Cell	121	3,31E-12	1,14E-10
WikiPathways	PDE4C	Calcium Regulation in the Cardiac Cell	49	2,29E-09	5,10E-08
WikiPathways	GJA4	Calcium Regulation in the Cardiac Cell	19	3,18E-08	5,81E-07
WikiPathways	ANKRD29	Calcium Regulation in the Cardiac Cell	26	9,50E-06	0,000103
WikiPathways	FSCN1	Calcium Regulation in the Cardiac Cell	184	2,00E-05	0,0002
WikiPathways	ADCYAP1R1	Calcium Regulation in the Cardiac Cell	113	3,26E-05	0,000309
WikiPathways	JAM2	Calcium Regulation in the Cardiac Cell	106	8,76E-05	0,000735
WikiPathways	MYH11	Calcium Regulation in the Cardiac Cell	159	0,000862	0,005272
WikiPathways	ANK2	Calcium Regulation in the Cardiac Cell	162	0,001024	0,006092
CHRM1	KEGG	Calcium signaling	121	1,57E-34	4,46E-32
ADCYAP1R1	KEGG	Calcium signaling	113	1,71E-13	7,01E-12
PDE4C	KEGG	Calcium signaling	49	7,98E-10	1,92E-08
ANKRD29	KEGG	Calcium signaling	26	5,17E-08	9,08E-07
F2RL3	KEGG	Calcium signaling	12	3,85E-07	5,70E-06
RCAN2	KEGG	Calcium signaling	88	6,66E-06	7,51E-05

PRX	KEGG	Calcium signaling	71	0,000484	0,003228
TIE1	KEGG	Calcium signaling	225	0,000818	0,005041
CCM2L	BioCarta	calcium signaling by hbx of hepatitis b virus	3	4,22E-06	4,99E-05
GFRA1	BioCarta	calcium signaling by hbx of hepatitis b virus	43	1,24E-05	0,000131
TIE1	BioCarta	calcium signaling by hbx of hepatitis b virus	225	0,000104	0,000852
ANKRD29	BioCarta	calcium signaling by hbx of hepatitis b virus	26	0,00045	0,003033
MYH11	BioCarta	calcium signaling by hbx of hepatitis b virus	159	0,000549	0,003593
SPRED3	BioCarta	calcium signaling by hbx of hepatitis b virus	55	0,001802	0,009766
RCAN2	PID	Calcium signaling in the CD4+ TCR	88	0,000509	0,003369
DMP1	PID	Calcium signaling in the CD4+ TCR	24	0,001071	0,006328
IPAVS	ANKRD29	Calcium_Signaling	26	7,70E-12	2,51E-10
IPAVS	PDE4C	Calcium_Signaling	49	1,14E-10	3,12E-09
IPAVS	GJA4	Calcium_Signaling	19	4,64E-05	0,000421
IPAVS	JAM2	Calcium_Signaling	106	9,94E-05	0,000821
IPAVS	FSCN1	Calcium_Signaling	184	0,000715	0,004498
IPAVS	TIE1	Calcium_Signaling	225	0,000773	0,004805
REACTOME	PDE4C	Calmodulin induced events	49	1,72E-23	2,06E-21
REACTOME	FSCN1	Calmodulin induced events	184	0,00036	0,002506
REACTOME	JAM2	Calmodulin induced events	106	0,000904	0,005488
REACTOME	PDE4C	CaM	49	1,72E-23	2,06E-21
REACTOME	FSCN1	CaM	184	0,00036	0,002506
REACTOME	JAM2	CaM	106	0,000904	0,005488
REACTOME	PDE4C	Cam-PDE 1 activation	49	1,74E-07	2,77E-06
REACTOME	CHRNB4	Highly calcium permeable nicotinic acetylcholine receptors	70	1,45E-18	1,08E-16

	REACTOME	CHRNB4	Highly calcium permeable postsynaptic nicotinic acetylcholine receptors	70	2,98E-23	3,51E-21
	REACTOME	CHRNB4	Highly sodium permeable acetylcholine nicotinic receptors	70	6,24E-14	2,70E-12
	REACTOME	JAM2	Highly sodium permeable acetylcholine nicotinic receptors	106	0,00103	0,006123
	REACTOME	PDLIM2	Ionotropic activity of kainate receptors	64	0,001038	0,006163
Amino acid metabolism						
	HumanCyc	ALDH2	β-alanine degradation	99	5,02E-05	0,000451
	WikiPathways	ALDH2	Alanine and aspartate metabolism	99	7,35E-05	0,000631
	INOH	ALDH2	Alanine Aspartate Asparagine metabolism	99	2,72E-05	0,000263
	HumanCyc	ALDH2	alanine biosynthesis/degradation	99	5,02E-05	0,000451
	WikiPathways	ALDH2	Amino Acid metabolism	99	8,21E-10	1,97E-08
	KEGG	ALDH2	Alanine, aspartate and glutamate metabolism	99	9,81E-11	2,72E-09
	REACTOME	ALDH2	Amino acid synthesis and interconversion (transamination)	99	0,001236	0,007132
	UniProt_Pathways	TMEM213	Amino-acid biosynthesis; S-adenosyl-L-methionine biosynthesis	7	0,001608	0,008884
	UniProt_Pathways	ALDH2	Amino-acid degradation; L-alanine degradation via transaminase	99	5,02E-05	0,000451
	KEGG	ALDH2	Arginine and proline metabolism	99	1,21E-06	1,62E-05
	SMPDB	ALDH2	Arginine and Proline Metabolism	99	0,000365	0,002537
	KEGG	ALDH2	Arginine biosynthesis	99	2,34E-07	3,63E-06
	INOH	ALDH2	Arginine Proline metabolism	99	3,79E-12	1,29E-10
	SMPDB	ALDH2	Arginine: Glycine Amidinotransferase Deficiency (AGAT Deficiency)	99	0,000365	0,002537
	REACTOME	TPPP	Asparagine N-linked glycosylation	124	0,001601	0,008856

	KEGG	PDE4C	Glutamatergic synapse	49	1,89E-10	5,00E-09
	KEGG	CHRM1	Glutamatergic synapse	121	6,45E-07	9,12E-06
	KEGG	PRX	Glutamatergic synapse	71	0,000352	0,00246
	KEGG	RCAN2	Glutamatergic synapse	88	0,000352	0,00246
	REACTOME	PDLIM2	Glutamate binding, activation of AMPA receptors and synaptic plasticity	64	1,79E-07	2,84E-06
	REACTOME	ANKRD29	Glutamate binding, activation of AMPA receptors and synaptic plasticity	26	2,10E-07	3,28E-06
	REACTOME	SHROOM4	Glutamate binding, activation of AMPA receptors and synaptic plasticity	27	2,23E-05	0,000221
	REACTOME	JAM2	Glutamate binding, activation of AMPA receptors and synaptic plasticity	106	0,000904	0,005488
	INOH	ALDH2	Glutamate Glutamine metabolism	99	7,28E-20	6,19E-18
	SMPDB	ALDH2	Glutamate Metabolism	99	5,00E-07	7,24E-06
	REACTOME	TPPP	Glutamate Neurotransmitter Release Cycle	124	2,27E-05	0,000224
	INOH	ALDH2	Histidine degradation	99	1,64E-18	1,21E-16
	KEGG	ALDH2	Histidine metabolism	99	2,19E-12	7,72E-11
	EHMN	ALDH2	Histidine metabolism	99	4,51E-06	5,29E-05
	REACTOME	CHRNB4	Histidine, lysine, phenylalanine, tyrosine, proline and tryptophan catabolism	70	0,001099	0,006468
	SMPDB	SPRED3	Leucine Stimulation on Insulin Signaling	55	2,96E-07	4,49E-06
	SMPDB	RTKN2	Leucine Stimulation on Insulin Signaling	44	1,58E-05	0,000162
	HumanCyc	CHRNB4	L-kynurenine degradation	70	0,001673	0,009186
	REACTOME	CHRNB4	Lysine catabolism	70	0,00142	0,008014
	INOH	ALDH2	Lysine degradation	99	7,55E-09	1,54E-07
	KEGG	ALDH2	Lysine degradation	99	0,000814	0,00502

EHMN	ALDH2	Lysine metabolism	99	0,000814	0,00502
REACTOME	ALDH2	Metabolism of amino acids and derivatives	99	7,33E-05	0,000629
Panther_Pathway	CHRM1	Metabotropic_glutamate_receptor_group_I	121	5,48E-05	0,000487
Panther_Pathway	TPPP	Metabotropic_glutamate_receptor_group_II	124	0,000258	0,001882
INOH	ALDH2	Methionine Cysteine metabolism	99	0,000718	0,004516
INOH	ALDH2	Valine Leucine Isoleucine degradation	99	2,15E-09	4,83E-08
INOH	HSPB2	Valine Leucine Isoleucine degradation	155	0,000272	0,001968
KEGG	ALDH2	Valine, leucine and isoleucine degradation	99	1,07E-06	1,44E-05
EHMN	ALDH2	Valine, leucine and isoleucine degradation	99	1,21E-06	1,62E-05
KEGG	HSPB2	Valine, leucine and isoleucine degradation	155	0,001301	0,007448
EHMN	HSPB2	Valine, leucine and isoleucine degradation	155	0,001406	0,007945
Metabolism of vitamins and minerals					
SMPDB	ANK2	Amiodarone Action Pathway	162	5,42E-14	2,37E-12
SMPDB	RCAN2	Amiodarone Action Pathway	88	3,59E-08	6,50E-07
SMPDB	HSPB2	Amiodarone Action Pathway	155	0,000345	0,002415
SMPDB	ANK2	Amlodipine Action Pathway	162	5,42E-14	2,37E-12
SMPDB	RCAN2	Amlodipine Action Pathway	88	3,59E-08	6,50E-07
SMPDB	HSPB2	Amlodipine Action Pathway	155	0,000345	0,002415
KEGG	ALDH2	Ascorbate and aldarate metabolism	99	3,73E-05	0,000348
KEGG	FSCN1	Choline metabolism in cancer	184	4,39E-06	5,16E-05
KEGG	GFRA1	Choline metabolism in cancer	43	8,87E-06	9,70E-05
KEGG	TIE1	Choline metabolism in cancer	225	2,88E-05	0,000276
KEGG	SPRED3	Choline metabolism in cancer	55	0,000613	0,003949
BioCarta	CHRM1	ion channels and their functional role in vascular endothelium	121	1,25E-07	2,04E-06

	BioCarta	CXCL12	ion channels and their functional role in vascular endothelium	273	2,32E-06	2,91E-05
	BioCarta	PDE4C	ion channels and their functional role in vascular endothelium	49	1,97E-05	0,000197
	BioCarta	ADCYAP1R1	ion channels and their functional role in vascular endothelium	113	3,46E-05	0,000325
	BioCarta	F2RL3	ion channels and their functional role in vascular endothelium	12	0,00075	0,004685
	REACTOME	ANK2	Ion homeostasis	162	1,20E-06	1,60E-05
	REACTOME	ANK2	Ion homeostasis	162	1,20E-06	1,60E-05
	REACTOME	HSPB2	Ion homeostasis	155	1,21E-05	0,000128
	REACTOME	HSPB2	Ion homeostasis	155	1,21E-05	0,000128
	REACTOME	ANKRD29	Ion homeostasis	26	0,000145	0,001143
	REACTOME	ANKRD29	Ion homeostasis	26	0,000145	0,001143
	REACTOME	ANKRD29	Ion transport by P-type ATPases	26	0,000163	0,001261
	REACTOME	ANKRD29	Ion transport by P-type ATPases	26	0,000163	0,001261
	UniProt_Pathways	ALDH2	Cofactor metabolism; retinol metabolism	99	0,00015	0,001173
	HumanCyc	ALDH2	retinoate biosynthesis I	99	2,85E-05	0,000274
	KEGG	ALDH2	Retinol metabolism	99	1,77E-08	3,38E-07
	REACTOME	ALDH2	Signaling by Retinoic Acid	99	1,94E-08	3,68E-07
	WikiPathways	ALDH2	Vitamin A and Carotenoid Metabolism	99	5,47E-07	7,85E-06
	WikiPathways	C2	Vitamin B12 Metabolism	275	4,08E-05	0,000376
	WikiPathways	HMGA2	Vitamin D in inflammatory diseases	499	2,23E-07	3,46E-06
Nitrogenous basis metabolism						
	HumanCyc	ALDH2	2,-deoxy-α-D-ribose 1-phosphate degradation	99	3,58E-08	6,48E-07

SMPDB	PDE4C	Adenine phosphoribosyltransferase deficiency (APRT)	49	5,38E-13	2,06E-11
REACTOME	SNX22	Cleavage of the damaged purine	8	0,00048	0,003205
REACTOME	SNX22	Depurination	8	0,00048	0,003205
WikiPathways	C2	Folate Metabolism	275	0,00155	0,008618
WikiPathways	ALDH2	Folate-Alcohol and Cancer Pathway Hypotheses	99	2,96E-07	4,49E-06
HumanCyc	PDE4C	guanosine nucleotides de novo biosynthesis	49	0,001293	0,007406
HumanCyc	PDE4C	guanosine ribonucleotides de novo biosynthesis	49	0,000818	0,005042
REACTOME	ALDH2	Metabolism of folate and pterines	99	0,000183	0,001399
REACTOME	PDE4C	Metabolism of nucleotides	49	1,79E-05	0,000182
REACTOME	HMGA2	Metabolism of nucleotides	499	0,000645	0,004125
REACTOME	RBMS2	Metabolism of RNA	104	1,91E-47	1,15E-44
REACTOME	RBPMS	Metabolism of RNA	446	4,16E-19	3,27E-17
REACTOME	HMGA2	Metabolism of RNA	499	7,09E-16	3,90E-14
KEGG	PDE4C	Purine metabolism	49	9,53E-71	1,58E-67
EHMN	PDE4C	Purine metabolism	49	2,00E-17	1,31E-15
SMPDB	PDE4C	Purine Metabolism	49	5,38E-13	2,06E-11
UniProt_Pathways	PDE4C	Purine metabolism; 3',5'-cyclic AMP degradation	49	1,35E-19	1,12E-17
UniProt_Pathways	PDE4C	Purine metabolism; 3',5'-cyclic GMP degradation	49	4,37E-08	7,79E-07
SMPDB	PDE4C	Purine Nucleoside Phosphorylase Deficiency	49	5,38E-13	2,06E-11
HumanCyc	PDE4C	purine nucleotides de novo biosynthesis	49	0,001174	0,006835
INOH	PDE4C	Purine nucleotides nucleosides metabolism	49	1,03E-74	1,97E-71
HumanCyc	SHCBP1	pyrimidine deoxyribonucleosides salvage	161	0,001475	0,008271
SMPDB	SHCBP1	Pyrimidine Metabolism	161	0,000164	0,001272

	KEGG	HMGA2	Pyrimidine metabolism	499	0,001154	0,006735
	WikiPathways	HMGA2	Pyrimidine metabolism	499	0,001202	0,006973
	INOH	HMGA2	Pyrimidine nucleotides nucleosides metabolism	499	2,14E-05	0,000212

7. DISCUSSION

In this study, we identified 13 potential novel miRNAs sequences, by *in silico* analyses of a large miRNA-Seq TCGA dataset of LUAD and LUSC samples. Target genes of novel miRNAs were negatively correlated in each histological subtype. To support our results, we showed structural analyses (GC content and length sequences), differential expression analyses (PCA, t-SNE, Kruskal-Wallis, and Mann-Whitney), and target mRNA analyses (target prediction, expression of target genes using RNA-Seq datasets, Spearman correlation analysis, genomic distribution of miRNAs and target genes, and miRNAs-mRNAs network).

The GC content and the length of sequences of the predicted novel sequences were calculated and compared with known miRNAs, showing similarities between them. The percentage of GC was supportive of specific and stable binding between miRNA-mRNA molecules. Our data agrees with the report by Carmel *et al*., which showed that the average GC content was higher in the seed sequence than in other regions in miRNA, reinforcing the essential role of the seed sequence in the binding and showing a possible association between the GC content and the repression caused by miRNAs (81).

Principal Component Analysis results have also verified that the novel miRNAs were able to differentiate between normal and tumor samples. In addition, t-SNE analysis was performed to refine the PCA results, showing that the novel miRNAs have tissue-specificity, thus demonstrating that sequences are exclusively expressed in lung tissues, instead of having widespread expression throughout other human tissues. Our results (novel miRNA discovery) confirm an important characteristic of known miRNAs, that is to show tissue specificity (19,22,23).

During all of the analysis steps, we observed a higher number of altered miRNAs and target genes in LUSC compared to LUAD. In LUSC, we identified 225 significantly deregulated novel miRNAs against 126 novel miRNAs in LUAD. These 225 novel miRNAs (LUSC) have 2,444 target genes, and the 126 novel miRNAs (LUAD) have 945 target genes. All validated in TCGA RNA-Seq datasets. Next, Spearman correlation analysis identified a much larger number of negatively

correlated miRNAs in LUSC than LUAD (66 novel miRNAs with 473 target genes in LUSC *versus* 18 novel miRNAs and 47 target genes in LUAD). Other studies have shown similar results between these tumor subtypes (82–84).

Twelve of the thirteen miRNAs showed a higher expression value in LUSC compared to LUAD. The majority (8 of 13, or ~62% of novel miRNAs) were significant ($p < 0.05$) higher expressed in LUSC than in LUAD: pred-nov-miR-10235-3p, pred-nov-miR-15587-5p, pred-nov-miR-15942-3p, pred-nov-miR-16806-5p, pred-nov-miR-18577-3p, pred-nov-miR-3878-5p, pred-nov-miR-603-5p, and pred-nov-miR-8583-5p.

Interestingly, 3 novel miRNAs (pred-nov-miR-10762-5p, pred-nov-miR-15942-3p, and pred-nov-miR-4423-5p) showed expression only in lung cancer tissues, but not in normal lung samples. A hypothesis for this finding is that these novel sequences, may represent oncofetal miRNAs, which are expressed in cancer and in fetal tissues but are not expressed in normal adult tissues. The role of oncofetal miRNAs has been described in NSCLC recapitulating the fetal expression patterns in tumor samples (70,85).

Considering that miRNAs negatively regulate gene expression, it is assumed that opposite expression levels are observed between a miRNA and its target gene(s). Regarding this question, Spearman correlation analysis demonstrated negative correlation between miRNA and target genes, both in LUAD and LUSC, with upregulated miRNAs targeting downregulated genes in these tumors.

miRNAs are commonly organized in clusters throughout the genome (22,24). Therefore, we determined the genomic localization of novel miRNAs identified in our data, and were able to show, computationally, that several novel miRNAs are mapped in regions near known miRNAs. This result supports their real existence as members of miRNA clusters or families.

As relevant as identifying novel miRNAs, it is to determine whether novel miRNAs have target genes interacting in biologically relevant pathways. In order to address this question, we generated miRNA-mRNA networks that show connections between novel miRNAs and genes in LUAD and LUSC.

The analysis of genes targeted by novel miRNAs showed that a subset of genes: *PDLIM2*, *CCDC141*, *SCN2B*, and *SUSD2*, play roles in the tumorigenesis processes of cell adhesion, migration, and cell cycle regulation, acting as tumor suppressor genes. Other genes in novel miRNAs-target gene networks have roles in metabolism pathways. Notably, our data showed genes regulating adenylate cyclase inhibition, phosphoinositide degeneration, potassium channel mediation, solutes and water transport, and calcium channel activity. Genes in metabolic pathways were *CHRM1*, *CLDN18*, *STAC*, *MYH11*, and *SCN2B*.

PDE4C and *CHRM1* were two target genes present in both histological subtypes and predicted in adenylate cyclase inhibition pathway. According to our data, these genes were down-regulated, which may implicate in an activation of this pathways in lung cancer. The activation of this pathways was associated with cellular proliferation (86) and metastasis (87) in NSCLC. Additionally, the inhibition of this pathway improved the radio sensitivity in prostate cancer cells (88).

Interestingly, Liu *et al*. showed that *HSPB2* is targeted by TP53 and suggested its role in regulating reactive oxygen species (ROS) levels and the Warburg effect, which are crucial aspects of cellular metabolism related to tumor biology (89).

Although targeting alterations in lung cancer metabolism holds promise as a relevant therapeutic approach, it is essential to acknowledge that the efficacy of this treatment may vary among patients, similar to other targeted approaches. A recent study highlighted distinct responses in lung adenocarcinoma and squamous cell carcinoma cell lines when glycolysis and/or glutaminase were inhibited. Furthermore, the authors postulated that the combination of radiation therapy and metabolic inhibition might only be beneficial to a subset of patients with non-small cell lung cancer. These findings emphasize the need for personalized treatment strategies to maximize the potential benefits of targeting cancer metabolism in NSCLC patients (90).

8. CONCLUSIONS

We identified 13 potential novel miRNAs in lung cancer and normal lung tissues; novel sequences have similar molecular characteristics regarding structure and composition when compared to known miRNAs: GC content and length of sequences.

Novel miRNAs were able to differentiate tumors from non-malignant tissues. Additionally, we predicted the target genes which are regulated by these potential novel miRNAs and related them to several metabolic pathways, which have been reported in the literature as deregulated in lung cancer. Such findings indicate that novel miRNAs identified in our study may be functional and with relevance to adenocarcinoma and squamous cell subtypes of lung cancer. These findings are important to support the real existence of these novel miRNAs.

Our results provide new insights about miRNAs involved in lung cancer and malignant metabolism.

9. FUTURE DIRECTIONS

In order to expand this study, experimental validation will be performed to demonstrate novel miRNA sequences in lung cancer. Validation experiments will be carried out in clinical samples of LUAD, LUSC, and histologically normal adjacent lung tissues, by quantitative real-time PCR.

REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021 May;71(3):209–49.
2. Instituto Nacional de Câncer - INCA. Instituto Nacional de Câncer - INCA, Estimativa 2023. 2023.
3. Rodak O, Peris-Díaz MD, Olbromski M, Podhorska-Okołów M, Dzięgiel P. Current landscape of non-small cell lung cancer: Epidemiology, histological classification, targeted therapies, and immunotherapy. Cancers (Basel). 2021 Sep 1;13(18).
4. Duma N, Santana-davila R, Molina JR. Non-Small Cell Lung Cancer: Epidemiology, Screening, Diagnosis, and Treatment. Mayo Clin Proc [Internet]. 2019;94(8):1623–40. Available from: https://doi.org/10.1016/j.mayocp.2019.01.013
5. Ho LJ, Yang HY, Chung CH, Chang WC, Yang S sen, Sun CA, et al. Increased risk of secondary lung cancer in patients with tuberculosis: A nationwide, population-based cohort study. PLoS One. 2021 May 1;16(5 May).
6. Qin Y, Chen Y, Chen J, Xu K, Xu F, Shi J. The relationship between previous pulmonary tuberculosis and risk of lung cancer in the future. Vol. 17, Infectious Agents and Cancer. BioMed Central Ltd; 2022.
7. Caramori G, Ruggeri P, Mumby S, Ieni A, lo Bello F, Chaminka V, et al. Molecular links between COPD and lung cancer: new targets for drug discovery? Expert Opin Ther Targets. 2019 Jun 3;23(6):539–53.
8. Denisov E v., Schegoleva AA, Gervas PA, Ponomaryova AA, Tashireva LA, Boyarko V v., et al. Premalignant lesions of squamous cell carcinoma of the lung: The molecular make-up and factors affecting their progression. Vol. 135, Lung Cancer. Elsevier Ireland Ltd; 2019. p. 21–8.
9. Herbst RS, Morgensztern D, Boshoff C. The biology and management of non-small cell lung cancer. Vol. 553, Nature. Nature Publishing Group; 2018. p. 446–54.
10. Noguchi M, Morikawa A, Kawasaki M, Matsuno Y, Yamada T, Hirohashi S, et al. Small Adenocarcinoma of the Lung Histologic Characteristics and Prognosis.
11. Hutchinson BD, Shroff GS, Truong MT, Ko JP. Spectrum of Lung Adenocarcinoma. Seminars in Ultrasound, CT and MRI. 2019 Jun 1;40(3):255–64.
12. Inamura K. Clinicopathological characteristics and mutations driving development of early lung adenocarcinoma: Tumor initiation and progression. Int J Mol Sci. 2018 Apr 23;19(4):1–17.
13. Succony L, Rassl DM, Barker AP, McCaughan FM, Rintoul RC. Adenocarcinoma spectrum lesions of the lung: Detection, pathology and treatment strategies. Vol. 99, Cancer Treatment Reviews. W.B. Saunders Ltd; 2021.
14. Butnor KJ. Controversies and challenges in the histologic subtyping of lung adenocarcinoma. Vol. 9, Translational Lung Cancer Research. AME Publishing Company; 2020. p. 839–46.
15. Souza VGP, Forder A, Brockley LJ, Pewarchuk ME, Telkar N, de Araújo RP, et al. Liquid Biopsy in Lung Cancer: Biomarkers for the Management of Recurrence and Metastasis. Int J Mol Sci. 2023 May 1;24(10).

16. Reis PP, Drigo SA, Carvalho RF, Lapa RML, Felix TF, Patel D, et al. Circulating miR-16-5p, miR-92a-3p, and miR-451a in plasma from lung cancer patients: Potential application in early detection and a regulatory role in tumorigenesis pathways. Cancers (Basel). 2020 Aug 1;12(8):1–16.
17. Zhang J, Fujimoto J, Zhang J, Wedge DC, Song X, Zhang J, et al. Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. Science (1979). 2014 Oct 10;346(6206):256–9.
18. Lee RC, Feinbaum RL, Ambros V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell [Internet]. 1993 Dec 3 [cited 2017 Jun 25];75(5):843–54. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8252621
19. Budakoti M, Panwar AS, Molpa D, Singh RK, Büsselberg D, Mishra AP, et al. Micro-RNA: The darkhorse of cancer. Cell Signal. 2021 Jul 1;83.
20. Brenda J. Reinhart, Frank J. Slack, Michael Basson, Amy E. Pasquinelli, Jill C. Bettinger, Ann E. Rougvie, et al. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature [Internet]. 2000 [cited 2022 Nov 20];403:901–6. Available from: https://www.nature.com/articles/35002607
21. Kozomara A, Birgaoanu M, Griffiths-Jones S. MiRBase: From microRNA sequences to function. Nucleic Acids Res. 2019 Jan 8;47(D1):D155–62.
22. Michlewski G, Cáceres JF. Post-transcriptional control of miRNA biogenesis. RNA. 2019;25:1–16.
23. Syeda ZA, Langden SSS, Munkhzul C, Lee M, Song SJ. Regulatory mechanism of microrna expression in cancer. Vol. 21, International Journal of Molecular Sciences. MDPI AG; 2020.
24. Gregorova J, Vychytilova-Faltejskova P, Sevcikova S. Epigenetic regulation of microrna clusters and families during tumor development. Vol. 13, Cancers. MDPI; 2021. p. 1–45.
25. Chipman LB, Pasquinelli AE. miRNA Targeting: Growing beyond the Seed. Vol. 35, Trends in Genetics. Elsevier Ltd; 2019. p. 215–22.
26. Souza CP, Cinegaglia NC, Felix TF, Evangelista AF, Oliveira RA, Hasimoto EN, et al. Deregulated micrornas are associated with patient survival and predicted to target genes that modulate lung cancer signaling pathways. Cancers (Basel). 2020 Sep 1;12(9):1–13.
27. Jiang M, Zhang P, Hu G, Xiao Z, Xu F, Zhong T, et al. Relative expressions of miR-205-5p, miR-205-3p, and miR-21 in tissues and serum of non-small cell lung cancer patients. Mol Cell Biochem. 2013 Nov 1;383(1–2):67–75.
28. Xu L bin, Xiong J, Zhang YH, Dai Y, Ren XP, Ren YJ, et al. miR-205-3p promotes lung cancer progression by targeting APBB2. Mol Med Rep. 2021 Aug 1;24(2).
29. Bao M, Pan S, Yang W, Chen S, Shan Y, Shi H. Serum miR-10a-5p and miR-196a-5p as non-invasive biomarkers in non-small cell lung cancer [Internet]. Vol. 11, Int J Clin Exp Pathol. 2018. Available from: www.ijcep.com/
30. Zhou Y, Yan J, Chen H, Zhou W, Xiao G, Zou H, et al. MiR-4652-5p Targets RND1 to Regulate Cell Adhesion and Promote Lung Squamous Cell Carcinoma Progression. Appl Biochem Biotechnol. 2022 Jul 1;194(7):3031–43.
31. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell [Internet]. 2000;100(1):57–70. Available from: https://pubmed.ncbi.nlm.nih.gov/10647931/

32. Hanahan D, Weinberg RA. Hallmarks of cancer: The next generation. Cell [Internet]. 2011;144(5):646–74. Available from: http://dx.doi.org/10.1016/j.cell.2011.02.013
33. Hanahan D. Hallmarks of Cancer: New Dimensions. Cancer Discov. 2022 Jan 1;12(1):31–46.
34. Wu KL, Tsai YM, Lien CT, Kuo PL, Hung JY. The roles of microRNA in lung cancer. Vol. 20, International Journal of Molecular Sciences. MDPI AG; 2019.
35. Alles J, Fehlmann T, Fischer U, Backes C, Galata V, Minet M, et al. An estimate of the total number of true human miRNAs. Nucleic Acids Res. 2019 Apr 23;47(7):3353–64.
36. Londin E, Loher P, Telonis AG, Quann K, Clark P, Jing Y, et al. Analysis of 13 cell types reveals evidence for the expression of numerous novel primate- and tissue-specific microRNAs. Proceedings of the National Academy of Sciences [Internet]. 2015;112(10):E1106–15. Available from: http://www.pnas.org/lookup/doi/10.1073/pnas.1420955112
37. Jima DD, Zhang J, Jacobs C, Richards KL, Dunphy CH, Choi WWL, et al. Deep sequencing of the small RNA transcriptome of normal and malignant human B cells identifies hundreds of novel microRNAs. Blood. 2010;116(23):e118-27.
38. Meiri E, Levy A, Benjamin H, Ben-David M, Cohen L, Dov A, et al. Discovery of microRNAs and other small RNAs in solid tumors. Nucleic Acids Res. 2010;38(18):6234–46.
39. Rock LD, Minatel BC, Marshall EA, Guisier F, Sage AP, Barros-Filho MC, et al. Expanding the Transcriptome of Head and Neck Squamous Cell Carcinoma Through Novel MicroRNA Discovery. Front Oncol. 2019;9:1305.
40. Barros-Filho MC, Pewarchuk M, Minatel B de C, Sage AP, Marshall EA, Martinez VD, et al. Previously undescribed thyroid-specific miRNA sequences in papillary thyroid carcinoma. J Hum Genet [Internet]. 2019;64(5):505–8. Available from: http://dx.doi.org/10.1038/s10038-019-0583-7
41. Pewarchuk ME, Barros-Filho MC, Minatel BC, Cohn DE, Guisier F, Sage AP, et al. Upgrading the repertoire of miRNAs in gastric adenocarcinoma to provide a new resource for biomarker discovery. Int J Mol Sci [Internet]. 2019;20(22):5697. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6888638/
42. Ryu S, Joshi N, McDonnell K, Woo J, Choi H, Gao D, et al. Discovery of novel human breast cancer micrornas from deep sequencing data by analysis of pri-microrna secondary structures. PLoS One [Internet]. 2011;6(2):e16403. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3035615/
43. Minatel BC, Martinez VD, Ng KW, Sage AP, Tokar T, Marshall EA, et al. Large-scale discovery of previously undetected microRNAs specific to human liver. Hum Genomics [Internet]. 2018;12(1):16. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29587854%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5870816%0Ahttps://humgenomics.biomedcentral.com/articles/10.1186/s40246-018-0148-4
44. Wake C, Labadorf A, Dumitriu A, Hoss AG, Bregu J, Albrecht KH, et al. Novel microRNA discovery using small RNA sequencing in post-mortem human brain. BMC Genomics [Internet]. 2016;17(1):3–11. Available from: http://dx.doi.org/10.1186/s12864-016-3114-3
45. Friedländer MR, Lizano E, Houben AJS, Bezdan D, Báñez-Coronel M, Kudla G, et al.

Evidence for the biogenesis of more than 1,000 novel human microRNAs. Genome Biol. 2014;15(4):1–17.
46. Pavlova NN, Thompson CB. The Emerging Hallmarks of Cancer Metabolism. Vol. 23, Cell Metabolism. Cell Press; 2016. p. 27–47.
47. Warburg O. On the Origin of Cancer Cells. Science (1979) [Internet]. 1956;123(3191):309–14. Available from: http://www.ncbi.nlm.nih.gov/pubmed/13298683
48. Azizi MIHN, Othman I, Naidu R. The role of micrornas in lung cancer metabolism. Vol. 13, Cancers. MDPI AG; 2021.
49. Sellers K, Allen TD, Bousamra M, Tan JL, Méndez-Lucas A, Lin W, et al. Metabolic reprogramming and Notch activity distinguish between non-small cell lung cancer subtypes. Br J Cancer. 2019 Jul 2;121(1):51–64.
50. Bamji-Stocke S, van Berkel V, Miller DM, Frieboes HB. A review of metabolism-associated biomarkers in lung cancer diagnosis and treatment. Metabolomics [Internet]. 2018;14(6):1–16. Available from: http://dx.doi.org/10.1007/s11306-018-1376-2
51. Chen YJ, Guo YN, Shi K, Huang HM, Huang SP, Xu WQ, et al. Down-regulation of microRNA-144-3p and its clinical value in non-small cell lung cancer: A comprehensive analysis based on microarray, miRNA-sequencing, and quantitative real-time PCR data. Respir Res. 2019 Mar 4;20(1).
52. Zhao X, Lu C, Chu W, Zhang B, Zhen Q, Wang R, et al. MicroRNA-124 suppresses proliferation and glycolysis in Non-Small cell lung cancer cells by targeting AKT-GLUT1/HKII. Tumor Biology. 2017 May 1;39(5).
53. Zhai S, Zhao L, Lin T, Wang W. Downregulation of miR-33b promotes non-small cell lung cancer cell growth through reprogramming glucose metabolism miR-33b regulates non-small cell lung cancer cell growth. J Cell Biochem. 2019 Apr 1;120(4):6651–60.
54. Muniyappa MK, Dowling P, Henry M, Meleady P, Doolan P, Gammell P, et al. MiRNA-29a regulates the expression of numerous proteins and reduces the invasiveness and proliferation of human carcinoma cell lines. Eur J Cancer. 2009 Nov;45(17):3104–18.
55. Miko E, Margitai Z, Czimmerer Z, Várkonyi I, Dezs B, Lányi Á, et al. MiR-126 inhibits proliferation of small cell lung cancer cells by targeting SLC7A5. FEBS Lett. 2011 Apr 20;585(8):1191–6.
56. Liu G, Li Y, Gao X. Overexpression of microRNA-133b sensitizes non-small cell lung cancer cells to irradiation through the inhibition of glycolysis. Oncol Lett. 2016 Apr 1;11(4):2903–8.
57. Fang R, Xiao T, Fang Z, Sun Y, Li F, Gao Y, et al. MicroRNA-143 (miR-143) regulates cancer glycolysis via targeting hexokinase 2 gene. Journal of Biological Chemistry. 2012 Jun 29;287(27):23227–35.
58. Feng G, Tong Y suo, Tao G zhou, Xu L. miR-206 regulates non-small cell lung cancer cell aerobic glycolysis by targeting hexokinase 2 Ke-gang. Available from: https://academic.oup.com/jb/advance-article-abstract/doi/10.1093/jb/mvz099/5631916
59. Chen D, Huang J, Zhang K, Pan B, Chen J, De W, et al. MicroRNA-451 induces epithelial-mesenchymal transition in docetaxel-resistant lung adenocarcinoma cells by targeting proto-oncogene c-Myc. Eur J Cancer. 2014;50(17):3050–67.

60. Xue X, Liu Y, Wang Y, Meng M, Wang K, Zang X, et al. MiR-21 and MiR-155 promote non-small cell lung cancer progression by downregulating SOCS1, SOCS6, and PTEN [Internet]. Vol. 7, Oncotarget. 2016. Available from: www.impactjournals.com/oncotarget/
61. Su C, Cheng X, Li Y, Han Y, Song X, Yu D, et al. MiR-21 improves invasion and migration of drug-resistant lung adenocarcinoma cancer cell and transformation of EMT through targeting HBP1. Cancer Med. 2018 Jun 1;7(6):2485–503.
62. Edmonds MD, Boyd KL, Moyo T, Mitra R, Duszynski R, Arrate MP, et al. MicroRNA-31 initiates lung tumorigenesis and promotes mutant KRAS-driven lung cancer. Journal of Clinical Investigation. 2016 Jan 4;126(1):349–64.
63. Raver-Shapira N, Marciano E, Meiri E, Spector Y, Rosenfeld N, Moskovits N, et al. Transcriptional Activation of miR-34a Contributes to p53-Mediated Apoptosis. Mol Cell. 2007 Jun 8;26(5):731–43.
64. Hermeking H. The miR-34 family in cancer and apoptosis. Vol. 17, Cell Death and Differentiation. 2010. p. 193–9.
65. Yingyi W, Zhao M, Liu J, Sun Z, Ni J, Liu H. MiRNA-125b regulates apoptosis of human non-small cell lung cancer via the PI3K/Akt/GSK3β signaling pathway. Oncol Rep. 2017 Sep 1;38(3):1715–23.
66. Li G, Li M, Hu J, Lei R, Xiong H, Ji H, et al. The microRNA-182-PDK4 axis regulates lung tumorigenesis by modulating pyruvate dehydrogenase and lipogenesis. Oncogene. 2017 Feb 16;36(7):989–98.
67. Mizuno S, Bogaard HJ, Gomez-Arroyo J, Alhussaini A, Kraskauskas D, Cool CD, et al. MicroRNA-199a-5p is associated with hypoxia-inducible factor-1α expression in lungs from patients with COPD. Chest. 2012;142(3):663–72.
68. Puisségur MP, Mazure NM, Bertero T, Pradelli L, Grosso S, Robbe-Sermesant K, et al. MiR-210 is overexpressed in late stages of lung cancer and mediates mitochondrial alterations associated with modulation of HIF-1 activity. Cell Death Differ. 2011 Mar;18(3):465–78.
69. Fehlmann T, Backes C, Kahraman M, Haas J, Ludwig N, Posch AE, et al. Web-based NGS data analysis using miRMaster: A large-scale meta-analysis of human miRNAs. Nucleic Acids Res. 2017;45(15):8731–44.
70. Cohn DE, Barros-Filho MC, Minatel BC, Pewarchuk ME, Marshall EA, Vucic EA, et al. Reactivation of multiple fetal mirnas in lung adenocarcinoma. Cancers (Basel). 2021;13(11):1–14.
71. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological) [Internet]. 1995;57(1):289–300. Available from: https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1995.tb02031.x
72. Bica-Pop C, Cojocneanu-Petric R, Magdo L, Raduly L, Gulei D, Berindan-Neagoe I. Overview upon miR-21 in lung cancer: focus on NSCLC. Cellular and Molecular Life Sciences [Internet]. 2018;75(19):3539–51. Available from: https://doi.org/10.1007/s00018-018-2877-x
73. Song K, Jiang Y, Zhao Y, Xie Y, Zhou J, Yu W, et al. Members of the miR-30 family inhibit the epithelial-to-mesenchymal transition of non-small-cell lung cancer cells by suppressing xb130 expression levels. Oncol Lett. 2020;20(4):1–8.
74. Metsalu T, Vilo J. ClustVis: A web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. Nucleic Acids Res.

2015;43(W1):W566–70.
75. Chen Y, Wang X. MiRDB: An online database for prediction of functional microRNA targets. Nucleic Acids Res. 2020;48(D1):D127–31.
76. Goldman MJ, Craft B, Hasti M, Repečka K, McDade F, Kamath A, et al. Visualizing and interpreting cancer genomics data via the Xena platform. Nat Biotechnol [Internet]. 2020;38(6):675–8. Available from: https://www.nature.com/articles/s41587-020-0546-8
77. Cheong WH, Tan YC, Yap SJ, Ng KP. ClicO FS: An interactive web-based service of Circos. Bioinformatics. 2015 May 25;31(22):3685–7.
78. Demsar J, Curk T, Erjavec A, Gorup C, Hocevar T, Milutinovic M, et al. Orange: Data Mining Toolbox in Python. Journal of Machine Learning Research. 2013;14:2349–53.
79. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Res. 2003;13(11):2498–504.
80. Rahmati S, Abovsky M, Pastrello C, Jurisica I. PathDIP: An annotated resource for known and predicted human gene-pathway associations and pathway enrichment analysis. Nucleic Acids Res. 2017;45(D1):D419–26.
81. Carmel I, Shomron N, Heifetz Y. Does base-pairing strength play a role in microRNA repression? RNA. 2012 Nov;18(11):1947–56.
82. Sun F, Yang X, Jin Y, Chen L, Wang L, Shi M, et al. Bioinformatics analyses of the differences between lung adenocarcinoma and squamous cell carcinoma using the Cancer Genome Atlas expression data. Mol Med Rep. 2017 Jul 1;16(1):609–16.
83. Yu H, Pang Z, Li G, Gu T. Bioinformatics analysis of differentially expressed miRNAs in non-small cell lung cancer. J Clin Lab Anal. 2021 Feb 1;35(2).
84. Petkova V, Marinova D, Kyurkchiyan S, Stancheva G, Mekov E, Kachakova-Yordanova D, et al. MiRNA expression profiling in adenocarcinoma and squamous cell lung carcinoma reveals both common and specific deregulated microRNAs. Medicine (United States). 2022 Aug 19;101(33):E30027.
85. Becker-Santos DD, Thu KL, English JC, Pikor LA, Martinez VD, Zhang M, et al. Developmental transcription factor NFIB is a putative target of oncofetal miRNAs and is associated with tumour aggressiveness in lung adenocarcinoma. Journal of Pathology. 2016 Oct 1;240(2):161–72.
86. Moody TW, Lee L, Jensen RT. The G Protein–Coupled Receptor PAC1 Regulates Transactivation of the Receptor Tyrosine Kinase HER3. Journal of Molecular Neuroscience. 2021 Aug 1;71(8):1589–97.
87. Tan M, Song X, Zhang G, Peng A, Li X, Li M, et al. Overexpression of adenylate cyclase-associated protein 1 is associated with metastasis of lung cancer. Oncol Rep. 2013 Oct;30(4):1639–44.
88. Appukuttan A, Flacke JP, Flacke H, Posadowsky A, Reusch HP, Ladilov Y. Inhibition of soluble adenylyl cyclase increases the radiosensitivity of prostate cancer cells. Biochim Biophys Acta Mol Basis Dis. 2014 Dec 1;1842(12):2656–63.
89. Liu S, Yan B, Lai W, Chen L, Xiao D, Xi S, et al. As a novel p53 direct target, bidirectional gene HspB2/αB-crystallin regulates the ROS level and Warburg effect. Biochim Biophys Acta Gene Regul Mech. 2014;1839(7):592–603.
90. Meijer TWH, Peeters WJM, Dubois LJ, van Gisbergen MW, Biemans R, Venhuizen

JH, et al. Targeting glucose and glutamine metabolism combined with radiation therapy in non-small cell lung cancer. Lung Cancer. 2018 Dec 1;126:32–40.