Original Research Article | J Adv Med Biomed Res. 2025; 33(161): 293-306

Category: Life Science | Peer Reviewed | Open Access

Journal of Advances in Medical and Biomedical Research | ISSN:2676-6264

Computational Identification of Dysregulated Genes and Pathways in Non-Small Cell Lung Cancer: A Systems Biology Approach

Aditya Kumar Jha[1], Mohammad Ali Abdullah Almoyad[2], Shadma Wahab[3], Garima Gupta[4,5], Khang Wen Goh[6], Amirhossein Sahebkar[7*], Prashant Kesharwani[8,9*]

1. *High-School Junior, VIBGYOR High School, Mumbai, India*
2. *Department of Basic Medical Sciences, College of Applied Medical Sciences, King Khalid University, Abha, Saudi Arabia*
3. *Department of Pharmacognosy, College of Pharmacy, King Khalid University, Abha, Saudi Arabia*
4. *Graphic Era Hill University, Dehradun, India*
5. *School of Allied Medical Sciences, Lovely Professional University, Phagwara, Punjab, India*
6. *Faculty of Business and Communication, INTI International University, Nilai, Malaysia*
7. *Department of Biochemistry, Saveetha Medical College and Hospital, Saveetha Institute of Medical and Technical Sciences, Chennai, Tamil Nadu, India*
8. *Next-Generation Translational Nanomedicine Laboratory, Department of Pharmaceutical Sciences, Dr. Harisingh Gour Vishwavidyalaya, Sagar, Madhya Pradesh, India*
9. *University Institute of Pharma Sciences, Chandigarh University, Mohali, Punjab, India*

Article Info

10.22034/jambr.33.161.293

Received: 2025/08/16;
Accepted: 2025/11/11;
Published Online: 12 Dec 2025;

Use your device to scan and read the article online

***Corresponding author:**
Amirhossein Sahebkar,
Department of Biochemistry, Saveetha Medical College and Hospital, Saveetha Institute of Medical and Technical Sciences, Chennai, Tamil Nadu, India
Email: amir_saheb2000@yahoo.com

Prashant Kesharwani,
Next-Generation Translational Nanomedicine Laboratory, Department of Pharmaceutical Sciences, Dr. Harisingh Gour Vishwavidyalaya, Sagar, Madhya Pradesh, India
Email: prashantdops@gmail.com

ABSTRACT

Background & Objective: Non-small cell lung cancer (NSCLC) stands as the predominant subtype of lung malignancy globally, which persists as a major driver of cancer-related deaths. This study systematically investigates bioinformatics-based investigations aimed at identifying differentially expressed genes (DEGs) and potential therapeutic targets associated with NSCLC.

Materials & Methods: Publicly available RNA expression datasets were analyzed using advanced in silico tools, including the enrichr function from the GSEApy package and R statistical software (v4.3.1), to uncover molecular pathways involved in disease progression. Key oncogenic signaling pathways such as *TGF-β* signaling, Wnt/β-catenin, PI3K-AKT-mTOR, and MAPK were found to be significantly associated with several genes, such as *AGER, ANK3, CSTA, FGG, AGR2, BRCA1, HDAC1, miR-577, TRIM29, PIK3CA,* and *JNK*. Pathway enrichment analysis further identified significant involvement in chronic myeloid leukemia, longevity regulation, thyroid hormone signaling, viral carcinogenesis, Epstein-Barr virus infection, and microRNAs in cancer.

Results: The correlation analysis revealed that *TRIM29* expression was higher in control samples, whereas *AGR2* was prominently expressed in NSCLC samples. Additionally, *HDAC1* and *BRCA1* were identified as promising diagnostic biomarkers.

Conclusion: These findings improve our comprehension of NSCLC disease mechanisms while identifying promising molecular candidates for formulating more potent, customized treatment approaches, potentially improving clinical outcomes for patients.

Keywords: Non-small Cell Lung Cancer (NSCLC), Bioinformatics Analysis, Differentially Expressed Genes (DEGs), Pathway Enrichment, RNA-seq, Therapeutic Targets

Copyright © 2025, This is an original open-access article distributed under the terms of the Creative Commons Attribution-noncommercial 4.0 International License which permits copy and redistribution of the material just in noncommercial usages with proper citation.

1. Introduction

1.1 Overview of Lung Cancer

Lung cancer stands as a leading contributor to cancer fatalities across the globe. In 2022, an estimated 2.2 million new cases of lung cancer and approximately 1.8 million deaths were recorded, accounting for approximately 18% of all cancer deaths (1). In many regions, lung cancer surpasses the collective deaths attributed to colorectal, breast, pancreatic, and prostate cancers, making it the foremost contributor to cancer mortality in both sexes (2). Tobacco smoking remains the dominant risk factor for lung cancer; risk increases substantially with duration and cumulative exposure and declines only gradually after cessation. Recent screening and prevention guidelines use long-term exposure thresholds (e.g., 20 pack-years) to define eligibility (3). At the national level, the United States reported 218,893 new lung cancer diagnoses in 2022 and 131,584 lung cancer deaths in 2023, underscoring the continuing domestic burden despite prevention and early-detection efforts (3). According to GLOBOCAN 2022, data revealed approximately 1.8 million cancer-related deaths worldwide in 2022. Lung cancer accounted for approximately 18.7% of these deaths, thereby constituting the most significant contributor to cancer mortality globally.

Non-small cell lung cancer (NSCLC) originates from pulmonary epithelial cells, encompassing histological variants including adenocarcinoma, squamous cell carcinoma, and large cell carcinoma. The development of NSCLC involves a series of genetic and epigenetic modifications that interfere with normal cellular homeostasis, ultimately causing unregulated cell division and malignant tumor development. The progression from a healthy state to a malignant lung cancer phenotype is believed to occur gradually through various genetic and epigenetic modifications, leading to invasive cancer through clonal expansion. As the primary cancer develops, additional genetic and epigenetic alterations accumulate during clonal expansion, influencing invasion, metastasis, and resistance to cancer treatments. TP53 gene defects rank among the most frequently detected genomic abnormalities in pulmonary malignancies. This gene encodes the p53 protein, which regulates cell cycle progression and prevents the accumulation of damaged DNA. Mutations in this gene result in the loss of its tumor suppressive function allowing tumor cells to proliferate uncontrollably. Moreover, research has revealed that *EGFR* (Epidermal Growth Factor Receptor) and *ALK* (Anaplastic Lymphoma Kinase) are key oncogenic drivers in specific subsets of non-small cell lung cancer (NSCLC) (4). Thus, integrating knowledge about a patient's tumor characteristics and genetics will significantly facilitate the optimal selection of treatments for individual patients.

1.2 mRNA & lncRNA Expression- A Promising Approach for Targeted Therapeutic of Cancers

Messenger RNA (mRNA) is a single-stranded ribonucleic acid generated through DNA transcription. It functions as an intermediary molecule that transports genetic information from the DNA in the nucleus to ribosomes in the cytoplasm, where it serves as a template for directing the sequential assembly of amino acids into functional proteins. Correlated mRNA and miRNA profiles within cancer signatures function together to regulate key signaling pathways driving malignancy. These gene regulatory networks operate in both cancer and non-cancer contexts, though disease status significantly influences the degree of molecular interactions within the network. Theoretically, mRNA-based therapeutic approaches enable the production of diverse proteins by harnessing endogenous cellular translation machinery in the transfected cell, both *in vitro* and *in vivo*. Compared to DNA-based drugs, mRNA transcripts demonstrate superior delivery efficiency and enhanced safety profiles as their activity occurs exclusively within the cytoplasm, eliminating the requirement for nuclear translocation. Critically, mRNA therapies pose no genomic integration risks, thereby preventing insertional mutagenesis or unintended infectious complications. Extensive research demonstrates that mRNA delivers improved cellular delivery efficiency and extended protein expression duration, surpassing both DNA and traditional protein drugs in therapeutic performance. By triggering immediate cytoplasmic translation without requiring genomic integration, mRNA minimizes insertional mutation risks which are frequently associated with DNA-based and protein therapeutic approaches (5).

Long non-coding RNAs (lncRNAs) are functional RNA molecules that are characterized by an RNA sequence exceeding 200 nucleotides that lack the ability to encode proteins (6). They are transcribed by RNA polymerase from various regions of the genome, including intergenic regions, introns of protein-coding genes, and antisense strands of coding genes. They can act as signals, decoys, guides, or scaffolds to regulate gene expression. LncRNAs are involved in various biological processes, such as development, metabolism, and disease (7). They have been implicated in diseases such as cancer, where they can function as oncogenes or tumor suppressors.

mRNA-based therapies show promising potential for protein deficiency disorders by facilitating persistent translation that produces stable protein levels, delivering enhanced efficacy compared to standard peptide therapeutics. Meanwhile, lncRNAs offer complementary therapeutic value through gene regulatory mechanisms (6). Utilizing expression data of lung cancer and control patients from publicly available datasets, i.e. Gene Expression Omnibus, TCGA, etc. and subsequently analyzing them through machine learning algorithms to

identify novel mRNA signatures after rigorous analysis may reveal novel therapeutic targets for lung cancer.

2. Materials and Methods

2.1 Study Objectives

This research aims to analyze *TGF-β* induced cancer progression and Wnt/β-catenin signaling pathway, PI3K-AKT-mTOR pathway, MAPK (mitogen-activated protein kinase) pathways for identification of a correlation between them and study *HDAC1* (protein histone deacetylase 1), *miR-577*, *TRIM29* (Tripartite motif-containing 29), *PIK3CA* (phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha), *JNK* (c-Jun N-terminal kinase) genes through in-silico methods to validate them as dysregulated genes in NSCLC with potential therapeutic applicability. Moreover, through this research, pathways with higher correlations would be prioritized and those with low correlations would eventually be sidelined for further analyses.

2.2 Gene Association and Pathway Enrichment Analysis

Pathway enrichment analysis is a computational method for detecting functional pathways that contain disproportionately high numbers of differentially expressed genes (DEGs) in a given dataset. In this study, pathway enrichment analysis helps in identifying the key signaling pathways and biological processes that are dysregulated in the disease. A previous study conducted pathway enrichment analysis using both upregulated differentially expressed genes (uDEGs) and downregulated differentially expressed genes (dDEGs) in NSCLC samples compared to control samples. Findings demonstrated that mitotic cell cycle regulation constitutes a critical molecular pathway driving NSCLC pathogenesis.

For this study, pathway enrichment analysis was performed using the GSEApy library on a gene list containing *'HDAC1'*, *'hsa-miR-577'*, *'TRIM29'*, *'PIK3CA'*, and *'JNK'*. The analysis identifies enriched pathways related to these genes and visualizes them based on their adjusted p-values. Additionally, the code retrieves and displays pathway information from the KEGG database for specific pathway IDs. Results from pathway enrichment analysis were visualized, with the top-enriched pathways displayed in the study.

Enrichment results were extracted and organized in tabular format in a DataFrame. The pathways were then sorted by adjusted p-values to identify the most significant associations. The GSEApy library performs Gene Set Enrichment Analysis (GSEA) using a variation of the Kolmogorov-Smirnov (KS) test to determine the enrichment score and p-value for each pathway. P-values were calculated through permutation testing, involving random shuffling of gene labels multiple times to establish statistical significance. The enrichment scores were recalculated for each permutation, with p-value representing the proportion of permutation exceeding the observed score. Lower p-values indicate stronger evidence of genuine enrichment, while higher values suggest results may be due to chance.

This study used the enrichr function to perform pathway enrichment analysis based on a list of input genes. A gene set database was used for enrichment analysis rather than importing raw gene expression data from individual studies. In this case, 'KEGG_2019_Human' represented the KEGG pathway database for humans. This identifies and visualizes biological pathways significantly associated with a list of input genes, aiding in the interpretation of gene-related research findings.

2.3 Bioinformatics Analysis

Differentially expressed genes analysis used R statistical software version 4.3.1 and the GEOQuery package. The `GEOquery` package was used to download gene expression data from the GEO dataset 'GSE7670', i.e, Data type- Expression profiling by array. This dataset focuses on molecular characterization of tumor biology to improve diagnosis, treatment approaches, and prognostic evaluation in lung cancer patients, comprising genetic information from 66 samples. The expression data were extracted using the `exprs` function and specific genes of interest were selected from the literature review such as "201209_at", "204369_at", "202504_at" among others. To visualize the expression patterns of these genes, the `pheatmap` function was employed from the `pheatmap` package. This function creates a heatmap, a graphical representation that allows us to visualize the relationships and patterns in gene expression data. The `cluster_rows` and `cluster_cols` parameters are set to `TRUE` to cluster both genes and samples within the heatmap for better pattern recognition. For analysis and comparison of the frequency of mutations in different transcript regions and regulatory regions in NSCLC, the gene expression data from the dataset were interpreted in the form of a boxplot for the gene set.

2.4 Literature Review

In this study, gene sets and microRNAs associated with the apoptosis pathway in lung cancer were identified from databases and published research. A comprehensive literature review was conducted using free search engines that provided access to a database of related references and abstracts in the life sciences and biomedical fields, such as Google Scholar, PubMed and ScienceDirect. Keywords associated with the targeted pathways and dysregulations were used to locate the most pertinent research studies, with emphasis on the gene sets identified in this study to establish their roles in pathway dysregulation and NSCLC pathogenesis.

3. Result

NSCLC-associated Pathways & RNA Signatures for Targeted Therapeutics in Lung Carcinomas

3.1 RNA Signatures in Lung Cancer

The reviewed studies focused on transcriptome mRNA signatures in non-small cell lung cancer to identify prospective biomarkers and therapeutic candidates. They used bioinformatic analysis and data visualization techniques to analyze gene expression profiles in NSCLC tissues and to compare them with those of healthy pulmonary tissues. A study conducted for an integrated analysis of miRNA and mRNA expressions in NSCLC identified that hsa-miR-96 was consistently up-regulated in all six NSCLC tissues, suggesting its potential as a noninvasive diagnostic marker (8). It also highlighted miRNA regulation in NSCLC development and its correlation with genes related to CpG islands. Another study compared the expression patterns of long noncoding RNAs (lncRNAs) and mRNAs in lung cancer. It found that differentially expressed lncRNA target genes encompassed almost all the differentially expressed genes from mRNA. Top downregulated lncRNAs, especially *FENDRR*, and mRNAs, i.e., *AGER*, *SFTPC*, *GKN2*, *CLDN18*, and *CLEC3B* in lung cancer showed remarkable fold changes (9). Researchers investigated histone deacetylases (HDACs) expression in lung cancer specimens. Analysis revealed elevated HDAC1 mRNA expression in stage III/IV disease and T3/T4 tumor classifications exhibiting a positive correlation with cancer progression. This evidence positioned HDAC1 as a viable molecular target for precision therapy in lung cancer. A customized treatment strategy incorporating *EGFR* mutations and *BRCA1* mRNA expression levels in advanced NSCLC revealed *RAP80* as a modulating factor influencing this tailored chemotherapy approach, wherein patients with EGFR mutations received erlotinib while mutation-negative patients underwent BRCA1-guided chemotherapy selection. Moreover, RNA sequencing analysis of miRNA expression between NSCLC specimens and normal lung tissue identified numerous dysregulated miRNAs specifically increased of *miR-577*, *miR-301b*, *miR-944, miR-891a,* and *miR-615-3p* expression and decreased *miR-338-3p* expression. A separate study compared gene expression patterns between two primary lung cancer types—Non-Small-Cell Lung Cancer (NSCLC) and Small-Cell Lung Cancer (SCLC) which revealed dysregulated genes in each type when compared with healthy lung specimens. The study highlighted five genes *TRIM29, ANK3, CSTA, FGG,* and *AGR2* as candidates for targeted therapy development (9).

In summary, all the research studies reviewed contribute to our understanding of RNA signatures in NSCLC. The findings from these studies suggest designing customized and enhanced therapeutic regimens for NSCLC patients, thereby improving therapeutic responses and extending patient survival.

3.2 Signaling Pathways in NSCLC

The pathways involved in Non-Small Cell Lung Cancer (NSCLC) development are key to oncology research. It is distinguished by various molecular alterations, which are orchestrated through complex signaling pathways that control cellular processes. Research in this domain assists in a variety of ways. For instance, understanding NSCLC pathways enables precision oncology by providing insight and allowing for the development of personalized therapies to maximize treatment effectiveness and reduce adverse reactions, resistance mechanisms are for comprehension into resistance mechanisms, which can then be used to develop strategies to combat treatment resistance, and pathway analysis for identification of biomarkers that can be used for early detection and diagnosis (10).

Ultimately, pathways improve precision medicine, promote therapeutic innovation, uncover resistance mechanisms, and promise biomarker-based diagnostics, providing hope for improved patient results. There are various pathways that are involved in the progression and development of NSCLC. Genetic signatures associated with these pathways are used for therapeutic target identification (Figure 1).

The table below sums up the conclusions from the reviewed research studies, giving a clear interpretation of the genes, their interaction pathway and their association with non-small cell lung cancer (Table 1).

Essential cellular processes such as cell division and death are controlled by intracellular signaling molecules that interact within regulatory networks. Lung cancer exhibits dysregulation of several pathways, specifically:

1) **RTK pathway-** The receptor tyrosine kinase (RTK) pathway includes genes such as *EGFR* and *ALK*. These genetic mutations produce sustained downstream pathway activation, particularly within RAS signaling (11).

2) **RAS pathway-** The RAS pathway, which is involved in cell division, survival, and proliferation, is essential in lung cancer. Aberrant activation of the RAS pathway can result from mutations in genes such as KRAS and NRAS (12).

3) **BRAF/MAPK pathway-** Genetic alterations in BRAF, particularly the V600E variant, trigger sustained MAPK cascade activation. This signaling network controls cellular expansion, division, and viability (13).

4) **PI3K pathway-** Significantly implicated in lung cancer pathogenesis, this signaling network becomes hyperactivated through PIK3CA and related gene mutations, driving cellular proliferation, survival, and metabolic processes essential for tumor growth (13).

5) **LKB1/AMPK pathway-** This cascade functions as a tumor-suppressive mechanism controlling cellular metabolic processes and proliferation. LKB1 functional impairment, typically resulting from genetic alterations, causes pathway deregulation that promotes lung malignancy development (14).

6) **TP53 pathway-** Lung cancer is often associated with changes in the TP53 pathway, which contains the tumour suppressor gene TP53. Mutations in TP53 can interfere with DNA repair and cell cycle

regulation, causing genomic instability and tumour growth (15).

7) **RB1/MYC pathway-** Another significant pathway in lung cancer is this one. Loss of cell cycle regulation and deregulation of the MYC oncogene, which is essential in cell survival and proliferation, can arise from mutations in the RB1 gene (16).

8) **Wnt/β-catenin pathway-** The onset and spread of lung cancer are linked to this route. Abnormal cell proliferation and differentiation can result from dysregulation of this pathway, which is frequently caused by mutations in genes like APC and CTNNB1 (5).

9) **Epigenetic pathways-** Lung cancer is also influenced by epigenetic processes, such as histone changes and DNA methylation. Histone alterations and abnormal DNA methylation patterns can change gene expression and aid in the growth of tumours (17).

10) ***TGF-β* pathway-** Performs a multifaceted role in lung cancer, helping to control the tumor microenvironment and promote tumor growth. TGF-β is regarded as a canonical tumour suppressor since it causes normal epithelial cells to undergo apoptosis and impede cell proliferation (18).

In this study, four (MAPK, PI3K, Wnt/β-catenin, *TGF-β*) of the ten major pathways associated with NSCLC have been reviewed and analyzed for further findings about their correlation with other pathways and if genes identified in previous research can potentially serve as a therapeutic target for lung cancer (Figure 2).

3.3 Correlation of Gene-sets with Pathways

3.3.1 *HDAC1* in Wnt/β-catenin signaling pathway

HDAC1, or histone deacetylase 1, has been shown to contribute to the Wnt/β-catenin signaling pathway in lung cancer. Studies showed that HDAC1 enhances β-catenin's stability and transcriptional activity by interacting with it and encouraging its deacetylation (18). This interaction between *HDAC1* and β-catenin enhances Wnt/β-catenin signaling activation, which exerts a key function in lung cancer progression and metastasis (19). Additionally, *HDAC1* has been shown to regulate the expression of Wnt pathway components, such as Wnt ligands and receptors, further contributing to the activation of the Wnt/β-catenin pathway. *HDAC1* can interact with and deacetylate other proteins involved in the pathway, such as TCF/LEF transcription factors, enhancing their transcriptional activity. Collectively, these observations indicate that HDAC1 functions play an important role in activating Wnt/β-catenin signaling within lung malignancies, underscoring its promise as a therapeutic intervention target for this disease.

3.3.2 *miR-577* in *TGF-β* induced cancer progression

It has been established that miR-577 contributes to the development of lung cancer caused by *TGF-β*. A study revealed that when *TGF-β* was administered to lung cancer cells, the expression of *miR-577* was markedly elevated. These findings suggest that *miR-577* contributes to *TGF-β* induced cancer progression in lung cancer by inhibiting the negative feedback regulation of *TGF-β* signaling through *SMAD7*. The downregulation of *SMAD7* by *miR-577* is specific to the protein level and not the transcript level, as confirmed by expression analysis of clinical samples. This finding highlights the importance of post-transcriptional regulation of *SMAD7* in coordinating *TGF-β* signaling in cancer cells (20).

3.3.3 *TRIM29* (Tripartite motif-containing 29) in Wnt/β-catenin signaling pathway

TRIM29, also known as ATDC (Ataxia Telangiectasia Group D Complementing), has been shown to play a role in the Wnt/β-catenin signaling pathway in lung cancer. It has been discovered that TRIM29 is elevated in lung cancer and linked to the growth and spread of the tumour. TRIM29 has been shown to activate β-catenin signalling, which in turn increases lung cancer cell motility and invasion. Specifically, *TRIM29* silencing decreases β-catenin transcriptional activity, whereas *TRIM29* overexpression increases it. These results imply that TRIM29 possesses an integral function in lung cancer via stimulating the Wnt/β-catenin signalling pathway.

3.3.4 *PIK3CA* (phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha) [(human)] in PI3K/AKT/mTOR pathway

PIK3CA is a gene that encodes the p110β ± catalytic subunit of the class I PI3K enzyme. Attenuation in *PIK3CA* is frequently found in human cancer, including lung cancer. These mutations lead to increased PI3K enzyme activity and activation of the PI3K/AKT/mTOR pathway. The PI3K/AKT/mTOR pathway represents a cellular signaling cascade that regulates proliferation, viability, and metabolic processes (21). This pathway's activation accelerates the growth of lung cancer tumours and leads to the emergence of treatment resistance. As a result, PIK3CA and the PI3K/AKT/mTOR pathway have been investigated for potential lung cancer treatment modalities.

3.3.5 *JNK* (c-Jun N-terminal kinase) in MAPK (mitogen-activated protein kinase) signaling pathway

JNK (c-Jun N-terminal kinase) belongs to the mitogen-activated protein kinase (MAPK) family and regulates diverse cellular functions, including proliferation, differentiation, viability, and programmed cell death (14). In lung cancer, *JNK* regulates LSCC (lung squamous cell carcinoma) development. Loss of *LKB1* (liver kinase B1) activity, a key suppressor gene of LSCC, leads to decreased expression of *MKK7* (MAPK kinase 7), which in turn reduces *JNK1/2* phosphorylation and lowers *JNK1/2* activities. Activation of *JNK1/2* signaling has been shown to induce cell apoptosis in multiple SCC (squamous cell carcinoma) cells, including lung SCC, suggesting that *JNK1/2* inactivation may promote or lead to SCC formation and progression. Additionally, a significant percentage of human LSCC has the JNK1/2 pathway deactivated, and JNK1/2 activity is favourably

correlated with the survival rates of patients with head and neck, cervical, and lung SCC. According to these results, JNK1/2 suppresses the growth of LSCC, and boosting JNK1/2 activity may be a useful treatment strategy for LSCC (22).

3.4 Pathway Enrichment Analysis of Genes

The KEGG Pathway enrichment results are shown in Figures 3 and 4. The common KEGG Pathways from the gene set were Chronic myeloid leukaemia, Longevity regulating pathway, Thyroid hormone signaling pathway, Viral carcinogenesis, Epstein-Barr virus infection and MicroRNAs in cancer for *'HDAC1', 'hsa-miR-577', 'TRIM29', 'PIK3CA'*, and '*JNK'* genes. All these pathways are associated with lung cancer. These results can be further interpreted for either inclusion or exclusion from studies related to lung carcinomas. A p-value of less than 0.05 is a frequently used cutoff for statistical significance. Pathways with adjusted p-values below 0.05 are often considered significant. As the p-value reduces, the stringency and significance of the associated pathway increase.

3.5 Differentially Expressed Genes (DEGs) Bioinformatics Analysis

Correlation heat maps of differentially expressed genes across samples provide valuable insights into NSCLC by identifying patterns of gene expression and relationships between genes and samples. In NSCLC, correlation heat maps reveal the similarity or dissimilarity of gene expression profiles between different samples, such as tumor and normal control samples. It also shows the correlation of gene expression levels between different experimental conditions or treatments, such as comparing the expression profiles of NSCLC cells treated with different compounds. Additionally, correlation heat mapping identifies clusters or groups of genes with similar expression patterns, which provide insights into potential gene regulatory networks or functional categories (23, 24). By visualizing the correlation between gene expression profiles, correlation heat mapping identifies genes that are co-expressed or co-regulated, providing valuable information about potential biomarkers or therapeutic targets in NSCLC. For the correlation heat-mapping in this study, nine genes, which were identified as dysregulated in lung cancer through literature review, and were *HDAC1, PIK3CA, TRIM29, AGER, ANK3, CSTA, FGG, AGR2, BRCA1* were used in the GSE7670 dataset from the GEO (Gene Expression Omnibus) database. The results from the correlation heatmap show strong correlation with the gene sets shortlisted (Figure 5 and 6).

The boxplot analysis revealed that *HDAC1* in the samples indicated higher gene expression in NSCLC patients than in controls, and interestingly reveals many individual data points outside the whiskers, requiring further investigation for the *HDAC1* gene. Similar results were observed for the *BRCA1* gene in the dataset. Except for *PIK3CA* and *AGER*, all the gene sets showed positive skewness in the gene expression profiles of the 66 samples.

Table 1. Top 6 genes shortlisted for pathway analysis and further detailed review.

Sr. No.	Gene/RNA	Pathway	Association/Context
1.	*HDAC1* (protein histone deacetylase 1)	Wnt/β-catenin signaling pathway	Disrupted *HDAC1* expression and enzymatic activity alters transcriptional regulation of Wnt pathway genes. Specific inhibitors may represent a potential therapeutic strategy to modulate the Wnt/β-catenin pathway in NSCLC.
2.	*miR-577*	*TGF-β* induced cancer progression	*TGF-β*, a multifunctional cytokine, plays a complex role in cancer development. Participates in multiple stages of malignancy, encompassing tumor initiation, advancement, and metastatic spread
3.	*TRIM29* (Tripartite motif-containing 29)	Wnt/β-catenin signaling pathway	Studied in the context of various cancer types, but its specific role in NSCLC and its association with the Wnt/β-catenin pathway requires further investigation through exploratory data analysis.
4.	*PIK3CA* (phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha) [(human)]	PI3K/AKT/mTOR pathway	Mutations trigger sustained PI3K-AKT-mTOR activation, facilitating tumor survival, growth, and apoptosis evasion. Aberrant PI3K-AKT-mTOR signaling, commonly driven by PIK3CA alterations, serves as a treatment intervention point in NSCLC.

Sr. No.	Gene/RNA	Pathway	Association/Context
5.	*JNK* (c-Jun N-terminal kinase)	MAPK (mitogen-activated protein kinase)	Enhanced *JNK* pathway activity correlates with tumor expansion and NSCLC advancement. Activated *JNK* has the potential for targeted therapies as inhibitors of *JNK* signaling can cease the pro-tumorigenic effects of this pathway in NSCLC cells.
6.	*AGER* (Advanced Glycosylation End-Product Specific Receptor)	NF-κB signaling pathway	Member of the immunoglobulin superfamily of cell surface receptors. Abnormally expressed in lung cancer.

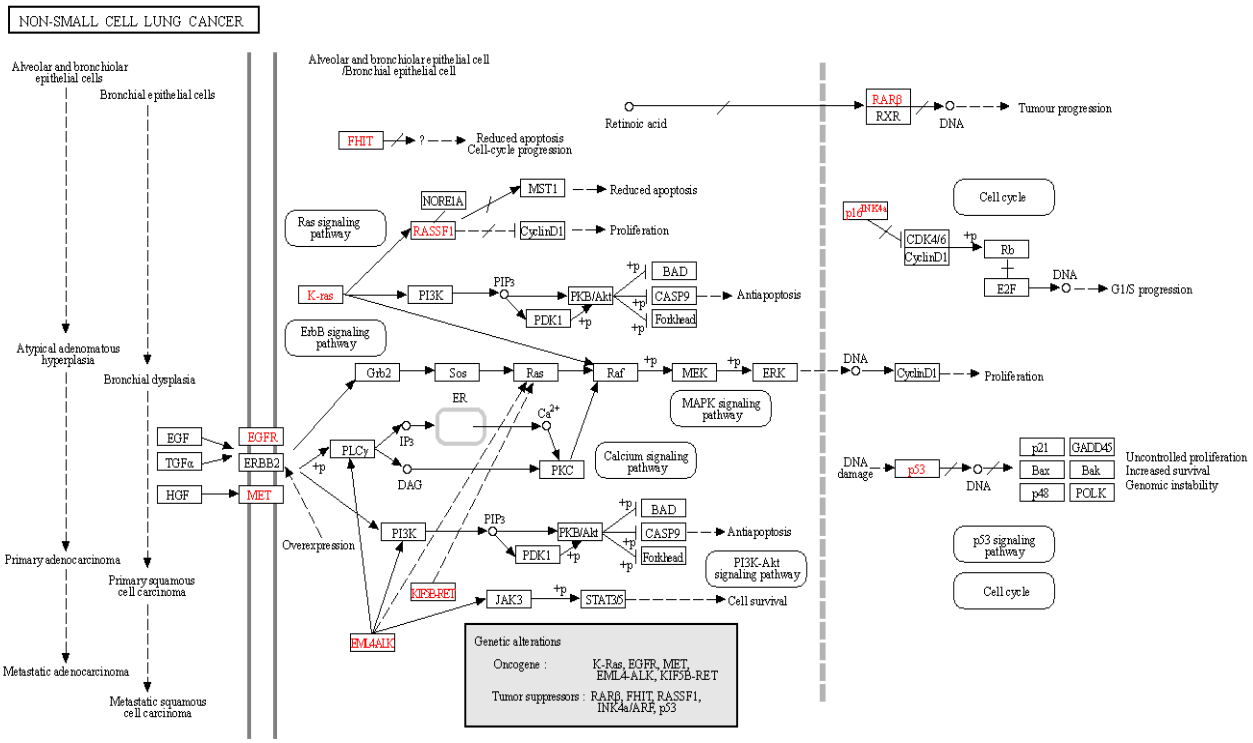

Figure 1. Schematic representation of the interaction between various stages of NSCLC pathways and genes in the human body. Source: KEGG- Bioinformatics resource for deciphering the genome (**Prepared by Authors, 2025**).

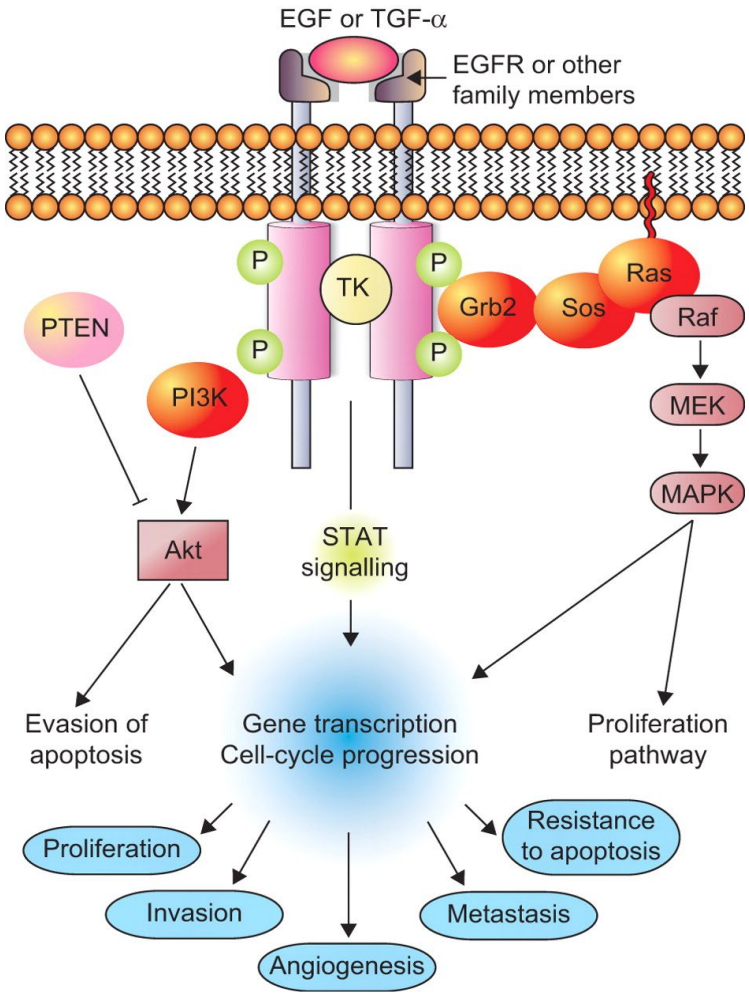

Figure 2. Epidermal growth factor receptor (*EGFR*) pathway. Ligands, such as epidermal growth factor (*EGF*), transforming growth factor (TGF)-α, or others, bind to the homo- and heterodimer kinase domain (TK), resulting in activation and receptor transphosphorylation. This creates docking sites for the adaptor proteins, Grb2 and Sos, which recruit Ras and phosphatidylinositol 3-kinase (PI3K), leading to the formation of two major signaling pathway branches, Ras/MAPK and PI3K/Akt. These networks result in, amongst others, proliferation, evasion of apoptosis and angiogenesis. MAPK: mitogen-activated kinase-like protein (**European Respiratory Journal**).

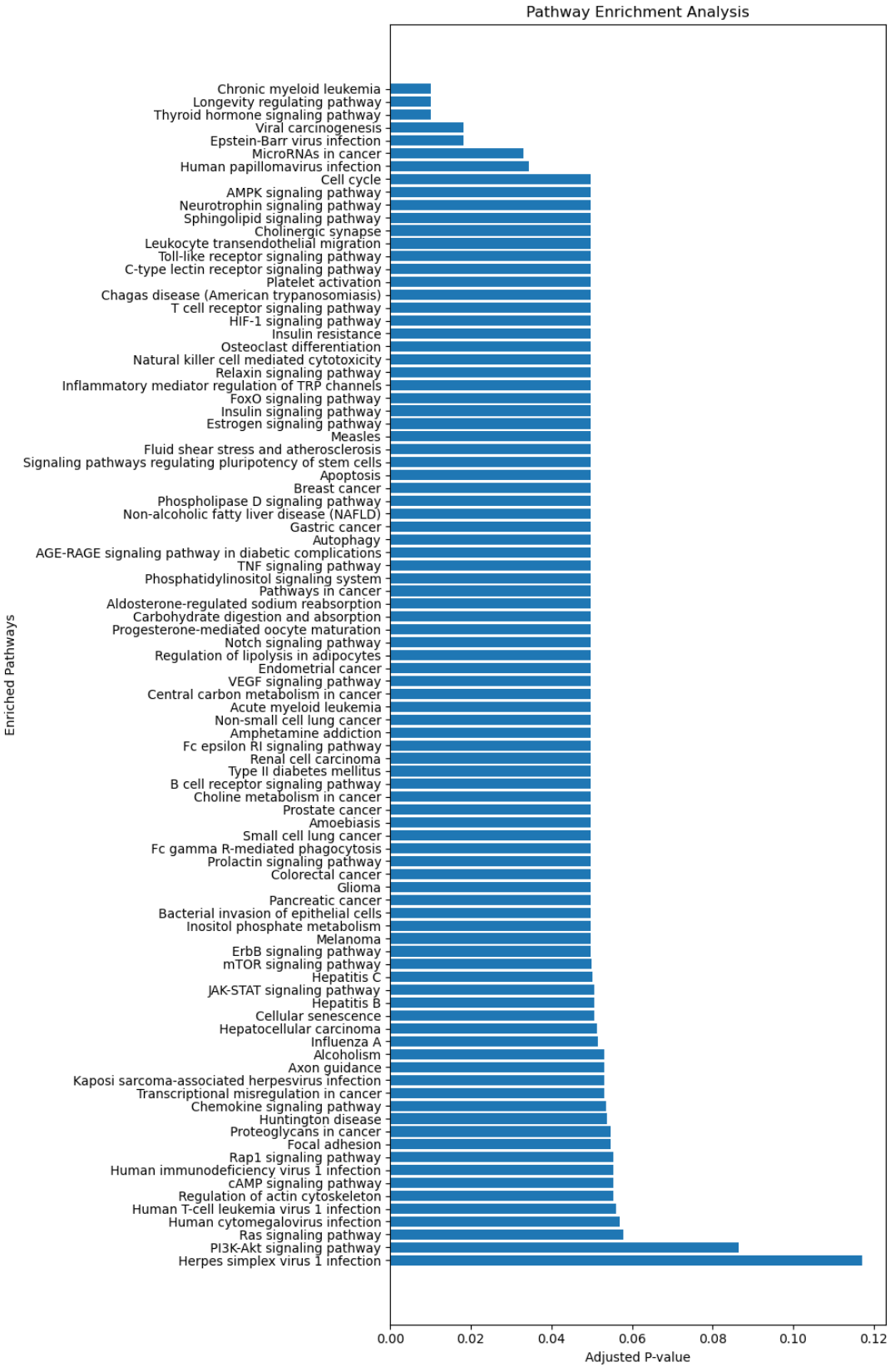

Figure 3. Pathway enrichment analysis for gene-set: *'HDAC1'*, *'hsa-miR-577'*, *'TRIM29'*, *'PIK3CA'*, *'JNK'*; enrichr function from GSEApy module for identification of relevant pathways associated with NSCLC leading to metastatic lung cancer (**Prepared by Authors, 2025**).

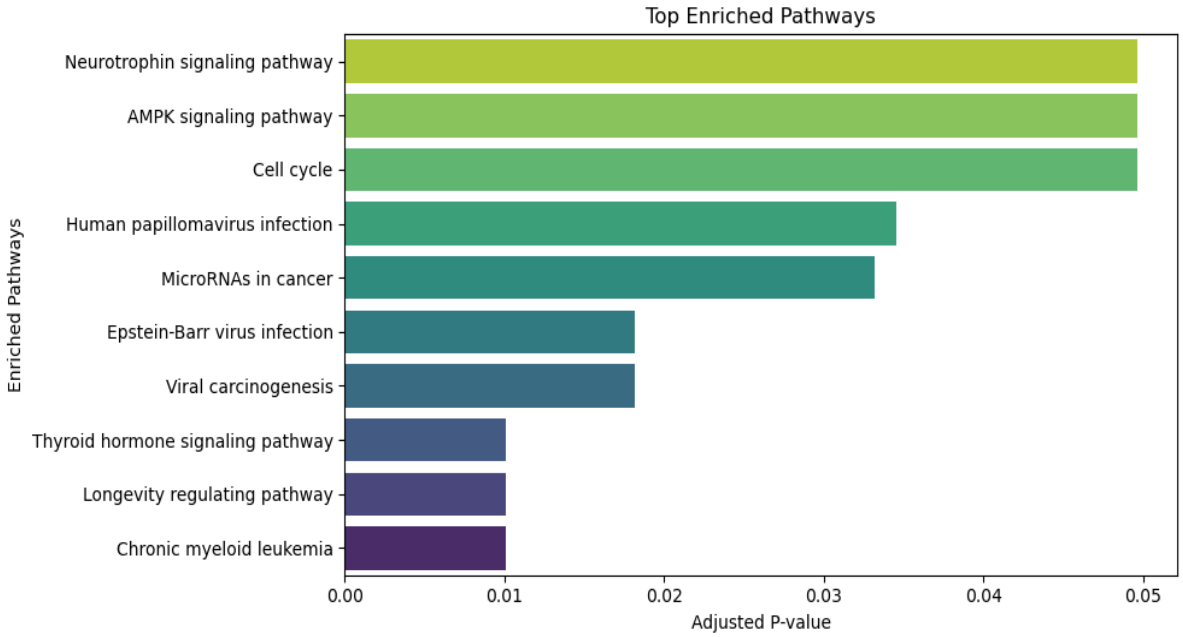

Figure 4. Top 10 enriched pathways from the pathway enrichment analysis (**Prepared by Authors, 2025**).

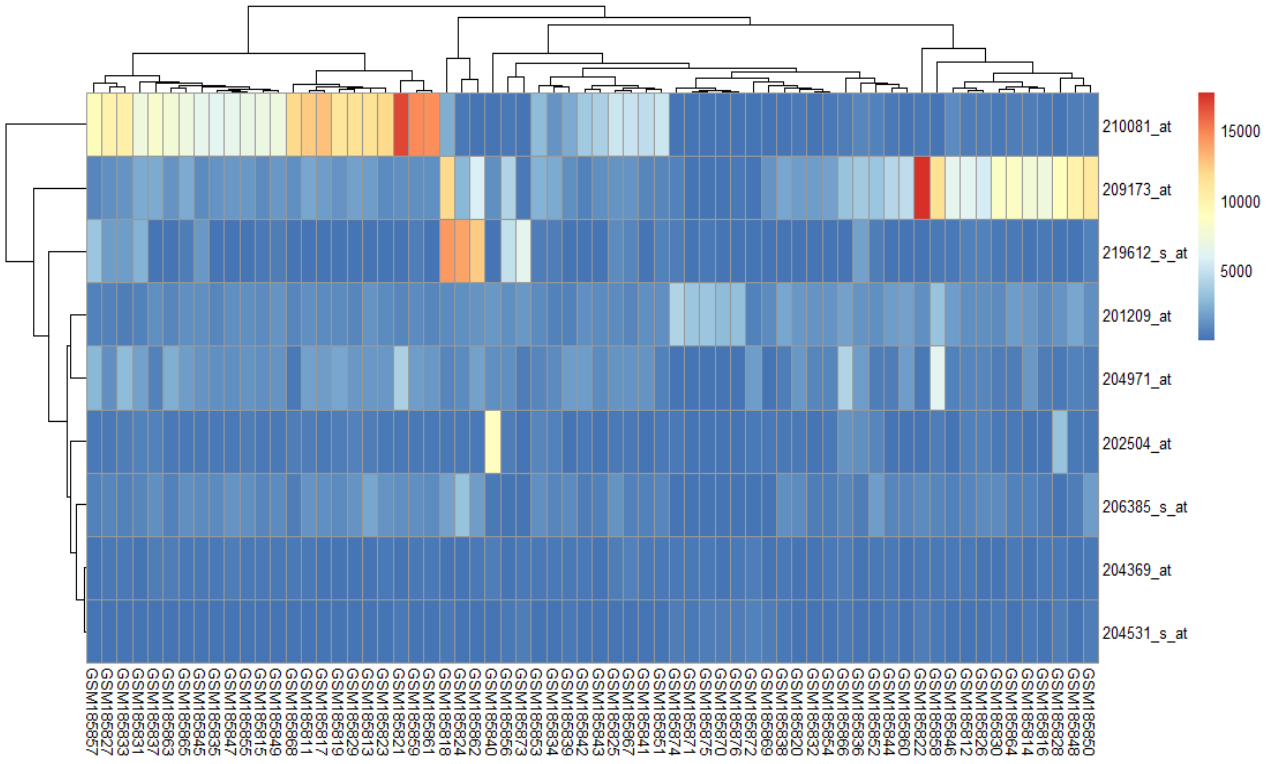

Figure 5. Correlation heatmap for Affymetrix probes '201209_at', '204369_at', '202504_at', '210081_at', '206385_s_at', '204971_at', '219612_s_at', '209173_at', and '204531_s_at' translating to gene symbols *HDAC1, PIK3CA, TRIM29, AGER, ANK3, CSTA, FGG, AGR2, BRCA1* respectively with 66 samples in GSE7670 dataset (**Prepared by Authors, 2025**).

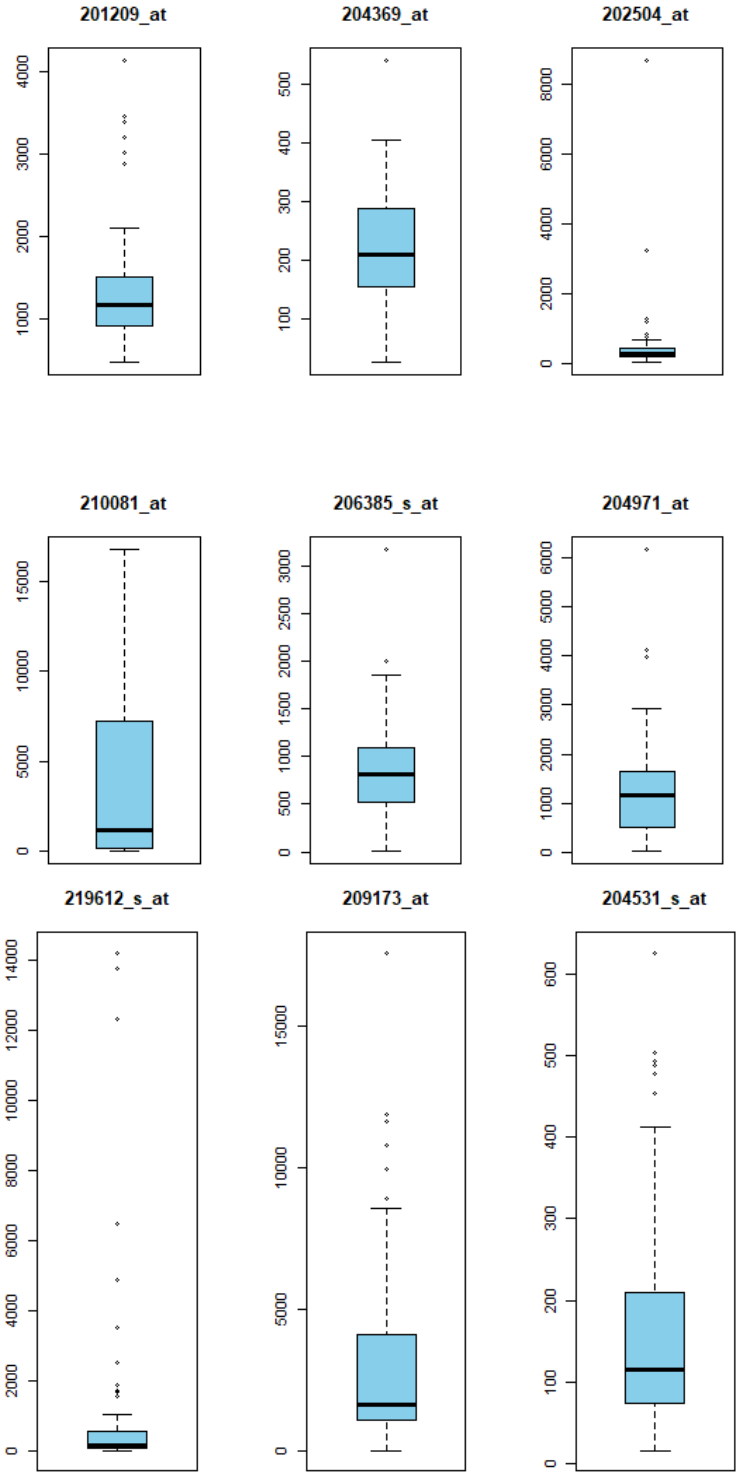

Figure 6. Box plot for gene symbols *HDAC1, PIK3CA, TRIM29, AGER, ANK3, CSTA, FGG, AGR2, BRCA1* for the samples in GSE7670 dataset. Source- Self procured) from the literature review. The hierarchical clustering was used to group genes by similar expression profiles. The highest transcript values are displayed as the reddest (hot), the lowest values are displayed as the bluest (cool), and intermediate values are a lighter color of either blue or red in different conditions. Conditions are the same as in Figure 6. A general overview shows that the correlation matrix reveals *TRIM29* as a highly correlated gene in the control patient samples in the cohort dataset whereas *AGR2* is correlated in the test patients (**Prepared by Authors, 2025**).

4. Discussions

The studies reviewed focused on RNA signatures in non-small cell lung cancer (NSCLC) to identify prospective biomarkers and therapeutic targets. The findings from these studies indicate that *HDAC1, miR-577, TRIM29, PIK3CA,* and *AGER* are potential therapeutic targets for NSCLC. These genes operate

across several key signaling pathways, notably the Wnt/β-catenin, TGF-β, PI3K/AKT/mTOR, and NF-κB pathways. Dysregulation of these pathways can lead to tumor growth, progression, and metastasis. Further research is needed to validate these genes' involvement in NSCLC and create specific treatments that may inhibit their activity. The findings from the reviewed studies provide new insights into the molecular mechanisms of NSCLC. Identifying potential therapeutic targets could lead to the development of more effective and personalized treatments for NSCLC patients.

The most enriched pathway in the pathway enrichment analysis was Chronic myeloid leukemia. Correlation heatmaps derived from literature-reported gene sets revealed strong inter-gene associations, indicating potential co-regulation. Expression profiles of NSCLC patients versus controls were visualized using box plots, emphasizing the diagnostic potential of HDAC1, BRCA1, and six other selected genes. These quantitative findings through bioinformatic analysis revealed the genes in the NSCLC pathway that are highly dysregulated and pave a path for further research in this field. This comprehensive and detailed approach, combining the existing literature on lung cancer, its transcriptomic data, and bioinformatic tools, provided a systematic framework for specific and focused therapeutic development for non-small cell lung cancer.

The heterogeneity of non-small cell lung cancer (NSCLC) also presents a significant challenge for treatment, as genetic mutations, tumor microenvironment variations, and acquired resistance mechanisms contribute to diverse patient responses. Mutations in oncogenes like *EGFR* and *KRAS* critically impact the efficacy of targeted therapies, yet their distribution varies among patients. For instance, mutations in *EGFR* can be used as predictive markers for sensitivity to *EGFR* inhibitors, whereas mutations in *KRAS* are often associated with resistance to such therapies, necessitating personalized therapeutic approaches **(25)**. Moreover, the heterogeneity of tumor microenvironment shaped by diverse immune cell populations and stromal constituents plays a critical role in the efficacy of therapies like chemotherapy and immunotherapy. To overcome this complexity, comprehensive molecular profiling is essential to identify individual genetic alterations, allowing clinicians to tailor treatment strategies more effectively. Combining targeted therapies with chemotherapy or immunotherapy may address multiple molecular pathways, potentially overcoming resistance. Continuous monitoring using techniques like imaging and liquid biopsy can help track tumor evolution and resistance, enabling timely adjustments to the treatment plan. By further research on the molecular heterogeneity of NSCLC, treatment can be better personalized to each patient's unique tumor characteristics, improving the chances of successful therapeutic outcomes.

Moreover, the cellular localization of genes associated with non-small cell lung cancer (NSCLC) is crucial for the development of targeted therapies. For instance, *AGER* (advanced glycosylation end-product specific receptor) is predominantly expressed in type II alveolar epithelial cells, playing a significant role in maintaining pulmonary homeostasis and modulating inflammatory responses. Alterations in *AGER* expression within these cells can disrupt the lung microenvironment, potentially affecting tumor progression and response to treatments. Similarly, *TRIM29* (tripartite motif-containing 29) is expressed in natural killer (NK) cells, which are integral to the body's innate immune response against tumors. *TRIM29* has been identified as a crucial negative regulator in immune responses to DNA viruses and cytosolic DNA, preventing potential damage caused by an overactive immune response **(26)**. The expression of *TRIM29* in NK cells may influence their cytotoxic behavior, potentially affecting the success of NK cell-based immunotherapies. Therefore, a comprehensive understanding of the cellular localization and function of genes like *AGER* and *TRIM29* is essential for optimizing immunotherapy strategies in NSCLC. Moreover, integrating non-invasive profiling techniques such as liquid biopsy and circulating biomarker analysis could enhance patient stratification and real-time monitoring of immune responses **(27)**.

5. Conclusion

This study performed an integrative reanalysis of multiple RNA expression datasets employing bioinformatics and computational approaches to identify dysregulated genes and pathways associated with NSCLC. Key genes *HDAC1*, *TRIM29,* and *PIK3CA* emerged as promising molecular candidates alongside their signaling pathway, including PI3K-AKT, MAPK signaling, and chromatin remodeling. This multi-layer evidence, which includes differential expression, pathway enrichment, and concordant expression patterns across datasets, provides convergent support for the aforementioned genes as a promising candidate for developing more effective, targeted, and personalized treatment strategies for NSCLC patients, thereby improving therapeutic responses and extending patient survival. It is important to acknowledge that this study is computational and retrospective. Therefore, further experimentation validation, including in vitro assays, in vivo models, and clinical studies, is imperative to substantiate these findings. Limitations of the analysis include reliance on existing gene expression datasets, susceptibility to batch and platform-related biases, and the lack of direct evidence for functional causation from transcriptomic associations. Future research should prioritize the functional characterization of the identified genes and pathways, the integration of multi-omics datasets, and the exploration of their roles in therapeutic resistance mechanisms. Overall, this investigation contributes novel insights into the molecular underpinnings of NSCLC and highlights promising targets for therapeutic intervention. The integration of systems biology approaches with experimental validation holds significant promise for the translation of computational discoveries into clinical applications.

6. Declarations

6.1 Acknowledgments

The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding them through large group Research Project under grant number RGP.2/264/45.

6.2 Ethical Considerations

Not applicable.

6.3 Authors' Contributions

Aditya Kumar Jha, Mohammad Ali Abdullah Almoyad, Shadma Wahab, Garima Gupta and Khang Wen Goh wrote the manuscript. Amirhossein Sahebkar and Prashant Kesharwani conceptualized, proofread and supervised it.

6.4 Conflict of Interest

The authors have no conflict of interest.

6.5 Fund or Financial Support

The authors extend their appreciation to the Deanship of Research and Graduate Studies at King Khalid University for funding this work through Large Research Project under grant number RGP2/346/46.

6.6 Using Artificial Intelligence Tools (AI Tools)

The authors were not utilized AI Tools.

References

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024;74(3):229-63. [DOI:10.3322/caac.21834] [PMID] [PMCID]
2. World Health Organization (WHO). Cancer fact sheet: World Health Organization; 2025. Available from: [https://www.who.int/news-room/fact-sheets/detail/cancer]
3. Centers for Disease Control and Prevention. U.S. Cancer Statistics: Lung cancer statistics: U.S. Department of Health and Human Services; 2025. Available from: [https://www.cdc.gov/lung-cancer/statistics/?CDC_AAref_Val=https://www.cdc.gov/cancer/lung/statistics]
4. Ess SM, Herrmann C, Frick H, Krapf M, Cerny T, Jochum W, et al. Epidermal growth factor receptor and anaplastic lymphoma kinase testing and mutation prevalence in patients with advanced non-small cell lung cancer in Switzerland: A comprehensive evaluation of real world practices. Eur J Cancer Care (Engl). 2017;26(6). [DOI:10.1111/ecc.12721] [PMID]
5. Groenewald W, Lund AH, Gay DM. The Role of WNT Pathway Mutations in Cancer Development and an Overview of Therapeutic Options. Cells. 2023;12(7):990. [PMCID] [DOI:10.3390/cells12070990] [PMID]
6. Qin S, Tang X, Chen Y, Chen K, Fan N, Xiao W, et al. mRNA-based therapeutics: powerful and versatile tools to combat diseases. Signal Transduct Target Ther. 2022;7(1):166. [DOI:10.1038/s41392-022-01007-w] [PMID] [PMCID]
7. Zhu J, Fu H, Wu Y, Zheng X. Function of lncRNAs and approaches to lncRNA-protein interactions. Sci China Life Sci. 2013;56(10): 876-85. [DOI:10.1007/s11427-013-4553-6] [PMID]
8. Zhang Y. LncRNA-encoded peptides in cancer. J Hematol Oncol. 2024;17(1):66. [PMCID] [DOI:10.1186/s13045-024-01591-0] [PMID]
9. Ma L, Huang Y, Zhu W, Zhou S, Zhou J, Zeng F, et al. An integrated analysis of miRNA and mRNA expressions in non-small cell lung cancers. PLoS One. 2011;6(10):e26502. [DOI:10.1371/journal.pone.0026502] [PMID] [PMCID]
10. Shriwash N, Singh P, Arora S, Ali SM, Ali S, Dohare R. Identification of differentially expressed genes in small and non-small cell lung cancer based on meta-analysis of mRNA. Heliyon. 2019;5(6):e01707. [PMCID] [DOI:10.1016/j.heliyon.2019.e01707] [PMID]
11. Malone ER, Oliva M, Sabatini PJB, Stockley TL, Siu LL. Molecular profiling for precision cancer therapies. Genome Med. 2020;12(1):8. [DOI:10.1186/s13073-019-0703-1] [PMID] [PMCID]
12. Du Z, Lovly CM. Mechanisms of receptor tyrosine kinase activation in cancer. Mol Cancer. 2018;17(1):58. [PMID] [PMCID] [DOI:10.1186/s12943-018-0782-4]
13. Giménez N, Martínez-Trillos A, Montraveta A, Lopez-Guerra M, Rosich L, Nadeu F, et al. Mutations in the RAS-BRAF-MAPK-ERK pathway define a specific subgroup of patients with adverse clinical features and provide new therapeutic options in chronic lymphocytic leukemia. Haematologica. 2019;104(3):576-86.

[DOI:10.3324/haematol.2018.196931] [PMID] [PMCID]

14. He Y, Sun MM, Zhang GG, Yang J, Chen KS, Xu WW, et al. Targeting PI3K/Akt signal transduction for cancer therapy. Signal Transduct Target Ther. 2021;6(1):425. [PMID] [DOI:10.1038/s41392-021-00828-5] [PMCID]
15. Shackelford DB, Shaw RJ. The LKB1-AMPK pathway: metabolism and growth control in tumour suppression. Nat Rev Cancer. 2009; 9(8):563-75. [DOI:10.1038/nrc2676] [PMID] [PMCID]
16. Mogi A, Kuwano H. TP53 mutations in nonsmall cell lung cancer. J Biomed Biotechnol. 2011;2011:583929. [DOI:10.1155/2011/583929][PMID][PMCID]
17. Kanwal R, Gupta S. Epigenetic modifications in cancer. Clin Genet. 2012;81(4):303-11. [DOI:10.1111/j.1399-0004.2011.01809.x] [PMID] [PMCID]
18. Baba AB, Rah B, Bhat GR, Mushtaq I, Parveen S, Hassan R, et al. Transforming Growth Factor-Beta (TGF-β) Signaling in Cancer-A Betrayal Within. Front Pharmacol. 2022;13: 791272. [DOI:10.3389/fphar.2022.791272] [PMID] [PMCID]
19. Billin AN, Thirlwell H, Ayer DE. β-Catenin-histone deacetylase interactions regulate the transition of LEF1 from a transcriptional repressor to an activator. Mol Cell Biol. 2000; 20(18):6882-90. [PMID] [PMCID] [DOI:10.1128/MCB.20.18.6882-6890.2000]
20. Wang L, Shen S, Wang M, Ding F, Xiao H, Li G, et al. Rho GTPase Activating Protein 24 (ARHGAP24) Silencing Promotes Lung Cancer Cell Migration and Invasion by Activating β-Catenin Signaling. Med Sci Monit. 2019;25:21-31. [PMID] [PMCID] [DOI:10.12659/MSM.911503]
21. Hu Y, He J, He L, Xu B, Wang Q. Expression and function of Smad7 in autoimmune and inflammatory diseases. J Mol Med (Berl). 2021;99(9):1209-20. [PMID] [PMCID] [DOI:10.1007/s00109-021-02083-1]
22. Chen J, Ye C, Wan C, Li G, Peng L, Peng Y, et al. The Roles of c-Jun N-Terminal Kinase (JNK) in Infectious Diseases. Int J Mol Sci. 2021;22(17):9640. [PMID] [PMCID] [DOI:10.3390/ijms22179640]
23. Liu J, Wang T, Creighton CJ, Wu S-P, Ray M, Janardhan KS, et al. JNK1/2 represses Lkb1-deficiency-induced lung squamous cell carcinoma progression. Nature Communications. 2019;10(1):2148. [PMCID] [DOI:10.1038/s41467-019-09843-1] [PMID]
24. Geng QS, Liu RJ, Shen ZB, Wei Q, Zheng YY, Jia LQ, et al. Transcriptome sequencing and metabolome analysis reveal the mechanism of Shuanghua Baihe Tablet in the treatment of oral mucositis. Chin J Nat Med. 2021;19(12):930-43. [DOI:10.1016/S1875-5364(22)60150-X] [PMID]
25. Massarelli E, Varella-Garcia M, Tang X, Xavier AC, Ozburn NC, Liu DD, et al. KRAS mutation is an important predictor of resistance to therapy with epidermal growth factor receptor tyrosine kinase inhibitors in non-small-cell lung cancer. Clin Cancer Res. 2007; 13(10):2890-6. [PMID] [DOI:10.1158/1078-0432.CCR-06-3043]
26. Li Q, Lin L, Tong Y, Liu Y, Mou J, Wang X, et al. TRIM29 negatively controls antiviral immune response through targeting STING for degradation. Cell Discovery. 2018;4(1):13. [DOI:10.1038/s41421-018-0010-9]
27. Kumar H, Mishra G, Sharma AK, Gothwal A, Kesharwani P, Gupta U. Intranasal Drug Delivery: A Non-Invasive Approach for the Better Delivery of Neurotherapeutics. Pharm Nanotechnol. 2017;5(3):203-14. [PMID] [DOI:10.2174/2211738505666170515113936]

How to Cite This Article:

Jha A K, Almoyad M A A, Wahab S, Gupta G, Goh K W, Sahebkar A et al . Computational Identification of Dysregulated Genes and Pathways in Non-Small Cell Lung Cancer: A Systems Biology Approach. J Adv Med Biomed Res. 2025;33(161):293-306.

Download citation:
BibTeX | RIS | EndNote | Medlars | ProCite | Reference Manager | RefWorks

Send citation to:
Mendeley Zotero RefWorks