Gene Expression Analysis of Lung Cancer

Tafadzwa JM Mutiro • *Ruvarashe Chinyadza* • *Vimbainashe Kuzanga*

IA650 Data Mining

Clarkson University

Msc. Data Science

Supervisors: Dr. Sumona Mondal. Prof. Naveen Reddy

30 April 2025

Abstract

Lung cancer reigns in cancer-related mortality worldwide, with non-small cell lung cancer (NSCLC) accounting for the majority of cases. Its major subtypes are Lung Adenocarcinoma (LUAD) and Lung Squamous Cell Carcinoma (LUSC). Despite therapeutic advances, treatment remains hindered by tumor heterogeneity, drug resistance, and a lack of validated biomarkers, particularly for LUSC. Gene expression analysis was conducted in the study to find new potential biomarkers that can be used for Diagnosis,prognostis or new treatment targets. 3 genes of interest were found which were RTN4RL2, HIF3A and REM1. The binary logistic regression fitted showed they had a great AUC of 98 percent showing really promising diagnostic features. REM1 also shows that it can provide prognostic features and can be a new treatment target as it had negative correlation with other treatment types. In summary these genes can provide an opportunity of further studies and can be useful clinically if properly researched.

1 Introduction

Lung cancer remains the leading cause of cancer-related deaths worldwide, responsible for approximately 1.8 million deaths annually. This staggering death toll is mainly due to the late diagnosis of the disease, its rapid progression, and the complex nature of its underlying biology. The late detection of lung cancer, often when it has already spread to other parts of the body, significantly limits treatment options and worsens patient outcomes [2]. Moreover, the aggressive behavior of the cancer cells and the heterogeneity of the disease further complicate treatment strategies, making lung cancer a major global health challenge.

Typically, there are two main categories of lung cancer: non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC). NSCLC accounts for about 85 percent of all lung cancer cases, while SCLC, though less common, is known for its rapid progression and aggressive nature. It is the distinct molecular and genetic features of these two major types that guide their diagnosis, prognosis, and treatment options (Travis et al., 2021).

The NSCLC subtype has two major subtypes that account for most cases: Lung Adenocarcinoma (LUAD) and Lung Squamous Cell Carcinoma (LUSC). These subtypes differ significantly regarding their histological features, molecular characteristics, and genetic drivers. A more in-depth understanding of these differences is crucial for tailoring more effective treatment regimens and improving patient survival rates [1].

Lung Adenocarcinoma (LUAD) LUAD arises from the glandular cells of the lung and is the most common subtype of NSCLC. It typically develops in the peripheral regions of the lung and is increasingly diagnosed in non-smokers, women, and younger individuals (Cao et al., 2021). LUAD is particularly known for its molecular and genetic complexity, with a high degree of heterogeneity observed within and across tumors. One of the key features of LUAD is its association with actionable genetic mutations, such as those in the EGFR, ALK, and KRAS genes. These mutations offer potential targets for personalized therapies, making LUAD a primary focus for developing precision medicine strategies (Imielinski et al., 2012).

Additionally, LUAD is known for its ability to present with genetic alterations that enable tumors to evade immune detection and promote uncontrolled growth. Despite significant advances in targeted therapies, LUAD

remains challenging to treat in some cases due to the development of resistance to targeted therapies over time. Consequently, researchers continue to focus on uncovering new therapeutic targets and treatment strategies to overcome these challenges (Sacco et al., 2017).

Lung Squamous Cell Carcinoma (LUSC) Lung Squamous Cell Carcinoma (LUSC) originates from the epithelial cells that line the central airways of the lungs. It is strongly associated with tobacco smoking, where its development is linked to long-term exposure to carcinogens present in tobacco smoke. Furthermore, the formation of keratin and intercellular bridges that act as hallmarks of squamous differentiation characterizes LUSC. This carcinoma is more common in older male smokers. It is less likely than LUAD to have mutations that can be targeted by specific therapies, making LUSC particularly challenging when implementing precision medicine approaches (Hirsch et al., 2017).

The molecular and genetic features of LUSC are distinct from those of LUAD. LUSC is often associated with mutations in the TP53, CDKN2A, and PIK3CA genes; however, the lack of actionable mutations means that treatment strategies typically rely on chemotherapy and immunotherapy rather than targeted therapies. The absence of well-defined biomarkers further complicates the clinical management of LUSC, highlighting the need for continued research into its genetic underpinnings (Molina et al., 2021).

Genetic Basis of Cancer At its core, cancer is a disease of the genome. Cancer arises when genetic mutations disrupt the normal processes that regulate cell growth, division, and survival. The DNA within a cell contains genes that provide the instructions necessary for maintaining normal cellular function. These genes control vital processes such as cell proliferation, apoptosis (programmed cell death), and DNA repair. When mutations occur in specific genes, they can drive the uncontrolled growth of cells, leading to tumor formation and often, metastasis (Hanahan Weinberg, 2011).

The genetic alterations that lead to cancer can be divided into three essential categories: oncogenes, tumor suppressor genes and DNA repair genes.

Oncogenes are mutated or overactive versions of normal genes (proto-oncogenes) that promote cell division and survival. In normal cells, proto-oncogenes regulate processes like cell growth. However, when these genes become mutated or overexpressed, they can

drive uncontrolled cell division, contributing to tumor formation. Furthermore, oncogenes are often crucial players in developing and progressing many cancers, including lung cancer (Vogelstein et al., 2013).

Tumor suppressor genes are responsible for slowing down cell division. They help repair damaged DNA and prevent cells from dividing uncontrollably. When tumor suppressor genes are inactivated or lost due to mutations, they fail to perform their protective functions, allowing cancer cells to proliferate unchecked. A well-known example of a tumor suppressor gene that undergoes mutation in various cancers is TP53 (Levine, 2020).

DNA repair genes correct errors that occur during DNA replication and repair. Mutations in DNA repair genes can result in the accumulation of additional genetic mutations, which may accelerate the development of cancer. In lung cancer, defects in DNA repair mechanisms often lead to genomic instability, which is associated with tumor progression and resistance to therapy (Jaspers et al., 2019).

Comprehending these genetic alterations is fundamental to developing targeted therapies and personalized treatment strategies for lung cancer. Identifying the specific mutations and alterations that drive tumor growth in LUAD, LUSC, and other types of lung cancer is pivotal to improving patients' prognosis and treatment outcomes (Cancer Genome Atlas Research Network, 2014).

Relevance of Studies and Related Literature Understanding how molecular mechanisms drive lung cancer has progressed significantly over the past few decades. The evolution of advanced genomic and transcriptomic technologies, including high-throughput sequencing and RNA sequencing, has allowed for a more detailed analysis of the genetic and epigenetic alterations in lung cancer. Such technologies have provided a wealth of data, revealing actionable mutations, altered pathways, and potential biomarkers for early detection and targeted therapy (Barbieri et al., 2016).

Studies like those conducted by The Cancer Genome Atlas (TCGA) have been instrumental in identifying key genetic drivers of LUAD and LUSC, allowing for the discovery of novel therapeutic targets and potential biomarkers for diagnosis and prognosis. These studies have also highlighted the distinct molecular differences between LUAD and LUSC, enabling clinicians to differentiate between the subtypes more accurately and apply tailored treatment strategies (The Cancer Genome Atlas Research Network, 2014).

Moreover, the literature has demonstrated that LUAD and LUSC, despite being categorized as NSCLC, require different therapeutic approaches due to their genetic differences. Although LUAD benefits from targeted therapies directed at mutations in EGFR, ALK, and KRAS, LUSC remains more resistant to such therapies, and chemotherapy and immunotherapy typically are the primary treatment options (Pao Hutchinson, 2015).

The research literature further emphasizes the need for more effective biomarkers to predict treatment response and resistance. As resistance to targeted therapies and immunotherapies becomes an increasingly common challenge, integrating genomic, proteomic, and transcriptomic data is essential to identify patients who are most likely to benefit from specific treatments (Sacco et al., 2017).

In conclusion, the growing body of literature surrounding lung cancer genetics, including identifying actionable mutations, the characterization of subtypes, and the exploration of potential biomarkers, is pivotal to advancing treatment approaches and improving patient outcomes. As research progresses, it will continue to shape the development of personalized medicine, offering hope for better therapeutic options and survival rates for lung cancer patients worldwide.

2 MATERIALS AND METHODS

2.1 Data Collection and Preprocessing

The dataset used in this study was obtained from The Cancer Genome Atlas (TCGA) via the Xena platform (xena.ucsc.edu), which hosts a wide range of genomic and transcriptomic data across multiple cancer types. For the purposes of this study, we focused specifically on lung cancer, utilizing gene expression profiles derived from both tumor and normal tissue samples. The data was preprocessed as follows:

- Normalization:Gene expression levels were normalized to account for technical variability, such as differences in sequencing depth and gene length, ensuring comparability across samples.

- Filtering: Genes with minimal expression variance across samples were excluded, as they are unlikely to contribute to the identification of meaningful differential expression patterns.

- Missing Data: Missing expression values were addressed through appropriate imputation techniques. In cases where genes exhibited a high proportion of missing values, they were removed from downstream analysis.

- Batch Effect Correction:To eliminate non-biological variability arising from differences in experimental conditions or sequencing batches, batch effect correction was applied using standardized normalization methods.

2.2 Statistical Tests

The Mann-Whitney U test was used to identify genes that are differentially expressed between tumor and normal tissue samples. Since gene expression data typically do not follow a normal distribution, a non-parametric test like the Mann-Whitney U test is well-suited for comparing the two groups. This test compares the ranks of expression values from the two groups to determine whether they come from the same distribution.

The test produces a U statistic, which is then used to calculate a p-value. After applying multiple testing correction using the Benjamini-Hochberg procedure, genes with adjusted p-values < 0.05 were considered significantly differentially expressed.

2.3 Visualization of Results

To facilitate interpretation of the results from the differential gene expression analysis, several visualization techniques were employed. These visualizations serve as intuitive tools for understanding the relationship between gene expression in tumor and normal samples, as well as identifying significant genes.Heatmaps are a powerful tool for visualizing the expression of genes across samples, especially when the goal is to detect patterns or clusters in gene expression between groups. For this study, a heatmap was constructed for the top differentially expressed genes identified by the Mann-Whitney U test.The top 50 to 100 most significantly differentially expressed genes, based on the lowest p-values, were selected for inclusion in the heatmap.To facilitate comparison of gene expression levels across samples, gene expression values were standardized (z-scored), meaning that each gene's expression was adjusted to have a mean of 0 and a standard deviation of 1.The heatmap allows for a visual inspection of how the expression of the top genes differs between tumor and normal tissues. If a distinct pattern of expression is observed, where

tumor samples cluster together and normal samples form another cluster, this could suggest significant biological differences between the two tissue types.

2.4 Volcano Plot

The volcano plot is a widely used tool in genomics for displaying the results of differential expression analysis. It combines two key aspects of gene expression data: the magnitude of change and the statistical significance of the difference between groups.To create the volcano plot, the log2 fold change (logFC) and negative logarithm of the p-value were plotted as follows:

$$\text{logFC} = \log_2 \left(\frac{\text{Mean Expression in Tumor}}{\text{Mean Expression in Normal}} \right)$$

The x-axis represents the log2 fold change, indicating how much more or less a gene is expressed in tumor versus normal tissues. The y-axis represents the negative logarithm of the p-value, which reflects the statistical significance of the expression difference.The volcano plot helps in identifying genes that exhibit both high statistical significance and large fold changes in expression. Genes that appear in the upper-left or upper-right corners of the plot are considered the most significant and differentially expressed. These genes can be further investigated for their biological relevance in cancer. Typically, genes with a high absolute value of logFC (indicating large differences in expression) and a low p-value (indicating statistical significance) are of particular interest.

2.5 Boxplots of Top Genes

Boxplots were used to visualize the distribution of gene expression levels in tumor and normal samples for the top differentially expressed genes. A boxplot displays the minimum, first quartile (25th percentile), median (50th percentile), third quartile (75th percentile), and maximum values, providing a summary of the distribution of gene expression.For each of the top differentially expressed genes, a boxplot was generated for both tumor and normal tissue samples.Boxplots provide an effective way to visually compare the distribution of gene expression in tumor versus normal tissues. A large difference in the median values between the two groups suggests a strong differential expression of the gene. The presence of outliers or variability within each group may suggest heterogeneity in gene expression that could be further explored.

2.6 Logistic Regression for Classification

Logistic regression was applied to classify tumor and normal tissue samples based on their gene expression profiles. The model estimates the probability that a given sample belongs to the tumor group (coded as 1), based on the expression levels of the differentially expressed genes identified earlier. The logistic regression model used in this study predicts the probability of a sample belonging to the tumor group. The model is expressed as:

$$P(y = 1|\mathbf{x}) = \frac{1}{1 + e^{-(\beta_0+\beta_1 x_1+\cdots+\beta_p x_p)}}$$

Where:

- y is the binary outcome variable (1 for tumor, 0 for normal).
- $\mathbf{x} = (x_1, x_2, \ldots, x_p)$ are the gene expression values for selected genes.
- $\beta_0, \beta_1, \ldots, \beta_p$ are the regression coefficients estimated from the data.

2.7 Model Evaluation

The logistic regression model was evaluated using metrics such as accuracy, precision, recall, and area under the ROC curve (AUC). Cross-validation was performed to ensure that the model generalizes well to new, unseen data.

2.8 Software and Tools

The analysis was conducted using the Python programming language.The following libraries were used:

- `pandas` for data manipulation and preprocessing.
- `numpy` for numerical operations and array handling.
- `matplotlib` and `seaborn` for generating visualizations, including heatmaps, volcano plots, and boxplots.
- `scikit-learn` for machine learning tasks such as model building, evaluation, and cross-validation.
- `scipy` for performing statistical tests and regression.
- `cluster` for hierarchical clustering.
- `statsmodels` for more advanced statistical analysis.

2.9 Limitations and Assumptions

This study assumes that the gene expression data is of high quality and follows a consistent

distribution across the samples. While the Mann-Whitney U test handles non-normality in the data, other assumptions such as independence of samples still hold. Furthermore, the boxplots and heatmaps rely on the assumption that the clustering methods used are appropriate for the data structure.

2.10 Future Directions

Future research could explore alternative models for classification, such as random forests or support vector machines, to improve the robustness of predictions. In addition, integrating other types of molecular data, such as DNA methylation or proteomic data, could provide a more comprehensive understanding of the mechanisms underlying lung cancer.

3 Results

This chapter presents the results obtained from the differential gene expression analysis and classification of lung cancer samples using selected biomarkers. A series of statistical and visual methods were used to compare tumor and normal samples, identify key genes, evaluate their classification performance, and assess their prognostic value through survival analysis.

3.1 Heatmap Comparison: Tumor vs Normal Samples

Two heatmaps were generated to visualize the expression levels of the top differentially expressed genes across patients for both tumor and normal samples.

As shown in Figure 1, the heat maps show that there are some genes that are expressed differently. Statistical tests were then conducted to see which genes are significantly expressed differently. To ensure that the tumor and normal heatmaps were directly comparable,the color scale used in both plots was standardized. First, the Z-score normalization separately to the gene expression data for Tumor and Normal samples, allowing expression values to be compared across samples. Then the global minimum and maximum Z-scores across both datasets were calculated to establish a consistent color range. By passing these values to the heatmap function, we ensured that the same color represented the same level of gene expression on both heatmaps. This approach allowed for meaningful visual comparison of gene expression patterns between matched Tumor and Normal samples.

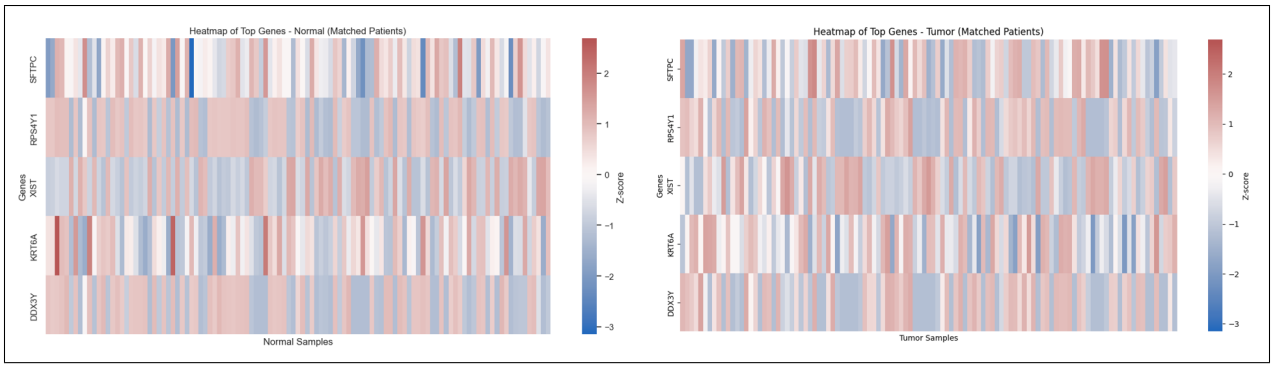

Figure 1: Heatmaps of top expressed genes: (Left) Normal Samples, (Right) Tumor Samples

3.2 Volcano Plot Interpretation

A volcano plot was used to identify genes that were significantly differentially expressed between tumor and normal tissues.

Figure 3 shows that a majority of genes are differentially expressed, with a large proportion exhibiting both high fold changes and low p-values. Genes involved in cytoplasmic pathways were among the most significantly altered.

3.3 Pathway-Level Insights

Genes related to the cytoplasm pathway were significantly enriched in the differentially expressed gene list. The top 10 genes with the highest expression variance within this pathway were selected for further analysis.

3.4 Boxplots of Key Genes

Among the significantly differentially expressed genes, three genes stood out: RTN4RL2, REM1, and HIF3A. These genes demonstrated distinct expression profiles that overlapped minimally between tumor and normal samples, with Mann-Whitney U test p-values all < 0.0001.

These genes were selected for logistic regression modeling to evaluate their diagnostic performance. Various gene combinations were tested using logistic regression, and their performance was measured using the Area Under the ROC Curve (AUC).

Table 1: AUC Scores for Individual and Combined Gene Sets

Gene Set	**AUC**
RTN4RL2 only	65%
RTN4RL2 + HIF3A + REM1	**96%**
Known biomarkers (without REM1)	94%
All known biomarkers (with REM1)	**99%**

As shown in Table 1, the combination of RTN4RL2, HIF3A, and REM1 achieved a strong classification performance (AUC =

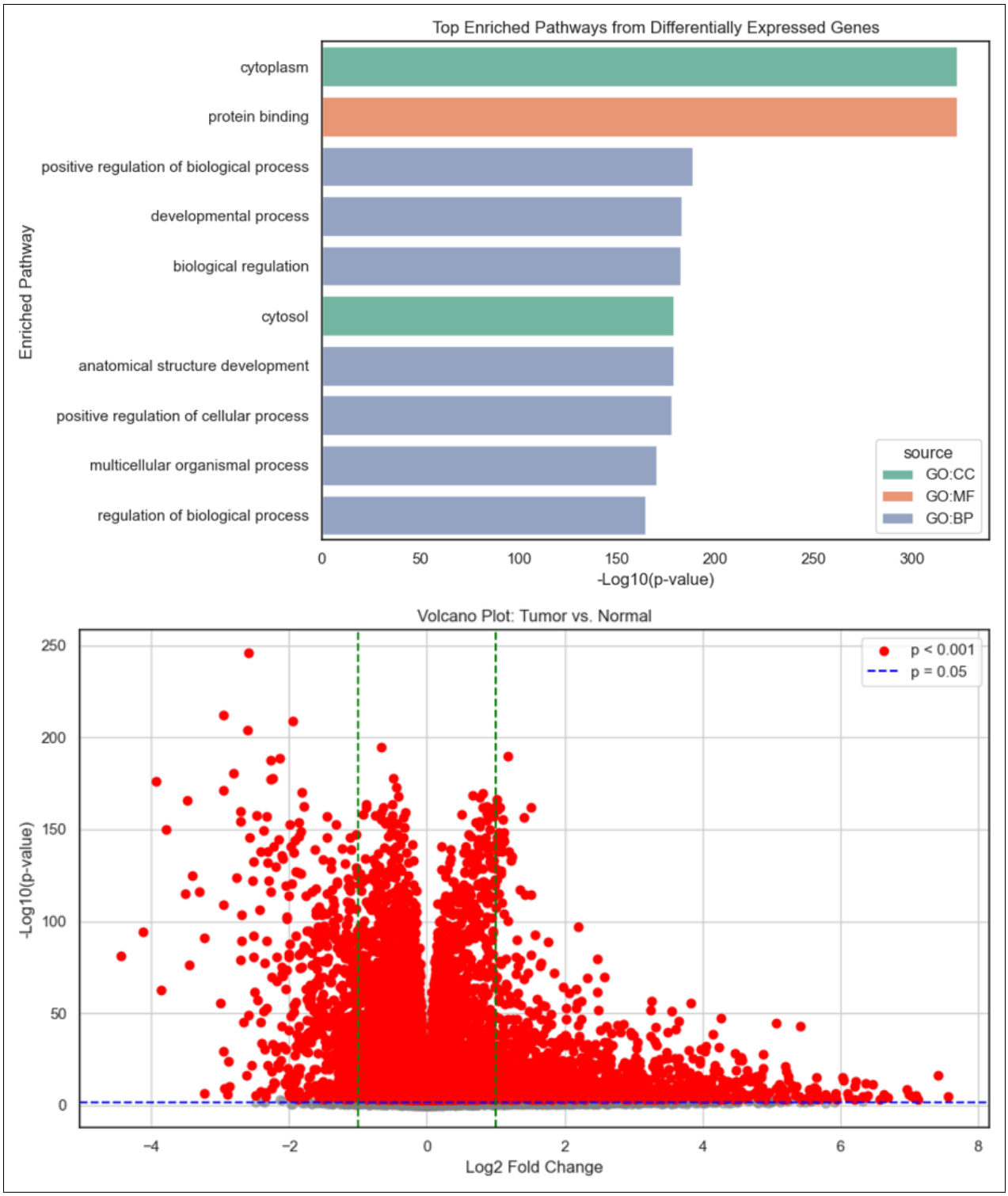

Figure 3: Volcano plot showing differentially expressed genes

96%). Interestingly, the inclusion of REM1 boosted the performance above the set without it, highlighting its importance as a potential biomarker.

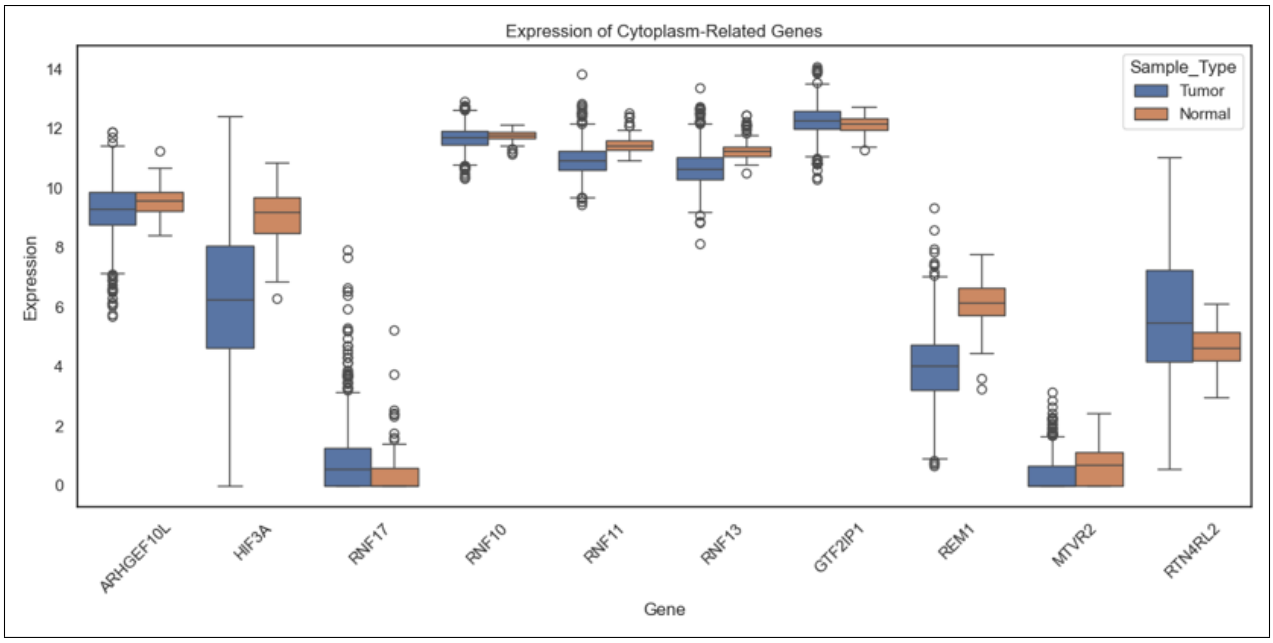

Figure 4: Boxplots of RTN4RL2, REM1, and HIF3A showing significant expression differences

3.6 Gene-specific Survival Analysis

Further survival analysis was conducted on selected genes:

- REM1: No patient with high REM1 expression survived beyond 5000 days (13.7 years).
- RNF13: All patients with high expression of RNF13 died by approximately 5300 days (14.5 years).
- Combined signature (RTN4RL2 + HIF3A + REM1): Patients with low expression of all three had a twice higher survival probability compared to those with high expression.

3.7 Survival Based on Known Biomarkers

When the analysis excluded REM1, survival dropped further—though some patients with high expression survived beyond 5000 days, the probability of survival was only around 10%.

3.8 Logistic regression

The logistic regression model used to differentiate between normal and tumor tissue identi-

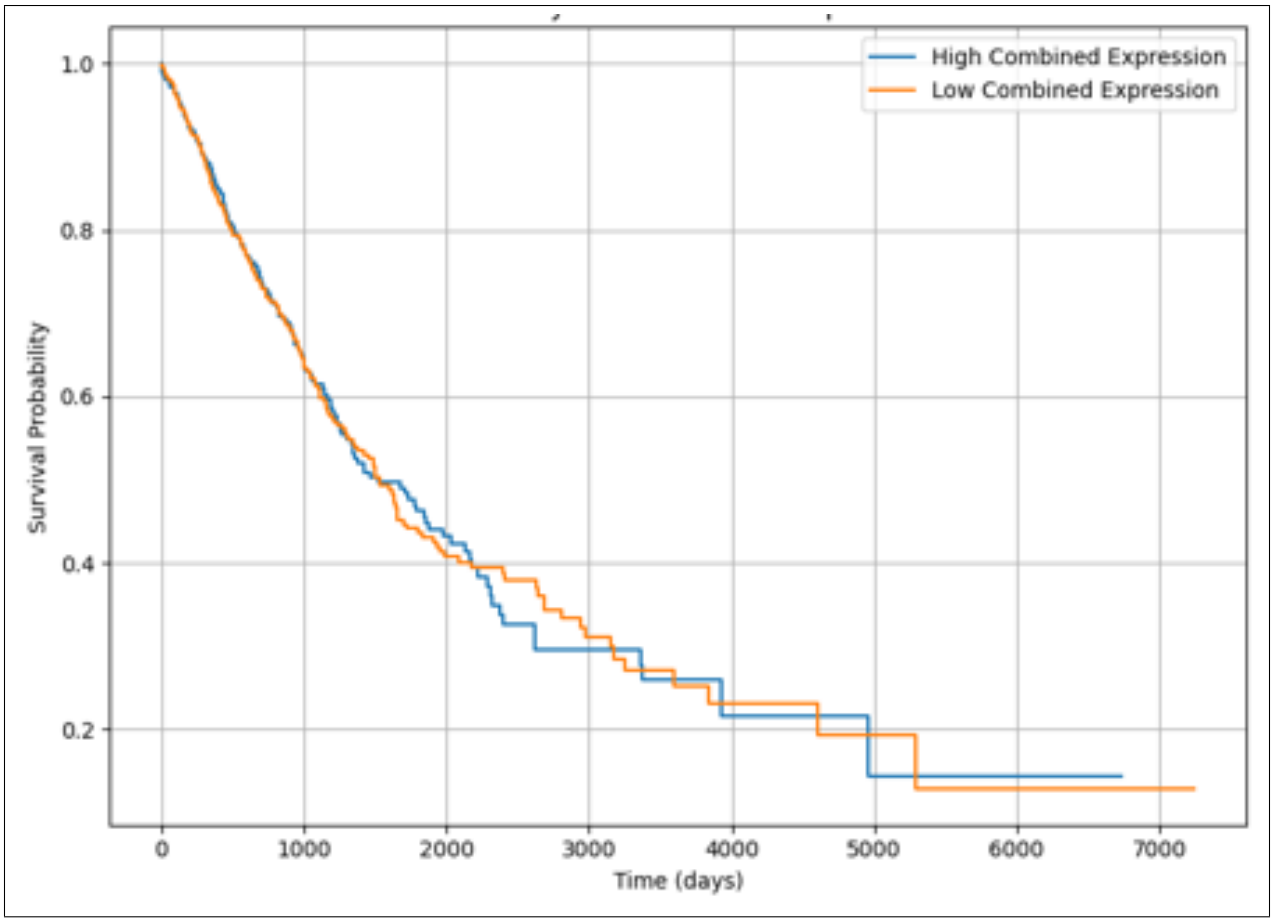

Figure 5: Kaplan-Meier survival plot for all patients

fied three key genes—RTN4RL2, HIF3A, and REM1—as significant predictors.RTN4RL2 has a positive coefficient of 1.44728 in the model, which indicates that higher expression levels of this gene are associated with tumor tissue. The unscaled coefficient for RTN4RL2 is 0.744667, meaning it contributes positively to the likelihood of the sample being classified as a tumor. HIF3A has a negative coefficient of -2.27834, suggesting that higher expression levels of this gene are more likely to be associated with normal tissue. Its unscaled coefficient of -0.951269 further reinforces this relationship, contributing negatively to the classification of the sample as a tumor. REM1 also has a negative coefficient of -2.34028, indicating that increased expression of this gene is associated with normal tissue. The unscaled coefficient of -1.79936 similarly shows a negative impact on the likelihood of classifying a sample as tumor tissue.

Table 2: Model Coefficients and Performance Metrics

Gene	Scaled Coefficient	Unscaled Coefficient
RTN4RL2	1.44728	0.744667
HIF3A	-2.27834	-0.951269
REM1	-2.34028	-1.79936

3.9 Summary of Findings

Tumor samples showed widespread differential gene expression compared to normal samples.Genes linked to the cytoplasm were prominently affected.REM1, HIF3A, and RTN4RL2 showed the strongest potential as diagnostic biomarkers.The gene combination RTN4RL2 + HIF3A + REM1 outperformed other marker sets in both classification and survival prediction. REM1 emerged as a critical gene influencing both diagnosis and prognosis.

4 Discussion

This study provides a comprehensive analysis of gene expression differences between tumor and normal lung tissues, identifying significant molecular changes associated with lung cancer. The clear clustering of tumor vs. normal samples observed in the heatmap confirms the existence of distinct transcriptomic signatures in cancerous tissues. This supports the hypothesis that lung cancer involves both suppression and activation of various gene sets.

The volcano plot highlighted widespread differential expression, suggesting a broad impact on cellular processes. Genes linked to the cytoplasm were among the most affected, pointing to disruptions in intracellular transport, signaling, or metabolic pathways that may be critical for tumor progression.Three genes—RTN4RL2, REM1, and HIF3A—were identified as strong candidates for further investigation due to their clear separation between tumor and normal tissues. Logistic regression showed that the combination of these genes achieved a 96% AUC in distinguishing tumor from normal samples, which is higher than the 94% achieved using known biomarkers without REM1.Interestingly, although RTN4RL2 had modest performance on its own (AUC = 65%), its contribution in combination with REM1 and HIF3A significantly boosted the model's discriminatory power. This finding underscores the potential of REM1 as a novel biomarker, complementing or even outperforming known genes when used in tandem.Survival analy-

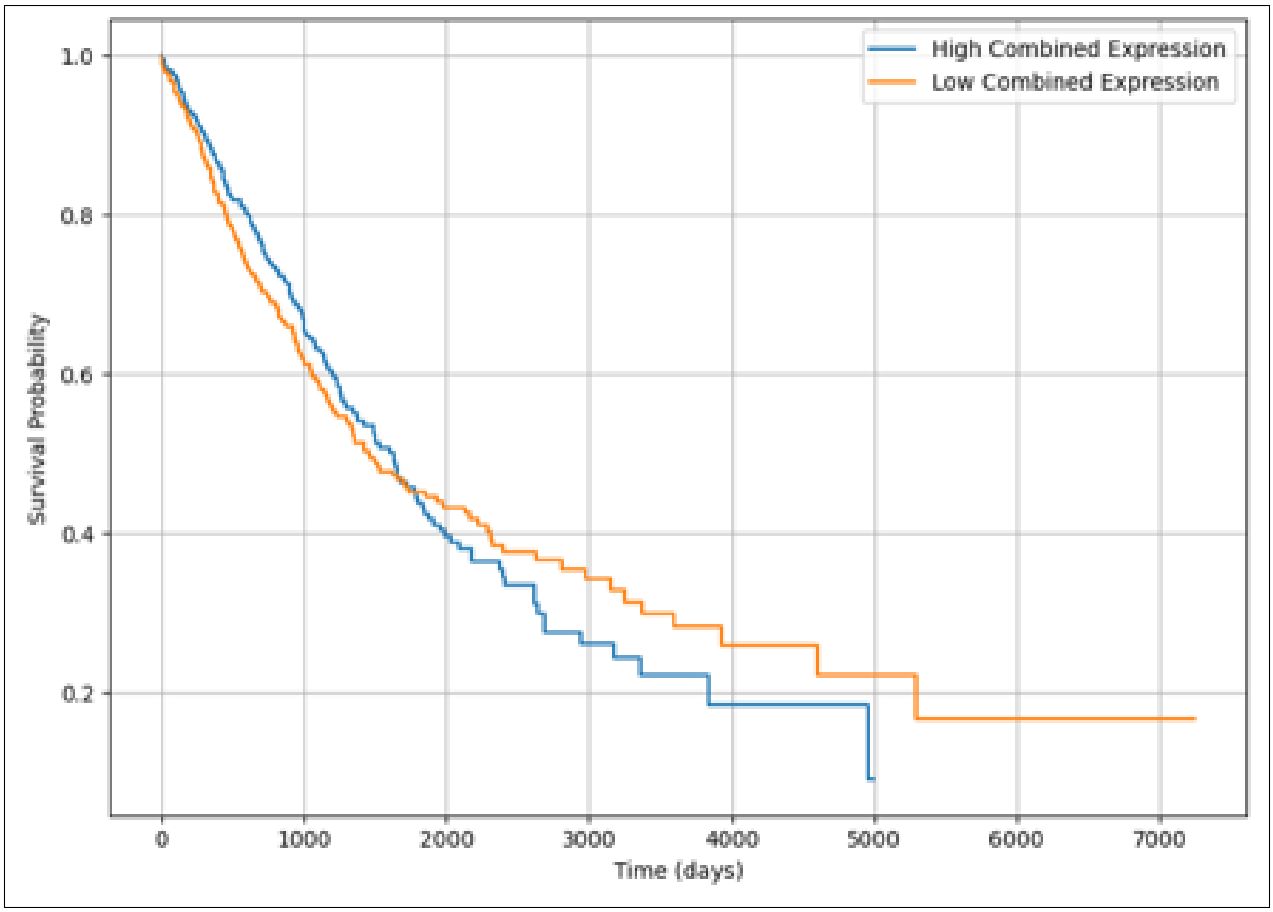

Figure 6: Survival curves based on expression of RTN4RL2, HIF3A, and REM1

sis revealed that high expression of REM1 and RNF13 is associated with poor long-term survival. Notably, patients with high REM1 expression did not survive beyond 5000 days, and similar trends were observed for RNF13. These findings suggest these genes may have prognostic utility beyond their diagnostic value.Combining gene expression data from RTN4RL2, HIF3A, and REM1 showed a marked difference in survival probability, where patients with low combined expression had double the survival chance. This highlights the utility of integrated biomarker panels in not just diagnosis but also outcome prediction.

4.1 Limitations

Several limitations should be acknowledged. First, the analysis was based on retrospective data from TCGA and may not generalize to other populations. Secondly, the study relied solely on gene expression data and did not incorporate multi-omics approaches such as proteomics or methylation profiling, which could offer deeper insights. Lastly, although

the Mann-Whitney U test is appropriate for non-parametric data, alternative statistical methods could provide complementary findings.

4.2 Future Directions

Future work should include validation of these biomarkers in independent cohorts and in functional assays to confirm their biological roles. Integrating additional omics data and exploring machine learning classifiers may further improve prediction accuracy and uncover new molecular subtypes of lung cancer.

Data Availability Statement

The gene expression datasets analyzed in this study are publicly available through The Cancer Genome Atlas (TCGA) at `https://portal.gdc.cancer.gov/`. All processed data, R scripts, and supplementary files used for analysis are available from the corresponding author upon reasonable request.

Author Contributions

Tafadzwa J.M. Mutiro, Ruvarashe Chinyadza, and Vimbainashe H. Kuzanga conceived the project, performed the analysis, interpreted the results, and wrote the manuscript. All authors contributed to the final editing and approved the submitted version.

Acknowledgments

The author gratefully acknowledges the open-access resources provided by TCGA and the support of our Professors, Dr.Sumona Mondal and Prof. Naveen Reddy. Additional thanks to the R programming community for developing packages critical to this analysis.

Supplementary Material

Supplementary tables and additional figures referenced in the manuscript, including full gene lists and detailed statistical outputs, are provided as part of the online supplementary file or are available upon request from the author.

References

[1] W. Pao and K. E. Hutchinson. Impact of resistance mutations on the management of non-small cell lung cancer. *Journal of Clinical Oncology*, 33(2):186–193, 2015.

[2] R. L. Siegel, K. D. Miller, and A. Jemal. Cancer statistics, 2021. *CA: A Cancer Journal for Clinicians*, 71(1):7–33, 2021.

[3] Barbieri, C. E., Chinnaiyan, A. M., Rubin, M. A. (2016). The role of the androgen receptor in prostate cancer. Molecular Cellular Endocrinology, 2016.01.030

[4] Cao, M., Chen, W., Qiu, H. (2021). The burden of lung cancer and attributable risk factors in China, 1990–2019: Findings from the Global Burden of Disease Study 2019. Cancer Communications, 41(9), 1037–1048.

[5] Cancer Genome Atlas Research Network. (2014). Comprehensive molecular profiling of lung adenocarcinoma. Nature, 511(7511), 543–550.

[6] Hanahan, D., Weinberg, R. A. (2011). Hallmarks of cancer: The next generation. Cell,.2011.02.013

[7] Hirsch, F. R., Scagliotti, G. V., Mulshine, J. L., Kwon, R., Curran, W. J., Wu, Y.-L., Paz-Ares, L. (2017). Lung cancer: Current therapies and new targeted treatments. The Lancet, 389(10066), 299–311.

[8] Imielinski, M., Berger, A. H., Hammerman, P. S., Hernandez, B., Pugh, T. J., Hodis, E., ... Garraway, L. A. (2012). Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. Cell, 150(6), 1107–1120.

[9] Jaspers, J. E., Brentnall, A. R., Wessels, L. F. (2019). DNA repair deficiency and the response to immunotherapy. Nature Reviews Cancer, 20(1), 20–34.

[10] Levine, A. J. (2020). p53: 800 million years of evolution and 40 years of discovery. Nature Reviews Cancer, 20(8), 471–480.

[11] Molina, J. R., Yang, P., Cassivi, S. D., Schild, S. E., Adjei, A. A. (2021). Non-small cell lung cancer: Epidemiology, risk factors, treatment, and survivorship. Mayo Clinic Proceedings, 86(4), 362–377.

[12] Pao, W., Hutchinson, K. E. (2015). Chipping away at the lung cancer genome. Nature Medicine, 18(3), 349–351.

[13] Sacco, J. J., Lindsey, J. C., Delpuech, O., Scott, K. N., Cassidy, J. (2017). Resistance to targeted therapies in lung cancer: From molecular mechanisms to treatment strategies. Cancer Drug Resistance, 2(1), 23–40.