Bioinformatics Analyses of Key Regulators of NF-κB Pathway in Relation with Small Cell Lung Carcinoma

By

Thinley Zangmo
20346071

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
July 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 have acknowledged all main sources of help.

Student's Full Name & Signature:

Thinley zangmo

Thinley Zangmo
Student ID: 20346071

Approval

The thesis titled “Bioinformatics Analyses of Key Regulators of NF-κB Pathway in Relation with Small Cell Lung Carcinoma” submitted by Thinley Zangmo (20346071) of Summer, 2020 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of Bachelor of Pharmacy (Hons) on July 2024

Supervised By:

Supervisor:

_____

Muhammad Asaduzzaman
Professor
School of Pharmacy
Brac university

Approved By:

Assistant Dean &
Program Director:

_____

Professor Dr.Hasina Yasmin
Assistant Dean and Program Director
Brac Universty

Dean:

_____

Professor Dr. Eva Rahman Kabir
Dean
School of Pharmacy
Brac University

Ethics Statement

No unethical activities was entertained during this thesis and no human or clinical trials are used during the process of research. The thesis was adhered to the ethical standards

Abstract

Small cell lung cancer is the most aggressive form of cancer and its progression is very fast and quick. The NF-kB pathway and small cell lung cancer are connected to each other. Despite all the findings and the advanced technologies, due to its challenges associated with its aggressive progressions of small cell lung cancer, only few clinical trials are currently available for SCLC. It has limited treatment options and lacks the targetable biomarkers (Patel & Das, 2023). The NF-kB pathway extensively regulates the pathogenesis of SCLC by promoting tumor development and progression. Additionally, the irregular activation of NF-κB can lead inflammation, development of the autoimmune diseases and malignant disorders, namely atherosclerosis and malignant neoplasm (Park & Hong, 2016). Therefore, it's important to understand and to identify how NF-κB pathway influences the progression of the cancer in the SCLC through the application of the various bioinformatics tools. This study highlights how NF-κB pathway influences the progression of the cancer. The present study also helps in the identification of the key genes that are responsible for SCLC in the NF-κB pathway using the spearman's correlation limit, finding the numbers of correlated genes and filtering the genes by using different bioinformatics techniques and approaches. *NFKB2, ITGB2, PRKCD, RELB, BCL3, STAT6* and *RIPK1* play critical role in the cellular growth and progression, cell survival, apoptosis, immune response and survival of the cancer patients. Finally, understanding the survival rate of the patient and pinpointing the critical pathways that can be targeted.

Keywords: Small cell lung cancer; NF-kB pathway; differentially expressed genes; Co-expressed genes; Gene set enrichment; SNPs

Dedication

Dedicated to all the cancer patients who have lost battles, to my beloved father for unwavering love and support, and to his majesty, the King Jigme Khesar Namgyal Wangchuk, whose wisdom and unwavering spirit have shaped my life.

Acknowledgement

I am immensely thankful for my Thesis Supervisor, Professor Muhammad Asaduzzaman, School of Pharmacy, BRAC University, for his tremendous support and direction in my completion of thesis. His expertise and relentless guidance throughout the studies have helped me to accomplish this project.

Additionally, I would like to express my gratitude to Wasima Taslim Raisa apu for her constant guidance, being helpful and inspiring along the journey.

I would also like to convey my heartfelt gratitude to our Dean, Professor Dr. Eva Rahman Kabir and Assistant Dean and Program Director, Professor Dr. Hasina Yasmin for including the thesis as the part of the curriculum and allowing me to submit the thesis.

Finally, to my family and friend Sunil Rai for their words of encouragement, love and motivation in my every aspect of life.

Table of Contents

List of Tables

List of Figures

List of Acronyms

NFKB2	Nuclear Factor Kappa B Subunit 2
ITGB2	Integrin Subunit Beta 2
PRKCD	Protein Kinase C Delta
RELB	REL Proto-Oncogene, NF-KB Subunit B
BCL3	B-cell CLL/lymphoma 3
STAT6	Signal Transducer and Activator of Transcription 6
RIPK1	Receptor Interacting Serine/Threonine Kinase 1
SCLC	Small cell lung cancer
SQSTM1	Sequestosome 1
RIPK1	Receptor interacting serine/threonine Kinase 1
RPL36	Ribosomal Protein L36
STAT5A	Signal transducer and activator of Transcription 5A
LGALS3BP	Galectin 3 Binding Protein
TRAF1	TNF Receptor associated factor 1
ZFP36	Ring Finger Protein

Chapter 1

Introduction

1.1 Lung cancer

Lung cancer is a type of cancer which starts due to abnormal and uncontrollable cell growth in the lungs. It's the most frequently occurring cancer worldwide and one of the leading causes of death (1.8 million death 8.7% of the total cancer deaths). The estimation of 20 million new cancer cases and 9.7 million deaths was recorded in the year 2022. It was estimated that 53.3 million people are expected to be alive within the 5 years of being diagnosed with cancer. In the lifetime, approximately 1 in 5 people are getting cancer, in Women-1 in 12 had died from the cancer and in Men-1 in 9 (WHO, 2024).

The two types of lung cancer are small cell lung cancer and non- small cell lung cancer. About 10-15% of the lung cancer accounts for the SCLC (Schild & Curran, 2012). While approximately, 85% of the lung cancer accounts the non- small cell lung cancer, (Basumallik & Agarwal, 2023). Surgery, chemotherapy, radiotherapy, and immunotherapy are currently available treatments. The alarming rates and frequency lay the foundation for further research into improving the public health, genetics additionally the novel treatment approaches.

Small cell lung cancer is the most aggressive form of cancer, the necrosis is extensive and its progression is very fast and quick. It usually arises from the epithelial cells and expresses 90% of TTF1, thyroid transcription factor. The solid tumor of the small cell lung cancer and the paraneoplastic syndromes are interconnected to each that is our body mistakenly attacks its own tissue due to the cancer. Syndrome of inappropriate antidiuretic, Ectopic Cushing

syndrome and Lambert-Eaton myasthenia syndrome are commonly occurring paraneoplastic syndrome, (Basumallik & Agarwal, 2023). As per the Thoracic Cancer, 2021, almost 1045 patients were diagnosed in Peking University Cancer Hospital for SCLC between 2008 and 2018. During their studies, from the total of 988 patients suffering from the small cell lung cancer, 603 patients (60.3%) had died and the age was between16-83. From the year 1983 to 2012, it has been found that the survival rate is median however the survival rate was only for 7 months.

1.2 Pathophysiology

Following are subtypes of small cell lung cancer-

1. Oat cell carcinoma and

2. Combined-SCLC.

The type of lung cancer that grows and spreads quickly, resulting in the fast spread of cancer to other parts of the body early is small cell carcinoma (oat cell cancer) it's the most common type of small cell lung cancer. The rare subtype of lung cancer is combined small cell carcinoma that contains both the characteristics of small cell lung cancer and non-small cell lung cancer in the same tumor (Basumallik & Agarwal, 2023).

1.3 Etiology and Risk Factors for Small Cell Lung Cancer

The primary cause of lung cancer is smoking and interestingly, the habit of smoking has been proven to be connected with all types of cancer which accounts for almost 85% of the cases. Although it is connected with all types of cancer, SCLC and squamous cell lung cancer are major ones (Basumallik & Agarwal, 2023).

1.3.2 Risk Factors for SCLC

Accordingly, to the American cancer of Society the risk factors for small cell lung cancer are those who are presently smoking cigarettes, pipes, or cigars or was in the past. Smoking is the major risk factor for small cell lung cancer. If the subjects are exposed to organic and inorganic carcinogens of cigarettes such as arsenic, chromium, beryllium, nickel, soot and tar. Or if the subjects are exposed to radiation such as radiation therapy to the breast and imaging tests. Being exposed to secondhand smoke/passive smoker, living in the place with heavily polluted air, has the family history of lung cancer, taking beta carotene supplement and infected with HIV human immunodeficiency virus are all subjected as risk factor.

1.4 Treatment and Management

Treatment of the small cell lung cancer depends on the stages; chemotherapy and the curative-intent radiation therapy are given those patients who are in the limited stage of small cell lung cancer. Palliation and the chemotherapy for the extensive stage patients.

In the case of limited space, adjuvant chemo radiation and stereotactic ablative radiotherapy (SABR) are given to the patient who are still suffering from the T1-2NO. In case of patients in the limited stage, early incorporation of RT and CT are done thus overall increasing the survival rate.

On the other hand, in the case of an extensive stage, systemic chemotherapy is the primary treatment for extensive stage SCLC patients, with consolidative thoracic RT and PCI recommended for improving overall survival and reducing brain metastasis, as well as novel agents like immunotherapies and targeted therapies. Example of immunotherapy, immune checkpoint inhibitors Nivolumab (PD-1 inhibitor antibody) and example of targeted therapy in clinical Rovalpituzumab tesirin (Basumallik & Agarwal, 2023).

1.5 Significance of SCLC

Compared to the other cancers and the type of lung cancer, the SCLC is not widely and extensively studied. So, in the hope and the curiosity of finding further more elaborate and detailed information this topic was chosen. The reason behind its being not extensively and widely studied as compare to the other cancer or the other lung cancer type like NSCLC is due to its prevalence, approximately only 15% corresponds to the SCLC while the NSCLC have much higher prevalence (Khurshid et al., 2023) so more clinical trials and detailed information are currently available, including the development of personalized medications unlike SCLC.

Due to the prevalence and the challenges associated with its aggressive progressions, only few clinical trials are currently available for SCLC. It has limited treatment options and lacks the targetable biomarkers (Patel & Das, 2023). Although it accounts for only a small percentage of lung cancer, there is a need to do further investigation and exploration of novel therapeutic approaches such as immunotherapy and combination therapies with the help of advanced technologies and knowledge in order to uplift the gap and to save the lives of people/ patients regardless of the rate of occurrences.

1.6 Pathway of SCLC

The NF-kB pathway plays an important role in the control of the various number of expressions of genes that regulate the immune response, cell survival and growth, proliferation, apoptosis, stress response, embryogenesis and the development of the various stimuli. It is activated by several inflammatory stimuli such as growth factors and infectious microbes and crucial for wholesome. The irregular activation of NF-κB can lead inflammation, development of the autoimmune diseases and malignant disorders, namely atherosclerosis and malignant neoplasm (Park & Hong, 2016).

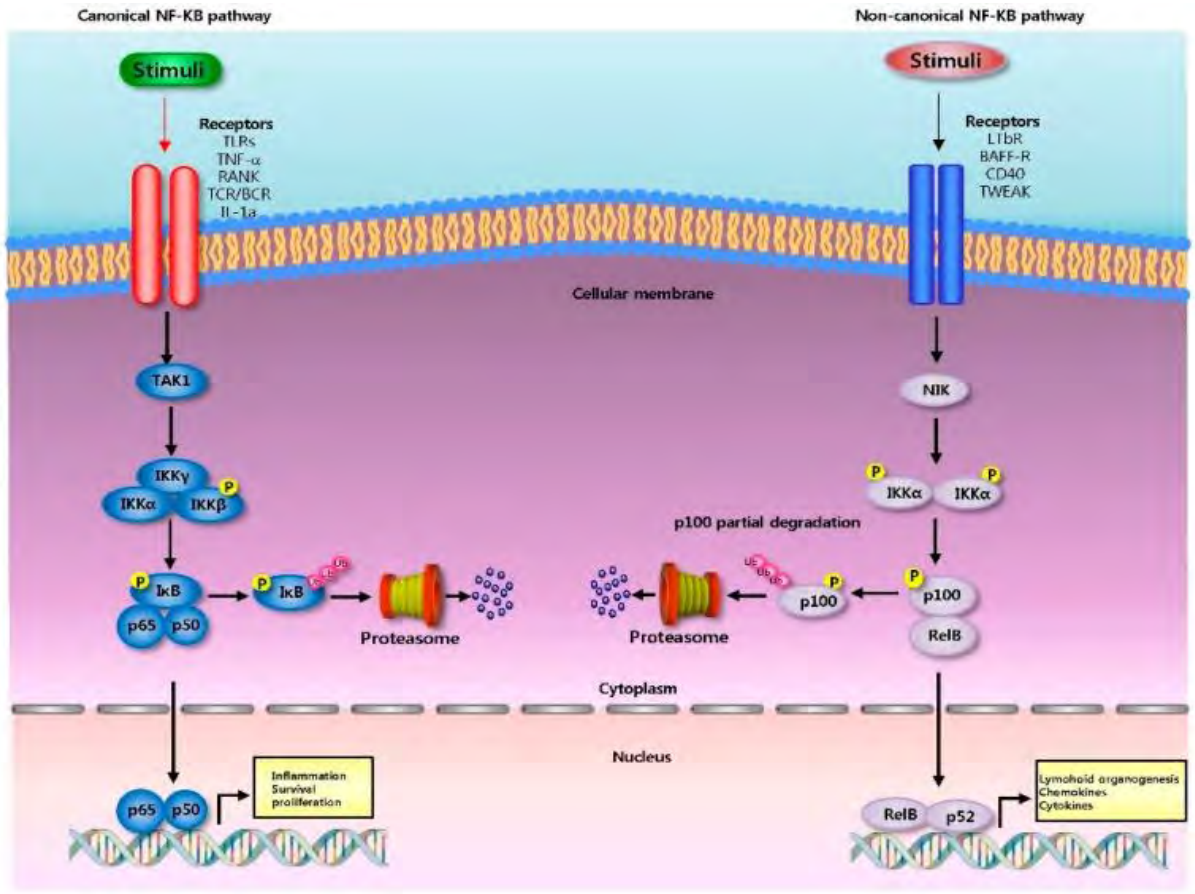


Figure 1.1: The NF-κB pathway ((Park & Hong, 2016).

The figure 1.1 outlines the two types of NF-KB signaling involving (canonical and non-canonical) involving the *NFKB1* gene, Figure 1.2 shows the genes involved in inflammation, proliferation, tumor promotion/progression, angiogenesis, metastasis and survival of SCLC correlated and co-regulated by *NF-KB1*.

The first one is the canonical NFKB pathway, where the stimuli such as antigens and toll-like receptors ligand (e.g., TNF) activities the complex IKKB subunit and this activation led to the phosphorylation of other IKB proteins. Consequently, degradation by the proteasome frees the NF-κB dimers, p50 (encoded by NFKB1) and p65. The p50 (encoded by NFKB1) and p65 are translocated in the nucleus, the regulation of the transcription of genes which are primarily

involved in cell survival, inflammation and cell division. On the other side, the non-canonical NF-κB pathway, which have does involve NFKB1 and is stimulated by B-Cell activating factor (BAFF), receptor activator of nuclear factor kappa-B ligand (RANKL), or lymphotoxin-α and depend on the NIK in order to activate the IKKα. Eventually results in the processing of the p100 into the p52. The p52 forms complexes with the ReLB and regulates the expression of genes which are crucial for lymphoid organogenesis, chemokines and cytokines (Park & Hong, 2016).

Therefore, it overall highlights the mechanism of NF KB activation which are responsible for the regulations, which respond to the different stimuli and play a crucial role in the various immune functions, development and the maintenance of the homeostasis.

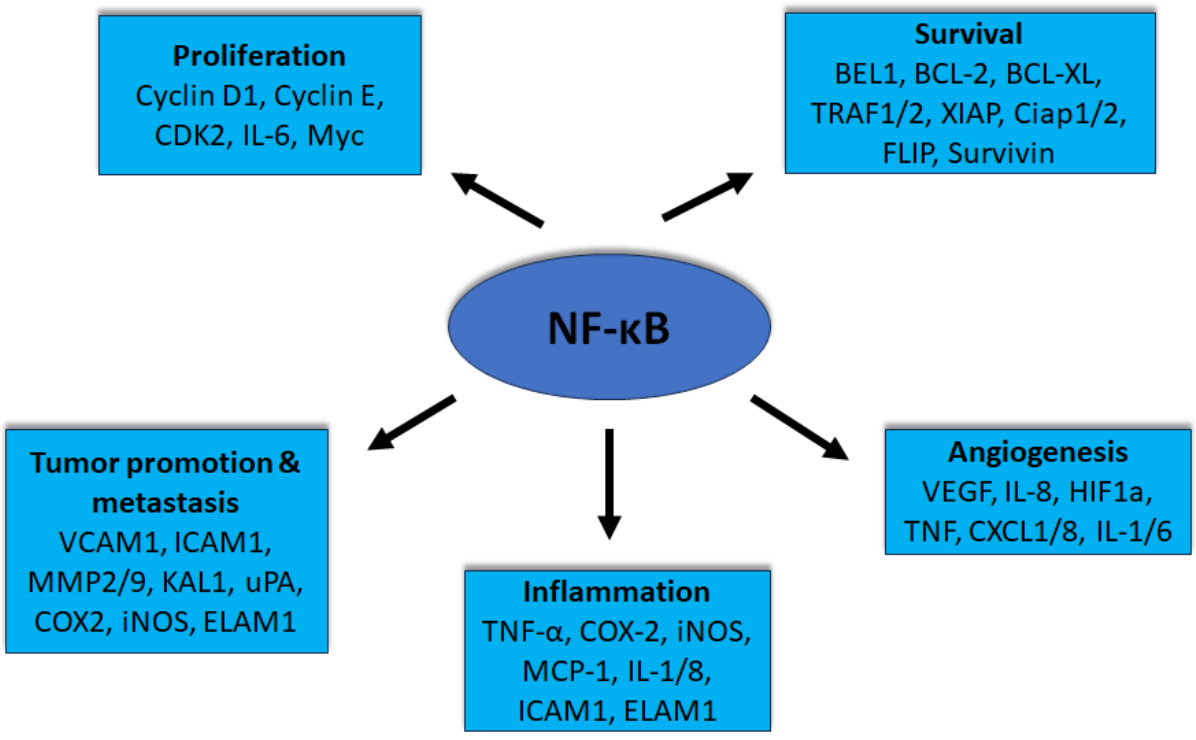


Figure 1.2: Schematic figure of NF-κB in Diseases

1.7 Relationship between the NF-kB pathway and SCLC

The NF-kB pathway and small cell lung cancer are connected to each other; the pathway NF-kB significantly regulates the pathogenesis of small cell lung cancer (SCLC) by promoting tumor development and progression. The scientist had conducted the various studies using the mice with the SCLC and found that blocking the key part of the NF-kB pathway, the NEMO/IKKY which is necessary for the activation of the canonical NF-κB signaling strongly delayed the growth as well as the onset of the SCLC thus increases the survival rate (Koerner et al., 2023).

They also found that ablation of the main NF-κB family member p65/ RelA somehow delays the onset of the tumor progression and the development of the cancer. Additionally, they concluded that even though IKK/NF-κB is already activated, it did not make the cancer worse or aggressive suggesting that endogenous levels of NF-κB are already enough to support tumor development. Moreover, the deficiency of the TNFR1 has not affected the development of the tumor thus representing that the TNF does not play an important role in the development of the tumor in the small cell lung cancer. Overall, it concludes how IKK/NF-κB signaling promotes SCLC and the importance of identifying those pathways and coming up with the ground breaking targeted therapeutic, (Koerner et al.).

1.8 Role of Bioinformatics in SCLC analysis

According to the NCBI (National Center for Biotechnology Information), the branch of NLM (National Library of Medicine) and NIH (National Institutes of Health), bioinformatics is the collection of data, analyzing, performing classifications, manipulating, the recovery process and storage including the visualizing all the necessary information with the help of advanced computerized technologies. Since bioinformatics have wide application in the genomics' and proteomics' field, the modelling the 3D structure of the proteins, analyzing the image and in the field of

designing the drug. So, with the utilization of bioinformatics and its applications, it helps in the finding of necessary data for small cell lung cancer, to collect that data and to discover and learn new knowledge, to uplift the gap on small cell lung cancer. Finally, to come up with the measures and strategies to prevent and to control the growth of tumor in the patient suffering from the small cell lung cancer and other infectious diseases.

Chapter 2

2.1 Aims and objectives

The aim of this project is to understand and to identify how NF-κB pathway influences the progression of the cancer in SCLC through the application of the bioinformatics, to come up with groundbreaking targeted therapy.

Major objectives of this project were as follows:

1. To identify the key genes responsible for SCLC in the NF-κB pathway.
2. To find the correlated genes responsible for SCLC by using different bioinformatics techniques and tools.
3. To identify various genes responsible for SCLC by using various signal transduction.
4. Survival analysis for individual genes regulating SCLC progression in relation with NF-KB pathway.
5. To learn and understand the survival rate of the patient

Chapter 3

Methodology

3.1 Data sources

For this research work, the following online databases, tools and resources were used:

Name of the Database	**purpose**	**URL**
KEGG	For selecting the pathway	https://www.kegg.jp/brite/query=04064&htext=br08901 keg&option=-
CBioPortal	For selecting the genes.	https://www.cbioportal.org/results/coexpression?cancer_study_list=sclc_ucologne_2015&Z_SCORE_THRESHOLD=2.0&RPPA_SCORE_THRESHOLD=2.0&profileFilter=mutations%2Cstructural_variants&case_set_id=sclc_ucologne_2015_sequenced&gene_list=NFKB1&geneset_list=%20&tab_index=tab_visualize&Action=Submit
Harmonizome 3.0	To find the gene set based on NFKB Pathway.	https://maayanlab.cloud/Harmonizome/gene_set/NFKB1/Pathway+Commons+Protein-Protein+Interaction
Venny 2.1	Filter the genes between cBioPoral and Harmonizome.	https://bioinfogp.cnb.csic.es/tools/venny/5.
Expression Atlas	For differential expression of genes.	https://www.ebi.ac.uk/gxa/home
Shiny Go	For gene enrichment.	http://bioinformatics.sdstate.edu/go/
Gene Mania	For co-expression and pathways.	https://genemania.org/search/homo-sapiens/ENSG00000077150/ENSG00000160255/ENSG00000163932/ENSG00000104856/ENSG00000077150/ENSG00000166888
SNPs	For differentially expression of genes	http://www.snps3d.org/modules.php?name=SnpAnalysis&locus_ac=4791

Kaplan-Meier Plotter	To illustrates the survival rate of patient.	https://www.kmplot.com/analysis/index.php?p=service

Table 3.1: List of Online Databases, Tools and Resources

3.2 Literature sources

Below is the list of literature sources which are used as references to collect the various information and data regarding lung cancer, especially small cell lung cancer and related required information.

- PubMed
- Springer nature.com
- Yale medicine
- National cancer institute
- National institute of health (NIH)
- World Health Organization (WHO)

3.3 Data collection method

Firstly, the pathway NF-κB was selected from the KEGG and from the cBioPortal the list of co-expressed genes which are involved in the NF-κB causing the SCLC was selected by using the spearman's correlation limit. Next the Harmonizome 3.0 was used to find the gene set based on NFKB Pathway which has common protein-protein interactions. Consequently, Venny 2.1 was used to find the common genes between the cBioPortal and Harmonizome 3.0.

After finding those common genes or the filtered genes, that list of filtered genes was again filtered using Venny 2.1 based on the 30-signal transduction from KEGG. Next, the differentially expressed list of genes was selected using the expression atlas and the enrichment method from the Shiny Go. The Gene Mania was used to find the networks of co-expression and pathway. SNPs databases was used for identifying differentially expressed genes and

Kaplan-Meier Plotter was used to assess the correlation between the expressions of all genes and SCLC survival rate among the patients.

Chapter 4

Primary data

4.1 NF-κB Correlated genes

For the NK-kB correlated genes, the Spearman's Correlation was taken till 0.3 including both positively and negatively correlated genes. Approximately around 1553 genes were found.

4.2 Common genes between the correlated NK-kB pathway and the other cell signaling

Table 4.1: Filtered genes

Sl.NO	Genes	Genes name
1	*ACTA2*	ACTA2 (Actin Alpha 2,
2	*BCL3*	BCL3 Transcription Coactivator
3	*FLNB*	Filamin B, Beta
4	*FYCO1*	FYVE and Coiled-coil domain autophagy adaptor 1
5	*IQSEC1*	IQ Motif And Sec7 Domain ArfGEF 1
6	*IRF2*	Interferon Regulatory Factor 2
7	*ITGB2*	Integrin Subunit Beta 2
8	*LGALS3BP*	Galectin 3 Binding Protein
9	*NSMAF*	Neutral Sphingomyelinase Activation Associated Factor
10	*PLD3*	Phospholipase D Family Member 3
11	*PRKCD*	Protein Kinase C Delta
12	*PRKCE*	Protein Kinase C Epsilon
13	*RBMX*	RNA Binding Motif Protein, X-Linked
14	*RELB*	RELB Proto-Oncogene, NF-κB Subunit
15	*RIPK1*	Receptor interacting serine/threonine Kinase 1
16	*RPL36*	Ribosomal Protein L36
17	*SQSTM1*	Sequestosome 1

18	*STAT5A*	Signal transducer and activator of Transcription 5A
19	*STAT6*	Signal transducer and Activator of Transcription 6
20	*TRAF1*	TNF Receptor associated factor 1
21	*ZFP36*	ZFP36 Ring Finger Protein
22	*NFKB2*	Nuclear Factor Kappa B Subunit 2

Table 4.2: Signal transductions

Pathways		
MAPK pathway yeast	Wnt pathway	TGF-beta pathway
ErbB pathway	Notch pathway	Hippo pathway
Ras pathway	Hedgehog pathway	Hippo pathway - fly
Rap1 pathway	Hedgehog pathway - fly	Hippo pathway - multiple species
VEGF pathway	JAK-STAT pathway	TNF pathway
Apelin pathway	NF-kB pathway	HIF-1 pathway
FoxO pathway	Calcium pathway	Phosphatidylinositol system
Phospholipase D pathway	Sphingolipid pathway	cAMP pathway
cGMP-PKG s pathway	PI3K-Akt pathway	AMPK pathway
Plant hormone signal transduction	mTOR pathway	MAPK pathway - plant

The table 4.1 represents the common genes from the (cBioPortal and Harmonizome). It was total of 22 genes and these 22 common genes were filtered with the 30-signal transduction in the table 4. 2 that are involved in the NF-kB pathway.

Chapter 5

Results

5.1 Co expressed gene list

When those 22 already filtered genes were again filtered to find the common genes using Venny 2.1, it was found that only eight signal transductions were present. Table 5.3 represents the eight signaling pathway that are involve in those 22 filtered genes. So based on that eight-signal transduction, the common genes were found.

Table 5.1: Signaling pathways

SL.NO	**Pathways**
1	Rap1 signaling pathway
2	TGF-beta
3	VGEF signaling pathway
4	NF-kB pathway
5	TNF signaling pathway
6	Sphingolipid pathway
7	PI3K-Akt pathway
8	mTOR signaling pathway

Table 5.2: cBioPortal and Gepia2

SI.NO	Pathway	Common gene(s)	Genes	Spearman correlation
1	Rap1 signaling pathway	1	ITGB2	0.405
2	TGF-beta	1	PRKCD	0.509
3	VGEF Signaling pathway	1	PRKCD	0.509
4	NF-kappa B signaling pathway	5	RELB BCL3 NFKB2 TRAF1 STAT6	0.448 0.542 0.408 — 0.498
5	TNF signaling pathway	5	BCL3 RELB RIPK1 SQSTM1 TRAF1	0.542 0.448 0.45 — —
6	Sphingolipid signaling pathway	1	NSMAF	—
7	PI3K-Akt signaling pathway	4	ITGB2 PRKCD STAT5A STAT6	0.405 0.509 0.498 —
8	mTOR signaling pathway	2	RBMX RELB	— 0.448

The table 5.2 is the result obtained which include the Number of common Genes, Genes, its spearman correlation and its eight pathways from the Table 5.1 through the application of the Venny.

5.2 Differential expression

Differential expression analysis is the process where we take the read count data that are normalized to statistically analysis in order to find the quantitative changes in the intensity of the expression among the experimental groups.

Parameters are

- **Log2FC** value indicates how much the gene or the transcription expression have changed when compared with the control group. The value is reported in the log scale to base 2.
- **Spearman's Correlation** helps to identify the significant expression of the gene as well as to determine how well the relationship among the genes can be described using the monotonic function. Is a non- parametric measure of the strength and the direction of the between the ranked variables.
- **P value** is a metric to assess the statistical significance of the genomic features. P stands for probability, it's the probability of seeing the results either as support of the genetic or other association as the observed result, if no association exists in fact.
- **P adjusted value** is the modified version of the p value or adjustment done to the p value in order to minimize the likelihood of selecting the false positive as statistically significant.

- **Positively correlated and negatively correlated** genes are those set of genes whether this set of set of genes are upregulated or enriched in the group and if the pathway is positively correlated with the group, it means that most of the genes of that particular pathway will be also positively correlated with the group and vice versa with the negatively correlated genes.

Table 5.3: Differentially expressed genes

Gene	Positively or negatively correlated	Log2fc value	Spearman correlation	P Value	Adjusted p value	Pathway(s)
NFKB2	positively correlated	-1.4	0.408	1.58E-04	$2.7538X10^{-4}$	NF-kappa B pathway
ITGB2	positively correlated	-1.3	0.405	1.77E-04	$4.5133X10^{-2}$	Rap1 pathway PI3K-Akt pathway
PRKCD	positively correlated	-1.9	0.509	1.24E-06	$5.1055X10^{-4}$	PI3K Akt pathway VGEF pathway TGF-beta
RELB	positively correlated	3	0.448	2.72E-05	$2.3372X10^{-17}$	NF-kB pathway TNF pathway mTOR pathway
BCL3	positively correlated	-1.5	0.542	1.71E-07	$7.6731X10^{-4}$	NF-kappa B pathway TNF pathway
STAT6	positively correlated	-1.7	0.498	2.20E-06	$1.2563X10^{-4}$	NF-kappa B pathway

						PI3K-Akt pathway
RIPK1	positively correlated	-1.4	0.45	2.51E-05	$5.095X10^{-4}$	TNF pathway

The table 5.3 represent the eight differentially expressed genes that are positively correlated with its spearman correlation, its log2fc value, P value, adjusted p value, and pathway that are involved in the its respective genes.

5.3 Gene set enrichment analysis

Gene set enrichment analysis is a computerized method, it helps in the determination of whether the pre -defined set of genes that belong to the specific pathway presents statistically significant agreement or differences between two biological states.

Parameters are

Enrichment FDR is the estimated probability for a given set of genes with its enrichment score representing the false positive finding.

Fold Enrichment is used to measure the magnitude of the enrichment, the higher the values the stronger the enrichment is and it's an important metric effect size.

The fold enrichment of a pathway is the ratio of the percentage of the genes in my list to the corresponding percentage in the background genes.

nGenes is the number of genes in the pathways which overlaps with our gene list

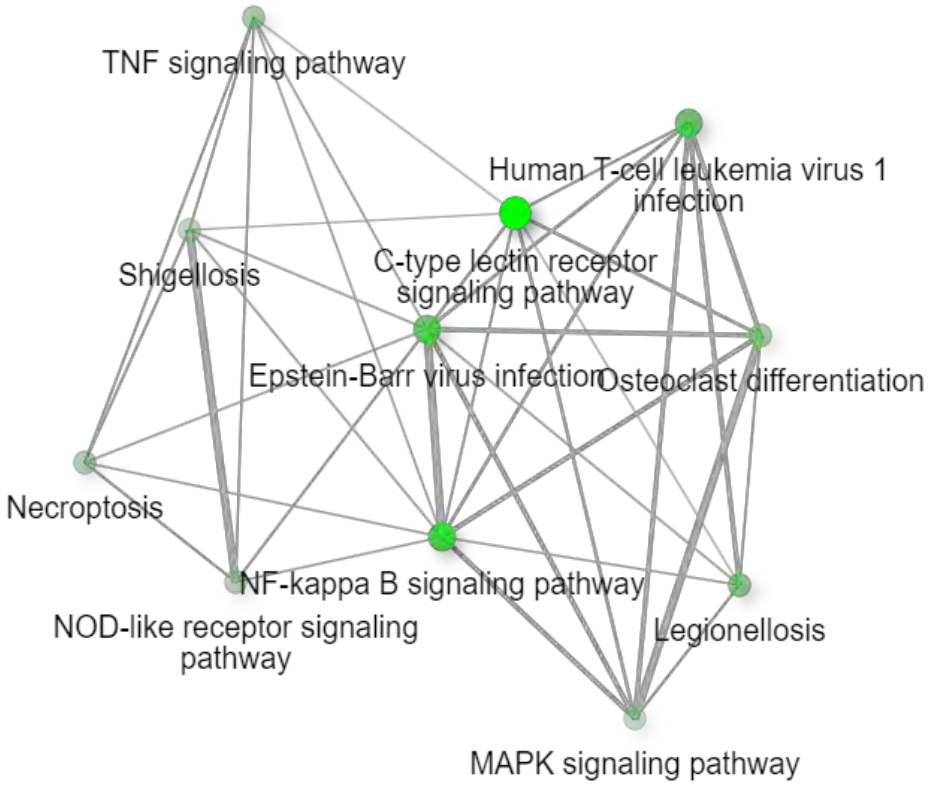


Figure 5.1: Schematic Network diagram showing pathway enriched in relation with NF-KB signaling.

The figure 5.1 represents the relationship between enriched pathways and how each pathways are related. The two nodes or pathways are connected to each other if they share more than 20% genes or default.

Table 5.4. Enrichment analysis

Pathway	**(nGenes)**	**Fold enrichment**	**Enrichment FDR**	**Genes**
C-type lectin receptor signaling pathway	4	125.7	8.4E-07	*NFKB2, PRKCD, REL, BCL3*
Legionellosis	2	114.7	1.5E-03	*ITGB2, NFKB2*
NF-kappa B signaling pathway	3	94.3	9.5E-05	*NFKB2, RELB, RIPK1*
TNF signaling pathway	2	58.4	4.9E-03	*BCL3, RIPK1*
Osteoclast differentiation	2	51.9	5.3E-03	*NFKB2, RELB*
Epstein-Barr virus infection	3	48.5	4.6E-04	*NFKB2, RELB, RIPK1*
Human T-cell leukemia virus 1 infection	3	44.2	4.6E-04	*ITGB2, NFKB2, RELB*
Necroptosis	2	41.1	7.4E-03	*STAT6, RIPK1*
NOD-like receptor signaling pathway	2	35.7	8.7E-03	*PRKCD, RIPK1*
Shigellosis	2	26.6	1.4E-02	*PRKCD, RIPK1*
MAPK signaling pathway	2	22.2	1.8E-02	*NFKB2, RELB*

The table 5.4 shows the eleven pathways and the number of genes involved based on the FDR and the fold enrichment.

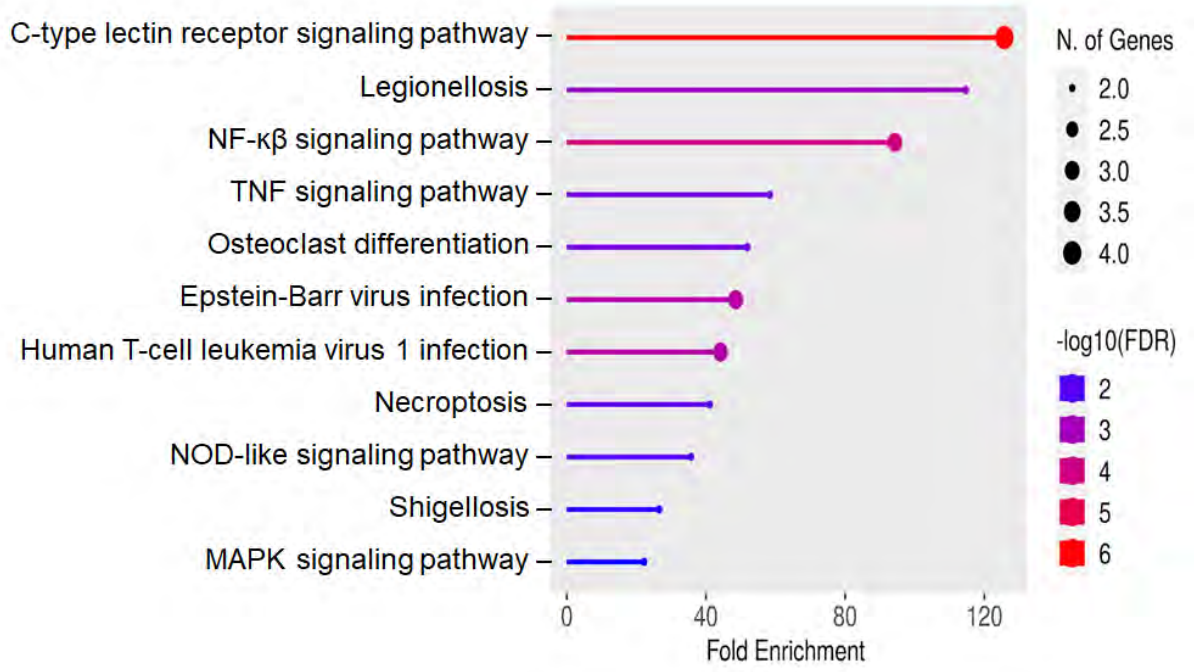


Figure 5.2: Chart of fold enrichment based on pathway

Figure 5.2 shows the chart of fold enrichment based on pathway, the red color represents the high and the blue color represents the low magnitude of the enrichment.

The pathway is sorted by fold enrichment, **fold enrichment** is Fold Enrichment it is used to measure the magnitude of the enrichment, the higher the values stronger the enrichment is and it's an important metric effect size.

The **fold enrichment of a pathway** is the ratio of the percentage of the genes in my list to the corresponding percentage in the background genes.

FDR false discovery rate is the guideline for deciding whether the genes are differentially expressed, if the rate at which features are called significant are truly null. It gives an idea of the expected number of the false positive hypotheses.

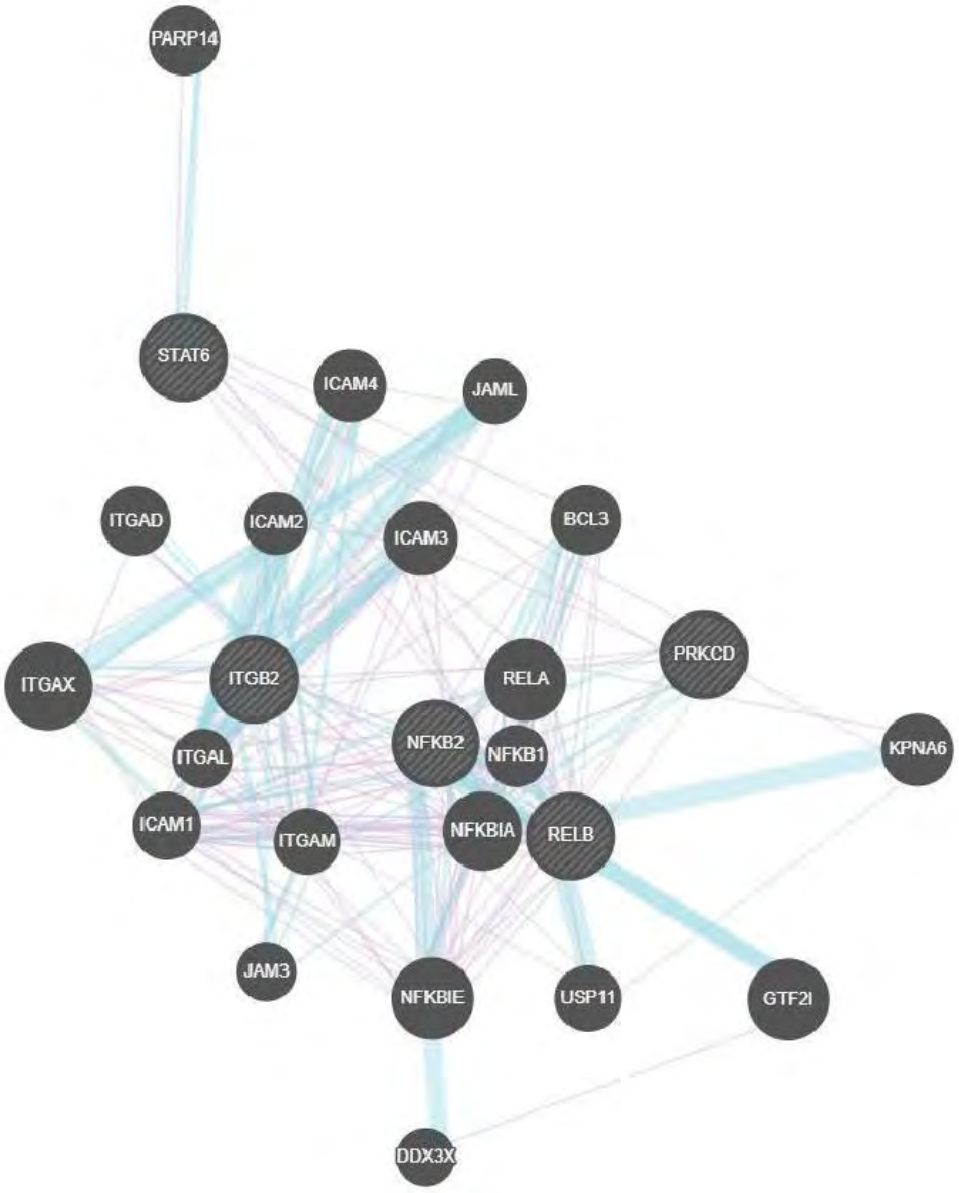


Figure 5.3: Networks of genes related to NF-KB pathway

The figure 5.3 represents the networks of genes, it was done by using the ensemble ID of differentially expressed genes from table 5.3. The purple color shows the Co-expression and the blue color shows the pathway

Table 5.5: SNPs analysis

Genes	SNPS variants	SNPs ID	SVM profile
NFKB2	I213T, R290L, G351R, A392G	3188993 11550592 45580031 11574848	-0.58 -1.88 1.32 0.49
ITGB2	Q354H R586W E630K	235330 5030672 2230531	0.31 0.10 -0.65
PRKCD	Y334C N348S F375S L410F C415Y R483W Y334C N348S F375S L410F C415Y R483W	34297342 33911937 1056998 34502209 35181307 35891605 34297342 33911937 1056998 34502209 35181307 35891605	3.11 -0.63 -1.11 -2.61 1.26 -0.11 3.11 -0.63 -1.11 -2.61 1.26 -0.11
RELB	P393R	34661029	-0.71
BCL3	M191T L313I	35980686 34585458	0.05 -0.3
STAT6	E149Q M181R L365F D419N	2626577 3024952 35416926 11172102	2.37 0.32 -1.36 -1.02
RIPK1	E234K Q398H A404S A438V A443V V646A	17548383 34457341 34872409 3173519 35722193 12194205	-1.06 -0.03 2.47 1.35 0.89 -1.12

The table 5.5 show the SNPs variants, SNPs ID and the SVM profile of those eight differentially expressed genes.

- **SNPs variants**, the SNPs stands for Single Nucleotide polymorphism, it's the most common type of genetic variation among the people and elevates the risk of development of certain diseases. It acts as biomarkers which helps in the prediction of the risk associated with the disease. At least 1% are present in the population.
- **SNPs ID** (Single Nucleotide Polymorphisms ID) are unique and uniquely used for identification of specific single nucleotide polymorphisms.
- The supervised learning algorithm that is best used for classification and regression problems is called **SVM profile (support vector machine)**. It also includes signal processing of the medical application, speech, image recognition additionally natural language processing.

From the table 5.5, it can be seen that the key genes of the NF-kB pathway can have the multiple SNP variants. These SNP variants may be responsible for the cancer cell growth, proliferation, angiogenesis, apoptosis, metastasis and also patient survival.

The following figures (figure 5.4-5.10) shows the KM plots and illustrates the survival rate among patients with respect to the expression of the individual genes.

5.4 Kaplan-Meier Plotters

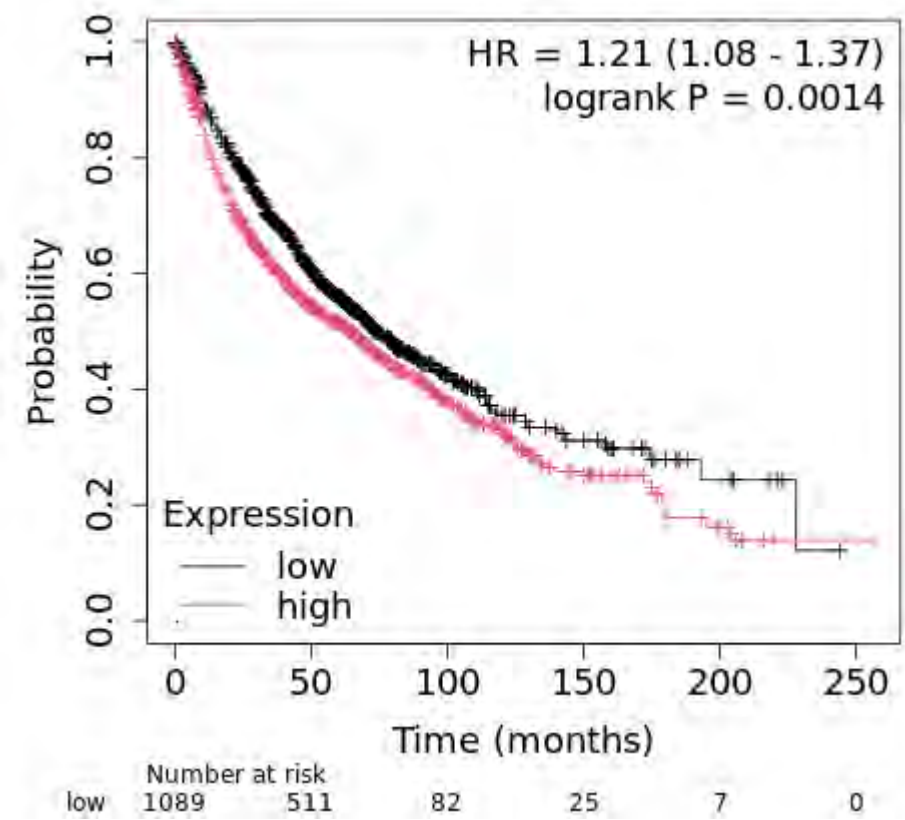
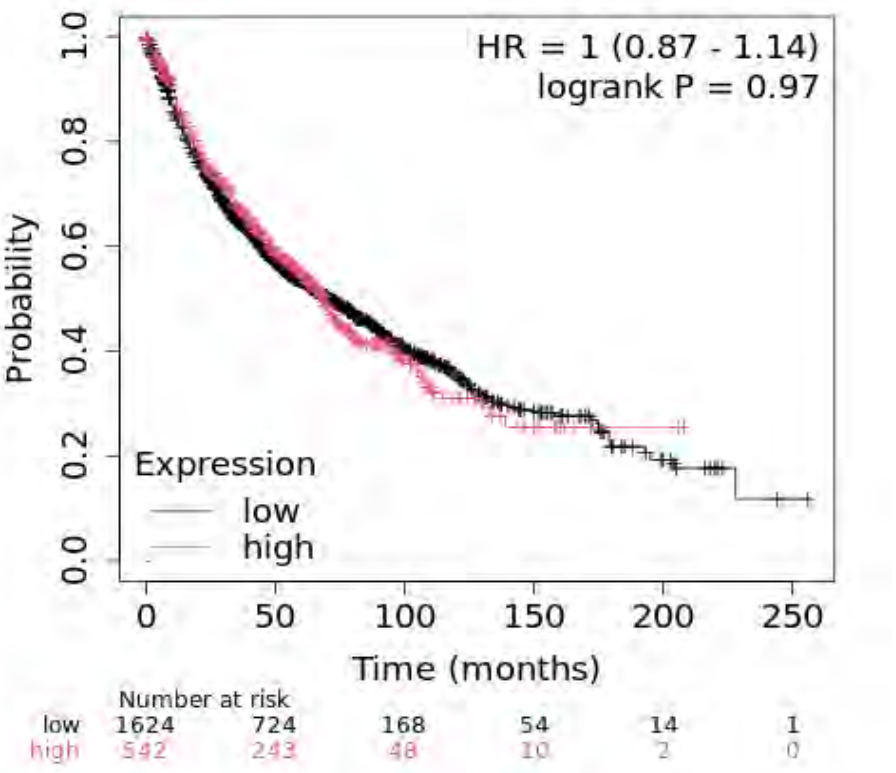


Figure 5.4: KM plot for NFKB2 gene. Here the Y axis represents the probability and the X Axis represents the Time (Month). The probability of a patient surviving, if this NFKB2 gene

(209636_at) is expressed at 50 months is 0.6 which means the rate of survival is 60%. If the gene is



expressed the rate of patient survival is very low.

Figure 5.5: KM plot for ITGB2 gene. Here the Y axis represents the probability and the X Axis represents the Time (Month). The probability of a patient surviving, if this ITGB2 gene (202803_s_at) is expressed at 100 months is 0.4, the rate of survival is 40%. If the gene is expressed the rate of patient survival is median and is not much affected when compared to NFKB2.

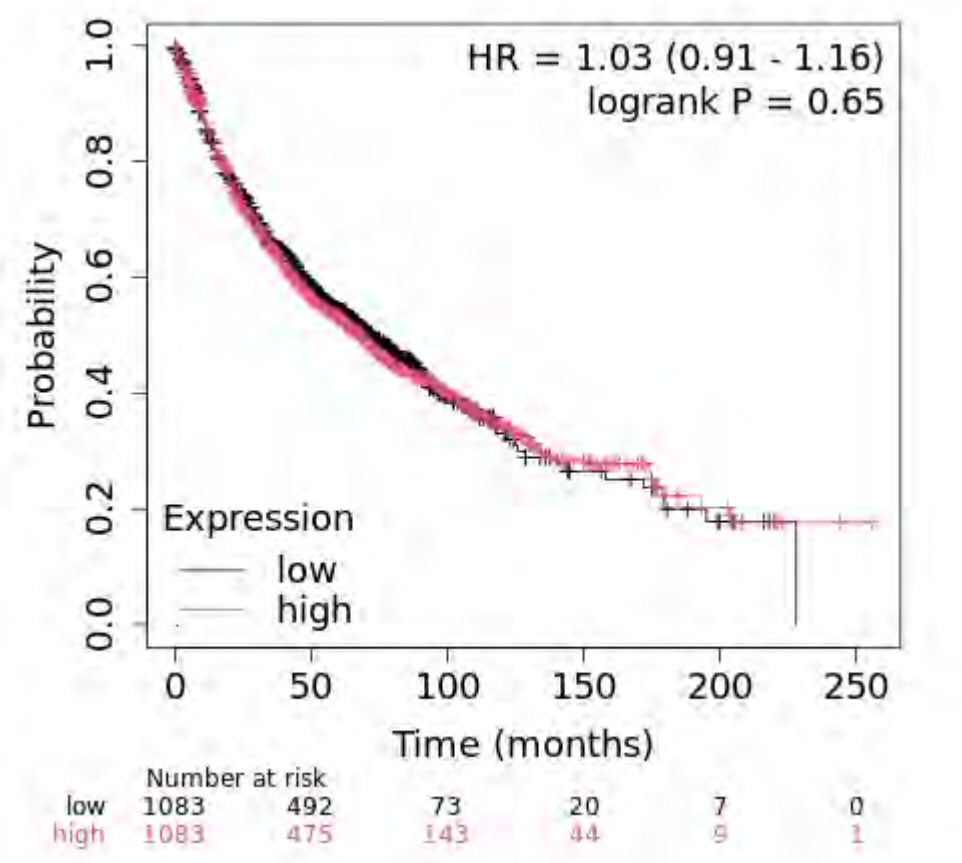


Figure 5.6: KM plot for PRKCD gene. Here the Y axis is the probability and the X axis is the Time (Month). The probability of a patient surviving, if this PRKCD gene (202545_at) is expressed at 50 months is 0.6, the rate of survival is 60%. If the expression of this gene is increased, the rate of patient survival is very low and high expression of this gene can cause death.

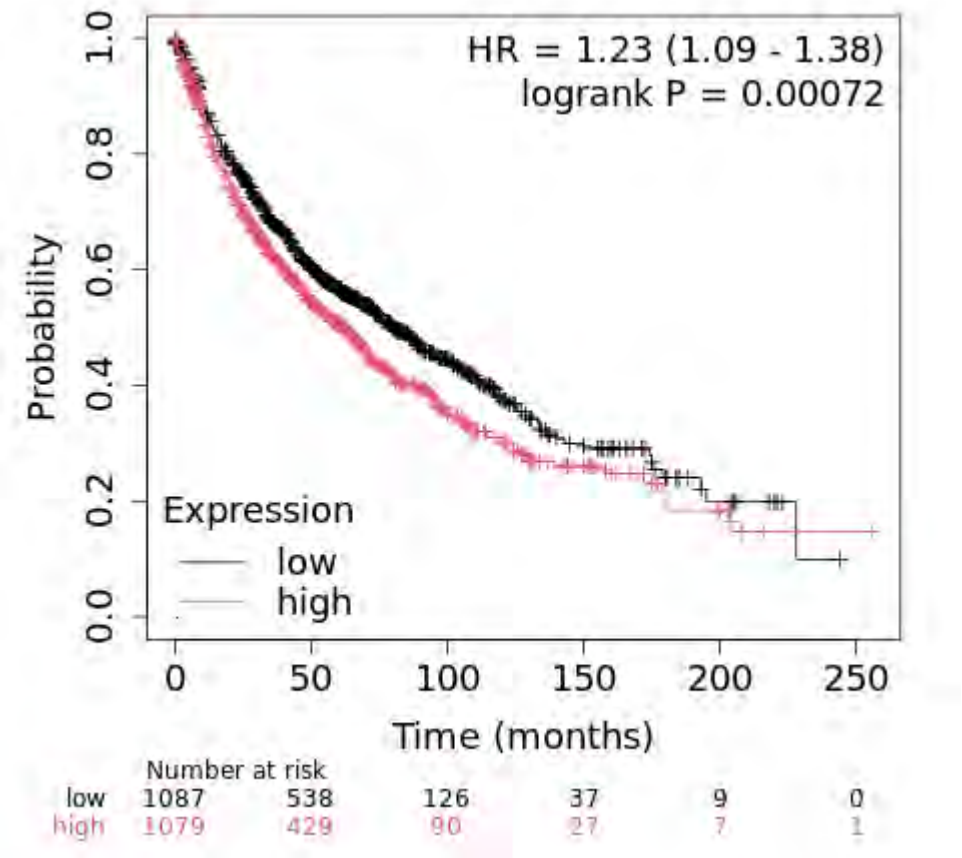


Figure 5.7: KM plot for REL-B gene. Here the Y axis is the probability and the X axis is the Time (Month). The probability of patient survival is approximately 60% at month 50. If the expression of the REL-B gene (205205_at) is increased in the patient, it significantly decreases the survival rate of the patient.

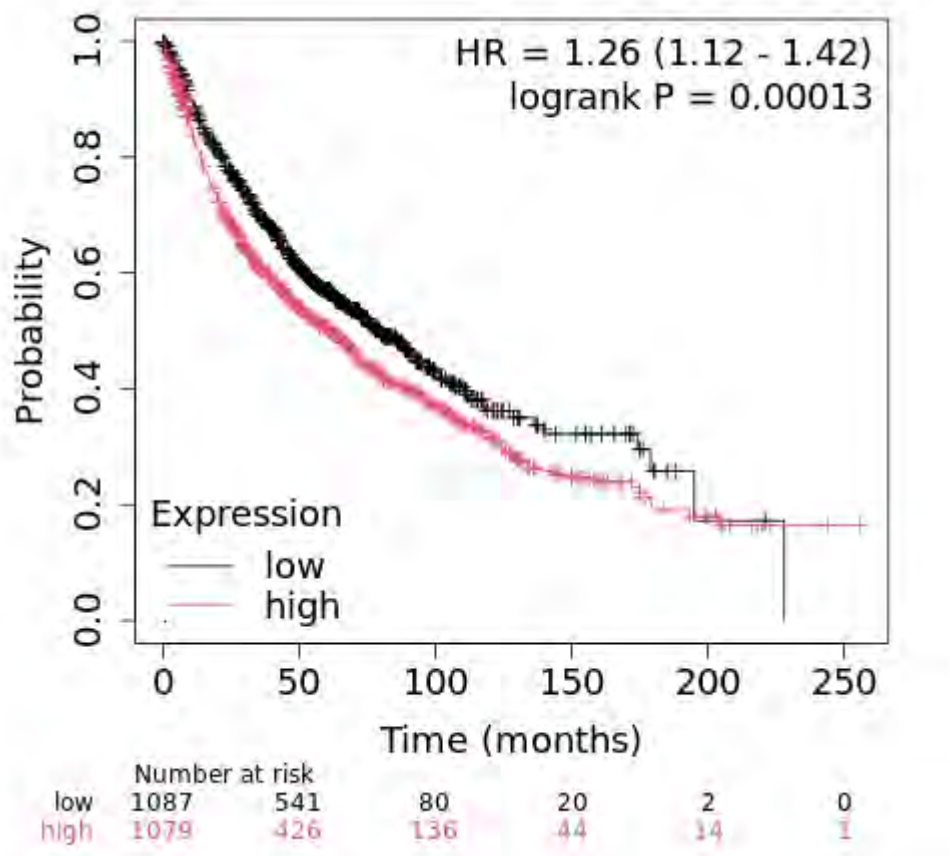


Figure 5.8: KM plot for BCL3 gene. Here the Y axis is the probability and the X axis is the Time (Month). The probability of patient survival are more than approximately around 65% at month 50. If the expression of the BCL3 gene (204907_s_at) is increased in the patient, it reduces the survival rate.

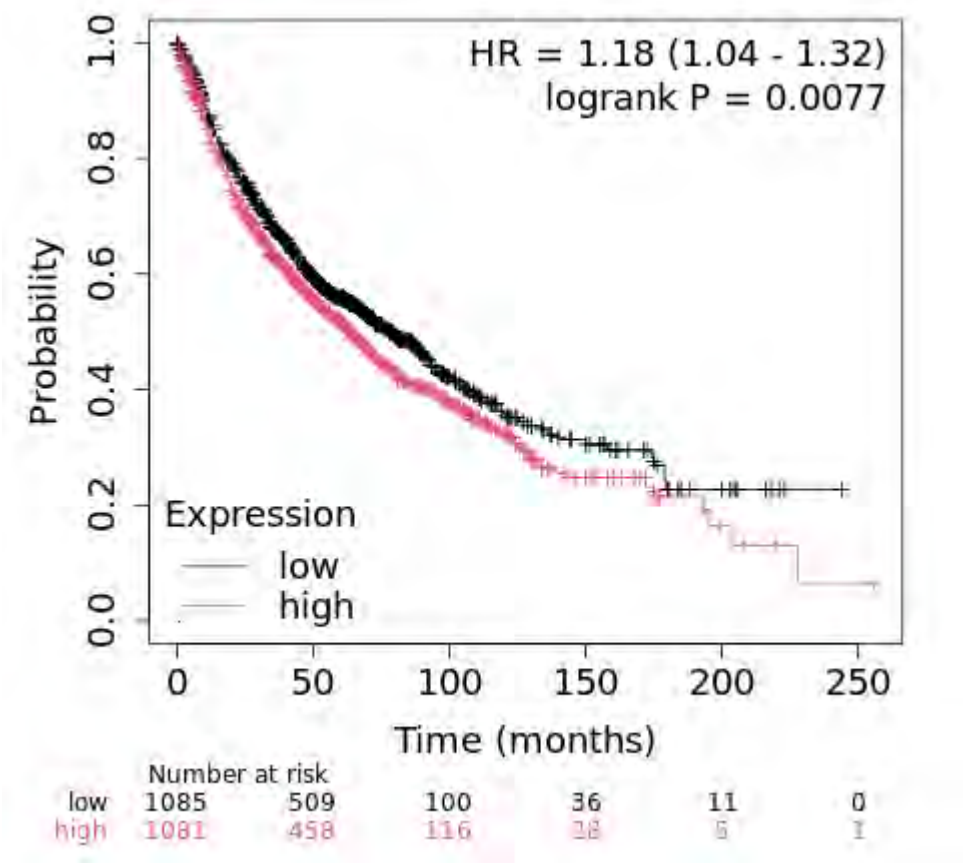


Figure 5.9: KM plot for STAT6 gene. Here the Y axis is the probability and the X axis is the Time (Month). The probability of patient survival is 60% at month 50. If the expression of the STAT6 gene (201332_s_at) is increased in the patient, it drastically decreases the survival rate of the patient.

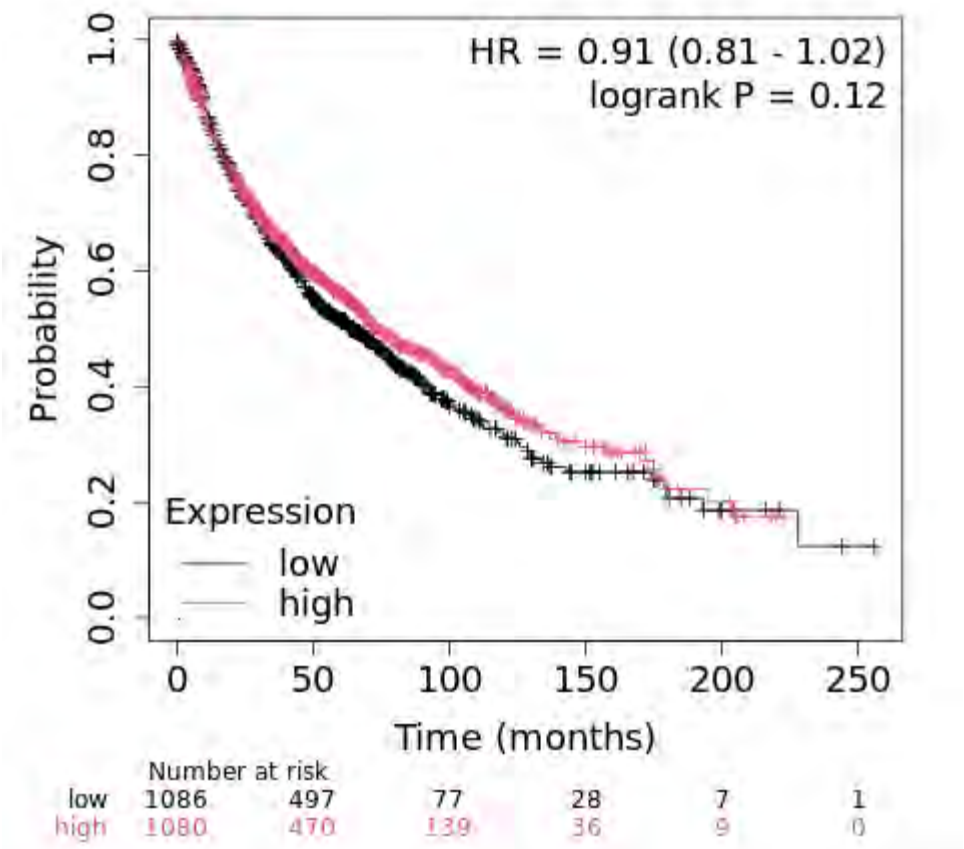


Figure 5.10: KM plot for RIPK1 gene. Here the Y axis is the probability and the X axis is the Time (Month). The probability of patient survival is approximately 40% at month 100. If the expression of the RIPK1 gene (209941_at) is increased in the patient, it highly decreases the survival rate of patient

Chapter 6

Conclusion

NF-κB regulates and controls various numbers of genes which are responsible for immune response, cell growth and proliferation, the cell's survival, apoptosis and development of the several stimuli. Thus, the irregular NF-κB activation leads to the development of the inflammatory, autoimmune and malignant disorders.

As found from our bioinformatics analyses, *NFKB2, ITGB2, PRKCD, RELB, BCL3, STAT6* and *RIPK1* are the key regulators in the NF-κB pathway and when driven by this gene the survival rate of patient is reduced extensively except in the *ITGB2,* the survival rate of patient is medium. Additionally, it showed the significant positive correlations with various pathways such as NF-κB, TNF, PI3K-Akt, mTOR, Rap1, and VEGF. Notably, the *PRKCD* and *STAT6* was reoccurring highlighting the significant role in various cellular processes, including signal transduction, immune response, and cell survival.

Based on the enrichment analysis of the pathways, the C-type lectin receptor signaling pathway, Legionellosis and NF-κB signaling pathway has the high fold enrichments that is 125.7, 114.7 and 94.3, respectively. These correlations underscore that *NFKB2, ITGB2, PRKCD, RELB, BCL3, STAT6* and *RIPK1* play important roles in regulating these pathways, potentially influencing disease mechanisms and cellular responses. In conclusion, it's important to understand and identify how NF-κB pathway influences the progression of the cancer in the SCLC based on the survival rate of the patient. So that personalized medications and strategies can be planned accordingly.

Future Aspect

Based on the findings of this research, future works can be designed in order to deepen the level of understanding regarding NF-κB pathway, the key regulators of NF-κB pathway in context of SCLC and its impact on the survival rates of the patient.

The expression levels of the key regulators of NF-κB Pathway, namely, *PRKCD, STAT6, BCL3, NFKB2 and RIPK1* need to be studied further to understand how the overexpression of these genes significantly reduces the survival rate of the patient. Experiments need to be designed and carried out using clinical samples in order to study the expression levels of these genes from the tumor mass by applying various laboratory techniques such as RT-q PCR, western blotting and Immunohistochemistry (ICH) staining. If the expression of genes are high and significantly reduces the survival rate of patient, it provides the potential to develop the small molecules inhibitors which can be used as drug or SiRNA as inhibitors and vaccine targeting the proteins eventually inactivating the proteins.

References

Albensi, B. C. (2019). What is nuclear factor Kappa B (NF-κB) doing in and to the mitochondrion? *Frontiers in Cell and Developmental Biology, 7*. https://doi.org/10.3389/fcell.2019.00154

Basumallik, N., & Agarwal, M. (2023). Small cell lung cancer. *In StatPearls.* https://www.ncbi.nlm.nih.gov/books/NBK482458/

Basumallik, N., & Agarwal, M. (2023). Small Cell Lung Cancer. *In StatPearls.* https://pubmed.ncbi.nlm.nih.gov/29494065/

CBioPortal, (2024). https://www.cbioportal.org/results/coexpression?cancer_study_list=sclc_ucologne_2015&Z_SCORE_THRESHOLD=2.0&RPPA_SCORE_THRESHOLD=2.0&profileFilter=mutations%2Cstructural_variants&case_set_id=sclc_ucologne_2015_sequenced&gene_list=NFKB1&geneset_list=%20&tab_index=tab_visualize&Action=Submit

False discovery rate. (2023). Columbia University Mailman School of Public Health. https://www.publichealth.columbia.edu/research/population-health-methods/false-discovery-rate

George, J., Lim, J. S., Jang, S. J., Cun, Y., Ozretić, L., Kong, G., Leenders, F., Lu, X., Fernández-Cuesta, L., Bosco, G., Müller, C., Dahmen, I., Jahchan, N. S., Park, K. S., Yang, D., Karnezis, A. N., Vaka, D., Torres, A., Wang, M. S., Korbel, J. O., Thomas, R. K. (2015). Comprehensive

genomic profiles of small cell lung cancer. *Nature*, *524*(7563), 47–53. https://doi.org/10.1038/nature14664

Expression Atlas, (2024). European Bioinformatics Institute, https://www.ebi.ac.uk/gxa/home

Gene Mania, (2024). https://genemania.org/search/homosapiens/

Global cancer burden growing, amidst mounting need for services. (2024). *World Health Organization.* https://www.who.int/news/item/01-02-2024-global-cancer-burden-growing--amidst-mounting-need-for-services

Győrffy, B. KM-plot (2024). https://www.kmplot.com/analysis/index.php?p=service

Harmonizome 3.0, (2024). https://maayanlab.cloud/Harmonizome/gene_set/NFKB1/Pathway+Commons+Protein-Protein+Interactions

Johnson, J. (2021). What is the life expectancy for small-cell lung cancer? *Medicalnewstoday.com.* https://www.medicalnewstoday.com/articles/small-cell-lung-cancer-life-expectancy#rates

Koerner, L., Schmiel, M., Yang, T., Peifer, M., Buettner, R., & Pasparakis, M. (2023). NEMO- and RELA-dependent NF-KB signaling promotes smallcell lung cancer. *Cell Death and Differentiation, 30*(4), 938–951. https://doi.org/10.1038/s41418-023-01112-5

Khurshid, H., Ismaila, N., Bian, J., Dabney, R., Das, M., Ellis, P., Feldman, J., Hann, C., Kulkarni, S., Laskin, J., Manochakian, R., Mishra, D. R., Preeshagul, I., Reddy, P.Saxena, A., Weinberg, F., & Kalemkerian, G. P. (2023). Systemic therapy for Small-Cell lung Cancer: ASCO-Ontario Health (Cancer Care Ontario) guideline. *Journal of Clinical Oncology, 41*(35), 5448–5472. https://doi.org/10.1200/jco.23.01435

KEGG, (2024).Kyoto Encyclopedia of Genes and Genomes. https://www.kegg.jp/brite/query=04064&htext=br08901.keg&option=-

Lung cancer. (2023). *World Health Organization.* https://www.who.int/news-room/fact-sheets/detail/lung-cancer?gad_source=1&gclid=CjwKCAjwydSzBhBOEiwAj0XN4P8f9AT8HBa0diPGoc1t5EZZMj6i1QNNzywU4XlPRtawm_4I6d0EfRoCSqAQAvD_BwE

Ma, X., Zhang, Z., Chen, X., Zhang, J., Nie, J., Da, L., Hu, W., Tian, G., Wu, D., Han, J., Han, S., Long, J., Wang, Y., & Fang, J. (2021). Prognostic factor analysis of patients with small cell lung cancer: Real-world data from 988 patients. *Thoracic Cancer, 12*(12), 1841–1850. https://doi.org/10.1111/1759-7714.13846

Park, M. H., & Hong, J. T. (2016). Roles of NF-KB in cancer and inflammatory diseases and their therapeutic approaches. *Cells, 5*(2), 15. https://doi.org/10.3390/cells5020015

Patel, S. R., & Das, M. (2023). Small cell lung cancer: Emerging targets and strategies for precision therapy. *Cancers, 15*(16), p.4016. https://doi.org/10.3390/cancers15164016

Rouillard, A. D., Gundersen, G. W., Fernandez,N.F., Wang, Z., Caroline, D., Michael, G., McDermott and Ma'ayan, A. (2016). The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. *The Journal of Biological DataBae and Curation, 2016*. https://doi.org/10.1093/database/baw150

Small cell lung cancer. (2022). *Yale Medicine.* https://www.yalemedicine.org/conditions/small-cell-lung-cancer

Small Cell Lung Cancer Treatment (PDQ®)–Patient Version. (2024). *National Cancer Institute.* https://www.cancer.gov/types/lung/patient/small-cell-lung-treatment-pdq

Schild, S. E., & Curran, W. J. (2012). Small cell lung cancer. *Clinical Radiation Oncology (Third Edition).* (pp. 795–804). Elsevier. https://doi.org/10.1016/b978-1-4377-1637-5.00041-9

Schmerker, J. (2023). SNP vs. SNV: What is the difference and why should you care?

Integrated DNA Technologies. https://sg.idtdna.com/pages/community/blog/post/snp-vs.-snv-what-is-the-difference-and-why-should-you-care

ShinyGO: a graphical gene-set enrichment tool for animals and plants. (2024). *South Dakota State University.* http://bioinformatics.sdstate.edu/go/

SNPs, (2024). http://www.snps3d.org/modules.php?name=SnpAnalysis&locus_ac=4791

Thomson, C. (2013). *Encyclopedia of Human Nutrition (Third Edition)*. (pp.186–192). Elsevier. https://doi.org/10.1016/b978-0-12-375083-9.00248-8

Tieri, P., Farina, L., Petti, M., Astolfi, L., Paci, P., & Castiglione, F. (2019). Network inference and reconstruction in bioinformatics. *Encyclopedia of Bioinformatics and Computational Biology*. (pp. 805–813). Elsevier. https://doi.org/10.1016/b978-0-12-809633-8.20290-2

Toh, C., Gao, F., Lim, W., Leong, S., Fong, K., Yap, S., Hsu, A. A., Eng, P., Koong, H., Thirugnanam, A., & Tan, E. (2007). Differences between small-cell lung cancer and non-small-cell lung cancer among tobacco smokers. *Lung Cancer, 56*(2), 161–166. https://doi.org/10.1016/j.lungcan.2006.12.016

Tak, P. P., & Firestein, G. S. (2001). NF-κB: a key role in inflammatory diseases. T*he Journal of Clinical Investigation, 107*(1), 7–11. https://doi.org/10.1172/jci11830

Venny 2.1, (2024). https://bioinfogp.cnb.csic.es/tools/venny/5.

Wilson, M. (2008). Support vector machines. *Encyclopedia of Ecology*. (pp. 3431–3437). Elsevier. https://doi.org/10.1016/b978-008045405-4.00168-3