Bioinformatics based exploration on STAT3 driven Molecular Pathways and Therapeutic Targets in Lung Adenocarcinoma

By

Ferdousi Islam
20346029

A thesis submitted to the School of Pharmacy in partial fulfillment of the requirements for the degree of Bachelor of Pharmacy

School of Pharmacy
Brac University
September 2024

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. I/We have acknowledged all main sources of help.

Student's Full Name & Signature:

Ferdousi Islam
20346029

Approval

The thesis titled “Bioinformatics based exploration on STAT3 driven Molecular Pathways and Therapeutic Targets in Lung Adenocarcinoma” submitted by Ferdousi Islam (20346029)], of Spring, 2024 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of Bachelor of Pharmacy.

Supervised By:

Muhammad Asaduzzaman, PhD
School of Pharmacy
Brac University

Approved By:

Dean:

A.F.M. Yusuf Haider, PhD
Acting Dean, School of Pharmacy
Professor, Department of Mathematics and Natural Sciences
Brac University

Ethics Statement

While performing this study, basic measures of ethical practices were upheld to the highest level. However, it should be made noted that this study did not entail on any experimentations that was performed on human beings or animals. The study was conducted solely for the analysis of data based on publicly available databases. Therefore, in an endeavor to make sure that the research was done responsibly and ethically so as to contribute to the generation of knowledge without a single ethical consideration being violated, efforts were made to make sure that none of the ethical considerations was injured.

Abstract

The research identified and analyzed genes related to the STAT3 pathway, which plays a crucial role in lung adenocarcinoma progression and treatment resistance. Thus, using BioPortal, Harmonizome, and KEGG PATHWAY several genes including *HRAS, SMAD2, IQGAP1*, and *NFKB1* were considered as significantly co-expressed with STAT3. These genes were also reported to affect different cancer pathways, which are Ras, TGF-beta and NF-kappa B signaling all of which are core to cancer cell proliferation and survival. The pathway analysis of STAT3 target genes demonstrated that the TGF-beta signaling and MAPK signaling pathways are most relevant to lung cancer cases proving their key role in STAT3 mediated tumorigenesis. These findings also point to possible new therapeutic targets to improve the treatment of lung adenocarcinoma, because the pathways detected could represent avenues for molecular therapies that the current standard treatment of chemotherapy and radiation does not fully address, and are likely to provide better survival outcomes.

Keywords: STAT3 Pathway, Lung Adenocarcinoma, TGF-beta Signaling, MAPK Signaling, Co-expressed Genes, Therapeutic Targets.

Dedication

I am privileged to dedicate this work to my dear parents and dear and respected supervisor Dr. Muhammad Asaduzzaman from School of Pharmacy, BRAC University for his kind essential support during the preparation of this project report and all the dear deaths of lung adenocarcinoma. The consolation and the opportunities that the given effort implies to those people who are diagnosed with such terrible diseases are that this would be a small contribution to the struggle against all these diseases on a continuous basis.

Acknowledgement

First of all, I would personally like to express my gratitude to Almighty God for endowing me with good health, energy, and tenacity in the course of this exercise. This is why I can say that it was by the divine graces of him that I was able to do this very assignment on time and forever I shall be grateful to him for that. I would like to express my gratitude to my supervisor, Dr Muhammad Asaduzzaman PhD from the School of Pharmacy at BRAC University for directing and assisting me and for what he has done for me, I owe much to have such opportunity to work under him. In this regard, his high level of commitment, professionalism and constant encouragement have been my source of inspiration during this whole journey. Besides all of this, I would like to extend my sincere gratitude towards the Active Dean, School of Pharmacy for facilitating and enabling the submission of our project work in due time. At this stage, I would like to come up with something most emotional; to say thanks to my parents. The biggest way they have supported me has always been provided by to them at every aspect of my life, the unconditional affection they shower on me, and the challenges they empower me to take. I did not believe in myself when they believed in me and for this, I shall be ever grateful to have had an opportunity to strive and achieve the remaining academic dreams. The unconditional love and such encouragement from the two made such a feat to be realized and I am very thankful to them. However, it might signal that it would have been impossible to do so in the absence of the resources.

Table of Contents

List of Tables

List of Figures

List of Acronyms

LUAD	Lung Adenocarcinoma
NSCLC	Non-Small Cell Lung Cancer
STAT3	Signal Transducer and Activator of Transcription 3
KEGG	Kyoto Encyclopedia of Genes and Genomes
GO	Gene Ontology
FDR	False Discovery Rate
HR	Hazard Ratio
HIF1	Hypoxia-Inducible Factor 1
PI3K	Phosphoinositide 3-kinase
MAPK	Mitogen-Activated Protein Kinase
VEGF	Vascular Endothelial Growth Factor
NF-kappa B	Nuclear Factor kappa-light-chain-enhancer of activated B cells
cAMP	Cyclic Adenosine Monophosphate
SNP	Single Nucleotide Polymorphism SVM - Support Vector Machine
RAGE	Receptor for Advanced Glycation End-products

Chapter 1

Introduction

Cancer that affects the lungs is the most common cause of mortality from cancer. Within the year 2012, there were roughly 1.8 million people who were diagnosed with lung cancer, and 1.6 million people lost their lives to the disease. Lung cancer is responsible for more deaths than breast cancer. Cancers affected the prostate and intestines. During the year 2018, there were more than 230,000 newly diagnosed cases of lung cancer. There is a lower incidence of lung cancer in women compared to men, and the chance of developing lung cancer increases with increasing age. Cigarette smoking is one of the primary contributors to the development of lung cancer on a global scale. Diagnostic costs for lung cancer are high, and the disease is typically discovered in its terminal stage. Chest radiography and sputum are both part of the screening process for lung cancer; nevertheless, this method does not succeed in reducing the mortality rate caused by disease. (Lee Ella A., 2022). During the average period of five years, the survival rate for lung cancer in the United States is fifteen percent. Small cell carcinoma and non-small cell carcinoma are the two variants of lung cancer that can be distinguished from one another. It is possible for adenocarcinoma, large cell carcinoma, and squamous carcinoma to be classified as non-small cell carcinoma. It has been determined that lung carcinoma can be classified in order to determine the prognosis and the type of tumor. Both the type of tumor and the extent of the metastasis can cause signs and symptoms that are distinct from one another. There are about 60%lung cancer possess oncogenic driven mutation and which is used in treatment response as well and it relates sometimes with different features(Sholl, 2015).

Lung cancer is characterized by a number of symptoms, the most frequent of which are the initial tumor (dyspnea, chest discomfort, cough), the intrathoracic spread (chest wall invasion, esophageal symptoms, Horner syndrome, and pleural effusion), and the extra thoracic spread (bone pain, fracture, disorientation, seizures, headache, weight loss, localized neurologic impairments). When attempting to diagnose cancer, it is necessary to do an examination of the patient's cancer type, including tissue diagnosis, evaluation of metastasis, staging, and functional patient evaluation. In the beginning of the evaluation of metastatic disease, it is dependent on the patient's medical history. Lung cancer is the greatest cause of death from cancer across the globe, and it has enormous implications for the health of people all over the world. 30-40%of lung cancer patient suffer from bone metastases and it correlate within 7months after it has been detected(Eastman et al., 2021).Lung cancer is subdivided by lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) and they are as non-small cell lung cancer (NSCLC). Though they are subtype of non-small cell lung cancer but they are treated in different way in medical science as a different cancer (J. W. Chen & Dhahbi, 2021).

1.1 Non-Small Cell Lung Cancer

It is a subtype of lung cancer that has the potential to result in death in patients with lung cancer. The overall cure and survival percentage is dismal, especially in cases when the disease has spread to other parts of the body. In light of this, research is currently focusing on developing new medications, and combination chemotherapy is urgently required. In addition, numerous types of treatment have been developed for this lung cancer, and the first exosome therapy for this non-small cell lung cancer is based on the natural exosomes that are produced by immune cells. (Chen et al., 2023) There are many categories of non-small cell lung carcinoma and lung adenocarcinoma is one of them. Oncology continues to face a tremendous obstacle in the shape of non-small cell

lung cancer (NSCLC) due to the high death rate associated with this disease and the complicated therapeutic landscape. Despite the fact that recent developments in molecularly targeted medicines and immunotherapies have demonstrated their potential, the prognosis for advanced non-small cell lung cancer (NSCLC) continues to be dismal, which calls for further research into innovative therapeutic approaches. Immunotherapy, and more specifically immune checkpoint inhibitors, has brought about a paradigm shift in the treatment paradigm by providing patients with long-lasting responses

1.2 Lung Adenocarcinoma

One of the types of lung cancer, known as lung adenocarcinoma, is the one that has the greatest morbidity and fatality rate, not just in China but internationally as well. Lung adenocarcinoma, the most frequent pathological type of lung cancer, accounts for around forty percent of all cases of lung cancer. The treatment for this condition is contingent on the size and kind of the tumor. Surgery is the primary method of treatment for patients who have been diagnosed at an early stage; however, those who have advanced stages or metastatic conditions require systematic therapy. The overall survival time of LUAD is still below average, despite the fact that major therapeutic improvements have been made in recent years. (Wei et al., 2023) Type II alveolar cells, also known as tiny airway epithelial cells, are the cells that are responsible for the development of LUAD. These cells release muscle and other chemicals. Mixed type lung adenocarcinoma (ADC) is characterized by the presence of a heterogeneous selection of histological development patterns. When it comes to clinical diagnosis, these histological characteristics and marker expression continue to serve as the foundation. The heterogeneity of the tumor is understood by sequencing technologies, which caused the LUAD to be separated into several molecular subsets. This type of mutation is referred to as driver mutation. In lung adenocarcinoma we seen 50 progressive

increased mutation with chromosomal aberrations. Also observed vulnerability of precancer to invasive LUAD(Hu et al., 2024)

Molecular modification, which is necessary for the growth of tumors and the beginning of their development, is represented by this. In many cases, it can be found in genes that regulate the continuation and proliferation of cells. Oncogene addiction is the term used to describe the phenomenon in which a tumor shows a response to the production of a single mutant oncogene in order to commence or prolong the growth and survival of the tumor. (Denisenko et al., 2018). This form of cancer is the histological subtype of non-small cell lung cancer that is seen the most frequently. An intricate tumor microenvironment (TME) is the classification that is given to it. There are times when the standard approach of defining stages is an inaccurate method. This research demonstrates the importance of developing a novel and reliable model of prognosis that will assist medical professionals in determining the level of danger that a patient poses when they are diagnosed with lung adenocarcinoma. As a result of this improved prediction, medical professionals and researchers are able to design new cancer treatments such as chemotherapy and immunotherapy (Zhang et al., 2023). Both histological and molecular heterogeneity are characteristics of lung adenocarcinoma, which is the most prevalent subtype of non-small cell lung cancer (NSCLC).

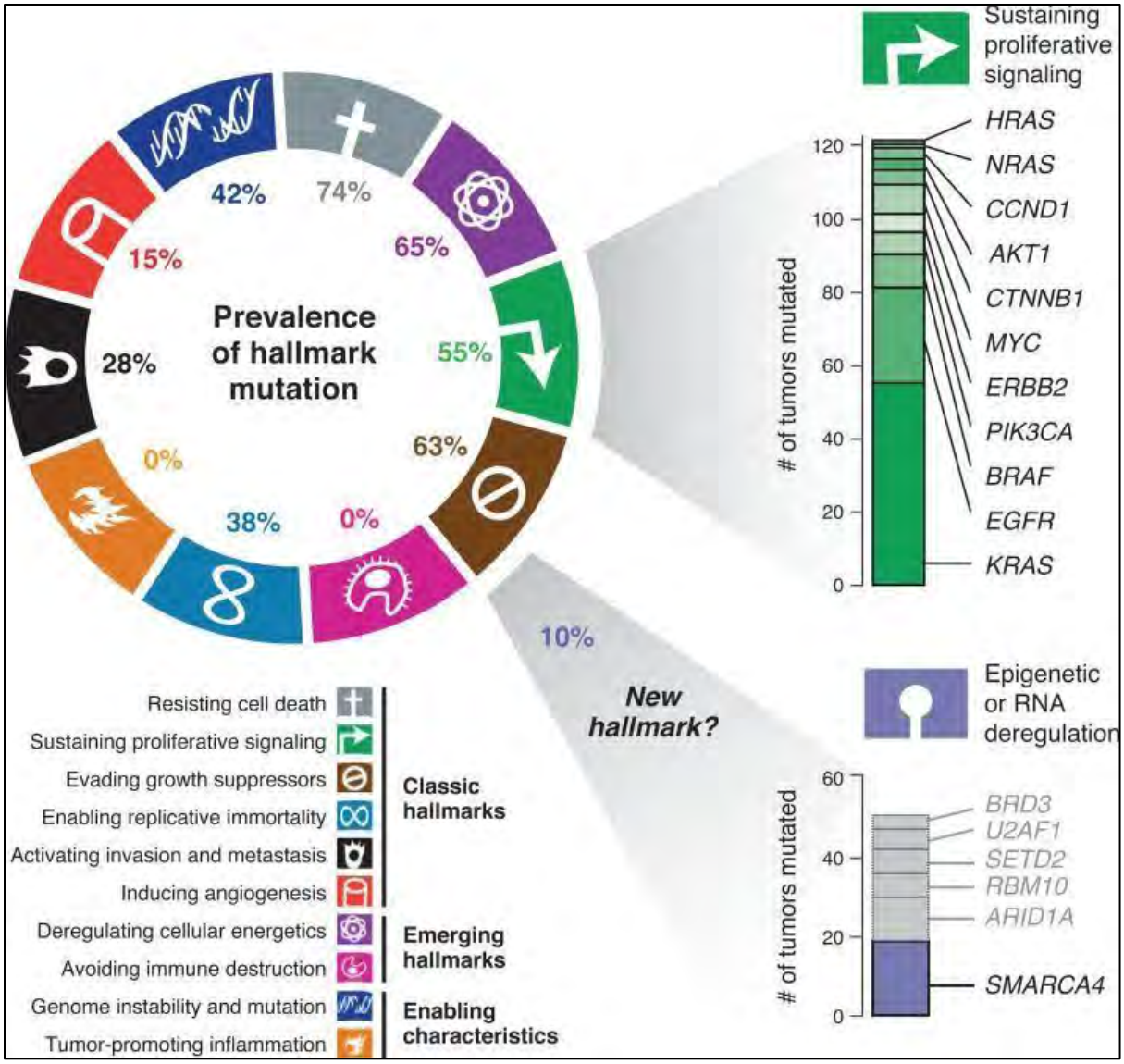

Figure 1: Hallmark of Lung Adenocarcinoma (Imielinski et al., 2013).

In targeting, the current image provides an overview of such a signature mutation in many cancers in particular focusing on lung adenocarcinoma. The proportion of cancers that present the mutations in each major cancer hallmark (figure 1) is as follows: The following are the hallmarks: proliferative signaling; the ability it demonstrated in 74 percent of the cases; insensitivity to growth suppressors where it responded in 42 percent; replicative immortality in 65 percent of the cases; and ability to induce angiogenesis in 63 percent of the cases. These mutations in lung adenocarcinoma are *KRAS, EGFR, BRAF,* and *TP53*. These mutations claimed to initiate the

beginning of the break of uncontrolled cell division, the ability to avoid the normal signals which tell a cell to cease growth, rights of division whenever they please and powers of encouraging the formation of other new blood vessels. Perhaps these bars may provide some indication of the comparative likelihood of some of these particular gene mutations for these traits relevant to these bars. This is also as applicable to a new possible attribute of focal epigenetic or RNA derangement with which the image also seems to suggest. With regard to the future prospective of the new hallmark, it is known that genes like the genes of *SMARCA4* are necessary for it. They are therefore applicable in construction of systemic focused molecular therapies and essentials of thought in the study of lung adenocarcinoma and are critical to the understanding of it.

1.3 Etiology

Smoking

lung adenocarcinoma is well known to be associated with smoking, with RR of 10 to 30 as compared with non- smokers. Second-hand smoking has been indicated to cause the death of thirty thousand people annually. Smoking of tobacco caries many chemicals which can be mutagenic. Moreover, it can be toxic and carcinogenic as well and all of this reason can be the cause of genotoxic effect. This effect occur double stranded DNA breaks and mutation of genes(Seguin et al., 2022)

Occupational risk factor

With regard to the values for relative risk, if asbestos is exposed, then the relative risk values correspond to six. This substance is also one of the most significant workplace-related lung cancer risk factors-known to man Some other workplace carcinogens include radon, arsenic, chromium,

nickel, vinyl chloride, ionizing radiation and the like. These are but a few; they may manifest themselves at the workplace as will be explained below.

Preexisting lung diseases

Factors which can raise the risk of lung adenocarcinoma are non- malignancy lung disease, idiopathic pulmonary fibrosis, tuberculosis, and chronic obstructive pulmonary disease.

Genetic polymorphism

Quantitative cancer risk for lung has been identified with the help of SNP used for knowing the gene variation where several different genetic polymorphisms are present for the same. This information was given in the paper that endeavored to find correlations between the whole genomes. They are other risk factors apart from smoking that encourage lung adenocarcinoma and these are for instance, it has been observed that the environmental substance known as radon, a colorless, odorless, radioactive gas was linked to lung cancer even in clients who were never smokers.

1.4 Treatment

Surgery

If the size of the tumor is such that it can be shaved off with a knife, then such an image results in the disconnecting of the given cancer cell by the surgeon. At times when a number of given tissues are to be reconstructed, it becomes possible to stray off normal cells; therefore, when this is done the existence of some cancer cells is ruled out. In some operations, they may be forced to remove the whole organ or some parts of it with a view of eliminating any possibility of the cancer reoccurring.

The ablation of radiofrequency

In this technique high energy radio waves with heat are applied so as to destroy cancer cells. Lung cancer is a disease classified into two types: the tumor type includes small cell lung cancer and non-small cell lung cancer, and the latter is qualified for radiofrequency ablation therapy. This kind of lung cancer is often diagnosed on the out surface of the lung or towards superolateral.

Treatment with radiation

SCLC and NSCLC also receive radiation therapy Traditional radiation therapy and concurrent chemoradiation for SCLC and NSCLC For the cancer cell, it is that the energy of the SR beam is far too high. It is also used in controlling of cancer pains with the intention to possibly reduce size of tumor cell and in many other circumstances it is given in case surgery is not feasible or is not contemplated at all.

Chemotherapy (CH)

Chemotherapy employs different categories of anticancer drugs and every one of such drugs acts the part of inhibiting the ability of cancer cells to divide, thereby causing the death of such cells. Nowadays, chemotherapy may equally be administered before surgery or after surgery. Where the tumor mass is quite big to be removed through surgery, then chemotherapy may be administered so as to shrink the tumor. The surgeon will be in a position to perform the surgery of removing the cancer cell in as much as the size of the tumor is reduced. When lung cancer is present the chemotherapy is administered intravenously. Hopefully the dosage of chemotherapy will also be modified bearing in mind the stage of cancer or the extent of malignancy.

Possibility of the applying of aimed medication

Selective therapy can be defined as the treatment that targets specific aspect of malignant cells. It is however, a rather more delicate form of treatment as compared to chemotherapy and hence comes bearing little or no harm. Chemotherapy in addition to killing a given cancer cell also kills some healthy cells while on the other hand targeted therapy kills only the cancer cell. They are in particular locked in a sort of way for the cancer cell alone. The anti-cancer drugs used to treat lung cancer include Osimertinib, crizotinib and bevacizumab to mention but a few.

Immunotherapy (IG)

The body does have cells which are capable of identifying and eliminating toxic matters; nevertheless, cancer cells excel in that they can mask such toxic matters and immune cells do not possess the means of perceiving the ill effects on the body if it is to be plagued by such cancer cells. Cancer cells do not possess such machinery hence do not destroy themselves; this is why immunotherapy ensures that the cancer cell becomes identifiable by the immune system of the body. Surgical or medical treatments that tend to make the patient's condition better in regard to some sign; operations or other procedure on which there are no theoretical grounds for hope that the patient will be improved, except possibly on the symptomatic level as defined by others. The tickle, pain and dyspnea that are associated with lung cancer are the only things that are alleviated by this type of cancer therapy technique. This kind of a treatment is offered to cancer patients who are in the terminal stages of the disease with the aim of improving palliation of conditions that arise from the illness.

Modern novel targeted therapies and immunotherapeutic agents

Both of these new categories of drugs are already being used in the primary management of lung adenocarcinoma. For instance, the molecular markers like EGFR as well as ALK have been proven to improve the survival of the patient suffering from the disorder by the use of TKIs. However, combination of immunotherapy with chemotherapy or with TKI is currently being considered as a way of dealing with the resistance mechanisms and increasing patients' lifespan

1.5 Rationale of the study

The rate of death from lung adenocarcinoma is quite high; this oncology is diagnosed most often in the world. I have decided to handle this specific type of lung cancer – based on the health incidence, the ratio of survival of people affected and the ratio of diagnosis of the disease – in this manner: Thus, lung adenocarcinoma is still a rather fatal cancer at the present stage, thus targeted therapy and new approaches are being investigated and there are a number of possibilities to investigate this malignant neoplasm. At any rate, there is debate over whether there are some issues; nonetheless, it cannot be discussed that whatever the etiology, lung adenocarcinoma primarily happens to smokers. Hence, on giving this sort of reasoning I made a decision in take Lung adenocarcinoma as my area of interest because I believed that there were greater opportunities for the subsequent researchers or pharmacists in terms of further research on the treatment and diagnosis of lung adenocarcinoma. And therefore, lung adenocarcinoma was chosen as the specific type of cancer to be investigated in the research – this kind of cancer is among the most widespread ones, has rather complex molecular tissue profile and is not very respond to treatments. In some way, Transporter of transcription activation, such as STAT3 for example, while it has at least offered a direction for targeting the therapy, using it has enhanced the disease progression for patients. Due to highly heterogeneous and plastic features of lung adenocarcinoma

which can be said to defining this malignancy as a set of challenges and opportunities that may be unrivaled in oncology in general this sub-discipline of oncology has been considered as highly important.

Chapter 2

2.1 STAT3 Pathway

STAT3 is expressed in all the cells in the body and has several functions including the development of the cells, immune response, and cell division. STAT3 is involved in the activation of other genes such genes as Bcl-2 and c-Myc; this means that STAT3 intervenes in the growth, invasiveness, and resistance to treatment of cancers. For this reason, STAT3 is one of the targets for intervention in cancer. Cancer cells are known to proliferate and researchers are working towards making drugs that will serve as potent inhibitors of the growth of such cells. This is to be attained by either antagonizing or regulating the STAT3 signal cascade. As a result, the targeting of STAT3 has emerged as a rather important issue in cancer research in an effort to enhance the treatment of the disease in the future. Few of the cancers like Lung adenocarcinoma have link with STAT3 signaling pathway (*Figure 2*) in the advancement of the malignancy. This pathway is known to be involved with several cellular proceedings. Certain of the RNA interference methods or small molecule inhibitors are still under investigation specifically for the use in lung adenocarcinoma. It is under investigation as a potential therapeutic intervention the STAT3 signaling.

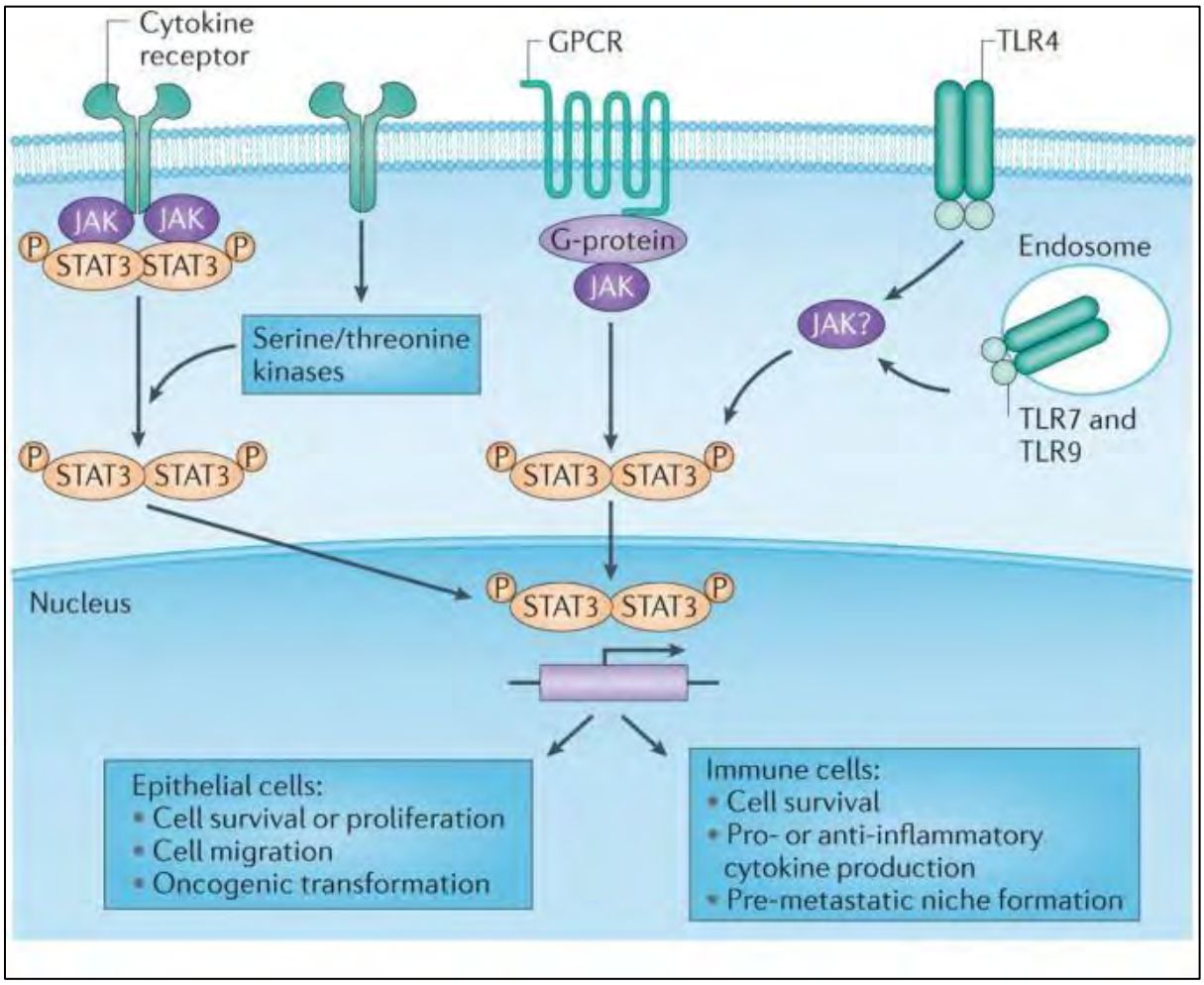

Figure 2: STAT3 Pathway in cancer. (Yu et al.,2014).

2.1.1 Mechanism of STAT3 pathway

The Janus kinases (JAKs) interfere with the activation of STAT proteins, and, when the latter are stimulated, they promote the development of malignant tumors and their metastasis. On this ground, both JAK and STAT3 are now considered as drug targets in cancer intervention. In cancer cells, activation of this JAK-STAT3 signaling also helped by IL-6 and the G-protein coupled receptor (Yu et al. ,2014). Among the phosphene interactions regulating cell activity depending on the signals it is possible to single out one of the most significant – activation of STAT3. Some of them bind themselves to receptors that are widespread on the outer membrane of the cell and this

may be the first step which occurs. For instance, the IL-6 /gp130 complex receptor is not an enzyme in its own right but this receptor will be capable of activating JAKs if the latter attaches itself to its specific ligands. These JAKs are then couple with a precise tyrosine cluster on STAT3 protein; the chain of the protein is at number 705. Nevertheless, some receptors for instance the VEGFR or the EGFR contain enzymatic domains themselves and interact with other enzymes. In addition, those receptors consequently induce the phosphorylation of STAT3 at position 705 Tyr in case their ligands occupy the corresponding sites. In like manner non-receptor tyrosine kinases like c-SRC can also phosphorylate JAKs and commence the series of events leading to STAT3 phosphorylation. When STAT3 proteins get phosphorylated, and despite the fact that the molecular mechanism of dimerization is not very clear, the SH2 domains of two STAT3 proteins can combine with each other in a process called dimerization. That is when these dimers, they relocate to the nucleus of the cell where they operate as a transcription factor. It is thought that the nuclear STAT3 dimers exhibit the aptitude to associate with designated segments of DNA which are acknowledged as promoters. These promoters are those promoters that are involved in the transcription or better still gene expression of genes. By this activation, the processes which are basic for the division and survival of cells, and other processes occurring in the cell, are initiated. However, it should be pointed out that there is a potential that this activation of STAT3 in cancer cells will serve to stimulate the cell to replicate and proliferate. When such mechanisms are understood one can form some conception of what might become targets for therapeutic intervention for diseases characterized by the dysregulation of cell proliferation such as cancers. (Lau et al. ,2019). Other signaling that has been linked to STAT3 is the PI3K/AKT and the RAS/MAPK and these have been uncovered as the laying bare of the mechanisms of this carcinogenic system. These connections have been described elucidating the challenges encountered when trying to target

STAT3 in cancer therapy and there should be extensive knowledge concerning the interaction of this molecule with other biochemical signaling pathways (Yu et al., 2014).

2.2 Relation between STAT3 pathway and lung adenocarcinoma

Lung adenocarcinoma is related to many pathways, one of which is called the STAT3; this refers to the signal transducer and activator of transcription 3. Lung adenocarcinoma is subtype of NSCLC, and the most frequently observed necrosis is of lung adenocarcinoma. When STAT3 pathway is on, the cancerous cell will divide a little faster, it will live longer and it is capable of moving to different organs of the body and cause invasiveness. There is another process known as angiogenesis whereby cancer cell grows through formation of blood vessels where by the tumor cell is supplied with oxygen foods. The revelation in the recent studies of lung cancer also indicate that the activation of the STAT3 pathway might be instrumental in making the tumor somehow immune to treatment. The activation of the route in effect thereby stifles the efficacy of chemotherapy and targeted therapy – the two major forms of cancer treatment and hence makes it hard to treat the cancer. The mechanism for this is again STAT-3 which can cause both cancer relapse and immunosuppression. This pathway makes the lung adenocarcinoma more lethal because it erases the chances of the immune system of the body to eliminate the cancerous growth. Also, the roles and affiliation of STAT3 with reference to the immunological microenvironment inside the tumor are being researched, which is predominantly focused on the development of the combination drugs intended to strengthen the immunological reaction against the tumor (Lee et al., 2022).

2.3 Objectives

Specific objectives of this work are to recognize the genetic shift, to biomarkers and precise molecular targets, the integration of STAT3 signaling, the genes of STAT3 signaling accountable for lung adenocarcinoma, comparative analysis, network analysis and gene-protein interaction, data analysis and interpretation and finally contribution made by the results to future investigation and therapies.

Chapter 3

Methodology

3.1 Data Sources

cBioPortal

cBioPortal is an internet site through which one can browse and query cancer genomics large datasets. It includes data from many sites: other working groups such as The Cancer Genome Atlas (TCGA), the International Cancer Genome Consortium (ICGC), and the like. In cBioPortal, the cancer genomics data can be visualized and explored by dynamic and essentially friendly instruments. Users can look for specific genes or types of cancer, make oncoprints, curves of patients' survival and networks of genomic changes and analyze how those have developed in the course of time across different studies. Further to this, we were also able to gather information on the cytoband of those genes alongside with the p-value, q-value and spearman's correlation.

https://www.cbioportal.org/results/coexpression?cancer_study_list=luad_tcga_pan_can_atlas_2018&Z_SCORE_THRESHOLD=2.0&RPPA_SCORE_THRESHOLD=2.0&profileFilter=mutatio

ns%2Cstructural_variants%2Cgistic&case_set_id=luad_tcga_pan_can_atlas_2018_cnaseq&gene_list=STAT3&geneset_li st=%20&tab_index=tab_visualize&Action=Submit

Harmonizome 3.0

Harmonizome 3. 0 is a whole experiment in bioinformatics, an integration of an enormous amount of information from different sources on genes and protein. All information from several databases is united and gives the complete vision of the work of genes and proteins, their interactions and the biological processes associated with them. The website allows researchers to use a set of tools to search for and study the relationship between genes, proteins and diseases and to visualize the relationships between these elements. Harmonizome 3. In terms of nominal values, national health spending as a share of GDP 0 is used to determine biomarkers, targets as well as the contributing factors to diseases.

https://maayanlab.cloud/Harmonizome/

Venny 2.1

Venny 2.1 is an online tool employed in the construction of the Venn diagrams that are useful in the presentation of the exchange between the dissimilar and the similar information. Using Venny 2.1 Therefore, one of the most obvious applications is that one list is compared with another list of genes, proteins or other biological molecules. It helps the datasets to clearly identify between what part they share and what makes them divergent. Users can input up to four lists and then by having the lists arranged in this manner the users can quickly identify which of the things in the intersection list belongs to and which do not belong to each list entered. Therefore, this type of visualization is useful for getting the idea of the relative similarities and differences between two

sets of such data by identifying features generic for both as well as those may be of further interest in genomics, proteomics, and other biology-related sciences.

https://bioinfogp.cnb.csic.es/tools/venny/

KEGG PATHWAY Database

KEGG or Kyoto Encyclopedia of Genes and Genomes is a very large description of the databases of the networks and reactions of the molecule in biological processes. It has the neatly labelled metabolic map, genetic signal transduction, signal transduction of the environment stimuli, cellular functioning, life hierarchy and mechanism, and people's ailments. With this technology, we are able to analyze close to 23 pathways and the genes associated with it, as described in the list below. Based on venny2. 1, we proceeded to exclude some of the enumerated so called 'pathways linked genes' with STATE3 pathways gene of lung adenocarcinoma. From this, we derived the second of our filtering genes.

https://www.genome.jp/kegg/pathway.html

Expression Atlas

Expression Atlas is a huge database which maps genes and proteins and their experience at any moment, in any organ or cell type and during embryonic and fetal stages. These data are obtained from either baseline or differential expression level data with respect to disease or drug. The Expression Atlas can be used to find out when and where genes are expressed, alteration of expression in different conditions and biological use of knowledge in association with the data, for instance, biomarkers identification or therapy targets. With this we were in a position to undertake the differential expression of the filtered genes as we are about to note henceforth. Thus, evaluating

up to here, it was possible to get the Id of the ensemble, Log2 FC value and the adjusted p value of the genes.

https://www.ebi.ac.uk/gxa/home

ShinnyGO 0.77

ShinyGO 0 is a web-based application developed and host on SHINY APPS banner. of the type that is implemented by the Shiny package of R. 77 was designed for GO enrichment analysis. In this way, one can submit a list of genes of interest and get fine-grained functional descriptions together with enrichment analysis. These assists them in identifying the GO terms and the overrepresented biological pathway in their data. The tool can also perform the pathway analysis if using the data from several, for example, KEGG, Reactome and others. Except for the advantage that is convenience, with the help of ShinyGO, it is possible to consider such features as functions of genes and their relationships.

http://bioinformatics.sdstate.edu/go77/

Gene MANIA

As of now, GeneMANIA is a very flexible and simple tool for inferring gene and gene set function from interactions. Of special importance is a vast set of genetic connections, pathways, and co-expression and co-localization and the homology of protein domains exists. For instance, in the GeneMANIA, the user enters the list of genes and the program will eventually spend time and demonstrate the connections between the multiple genes and how they work. This tool is particularly useful when trying to pinpoint the function of a given gene, or when identifying other members of a particular pathway or complex or indeed, what a given gene is up to. As it is currently

designed it includes live graphics and the researchers can look and understand what the given gene sets are biologically.

https://genemania.org/

SNPs3D

SNPs3D is statistical program that superimposes SNPs on 3D structures of proteins to ascertain effects of polymorphism. It includes SNP annotation, SNP annotation is necessary to predict how such genetic differences impact on 3D protein structures, stability and interactions, SNPs3D can guess when SNP locations are expressed as protein structures. The outcome, positive or negative, of the SNs being implemented can also be understood with the help of images that are available to the researchers. It is also very useful in relating differences in genes to molecular consequence. It gives us a worthy knowledge for surveying diseases, NTICWE and prognosis of drug reactions and for the individual therapeutic formations.

Kaplan-Meier plotter

Kaplan-Meier Plotter is an online resource that gives an insight into the role of genes in influencing existence in different types of cancer. Large databases TCGA, gene expression information, Kaplan-Meier survival curves are obtained from clinical information. Indeed, the tool permits the user to enter one or more genes and to choose the type of cancer they want to have information about. This then yields a survival plot that shows on the basis of gene expression value how patient survival is related. This helps the researcher to establish biomarkers for prognosis and also to determine how the levels of a particular gene affect prognosis of cancer. This makes such things as personalized medicine and the concept of treating the patient rather than the disease much more employable in the medical practice.

https://kmplot.com/analysis/

3.2 Literature Search

PubMed

To date, PubMed is an exclusively web-based searching tool that can search within the MEDLINE to and provide references and papers on biomedical as well as on the life science fields By it, we can find such things as, for example, an option for an advanced search for articles and filters to find articles that are scholarly and peer-reviewed. It thus becomes a very useful tool in searching for current data that may be useful to researchers, clinicians or student in many areas of health and biomedical science.

https://pubmed.ncbi.nlm.nih.gov/

Google scholar

It is a web search engine that allows for the searching of the full text or full metadata of the scholarly content of any discipline and type of publications. It has papers, theses, books, preprints, abstracts, technical reports and other scholarly works which have been reviewed by other scholars. These are obtainable from academic publishers, professional associations, universities and other scholarly societies. Google Scholar allows users to search or specific articles or authors or specific publication.

https://scholar.google.com/

Scopus

Scopus originated in Norway and is one of the biggest indexes of abstracts and citations in a cut number of fields. They cover subjects in arts and humanities; sciences, technology, medicine and social sciences; business administration. Is subscribed with Elsevier; processes content from over 24000 peer reviewed journals; over 5000 publishers globally. It contains features for monitoring, analyzing and representing the results of the Studies it contains.

https://www.scopus.com/standard/marketing.uri

ResearchGate

ResearchGate is a web service based on various features for scientists and researchers, and the popularity intention of the audience of working scientists. The registered users can upload documents and other study materials, post and reply to messages in the forum and freely communicate with other users from all over the planet. It enables researchers to: – upload and post the work they have done – surf through the work of other researchers who have published their work on the site – handle other matters concerning science. to which it has permeated the said domain.

https3.3://www.researchgate.net/search

3.3 Data Collection Method

To begin with, we employed several tools for the purpose of searching for data, for example, STAT3 related genes. To this end, we utilized cBioPortal in performing the analyses that are mentioned next. Through cBioPortal we have seen the connected genes of lung cancer. More so, the reports included cytoband of the genes, p, q, and spearman's correlation factors. Then, we use

Harmonizome 3 thereafter. 0, required as a starting step to determine the gene set that has been mentioned to be associated with the STAT3 pathway. Venney2. 1 is the sieve that we use to filter the genes in cBioPortal and Harmonizome 3 that are available to be investigated. 0. The tumor genes that we have identified from the cBioPortal are as follows: List 1 – Tumor genes table of this tool. Following that, the genes that were identified by Harmonizome 3 are listed below the genes that have been previously included. This will mean that if the parameter is set at 0 to list 2 then it will compare the two sets of genes to be list 2. In other words, it can filter out the overlapping regions, that is,' the genes that are present in both. These are the first genes that we have knocked out, and among all, all interact with the STAT3 pathway. Further, they are linked to lung adenocarcinoma. The fourth type of data was obtained from the KEGG PATHWAY database: these pathways were examined to assess the combined information from the metabolic and signaling pathways. According to KEGG, we carry out an assessment of 23 pathways and then we check for similarity between the contained genes of the pathways and the filtered genes. Again, venny2. 1 only the authorized entity known as 1 is permitted to conduct this screening. We, at last, were given our genes that were screened 22 seconds earlier. These filtered genes for differential expression were taken for further analysis on Expression Atlas and we recorded the ensemble id, Log2 FC and the adjusted P value as follows; As part of this project, we also used ShinnyGo 0. had to change the selected genes and using this method we obtained FDR, number of genes, pathway and fold enhancement data. Co-expression and route analysis enable us to reveal the relationship that Gene MANIA has developed in the interaction network.

Chapter 4

Primary Data

4.1 STAT3 Corelated Genes

In the real language of the researcher Stat3 correlated genes are genes that have up increased activity when interacting with Stat3/ or the genes whose generation is regulated by the Stat3 protein. It may directly regulate them or do something else to the level of some genes' expression. Thus, when one wish to regulate some certain genes which we refer to as target genes by Stat3, this is done through Stat3 response elements; regions of DNA that are specific sequences of DNA located in the promoter region of the target genes. That, in essence, is what makes this binding functional – its capability to interact with the target genes in a manner that should probably affect their transcriptions. We could, therefore, perhaps for the first time, by the use of BioPortal, reveal the genes that are linked to STAT3 – the marker of lung cancer. For the Harmonizome 3.0 here we have also presented new genes associated with the STAT3 signaling pathway.

4.2 Common genes between STAT3 and Other pathways

The situation remains quite similar in analysis of lung cancer, and it can be stated that the level of convergence and cross-talk between genes of different pathways is possible to describe as high, for instance, in relation to the STAT3 pathway and other pathways, genes from which seem to interact with genes of the STAT3 pathway. This also confirms that these pathways are somehow related, complementing and cooperating with each other and with other pathways in lieu of the rest in favor of tumor promotion and survival. Aurin such interactions demonstrate the potential application of the so-called 'polypathway' treatment in which several pathways are, in essence,

switched off to promote cancer therapy. This is due to the fact that the genes involved in both STAT 3 and other pathways are the most frequently involved genes in the overlapping duties that concern cell living and division. As such, this connectivity opens thematic probabilities of attenuating critical signaling nodes across these tracks.

4.2.1 Parameters

Positive or negative correlation

A positive correlation refers to a relationship between two or more genes where an increase in the expression of one gene is associated with an increase in the expression of the other genes. On the other hand negative coefficient mean that as the gene expression level of the first gene rises, the gene expression level of the other declines. This lends support to the hypothesis that the two genes could be having different roles in physiological functions or pathways. Such knowledge is for this number of reasons useful, to mention but the molecular etiology of disease, advanced drug design with specific reference to the individual and how genes interact.

Spearman correlation

The technique is an ordinal level non-parametric statistic used to determine, the strength and direction of linear relationship between two ranked variables such as extent of gene expression and cancer progression. As a result, Spearman coefficient (Table 1) seems to be the most appropriate in non-linear and ordinal data analysis of the complicated biological contexts in contrast to Pearson's correlation. This is so because it employs ranks rather than data values; it integrates all the different values and provides a rank of these rather than its value. For example, there are studies where one investigates, for example, whether elevated levels of specific gene are indicative of is or investigate relations between the genes and patients' outcomes of treatment. If

the data does not appear to be normally distributed or if its distribution is ignored because its outliers are considered anomalous, then the reliability of the results won't be very high; this method can be quite beneficial in such situations p-value. P-value stands for 'probability value', which represents the probability that the observed level of association between the gene expressions just occurred by chance. Where the outcome of analysis leads to a low p-value and in most cases equal to or lesser than 0. 05 while the others are less than 005 which is significant as it is less than 05 or . The hypotheses retain as a result the association is statistically significant.

q-value

The control of the FDR is done by use of the q-value which is a statistic that is used in the detection of significant relationships between genes in different comparisons. It does this by modifying the p-values that are output by tests of Gene Co-expression. This is done by use of dependent samples which reduces the number of subjects and the influence of the multiple comparisons on increase of type I error. This implies that since a lower q-value points that it has a greater likelihood that the co–expression association between genes is not a false positive discovery, another useful measure of the result is the q-value (Table 1)

Table 1 Common genes between STAT3 and other pathways.

SL NO.	Common Genes	Pathways	Correlation	Spearmen's correlation	p-value	q-value
1	HRAS	Ras, Rap1, VEGF, MAPK	Negatively	-0.15	0.032	0.32
2	SMAD2	Ras, TGF-beta, VEGF, PI3K	Positively	0.29	0.005	0.05
3	IQGAP1	Rap1, VEGF	Positively	0.12	0.04	0.32
4	NCL	ErbB, VEGF	Positively	0.18	0.017	0.17
5	SMAD1	TGF-beta, PI3K	Positively	0.33	0.009	0.09
6	RANBP3	TGF-beta, PI3K	Negatively	-0.07	0.028	0.28
7	NFKB1	Hedgehog, VEGF, TNF, HIF1, FoxO, MAPK	Positively	0.25	0.004	0.04
8	NFKB2	TNF KAPPAB, MAPK	Positively	0.22	0.023	0.23
9	KHSRP	TNF	Negatively	-0.1	0.055	0.41
10	EP300	HIF1, PI3K, AMPK	Positively	0.17	0.037	0.3
11	ARFGEF2	cAMP	Positively	0.21	0.014	0.14
12	CKAP5	cAMP	Negatively	-0.14	0.045	0.34
13	ARFGEF3	cAMP	Negatively	-0.12	0.061	0.42
14	CKAP6	cAMP	Positively	0.19	0.021	0.21
15	EIF4G1	PI3K, MTOR	Positively	0.15	0.029	0.28
16	GBF1	AMPK	Negatively	-0.08	0.04	0.32
17	TAOK1	MAPK	Negatively	-0.05	0.065	0.43
18	TAOK2	MAPK	Positively	0.12	0.05	0.36

Chapter 5

Results

5.1 Co-Expressed Genes

Co-expressed genes may belong to the same biochemical pathway in the cell or may have a presumed functional relationship. Some of the genes are found to be co-expressed and co-

expression data (Table 2) are important in understanding the functional associations and interactions and in studying the pathways.

Table 2 Co-expressed genes of STAT3 pathway.

SL NO.	Genes	Pathways	Correlation	Spearmen's correlation	p-value	q-value	Pearsons's correlation coefficient
1	HRAS	Ras, Rap1, VEGF, MAPK	Negatively	-0.15	0.32	0.32	-0.12
2	SMAD2	Ras, TGF-beta, VEGF, PI3K	Positively	0.29	0.05	0.05	0.31
3	IQGAP1	Rap1, VEGF	Positively	0.12	0.32	0.32	0.1
4	NCL	ErbB, VEGF	Positively	0.18	0.17	0.17	0.2
5	SMAD1	TGF-beta, PI3K	Positively	0.33	0.09	0.09	0.35
6	RANBP3	TGF-beta, PI3K	Negatively	-0.07	0.28	0.28	-0.09
7	NFKB1	Hedgehog, VEGF, TNF, HIF1, FoxO, MAPK	Positively	0.25	0.04	0.04	0.27
8	NFKB2	TNF KAPPA B, MAPK	Positively	0.22	0.23	0.23	0.24
9	KHSRP	TNF	Negatively	-0.1	0.41	0.41	-0.08
10	EP300	HIF1, PI3K, AMPK	Positively	0.17	0.3	0.3	0.15
11	ARFGEF2	cAMP	Positively	0.21	0.14	0.14	0.22
12	CKAP5	cAMP	Negatively	-0.14	0.34	0.34	-0.13
13	ARFGEF3	cAMP	Negatively	-0.12	0.42	0.42	-0.11
14	CKAP6	cAMP	Positively	0.19	0.21	0.21	0.18
15	EIF4G1	PI3K, MTOR	Positively	0.15	0.029	0.28	0.13
16	GBF1	AMPK	Negatively	-0.08	0.04	0.32	-0.1
17	TAOK1	MAPK	Negatively	-0.05	0.065	0.43	-0.07
18	TAOK2	MAPK	Positively	0.12	0.05	0.36	0.14

5.2 Differential Expression Analysis

Differential expression analysis (Table 3) is a way of trying to work out how many statistical differences exist between genes which display notably different expression levels in one or more conditions for instance cancerous and non-cancerous tissues. Therefore, the analysis of differential expression is of extraordinary significance in lung cancer and is employed to find out which genes change their expression in cancerous and healthier tissues. This helps in the discovery of biomarkers that can be used in diagnosing, predicting and certainly identifying targets for treating the illness. To this end, this study adds its voice in indicating what genes are turned on or off in cancer in the knowledge that diseases exist and the intention of discovering disease-specific pathways in order to come up with treatment that boosts the efficiency of treatments.

Table 3 Differential expression analysis.

SL NO	Genes	Ensemble ID	Correlation	Spearmen's correlation	Pearsons's correlation coefficient	p-value	Log2FC Value	P adjusted value
1	HRAS	ENSG00000174775	Negatively	-0.15	-0.12	0.032	-1.3	2.3298×10^{-6}
2	SMAD2	ENSG00000049759	Positively	0.29	0.31	0.005	-2.5	3.8637×10^{-17}
3	IQGAP1	ENSG00000140575	Positively	0.12	0.1	0.04	-1.3	1.5706×10^{-3}
4	NCL	ENSG00000203697	Positively	0.18	0.2	0.017	1.4	1.8712×10^{-3}
5	NFKB1	ENSG00000109320	Positively	0.25	0.27	0.004	1.4	1.3616×10^{-5}
6	NFKB2	ENSG00000077150	Positively	0.22	0.24	0.023	1.7	1.1085×10^{-5}
7	KHSRP	ENSG00000088247	Negatively	-0.1	-0.08	0.055	2.2	2.312×10^{-9}
8	EP300	ENSG00000100393	Positively	0.17	0.15	0.037	1.1	3.9118×10^{-3}
9	ARFGEF2	ENSG00000124198	Positively	0.21	0.22	0.014	1.6	3.3367×10^{-4}
10	CKAP5	ENSG00000175216	Negatively	-0.14	-0.13	0.045	1.1	1.9016×10^{-5}
11	ARFGEF3	ENSG00000112379	Negatively	-0.12	-0.11	0.061	-1.2	3.2832×10^{-9}
12	EIF4G1		Positively	0.15	0.13	0.029		
13	GBF1	ENSG00000107862	Negatively	-0.08	-0.1	0.04	1.7	3.2117×10^{-6}
14	TAOK1	ENSG00000160551	Negatively	-0.05	-0.07	0.065	-6.9	5.8215×10^{-10}

Log2FC Value

This is known as Signal Silhouette which defines the directionality of the observed signal, while Log2FC is the degree of change of the amounts of gene or protein from one state to a different state. It is defined in terms of log to the base of 2 of the degree of presence of a gene in one sample relative to another. When the number is greater than zero it is an indication that that particular expression has increased whereas when the number is below zero it indicates that it has reduced; a number being equal to zero indicates that it has not changed. With the aid of this scale, one is in a position to identify how to contextualize differences of expression.

P adjusted value

Another reason the adjusted p-value cuts down the number of false positives when trying to test a theory is because it is brought in in order to deal with the multiple comparisons. It regulates the p-values to such a degree in a way that they will not be less than the intended level of significance.

5.3 Enrichment Analysis

The enrichment analysis is used to find out whether, in a given disease or condition, specific groups of genes or pathways are overrepresented. It can be used to **analyze** which biological procedures or molecular complexes are more active in the condition that is under investigation. The reason why enrichment analysis is useful in lung cancer is that it shows that there are some of the biological process pathways, which are up- or downregulated among the differentially expressed genes. It helps the researchers come across the ways that cancer spreads, biomarkers for its detection or identification and treatment options.

5.3.1 FDR (False Discovery Rate)

This is the number of all the false positives that we have expected to obtain on the list of the pathways or the categories which were seen to be significantly enriched in the data. This is especially useful should you be undertaking many statistical tests because it helps in putting a ceiling on as many wrong signals as possible.

5.3.2 Fold Enrichment

Fold enrichment analysis (Table 4) is usually measuring the folded times, in other words the number of times a specific pathway or category is found to be overlaid in the gene list of interest relative to another set. This shows the level of association of pathway with the condition under study or being investigated on.

Table 4 Enrichment analysis of STAT3 pathway genes

SL NO.	FDR	n Genes	Pathway Genes	Fold enrichment	Pathways
1	8.80E-03	2	57	61.5	Legionellosis
2	8.80E-03	2	65	54	Inflammatory bowel disease
3	9.20E-03	2	71	49.4	Adherens junction
4	9.40E-03	2	76	46.1	Pancreatic cancer
5	1.00E-02	2	93	37.7	TGF-beta signaling pathway
6	1.00E-02	2	97	36.2	Prostate cancer
7	1.00E-02	2	100	35.1	AGE-RAGE signaling pathway in diabetic complications
8	1.00E-02	2	101	34.7	Chagas disease
9	1.00E-02	2	104	33.7	NF-kappa B signaling pathway
10	1.00E-02	2	104	33.7	C-type lectin receptor signaling pathway
11	1.00E-02	2	108	32.5	Th17 cell differentiation
12	1.00E-02	2	109	32.2	HIF-1 signaling pathway
13	5.90E-04	4	222	31.6	Human T-cell leukemia virus 1 infection
14	1.20E-02	2	126	27.8	Cell cycle
15	1.20E-02	2	126	27.8	Osteoclast differentiation
16	1.20E-02	2	129	27.2	Relaxin signaling pathway
17	6.30E-03	3	203	25.9	Viral carcinogenesis
18	8.80E-03	3	252	20.9	Endocytosis
19	8.80E-03	3	294	17.9	MAPK signaling pathway
20	6.30E-03	4	530	13.2	Pathways in cancer

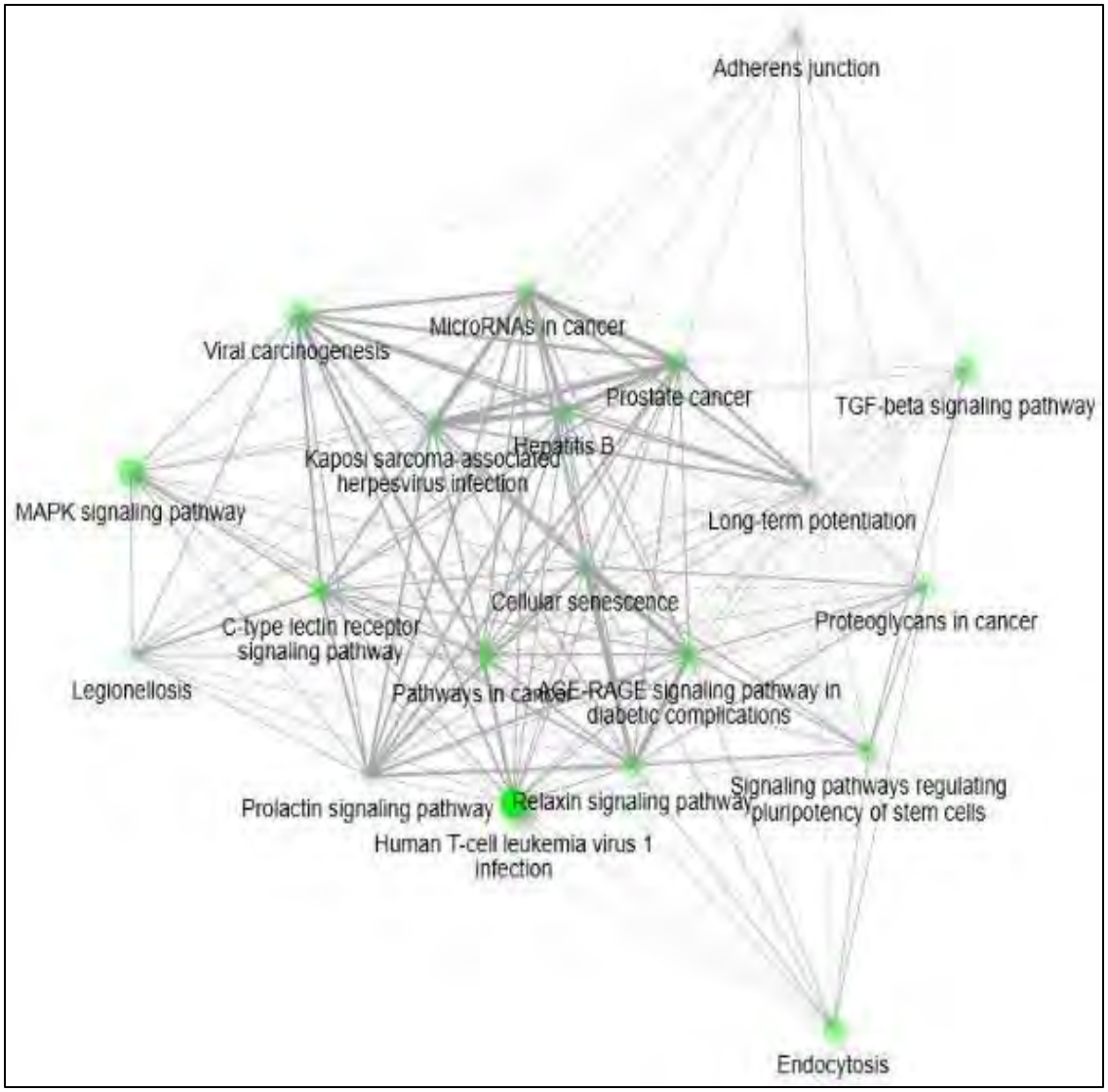

Figure 3: Network diagram showing TGF-β corelated pathways.

The picture depicts a network of illnesses; and the body's biological processes, acknowledged as a complex system. Each green node is a separate disease or pathway such as the 'MAPK signaling pathway', 'TGF-beta signaling pathway', or 'prostate cancer'. The linking (edges) shows which routes are connected to these diseases, and in return how those are affected by the diseases. Among those molecular pathways which are involved in the map are MAPK signaling and TGF-beta signaling and since they are in some ways related to many diseases they result in many diseases. That is why the central nodes, such as cellular senescence and cancer-related pathways, contain information about these topics being worked on and associated with numerous subjects. Which goes to explain how vital they are in the process of searching about diseases and possibly, treatment. Over everything, the diagram gives an obvious picture of how closely related everything in the biological world and pathological conditions are. This makes it even more important to study this connection in a bid to proffer good solutions. In particular when it comes to lung cancer, the picture is useful. It expresses the important route most especially the MAPK and TGF-beta signaling route using the STAT3 signaling route. In lung cancer, STAT is therefore considered to be an important factor which supports the survival, growth and migration of tumor cells. Which means that the cellular senescence, and the other for it, such as of leading to cancer, are essential and closely related to the other pathways. For this reason, it is important to know about its functioning and notably the STAT 3 pathway with the view of embarking on personalized therapy for the treatment of lung adenocarcinoma. This network map also portrays the aspect of biology of these pathways and the realization that more has to be discovered on how these pathways link up if there is improvement on the results of the treatment to be made.

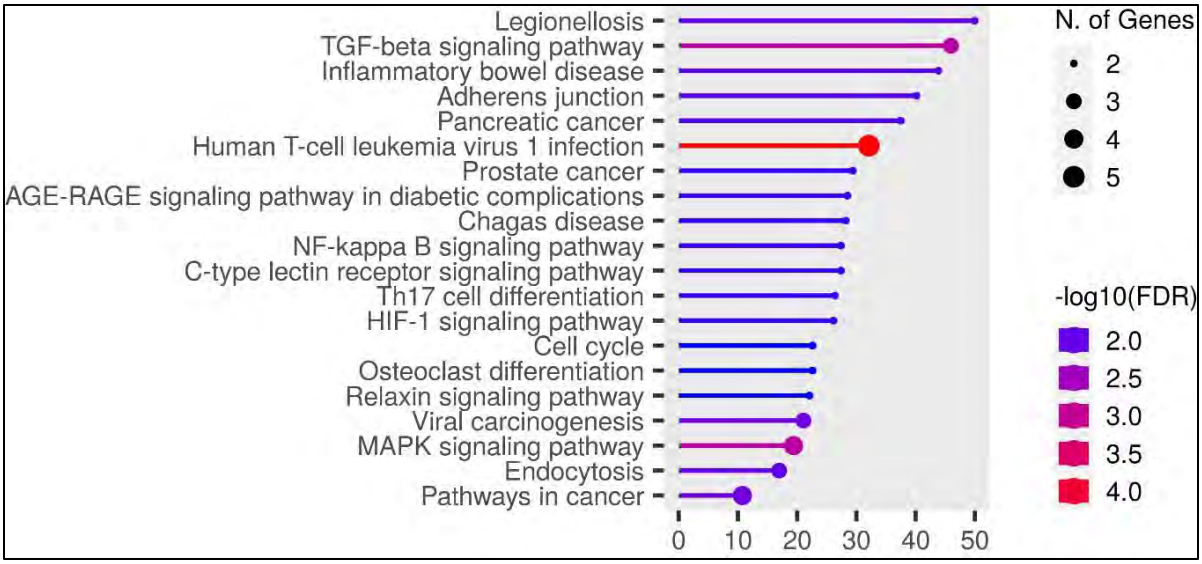

Figure 4: Fold Enrichment

It exists in the form of the bar graph which shows the names of various biological pathways and diseases and it gives information about to what extend these sets are ‘‘enriched’ for the particular assortment. Fold enrichment value is drawn as bar graph and for each and every pathway and disease which has been studied there exist a single bar figure related to it. Numbers of swipes and grams; touch time; For these – fewer equals more statistically significant and therefore a darker red is given to the bar (figure 4). Regarding the numbers of genes, these are presented by the sizes of the dots located in the extremities of the bars presented in each of the pathways. Big dots mean principally that in this area we have very many genes. Some routes are complex and contact with heavy; as the Legionellosis and TGF-beta signaling routes and Inflammatory Bowel Disease and so on. Similarly, other pathways such as the MAPK signaling system, prostate cancer, pancreatic cancer recorded sub percentile and whole percentile levels of enhancement, explained how vitally important they were to the probe. Regarding the paths connected with STAT3 activity and lung adenocarcinoma, the following are among the paths that have been identified. Of these, the TGF-beta signaling pathway and the MAPK signaling pathway are few of the significant ones. These two pathways are exerting effect on STAT3 and of great importance in lung adenocarcinoma due

to the survival, proliferation and metastasis they grant. The same graph thus defines a big category path ways in cancer that in its own turn has a number of signaling path ways to do with cancer. As STAT3 is taken to be involved in various cancers, it with a very high certainty will be classified here. And, for enhancement to this revelation it would only be proper to point out that especially, the STAT3 pathway as well as all the other signaling pathways are mandatory for the development of lung cancer. It is for this reason that we are informed on the enlargement of these pathways and the complex of STAT3 with them with an intention to seek for the treatment which in turn improves the therapy of lung adenocarcinoma.

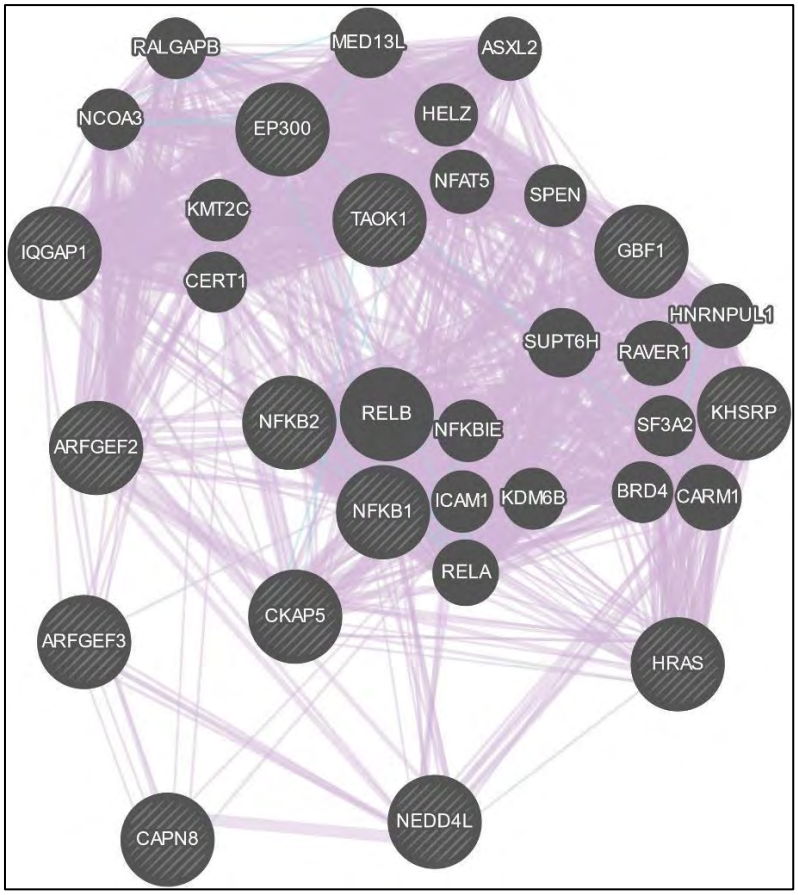

Figure 5: Network diagram based on co-expression and pathway.

This is in the form of a Network and the manner in which the genes and the proteins have been connected to show the researcher how they are related. These include nodes which represent the

genes and proteins and the edges connecting the two which we also term as links. Each and every node is a gene or protein that is connected in the profound metabolic activities that occur in an organism's body. These links show how these parts are connected. The nodes that are located more or less in the center of the screen such as *RELB, NFKB1* and *RELA* are in fact involved in regulatory pathways such as the NF-kappa B signaling pathway which is involved in inflammation and immune response. Certain nodes, namely *EP300* and *KMT2C,* act to either load or unload RNA-pol II, or to modulate the structure of chromatin. Some of them are associated with pathways which regulate cell proliferation and differentiation- for example *HRAS*. This is the layout in which nodes are unbelievably compact and there are many links which will show how tangled the biological relations are. Here, both very complex interactions of these genes and proteins can be seen thus the title of the image, network map. This is especially valuable regarding the explanation of how the cells work and the search for the treatment in such diseases as cancer.

5.4 SNPs Analysis

At the individual level the variation features the DNA sequences and these are studied under the Single nucleotide polymorphism (SNP). This unit also has coding and non-coding regions within a given genome. SNPs may change genes by some form of regulation and may also lead to changes in the function of a protein. Since SNPs also informs where diseases are, how genetic variation manifests in a gene, and where the genes that are associated with specific traits are, it is much easier to isolate disease genes from SNPs. SNP research is of immense importance in the field of genomics, pharmacogenomics, and evolution since it provides a hint of how disease, drug metabolism, and the distribution of population is aligned to genes. Such research could help finding potential candidates for disease genes that might bolster disease prevention as well as augment the types of therapies that are being developed.

Table 5 SNPs Analysis

Genes	SNPs Variants	SNPs ID	SVM Profile
HRAS	Q61K	28933406	-0.71
	Q61K	28933406	0.06
IQGAP1	S256A	12324924	0.97
NFKB1	R579K	4648086	
	D608H	11556612	
	H712Q	4648099	
NFKB2	I213T	3188993	-0.58
	R290L	11550592	-1.88
	G351R	45580031	1.32
	A392G	11574848	0.49
KHSRP	P532T	11546542	0.38
	A603T	13343528	1.88
EP300	M289V	45528742	0.93
	I997V	20551	1.89
	P1222H	7285319	-0.27
	T2174S	5758252	1.70
	P2221Q	28937578	1.50
	Q2223P	1046088	1.89
ARFGEF2	I432V	41296207	-0.12
	A527V	6063343	-0.45
EIF4G1	R5H	34838305	-2.92
	Y115C	16858632	-0.03
	M236V	2178403	1.59
	S905L	11559212	0.51
	R1020H	34086109	-0.43
	A1033P	35629949	2.21
	L1037P	2230570	1.95
	A1145P	11559216	-1.47
TAOK1	A855T	34151057	0.02

The following table lists out genes which contain specific single nucleotide polymorphism (SNP) variations as well as the SNP numbers and/or SVM profile numbers for these genes: These scores show how these SNPs are likely to affect gene function, especially in lung cancer. For example, *HRAS* gene with Q61K variant (Clin var rs28933406) has SVM of -0. 71 that is, it may do harm to have a beneficial effect, which on the one hand promotes the growth and on the other preserves survival of cancer cells. *NFKB1* and *NFKB2* genes are of huge importance to various mechanisms of the immune system as well as inflammation within the body. It contains many SNPs, for instance, R579K, D608H, and I213T; furthermore, the SVM scores move from - 1 to 1. 88 to 1. 32. These variations may in some ways change the function of genes, and thus the development and progression of lung adenocarcinoma tumor may be affected by changes. It functions in control of transcription as well as the reengineering of chromatin. It has SNPs; M289V and T2174S which have scores that range from -0. 27 to 1. 89 to demonstrate that it acts otherwise on gene expression and tumor development. Understanding these genetic differences and how they might manifest in persons is imperative to understanding how lung adenocarcinoma functions at the molecular level, to identify biomarkers for the earliest possible diagnosis of the disease, and to erect individualized survivorship plans to improve patient's outcomes.

5.5 Survival analysis

Survival analysis is thus a statistical technique in the field of biostatistics to create a 'glimpse' of how long it is going to take for a certain event to happen for instance death, relapse or recovery. One of the functions includes the survival function, of probability of a person being alive after a given time in the calendar while the other is hazard function which defines the dangerous of occurrence of an event at the current time. Two methods used frequently in survival analysis that are used with survival data are the Kaplan Meier, which estimates the survival probability and

the Cox proportional hazard which estimates the hazard ratio. The method can also deal with censored data: an event in a study did not take place in some of the study population throughout the study duration. In his opinion, all in all, the utility of the concept of survival analysis cannot be overestimated if the time-to-event data have to be analyzed on the terrain of the social sciences, engineering or in the medical study. They help to define some possible risk indicators and to evaluate effectiveness of a treatment.

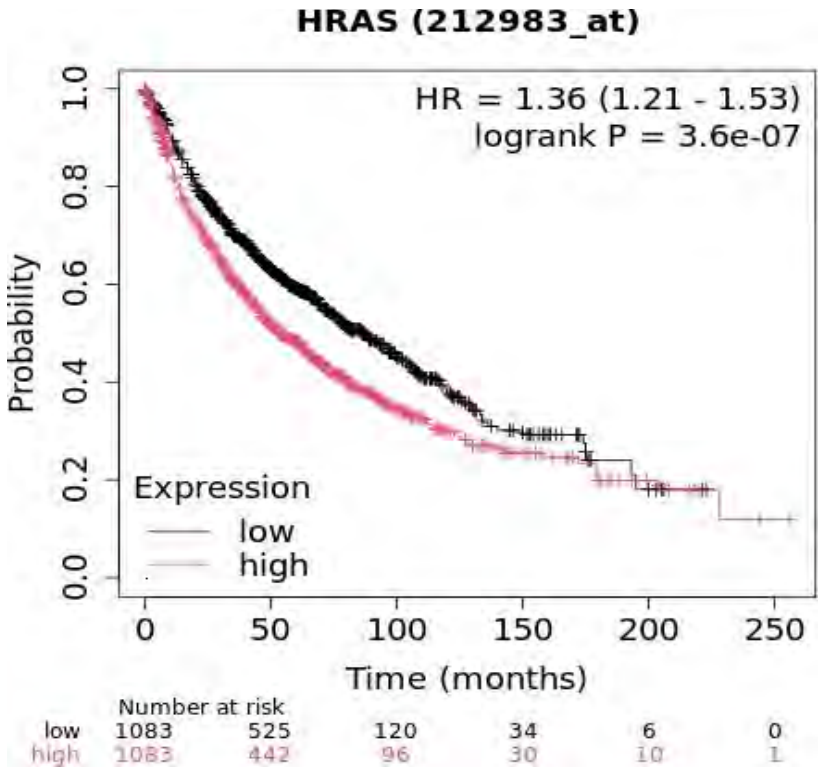

Figure 6: Survival plot of HRAS.

Survival curves using Kaplan-Meier estimation for lung adenocarcinoma reveal effect of gene HRAS on survival of patient. The plot shows two groups that tumor cells with low HRAS expression (black line) grew more slowly than tumor cells with high HRAS expression (pink line). Consistent with the results observed in the first set of analyses, patients with low HRAS-status have a substantially higher survival probability over time when compared to patients with high HRAS expression, thus supporting our hypothesis that high HRAS is linked to poor survival

prognosis. For HRAS, high expression patients have 1.36 times higher hazard rate (95%CI = 1.26–1.48) and higher mortality rate than the low expression patients. This difference is significant with the log rank p-value of 0.0000036. Therefore, HRAS is a potential candidate gene for targeted therapy in lung adenocarcinoma, because HRAS downregulation might be associated with more favorable survival prognosis in patients with high HRAS expression.

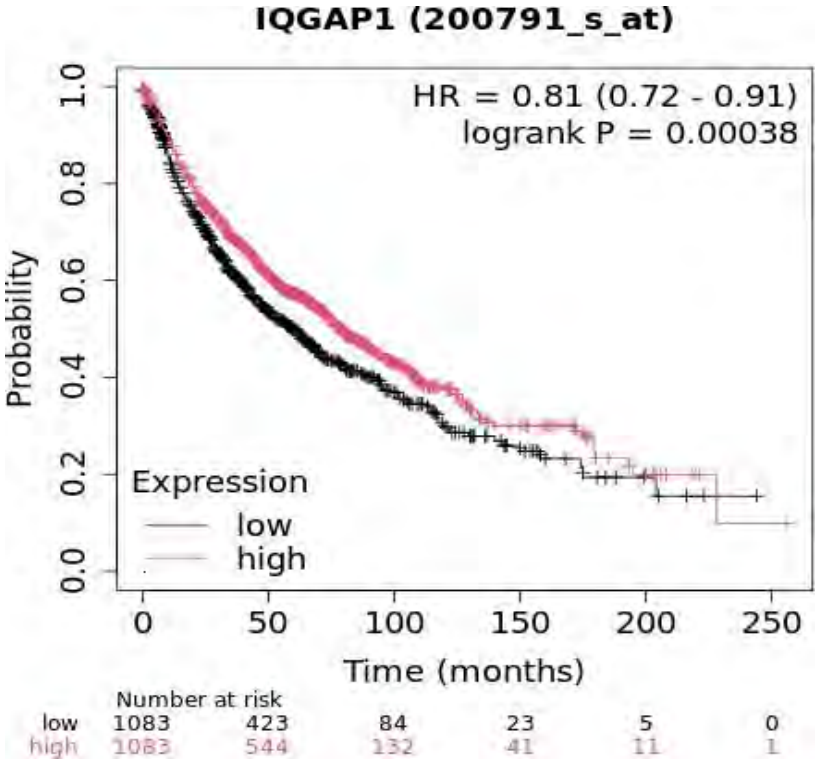

Figure 7: Survival plot of IQGAP1.

The graph proves that the likely survival rate of patients with high IQGAP1 expression probability is higher than that of patients with low IQGAP1 expression probability. In light of this, one gets the impression that high levels of IQGAP1 signify improved survival in lung adenocarcinoma. The above data also demonstrate that patients with high IQGAP1 expression have a 19% reduction in the risk of death as compared with those with low SQ expression (HR = 0.81) that supports the protective nature of IQGAP1. This survival difference is also revealed by a statistically significant

p-value on the logrank P = 0.00038. Therefore, IQGAP1 could be viewed as having a tumor suppressive role in lung adenocarcinoma and might be used for predicting favorable prognosis.

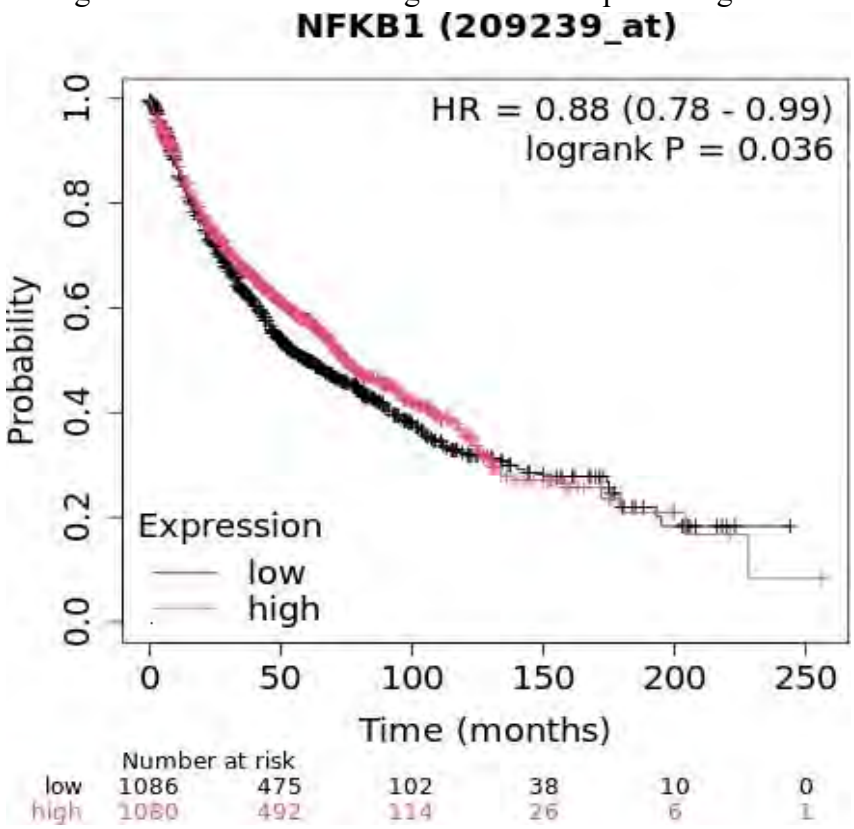

Figure 8: Survival plot of NFKB1.

The Kaplan-Meier survival plot for lung adenocarcinoma, associated with the NFKB1 gene, reveals that patients with high expression of NFKB1 have slightly better survival rates than patients with low expression. The results regarding risk associated with the death of patients, the high expression of NFKB1 decreases the hazard ratio (HR = 0.88) translating to a 12% decreased risk thus making it protective. The above results were statistically proved in this study as indicated by the logrank P = 0.036 which suggests a survival difference between males with gastric cancer treated with neoadjuvant chemotherapy and chemoradiotherapy compared to S scavenger production. Since increased NFKB1 has been associated with better prognosis, it may not be wise to attempt to knock it down, because that may diminish its protective effect. Therefore, NFKB1

could be useful to determine prognosis and enroll more effective therapeutic interventions for lung adenocarcinoma.

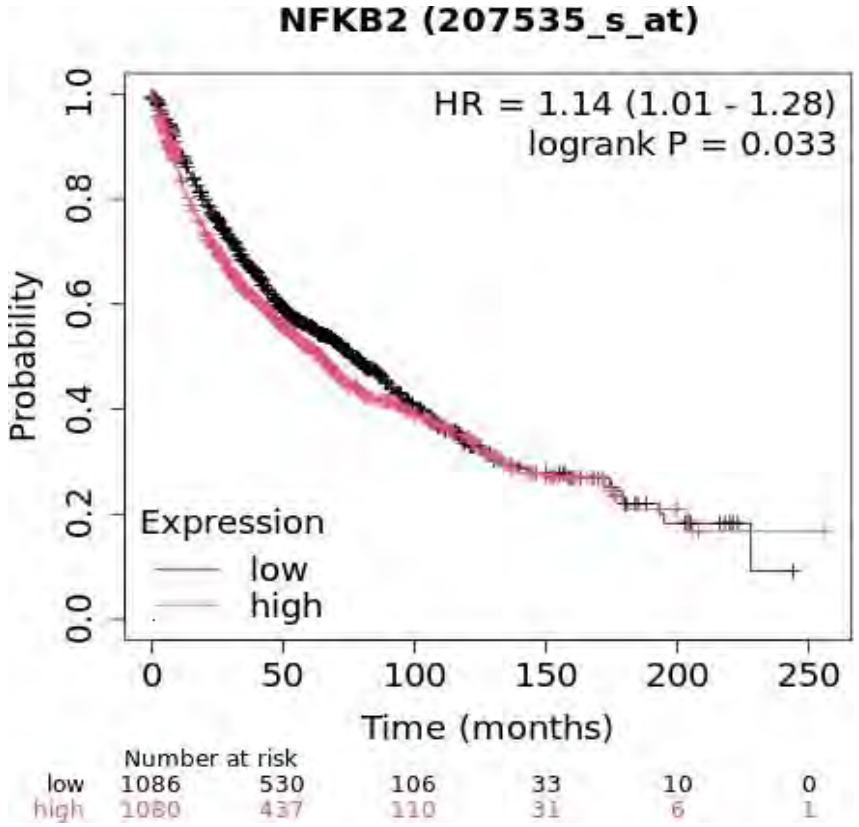

Figure 9: Survival plot of NFKB2.

Likewise, from the Kaplan-Meier survival plot of lung adenocarcinoma and NFKB2 gene, high expression of NFKB2 (pink line) is associated with slightly lesser survival probability than that of low expression group (black line). The hazard ratio (HR = 1.14) reveals that cancer tissues with high NFKB2 expression have a 14% greater likelihood of death than tissues with low NFKB2 expression; therefore, NFKB2 might be involved in poor prognosis. The logrank P = 0.033 shows that there is statistically significant difference in the survival between the two groups. Since high NFKB2 expression is associated with poor survival in LAD patients, NFKB2 could be a candidate for therapeutic intervention in this disease, since its down-regulation could be beneficial to the patient.

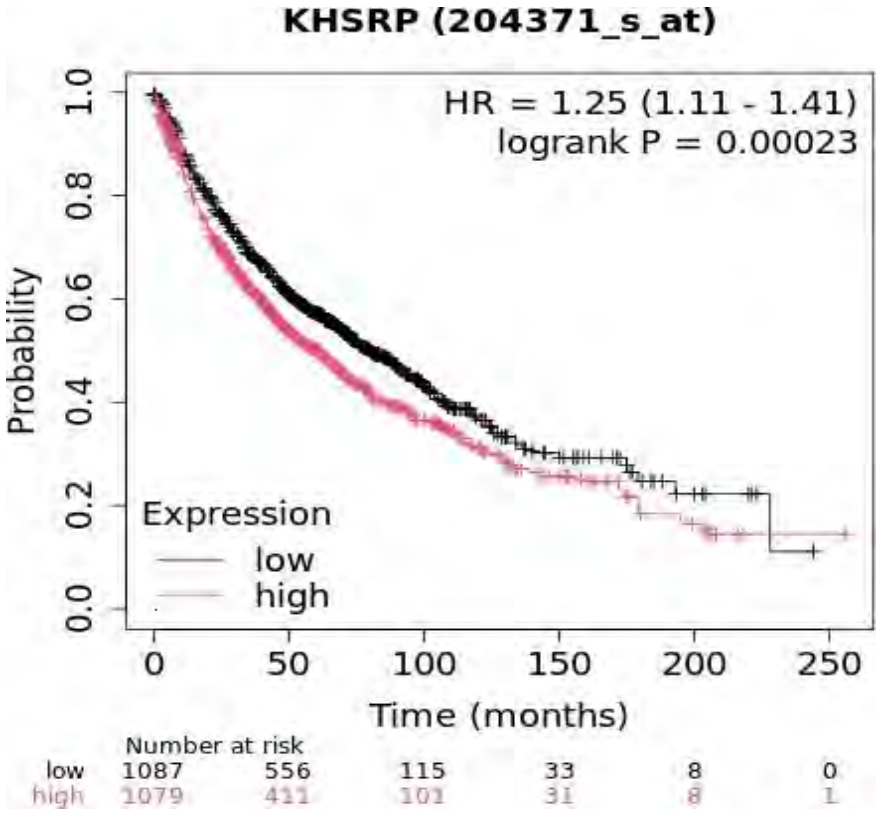

Figure 10: Survival plot of KHSRP.

The Kaplan-Meier survival plot for lung adenocarcinoma, shows that patients with high expression (pink line) have less survival compared to those with low expression (black line). Moreover, hazard ratio of *KHSRP i*s 1.25 which basically indicates that patients with high KHSRP expression have a 25% higher risk of death compared to those with low expression, highlighting a negative impact on survival.**)** The p-value (logrank P = 0.00023) which confirms that the survival difference is statistically significant. Finally, we can find out that *KHSRP* might be associated with worse survival outcomes in lung adenocarcinoma. So, KHSRP can be a potential candidate for lung adenocarcinoma and the inhibition of this gene can be beneficial in the treatment of lung adenocarcinoma.

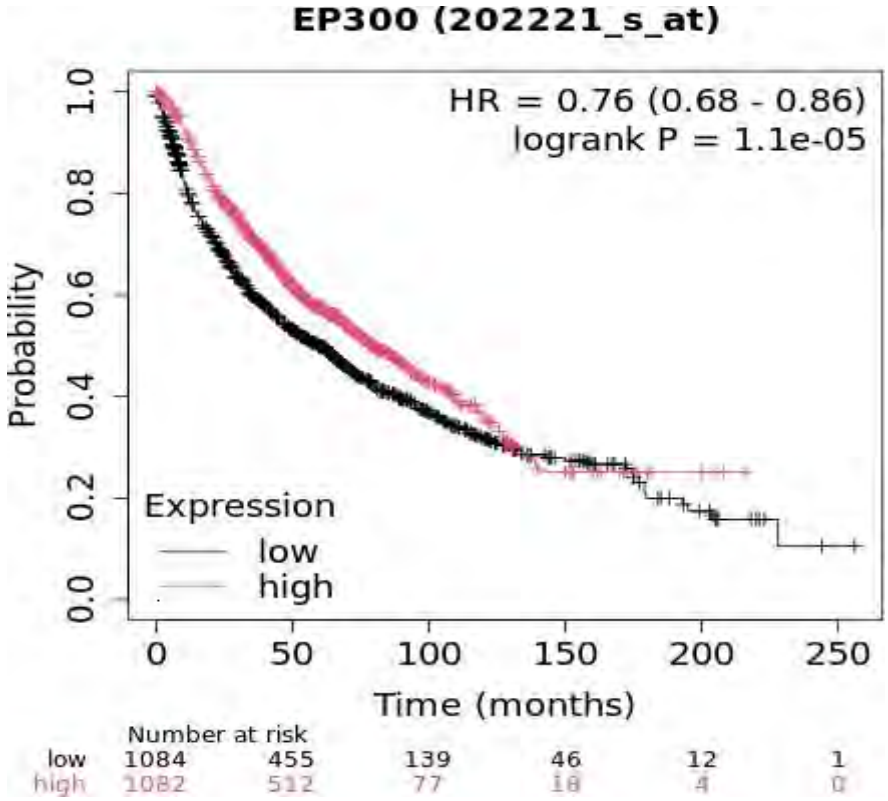

Figure 11: Survival plot of EP300.

The Kaplan-Meier survival plot for lung adenocarcinoma: an analysis of the EP300 gene reveals that patients with high EP300 expression ranked with pink color score higher on the scale of survival probability as compared to low rank scored with black color. The results of this study shows that high EP300 expression is significantly and inversely related to death with a hazard ratio of 0.76; the implication being that as opposed to low EP300, high expression appears to offer some protection. The survival curve difference comparing the two groups is statistically significant according to p-value (logrank P = 1.1e-05). As the high expression of EP300 is related to better prognosis, EP300 may act as a tumor suppressor in lung adenocarcinoma and can be used for identifying favorable prognosis patients. Instead of therapies directing focus to suppress it, they could also support the preservation or boost to its presence as a means of promoting survival

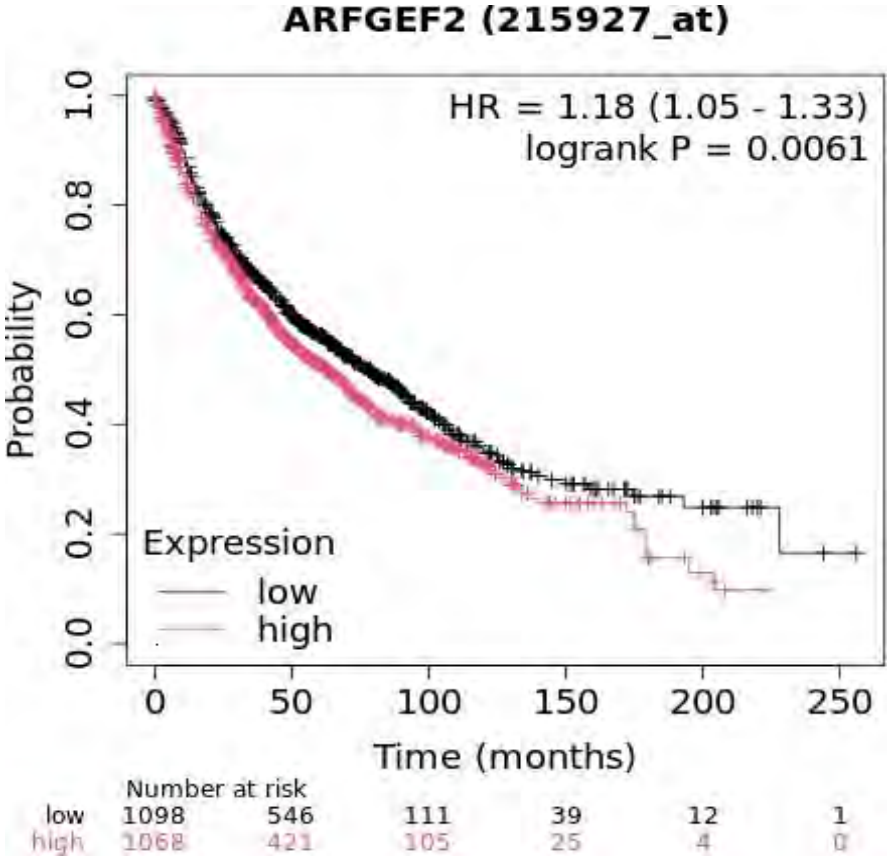

Figure 12: Survival plot of ARFGEF2.

In this figure of Kaplan-Meier survival plot of lung adenocarcinoma and ARFGEF2 gene, it is noted that there is less survival probability in high ARFGEF2 group than the low group. Furthermore, the results also show that patients with high ARFGEF2 overexpression have 18% higher risk of death or poor survival in AMI, as calculated from the hazard ratio (HR = 1.18). This difference in survival is statistically significant as given by p-value (log rank P = 0.0061). Thus, ARFGEF2 may negatively affect survival in lung adenocarcinoma and could be considered as potential therapeutic target If the expression of this gene will be inhibited, the survival of the patients with lung adenocarcinoma can be improved.

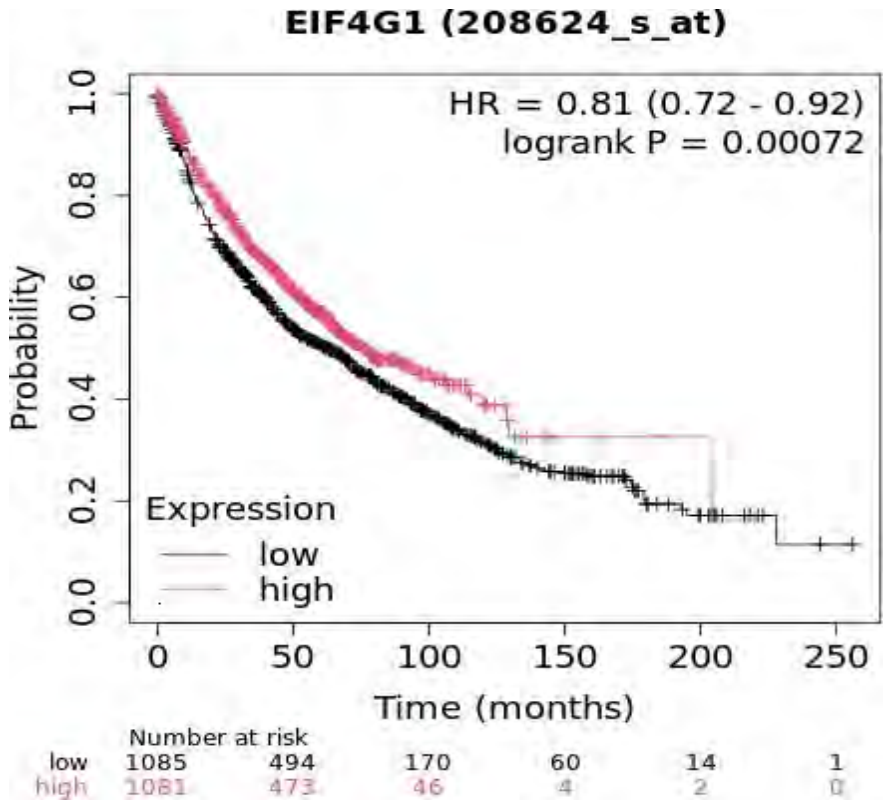

Figure 13: Survival plot of EIF4G1.

The Kaplan's estimator of survivor's probability of the lung adenocarcinoma patients with reference to the gene EIF4G1 depicts that, for the patients possessing high level of EIF4G1 (pink line) the probability of survival is higher that the chance of patients having low level of expression for the same gene (black line). These findings are supported by the hazard ratio, which at 0.81, show that high EIF4G1 expression correlates with a moderative survival advantage, making it a protective factor. The p-value (logrank P = 0.00072) argues to the effect that this survival difference is statistically significant. Thus, high EIF4G1 expression was associated with better survival, and EIF4G1 might be a biomarker of favorable prognosis in lung adenocarcinoma. Targeting other pathways that may promote EIF4G1 expression may also be of potential therapeutic value.

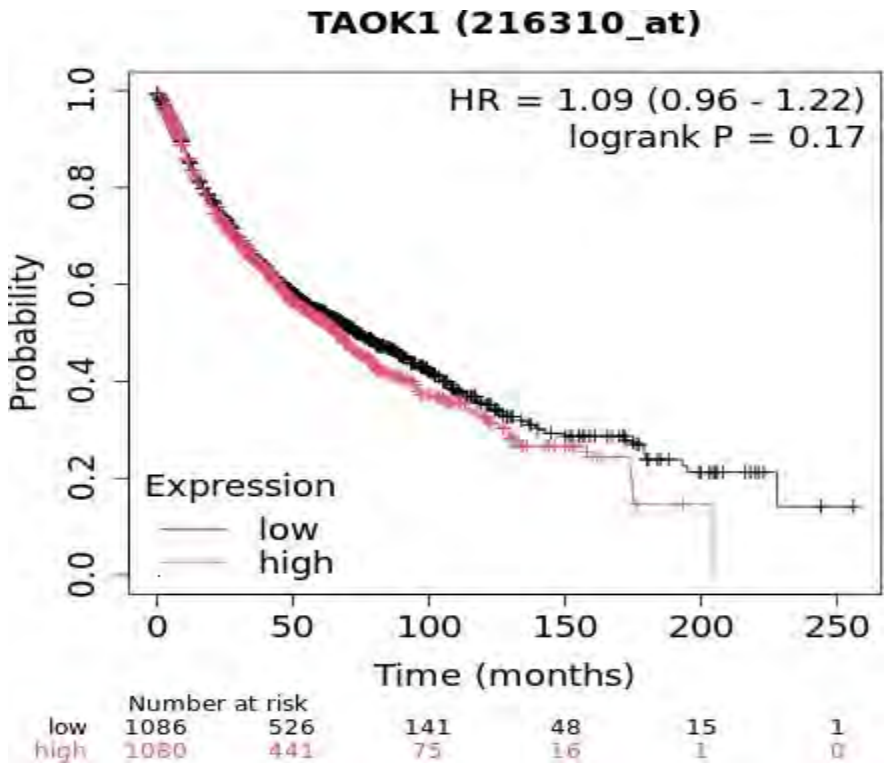

Figure 14: Survival plot of TAOK1.

Kaplan-Meier survival curve of lung adenocarcinoma is plotted as regards the TAOK1 gene; it indicates that the survival probability of the specimens with high TAOK1 expression (pink line) is slightly lower compared to the low expression (Black line). Moreover, it is possible to state that the indicated gap between the two groups as not very large. Data on the prognosis indicate some increase in death rates among patients with high TAOK1 expression (HR = 1.09), but this risk is not particularly high. However, the indicated survival variation is not statistically proven (p = 0.17, logrank P = 0.17). Therefore, TAOK1 may not able to work as an independent prognostic factor for lung adenocarcinoma patients' survival, and it cannot be considered a therapeutic target based on the obtained data.

Chapter 6

Conclusion

Overall, the findings of this study suggest that the STAT3 pathway is involved in the events that determine the progression of lung adenocarcinoma as well as the ability of the disease to resist treatment. Through an extensive bioinformatics analysis, we were therefore able to detect genes of relevance to the STAT3 pathway. Some of these are *HRAS, SMAD2, IQGAP1, NCL, SMAD1, RANBP3, NFKB1* and *NFKB2*, and *EP300.* These genes are involved in a variety of the oncogenic pathways including Ras, rap1, VEGF, MAPK, TGF-beta, PI3K, Hedgehog, TNF, HIF1 and FoxO, AMPK signaling. The natural next step of the differential gene expression study was to highlight the fact that all of these genes had established links with the development of lung adenocarcinoma and patient prognosis, which means that they may also be potentially valuable as targets for the development of therapies. In our present enrichment study, we identified pathways that are relatively enriched in lung adenocarcinoma. These include the TGF- β signaling system, the MAPK signaling pathway, and the NF kappa B signaling pathway. The network analysis also offered further information on the interacting events between these pathways and STAT3 and this therefore embraces the interrelated nature of the molecular processes that are attributable to causes of this malignancy. The results of Kaplan-Meier survival plots demonstrated that the potential target for lung adenocarcinoma is *HRAS, ARFGEF2, KHSRP*; whereas, *IQGAP1, EP300*, and *EIF4G1* are associated with better survival could be potential biomarkers. *NFKB1* was an independent protective factor while *NFKB2* was an independent risk factor and for *TAOK1* the risk in was slightly increased but non-significant. These findings stress the need to target the STAT3 pathway as well as the associated genes; these approaches can now be suggestive of new

therapeutic strategies and molecular therapies that if integrated with traditional treatments including chemotherapy as well as radiation may enhance the overall survival rates for patients diagnosed with lung adenocarcinoma. These results thus stress the need for other studies that would focus on the STAT3 inhibitors, and other combo drugs that target these essential pathways in order to conquer the drug resistance and prolong the span of the patients' lives. As such, this study holds a good promise for the subsequent works, envisioning the molecular ties and investigating potential therapeutic targets to improve the methods of lung adenocarcinoma treatment and, therefore, enhance patients' prognosis and survival.

References

Akin, S., & Ulcay, D. (2023). Lung cancer. In *Medical Nursing*. https://doi.org/10.36106/paripex/2008367

Chen, J. W., & Dhahbi, J. (2021). Lung adenocarcinoma and lung squamous cell carcinoma cancer classification, biomarker identification, and gene expression analysis using overlapping feature selection methods. *Scientific Reports*, *11*(1), 13323. https://doi.org/10.1038/s41598-021-92725-8

Chen, J., Wu, F., Hou, E., Zeng, J., Li, F., & Gao, H. (2023). Exosomal microRNA Therapy for Non-Small-Cell Lung Cancer. *Technology in Cancer Research and Treatment*, *22*(10). https://doi.org/10.1177/15330338231210731

Collins, L. G., Haines, C., Perkel, R., & Enck, R. E. (2007). Lung cancer: Diagnosis and management. *American Family Physician*, *75*(1), 56–63.

Denisenko, T. V., Budkevich, I. N., & Zhivotovsky, B. (2018). Cell death-based treatment of lung adenocarcinoma article. *Cell Death and Disease*, *9*(2). https://doi.org/10.1038/s41419-017-0063-y

Eastman, D., Messerly, C., Udupa, T., & Scott, R. T. (2021). Non-small cell lung adenocarcinoma (NSCLA) metastasizing to the talus: A case study. *Foot & Ankle Surgery: Techniques, Reports & Cases*, *1*(3), 100054. https://doi.org/10.1016/j.fastrc.2021.100054

Hu, X., Zhu, B., Vokes, N., Fujimoto, J., Rojas Alvarez, F. R., Heeke, S., Moreira, A. L., Solis, L. M., Haymaker, C., Velcheti, V., Sterman, D. H., Pass, H. I., Cheng, C., Lee, J. J., Zhang, J.,

Wei, Z., Wu, J., Le, X., Ostrin, E., … Zhang, J. (2024). The evolution of lung adenocarcinoma precursors is associated with chromosomal instability and transition from innate to adaptive immune response/evasion. *Research Square*, *5*(05–24), rs.3.rs-4396272. https://app.dimensions.ai/details/publication/pub.1171656299

Imielinski, M., Berger, A. H., Hammerman, P. S., Pugh, T. J., Hodis, E., Cho, J., Suh, J., Sivachenko, A., Sougnez, C., Auclair, D., Lawrence, M., Stojanov, P., Cibulskis, K., Choi, K., Waal, L. De, Sharifnia, T., Brooks, A., Greulich, H., Banerji, S., … Pao, W. (2013). *NIH Public Access*. *150*(6), 1107–1120. https://doi.org/10.1016/j.cell.2012.08.029.Mapping

Kuhn, E., Damiani, S., Cavazza, A., Comin, C. E., Morbini, P., & Cancellieri, A. (2018). Adenocarcinoma classification: patterns and prognosis. *Pathologica*, *110*, 5–11. pmid: 30259909

Lee Ella A., E. K. (2022). Lung Cancer Screening. *Semin Respir Crit Care Med*, *43*(06), 839–850. https://doi.org/10.1055/s-0042-1757885

Seguin, L., Durandy, M., & Feral, C. C. (2022). Lung Adenocarcinoma Tumor Origin: A Guide for Personalized Medicine. *Cancers*, *14*(7). https://doi.org/10.3390/cancers14071759

Sholl, L. M. (2015). Biomarkers in lung adenocarcinoma: A decade of progress. *Archives of Pathology and Laboratory Medicine*, *139*(4), 469–480. https://doi.org/10.5858/arpa.2014-0128-RA

Wei, X., Li, X., Hu, S., Cheng, J., & Cai, R. (2023). Regulation of Ferroptosis in Lung Adenocarcinoma. *International Journal of Molecular Sciences*, *24*(19). https://doi.org/10.3390/ijms241914614

Zhang, P., Liu, J., Pei, S., Wu, D., Xie, J., Liu, J., & Li, J. (2023). Mast cell marker gene signature: prognosis and immunotherapy response prediction in lung adenocarcinoma through integrated scRNA-seq and bulk RNA-seq. *Frontiers in Immunology*, *14*(May), 1–19. https://doi.org/10.3389/fimmu.2023.1189520