Journal of Computational Intelligence in Bioinformatics
ISSN 0973-385X Volume 18, Number 2 (2025) pp. 91-111

https://dx.doi.org/10.37622/JCIB/1.8.25.91-111

A Comprehensive Gene Expression and Regulatory Network Analysis Reveals Biomarkers and miRNA Regulation in ALK-Positive Non-Small Cell Lung Cancer.

Neha Sharma[1*], Vaishnavi Kanungo[2] , Atharva Mandloi[3], Lakshmi Pillai[4], Payal Puri[5]

[1]Research Associate (Bioinformatics), Inner Wellness Global Research Foundation, 51, London Villas, Indore, Madhya Pradesh, India
E-mail: iwgrfoundation@gmail.com
[2]Research Scholar (Bioinformatics), Institute for Innovative Learning and Research Academy., 1750, Scheme No 114, Indore, Madhya Pradesh, India
E-mail: iilracademy@gmail.com
[3]PGY1 Resident Community Medicine, College- Department of Community Medicine at MGM Medical College, Indore, MP.
E-mail- akm4497@gmail.com
[4] Professor, Institute of sciences, Department of Bioscience, SAGE University,
Kailod , kartal Road, Rau Bypass Road, Indore, Madhya Pradesh
Email: lakshpillai1@gmail.com
[5]Principal Scientist, Institute for Innovative Learning and Research Academy. 1750, Scheme No 114, Indore, Madhya Pradesh, India
E-mail: iilracademy@gmail.com
[*]Corresponding author

Abstract

This study investigates gene expression profiles and regulatory mechanisms in ALK-positive non-small cell lung cancer (NSCLC) using the GSE31210 dataset. A total of 11 ALK-positive and 20 normal lung samples were analyzed. Differential gene expression analysis (p-value < 0.05; |logFC| ≥ 1.5) identified 1,215 overexpressed and 1,918 underexpressed genes. Protein–protein interaction networks were constructed using STRINGDB and visualized in Cytoscape, with the top 10 hub genes from each group identified via the CytoHubba plugin. Functional enrichment analysis using Enrichr revealed significant pathways, including cell cycle regulation, DNA replication, and immune-related signaling.Regulatory network analysis with miRNet identified key interactions between hub genes, miRNAs, and transcription factors. Overexpressed genes such as CCNB1, TOP2A, and EGFR were targeted by miRNAs including hsa-miR-16-5p and hsa-miR-34a-

5p, and regulated by transcription factors such as TP53 and STAT3. Underexpressed genes including TTN, FGF2, and CAV1 were linked to miRNAs such as hsa-miR-19a-3p and TFs like CREB1 and RELA. Notably, TP53, STAT3, and hsa-miR-34a-5p were shared across both networks, underscoring their central roles in ALK-positive NSCLC. This comprehensive analysis highlights potential biomarkers and therapeutic targets, providing a foundation for further research.

Keywords: *ALK-positive non-small cell lung cancer (NSCLC), Gene Expression Analysis, Functional Enrichment Analysis, Network clustering, miRNA*

Introduction

About 85% of all instances of lung cancer are non-small cell lung cancer (NSCLC), making it the most common kind of the disease. Among its various subtypes, anaplastic lymphoma kinase (ALK)-positive NSCLC is particularly notable due to its distinct genetic profile and responsiveness to targeted therapies. ALK rearrangements[21], involving the echinoderm microtubule-associated protein-like 4 (EML4) gene, result in the formation of oncogenic fusion proteins. These fusion proteins activate downstream signaling pathways such as the PI3K/AKT and RAS/RAF/MEK/ERK pathways, promoting cell proliferation, survival, and tumorigenesis while inhibiting apoptosis. This is particularly prevalent in younger patients, non-smokers, and those with adenocarcinoma histology. EML4-ALK fusion gene was discovered in 2007 [1], since then numerous studies have reported varying prevalence rates. For instance, Shaw et al. (2009) [2] found ALK rearrangements in approximately 5% of NSCLC patients, while more recent studies, such as those by Kwak et al. (2010) [3], have identified it in about 3-5% of cases. The variation in prevalence rates underscores the importance of genetic screening in diagnosing and treating NSCLC. Patients with ALK-positive NSCLC often present with symptoms similar to those of other lung cancer types, including persistent cough, chest pain, shortness of breath, and unexplained weight loss[4]. Additionally, some patients may experience more specific symptoms such as fatigue, hoarseness, and hemoptysis (coughing up blood). Due to the aggressive nature of ALK-positive NSCLC, early detection and targeted treatment are crucial for improving patient outcomes.

Identifying ALK rearrangements has significant therapeutic implications, as patients with ALK-positive NSCLC often respond well to ALK inhibitors, such as crizotinib, ceritinib, and alectinib [5]. These targeted therapies have shown remarkable efficacy in improving progression-free survival and overall survival rates compared to conventional chemotherapy[6].

Identifying and characterizing differentially expressed genes (DEGs) in ALK-positive NSCLC are critical for understanding the molecular mechanisms of the disease and developing targeted therapies. Despite the advancements in targeted therapies, resistance to ALK inhibitors remains a challenge, necessitating ongoing research to understand the molecular mechanisms driving ALK-positive NSCLC. Gene

expression profiling and bioinformatics analyses provide powerful tools to uncover these molecular alterations and to identify potential biomarkers and therapeutic targets.

In the present study, the microarray dataset GSE31210 from the Gene Expression Omnibus (GEO) database was analyzed, which comprises gene expression profiles from multiple lung cancer subtypes and healthy lung tissues. Adenocarcinoma (ADC), the predominant subtype of non-small cell lung cancer (NSCLC), accounting for about 40% of all lung cancers, arises from the glandular tissues of the lung. The GSE31210 dataset represents a comprehensive collection of gene expression data specifically from ADC cases, making it highly relevant for studying the molecular characteristics of ALK-positive NSCLC. By employing a combination of differential gene expression analysis, network analysis, and functional enrichment, we aimed to gain insights into the biological processes and pathways that are dysregulated in ALK-positive NSCLC. This study aims to provide a comprehensive overview of the gene expression landscape in ALK-positive NSCLC, identifying key genes and pathways that could serve as biomarkers or therapeutic targets[20,22,23]. By integrating various bioinformatics tools and resources, we offer a detailed analysis that enhances our understanding of this specific subtype of lung cancer, paving the way for future research and clinical applications.

Methods

Data Collection and Preprocessing

In this research, we utilized the microarray dataset under the accession number GSE31210 from the GEO database [7,8,9], which includes gene expression profiles of 246 samples comprising ALK-positive, EGFR-positive, KRAS-positive, triple-negative, and healthy lung tissues. For this study, we selected the subset of 11 ALK-positive and 20 healthy lung samples to identify differentially expressed genes and elucidate their roles in the pathogenesis of ALK-positive NSCLC. The raw data were pre-processed using GEO2R, an interactive tool that allows users to compare gene expression across multiple experimental conditions.

Processing of Data and Selection of Differentially Expressed Genes

Differential expression analysis was performed using the GEO2R tool [10], which allows for the comparison of two or more groups of samples in a GEO Series. The data were force normalized within GEO2R to ensure consistency across samples GEO2R creates a text file with a list of differentially expressed genes in disease vs. control format. In the present study, differential gene expression analysis was performed to identify significantly over-expressed or under-expressed genes in ALK-positive NSCLC samples compared to healthy lung samples. Genes with a p-value lower than 0.05 were considered statistically significant. The genes were then sorted based on their log fold change (logFC) values, 1215 overexpressed (logFC > 1.5) and 1918 under-expressed (logFC < -1.5) genes.

DEG Identification and Network Analysis

The selected 1215 over-expressed and 1918 under-expressed genes were separately submitted into STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) to construct protein-protein interaction (PPI) networks[11] with a confidence score threshold of 0.7 to ensure high-confidence interactions. The resulting networks were exported and visualized using Cytoscape, a network analysis tool. [12]
Within Cytoscape, the CytoHubba plugin [13] was used to identify hub genes based on their degree of connectivity. The top 10 hub genes from both gene sets were selected for further analysis.

Functional Enrichment Analysis and miRNet Analysis
The top 10 hub genes from the overexpressed and under-expressed gene sets were subjected to functional enrichment analysis using Enrichr [14]. This tool provided insights into the involvement of these genes in various biological processes, molecular functions, cellular components, and KEGG pathways. The enrichment results, including bar graphs, were downloaded and analyzed to identify key pathways and processes.
To further investigate the regulatory network, miRNet analysis was performed on the top 10 hub genes [16]. Significant interactions were identified for hub genes of over-expressed and under-expressed genes, along with miRNA and transcription factors [17].

Data Visualization
Various plots, including volcano plots, box plots, and mean difference plots, were generated using GEO2R to visualize the distribution and significance of the differentially expressed genes. Network diagrams of the PPI networks and enrichment results were created using Cytoscape and Enrichr to illustrate the interactions and functional relevance of the identified genes. The network of interaction among DEGs, miRNAs, and transcription factors (TFs) was generated using miRNet.

Results
Differential gene expression analysis was performed using GEO2R to identify significantly overexpressed or under-expressed in ALK-positive NSCLC samples compared to healthy lung samples. The data were normalized using force normalization to ensure comparability across samples.
Volcano plots, box plots, and mean difference plots were generated to visualize the distribution and significance of the DEGs. The volcano [Fig. 1] plot highlighted the most significantly upregulated and downregulated genes, while box plots [Fig. 2]and mean difference plots [Fig. 3] provided insights into the expression patterns across different sample groups.

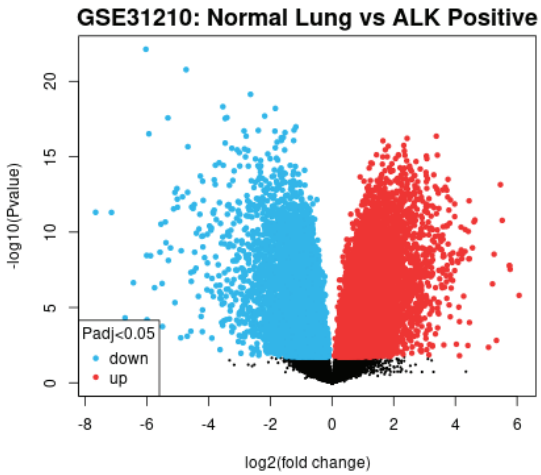


Fig. 1 The volcano plot displays the relationship between the fold change (logFC) and the statistical significance (p-value) of genes, highlighting the most differentially expressed genes in ALK-positive NSCLC samples from GSE31210.

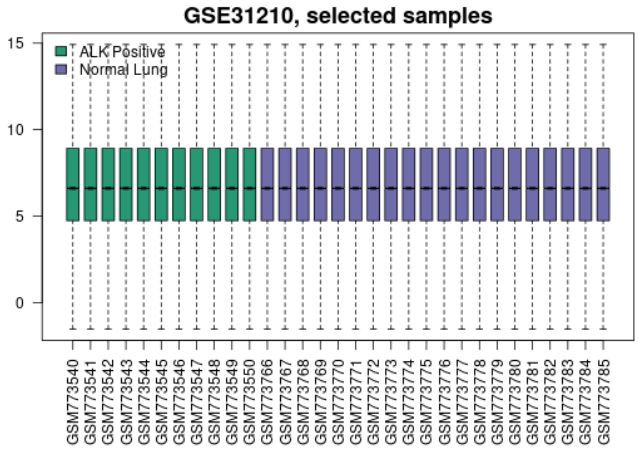


Fig. 2 Gene expression value distribution for GSE31210. There are no units for the y-axis. The gene expression levels of a single patient sample are displayed in each box plot.

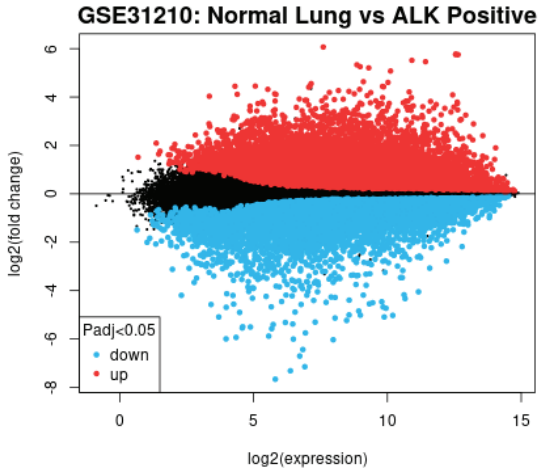


Fig. 3: The mean difference plot illustrates the average expression differences of

genes between ALK-positive NSCLC samples and the control group from GSE31210, highlighting genes with the most significant changes.

Protein-Protein Interaction Network Construction

In the dataset GSE31210 out of 54676 genes, 1215 genes were found to be over-expressed or upregulated, and 1918 genes were under-expressed or downregulated, which passed the set cut-off of p-value 0.05 and logFC value of |1.5|.

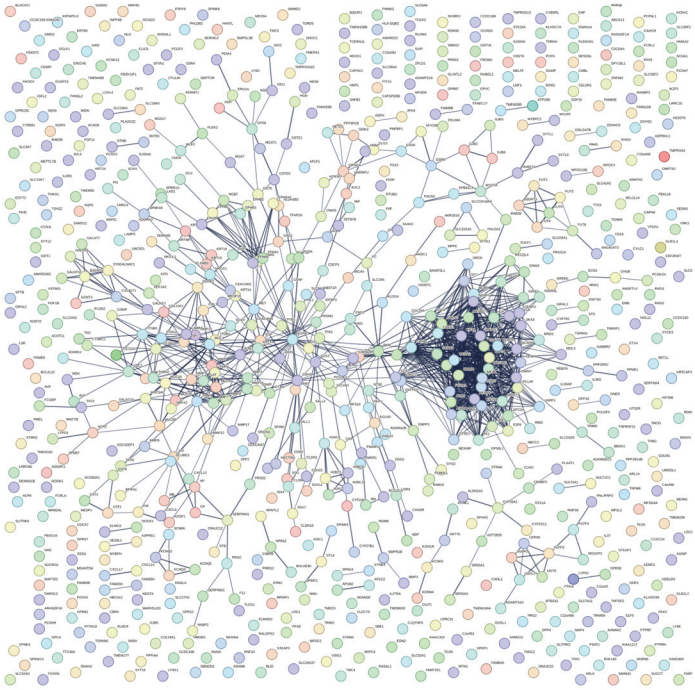


Fig. 4 The MCL (Markov Cluster Algorithm) Clustering method was applied to create the protein-protein interaction (PPI) network. In this network, 1215 overexpressed genes in ALK-positive NSCLC were used and the interaction score was set at high confidence (0.700).

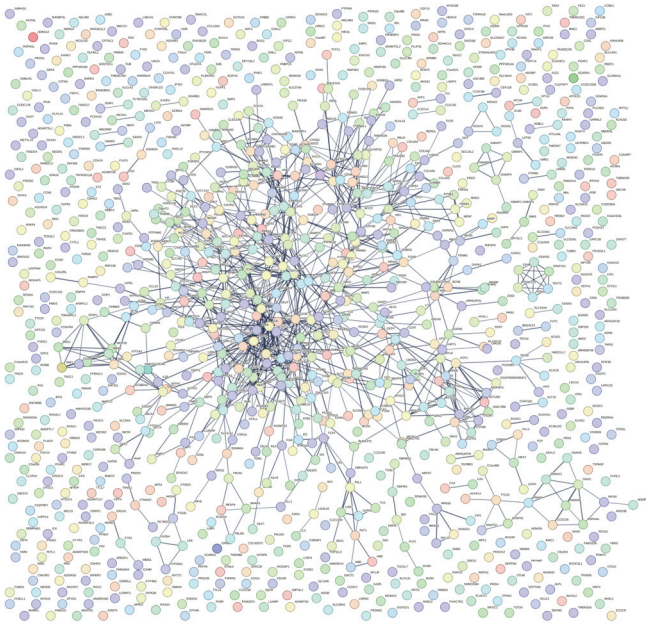


Fig. 5 The MCL (Markov Cluster Algorithm) Clustering method was applied to create the protein-protein interaction (PPI) network. In this network, 1918 underexpressed genes in ALK-positive NSCLC were used and the interaction score was set at high confidence (0.700).

Overexpressed and Under-expressed genes were analyzed using STRING [11] to construct protein-protein interaction (PPI) networks with a confidence score threshold of 0.7. The resulting networks were visualized using Cytoscape. The overexpressed gene network comprised key nodes that indicated potential interactions relevant to tumorigenesis and cancer progression. The PPI network has proteins (which do not interact at all), while there are interesting connections among many proteins in the network (which have connections to other proteins). Fig.4 and Fig.5 show the PPI network of overexpressed and underexpressed genes respectively.

Identification of Hub Genes

After that, the networks of protein-protein interactions were exported and visualized in Cytoscape. Within Cytoscape, the CytoHubba plugin was used to identify the top 10 hub genes based on their degree of connectivity in the PPI networks. For the overexpressed gene set, the top 10 hub genes included CCNB1, CCNA2, TOP2A, CCNB2, BUB1B, BIRC5, DLGAP5, CDC45, EGFR, and MELK. Similarly, the under-expressed gene set yielded 10 hub genes, including IL6, FGF2, VWF, PECAM1, TLR4, TTN, CAV1, BDNF, CDH5, and KIT. . Fig. 6 and Fig. 7 show the network of the top 10 hub genes obtained from overexpressed and under-expressed gene sets, respectively. Tables 1 and 2 depict the rank and degree score of the top ten hub genes of overexpressed and under-expressed genes, respectively. It was seen that in the top three hub genes of overexpressed genes were Cyclin B1 (CCNB1) with a degree of 47, Cyclin A2 (CCNA2) with a degree of 43, and DNA Topoisomerase II Alpha (TOP2A), Cyclin B2 (CCNB2), BUB1 Mitotic Checkpoint Serine/Threonine Kinase B (BUB1B) and Baculoviral IAP Repeat Containing 5 (also known as

Survivin) (BIRC5) with a degree of 39. The top three hub genes of under-expressed genes were Interleukin 6 (IL6) had the highest degree (Degree 40) followed by the Fibroblast Growth Factor 2 (FGF2) with degree 23 and Von Willebrand Factor (VWF) with a degree of 22. These hub genes play critical roles in the molecular mechanisms underlying ALK-positive NSCLC.

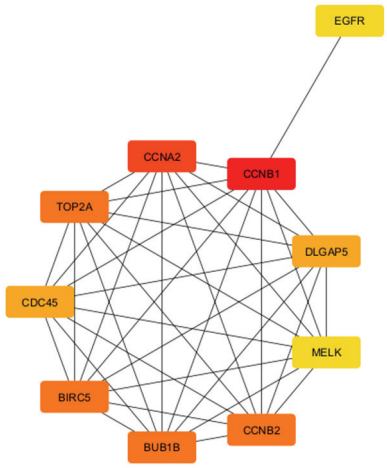


Fig. 6 Network of top ten hub genes (from overexpressed gene set) obtained using Cytoscape. The color code represents the degree here, the red color denotes the highest degree of centrality, orange denotes the medium degree and yellow denotes the lowest degree of centrality.

Rank	Name of Hub genes	Degree Score
1	CCNB1	47
2	CCNA2	43
3	TOP2A	39
3	CCNB2	39
3	BUB1B	39
3	BIRC5	39
7	DLGAP5	38
7	CDC45	38
9	EGFR	37
9	MELK	37

Table 1 Top 10 in network STRING network ranked by Degree method (for Over-expressed genes)

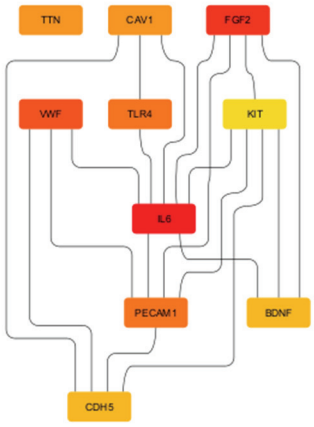


Fig. 7 Network of top ten hub genes (from underexpressed gene set) obtained using Cytoscape. The colour code represents the degree here, the red colour denotes the highest degree of centrality, orange denotes medium degree and yellow denotes the lowest degree of centrality.

Rank	Name of Hub gene	Degree Score
1	IL6	40
2	FGF2	23
3	VWF	22
4	PECAM1	21
4	TLR4	21
6	TTN	20
6	CAV1	20
8	BDNF	19
8	CDH5	19
10	KIT	18

Table 2 Top 10 in network STRING network ranked by Degree method (for Underexpressed genes)

Gene Ontology processes and KEGG pathways Analysis for the DEGs

Functional enrichment analysis was performed on the top 10 hub genes identified from the over-expressed and under-expressed gene sets This analysis was conducted

using Enrichr, which allowed us to explore various biological processes, cellular components, molecular functions, and KEGG pathways in which these genes are involved. Detailed results of these enrichment analyses are provided below:

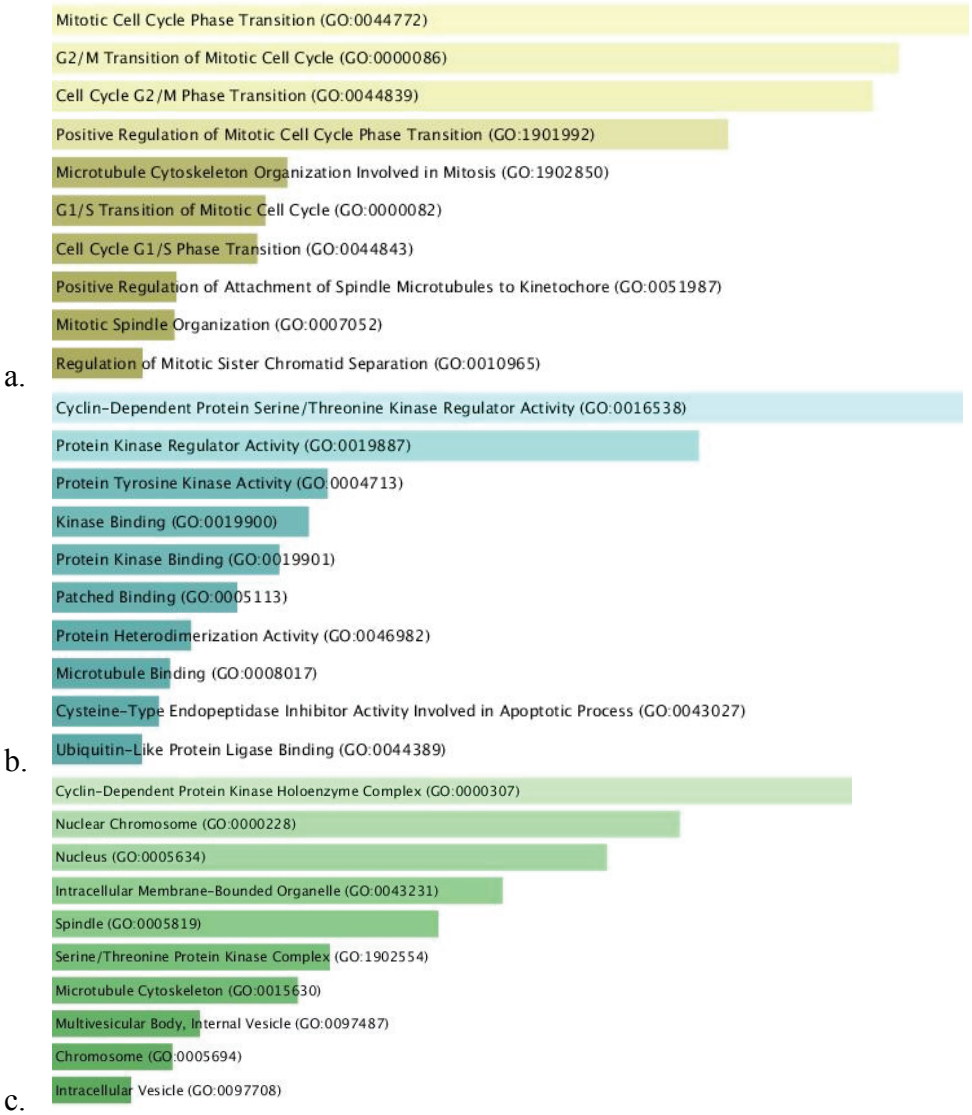
Functional enrichment analysis of overexpressed genes (GO Biological processes)
Gene Ontology (GO) enrichment analysis of overexpressed genes in ALK-positive NSCLC revealed strong association with cell cycle regulation, particularly mitotic transitions [Fig.8a]. Key enriched processes include G2/M and G1/S transitions, mitotic spindle organization, and chromatid separation, which are critical for proper cell division. Dysregulation of these checkpoints is a hallmark of cancer, enabling uncontrolled proliferation. Processes like positive regulation of mitotic transitions and microtubule cytoskeleton organization further underscore the role of these genes in ensuring fidelity of mitosis. These findings suggest that ALK activation may drive tumor progression through enhanced or aberrant mitotic activity. The enrichment of these biological processes supports the identification of cell cycle-related biomarkers and miRNA regulators, offering mechanistic insights and potential therapeutic targets in ALK-positive non-small cell lung cancer.

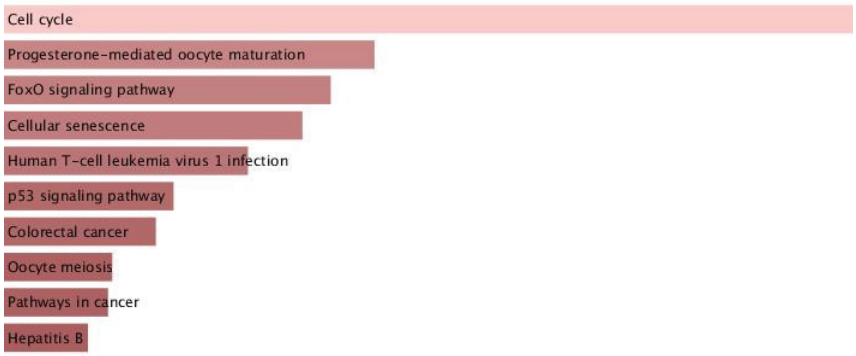
Functional enrichment analysis of overexpressed genes (GO Molecular Functions)
The GO molecular function enrichment analysis revealed that the overexpressed genes in ALK-positive NSCLC are predominantly involved in kinase regulation and binding activities, particularly cyclin-dependent protein serine/threonine kinase regulator activity, protein tyrosine kinase activity, and protein kinase binding[Fig.8b]. These functions are essential in controlling signaling cascades that drive cell cycle progression, proliferation, and survival, all of which are commonly dysregulated in ALK-driven lung cancers. Additional enriched functions, including microtubule binding and ubiquitin-like protein ligase binding, point to roles in cytoskeletal dynamics and protein degradation pathways, which influence mitosis and apoptosis. The presence of cysteine-type endopeptidase inhibitor activity suggests suppression of apoptotic signaling. These functional categories highlight key oncogenic mechanisms and further support the relevance of these genes as biomarkers or therapeutic targets in ALK-positive NSCLC.

Functional enrichment analysis of overexpressed genes (GO Cellular Component)
The Gene Ontology (GO) Cellular Component enrichment analysis suggests The most enriched components include the cyclin-dependent protein kinase holoenzyme complex, nucleus, nuclear chromosome, and spindle, suggesting a strong involvement in cell cycle regulation and mitosis, processes often dysregulated in cancer [Fig.8c]. Components like the microtubule cytoskeleton and serine/threonine protein kinase complex indicate roles in cell division and signal transduction. The presence of intracellular membrane-bounded organelles, multivesicular bodies, and intracellular vesicles points to alterations in intracellular trafficking and communication. This enrichment suggests that ALK-positive NSCLC overexpresses genes localized in structures critical for proliferation, providing insight into potential biomarkers and therapeutic targets.

Functional enrichment analysis of overexpressed genes (KEGG Pathway)

The KEGG pathway enrichment analysis of overexpressed genes in ALK-positive non-small cell lung cancer (NSCLC) revealed significant involvement in pathways such as Cell cycle, p53 signaling, FoxO signaling, and cellular senescence [Fig.8d]. These pathways are critically associated with cancer hallmarks including uncontrolled proliferation, evasion of apoptosis, and altered cellular aging—all of which are characteristic of ALK-driven tumorigenesis. The enrichment of broader oncogenic pathways such as Pathways in cancer, Colorectal cancer, and Hepatitis B indicates shared molecular mechanisms across cancer types. These results suggest that the identified genes may play central roles in tumor progression and response to cellular stress. Furthermore, their regulation by specific miRNAs highlights potential biomarker candidates and therapeutic targets in the context of ALK-positive NSCLC.





d.

Fig. 8 The enrichment analysis results in a graphical format. The x-axis represents the number of hub genes (over-expressed) and the y-axis represents a) biological processes, b) molecular functions, c) cellular components, and d) the enriched KEGG pathways

Functional enrichment analysis of under-expressed genes (GO Biological Functions)

The GO Biological Process enrichment analysis of under-expressed genes revealed suppression of key regulatory functions critical to cancer signaling and immune modulation[Fig.9a]. Notably, downregulation was observed in positive regulation of intracellular and general signal transduction, including pathways like MAPK and JAK-STAT, indicating impaired oncogenic signaling. Reduced expression in genes maintaining the blood-brain barrier and regulating cell motility and migration suggests compromised barrier integrity and altered metastatic potential. Additionally, the negative regulation of apoptosis being enriched implies decreased anti-apoptotic signaling, which may affect tumor survival mechanisms. Suppressed regulation of cytokine production involved in inflammatory responses points to altered tumor-immune interactions. Together, these findings highlight disrupted signaling, immune evasion, and metastatic control mechanisms in the ALK-positive NSCLC gene expression landscape.

Functional enrichment analysis of under-expressed genes (GO Molecular Functions)

The GO Molecular Function enrichment analysis of under-expressed genes highlights reduced activity in several critical receptor-binding and signaling functions[Fig.9b]. Notably, Growth Factor Activity, VEGF receptor binding, and BMP receptor binding were enriched, indicating downregulation of pathways involved in angiogenesis, cell proliferation, and tumor growth regulation. Suppression of IL-6 receptor binding and Nitric Oxide Synthase binding suggests impaired immune modulation and oxidative signaling, both essential in tumor–immune interactions. Reduced Patched binding points toward dysregulation in Hedgehog signaling, which is often altered in cancer. These findings collectively suggest that tumor cells may silence these regulatory molecular functions to evade immune surveillance, control angiogenesis differently, or bypass growth constraints — shedding light on molecular vulnerabilities and potential therapeutic targets in ALK-positive NSCLC.

Functional enrichment analysis of under-expressed genes (GO Cellular Component)

The GO Cellular Component enrichment analysis of the underexpressed hub genes

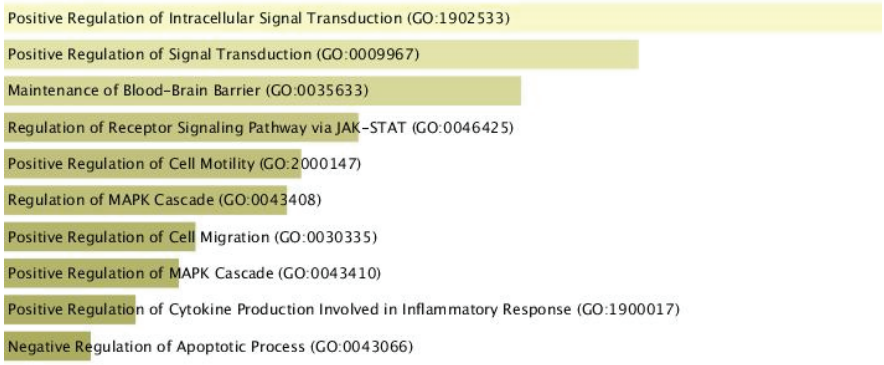
revealed significant enrichment in compartments such as the Platelet Alpha Granule, Cytoplasmic Vesicle Membrane, and Endoplasmic Reticulum Lumen, indicating downregulation of pathways involved in protein storage, secretion, and vesicle trafficking. Components like the Cell-Cell Junction, Catenin Complex, and Caveolae suggest possible disruptions in cell adhesion, signaling, and membrane organization[Fig.9c]. Enrichment in the Endosome Membrane and Bounding Membrane of Organelle reflects suppressed endocytic and intracellular transport processes. Collectively, these results suggest that key cellular processes related to intercellular communication, transport, and structural organization are compromised, which may contribute to the pathogenesis and progression of ALK-positive NSCLC by affecting tumor microenvironment interaction and immune evasion.

Functional enrichment analysis of under-expressed genes (KEGG Pathway)

The KEGG pathway enrichment analysis of under-expressed genes reveals several critical cancer-associated signaling pathways that are potentially suppressed in ALK-positive NSCLC[Fig.9d]. The most prominent is the PI3K-Akt signaling pathway, which is essential for cell survival, growth, and metabolism. Its underexpression may indicate a shift in signaling dependency, possibly due to the presence of the ALK fusion oncogene overriding normal control mechanisms.

Additional pathways such as the MAPK and Ras signaling pathways, also central to proliferation and differentiation, are downregulated, suggesting suppression of canonical tumor signaling routes in favor of alternative, possibly mutated, pathways. Pathways related to proteoglycans in cancer and fluid shear stress point to changes in the tumor microenvironment and metastasis potential. Collectively, these results reflect a rewiring of signaling networks in tumor cells, aiding in immune evasion and therapy resistance.

a.



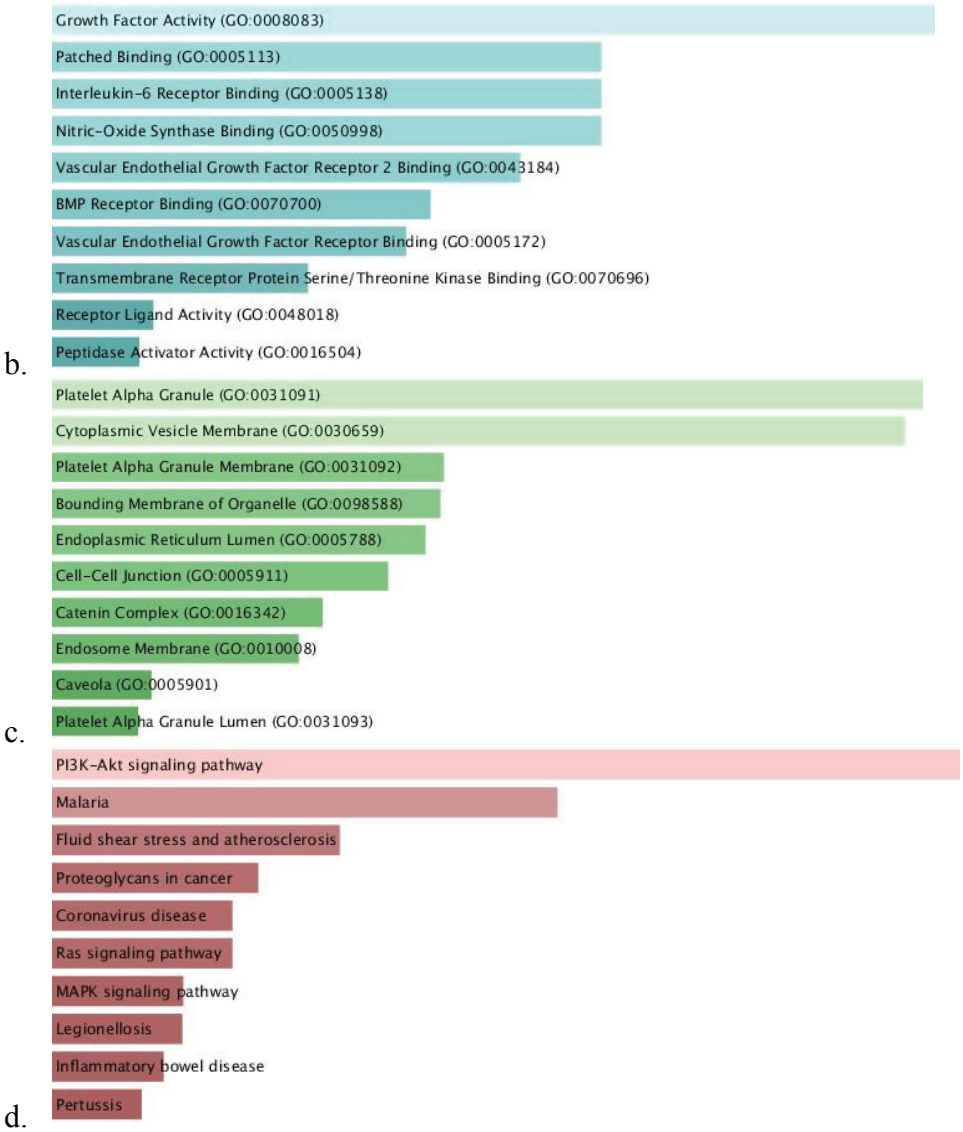


Fig. 9 The enrichment analysis results in a graphical format. The x-axis represents the number of hub genes (under-expressed) and the y-axis represents a) biological processes, b) molecular functions, c) cellular components, and d) the enriched KEGG pathways.

Identification of miRNAs and TFs Associated with Top overexpressed Genes Using miRNet

The regulatory network analysis of overexpressed hub genes in ALK-positive non-small cell lung cancer (NSCLC) revealed several key oncogenic drivers and their upstream regulators. Central genes such as CCNB1, CCNA2, TOP2A, MELK, BIRC5, and EGFR showed high connectivity within the network, indicating their prominent roles in promoting uncontrolled cell proliferation, cell cycle progression, and tumor survival[Fig.10a]. These genes were found to be targeted by multiple tumor-suppressive miRNAs, including hsa-miR-34a-5p, hsa-miR-16-5p, hsa-miR-

124-3p, and hsa-miR-107, whose reduced activity may contribute to the overexpression of oncogenic targets. Additionally, transcription factors such as TP53, STAT3, MYC, SP1, and E2F1 were involved in regulating both the overexpressed genes and their interacting miRNAs, indicating a complex transcriptional and post-transcriptional regulatory network. This integrated analysis underscores how the combined dysregulation of miRNAs and transcription factors facilitates the overexpression of key oncogenes, thereby contributing to the pathogenesis of ALK-positive NSCLC and identifying potential therapeutic targets.

Names of DEGs, miRNAs, and TFs	**Degree**
CCNB1	287
TOP2A	277
EGFR	254
CCNA2	239
BIRC5	200
MELK	169
DLGAP5	151
BUB1B	128
CCNB2	124
hsa-mir-16-5p	10
hsa-mir-34a-5p	10
hsa-mir-103a-3p	10
hsa-mir-107	10
hsa-mir-124-3p	10
hsa-mir-129-2-3p	10
hsa-mir-147a	10
hsa-let-7d-5p	10
hsa-miR-103a-3p	10
hsa-miR-106b-5p	10
hsa-miR-107	10
hsa-miR-124-3p	10
hsa-miR-129-2-3p	10
hsa-miR-147a	10

hsa-miR-15a-5p	10
hsa-miR-16-5p	10
hsa-miR-195-5p	10
hsa-miR-19a-3p	10
hsa-miR-19b-3p	10
hsa-miR-34a-5p	10
TP53	5
E2F1	3
NFKB1	3
RELA	3
SP1	3
YBX1	3
BRCA1	2
E2F3	2
E2F4	2
EP300	2
ESR1	2
HDAC1	2
KLF4	2
KLF5	2
MYC	2
STAT3	2
TFAP2A	2

Table 3 List of top miRNAs, transcription factors, and top DEGs (over-expressed) with their degree of centrality.

Identification of miRNAs and TFs Associated with Top under-expressed Genes Using miRNet

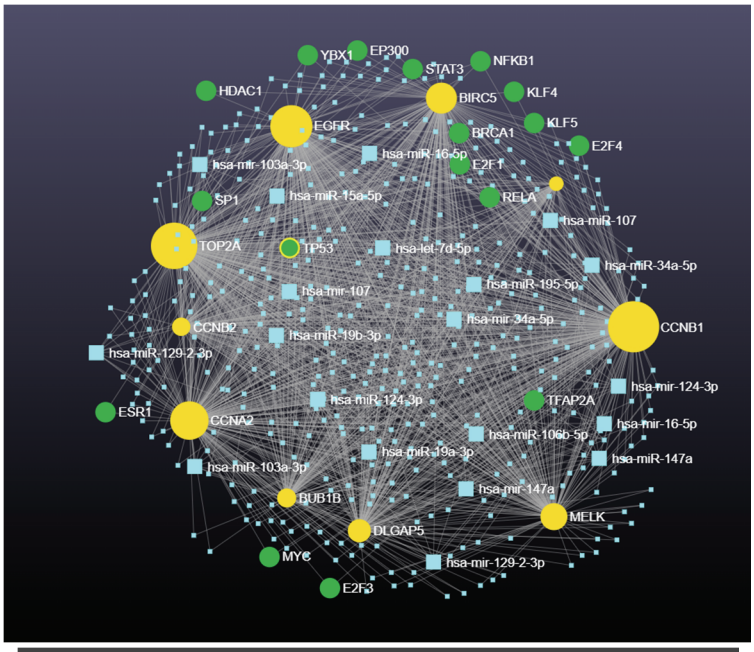
The miRNA–TF–gene interaction network constructed from the top under-expressed genes revealed critical insights into the regulatory mechanisms potentially suppressed in ALK-positive NSCLC. Key hub genes such as FGF2, TTN, IL6, VWF, BDNF, and CAV1 exhibited strong connectivity, indicating their significant but silenced roles in maintaining normal cellular signaling, angiogenesis, inflammation, and tissue

remodeling [Fig.10b]. These genes were found to be regulated by several important tumor-suppressor miRNAs, including hsa-miR-34a-5p, hsa-miR-210-3p, hsa-miR-27a-3p, and hsa-miR-19a-3p, suggesting a complex post-transcriptional suppression network. Transcription factors such as TP53, STAT3, NFκB1, RELA, STAT1, and CREB1 were also involved in modulating these genes and miRNAs, indicating a disrupted transcriptional regulation landscape. The suppression of this network may compromise immune response, cellular communication, and tumor microenvironment regulation, contributing to disease progression and offering potential targets for therapeutic restoration in ALK-positive NSCLC.

Names of DEGs,miRNAs and TFs	**Degree**
TTN	217
FGF2	207
CAV1	206
VWF	143
IL6	128
BDNF	122
KIT	102
hsa-miR-19a-3p	10
hsa-miR-210-3p	10
hsa-miR-27a-3p	10
hsa-miR-34a-5p	10
CREB1	2
EGR1	2
GATA6	2
KLF4	2
NFKB1	2
RELA	2
STAT1	2
STAT3	2
TP53	2

Table 4 List of top miRNAs, transcription factors, and top DEGs (underexpressed) with their degree of centrality.

10a.



10b.

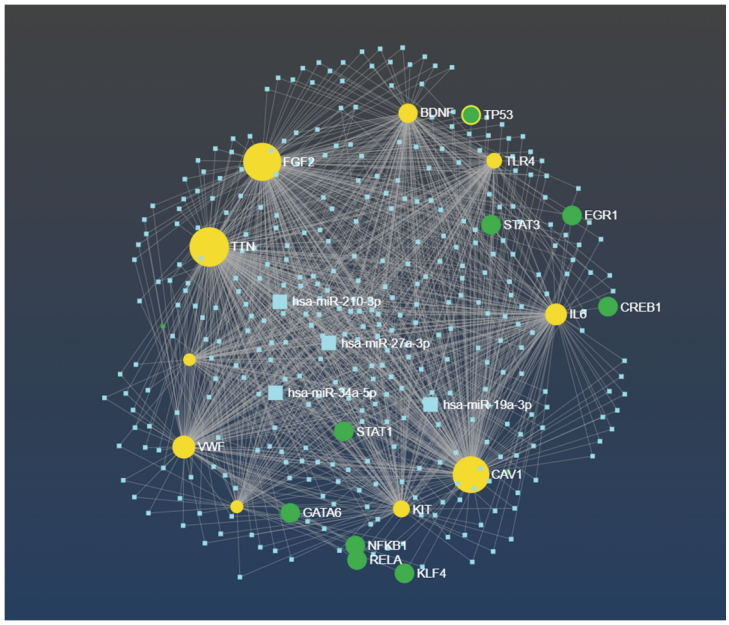


Fig. 10 a) The network was generated for the over-expressed hub genes. The genes CCNB1 and TOP2A have maximum connectivity (degrees 287 and 277 respectively) and the miRNA with maximum connectivity (degree 10) and transcription factors with maximum connectivity (degree 2, 3 and 5). b) The network generated using miRNet shows the interaction among DEGs (underexpressed), miRNAs, and transcription factors (TFs). The genes are represented in yellow, the miRNAs in blue, and the transcription factors in green. The genes TNN and FGF2 have maximum connectivity (degrees 217 and 207 respectively) and the miRNA hsa-miR-19a-3p, hsa-miR-210-3p, hsa-miR-27a-3p, hsa-miR-34a-5p are the ones with maximum connectivity (degree 10). All the transcription factors have the same connectivity (degree 2).

Regulatory Network Reveals Shared miRNA and Transcription Factor Control
Regulatory network analysis using miRNet revealed that the top hub genes are regulated by a complex network of miRNAs and transcription factors. Overexpressed genes such as *CCNB1*, *TOP2A*, and *EGFR* were found to be targeted by tumor-suppressive miRNAs including *hsa-miR-16-5p* and *hsa-miR-34a-5p*, and regulated by transcription factors such as *TP53* and *STAT3*. Conversely, underexpressed genes like *TTN*, *FGF2*, and *CAV1* were linked to miRNAs such as *hsa-miR-19a-3p* and TFs including *CREB1* and *RELA*. Notably, *TP53*, *STAT3*, and *hsa-miR-34a-5p* were present in both networks, suggesting their role as central regulators across multiple pathways. This highlights potential feedback dysregulation in ALK-positive NSCLC and supports the idea of shared regulatory mechanisms contributing to both oncogene activation and tumor suppressor silencing.

Conclusion

In conclusion, this study elucidates the complex molecular and regulatory landscapes of ALK-positive NSCLC. Integrating gene expression analysis, network analysis, and regulatory network analysis provides a holistic view of the pathways and processes driving this subtype of lung cancer. These findings pave the way for future research and potential therapeutic developments to target the specific molecular abnormalities in ALK-positive NSCLC. Further functional studies and clinical investigations are warranted to validate these potential biomarkers and therapeutic targets and to develop effective treatment strategies.

References

[1] Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, et al. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. Nature. 2007; 448: 561–6.

[2] Shaw, A. T., Yeap, B. Y., Mino-Kenudson, M., Digumarthy, S. R., Costa, D. B., Heist, R. S., ... & Iafrate, A. J. (2009). Clinical features and outcome of patients with non–small-cell lung cancer who harbor EML4-ALK. Journal of clinical oncology, 27(26), 4247-4253.

[3] Kwak, E. L., Bang, Y. J., Camidge, D. R., Shaw, A. T., Solomon, B., Maki, R. G., ... & Iafrate, A. J. (2010). Anaplastic lymphoma kinase inhibition in non–small-cell lung cancer. New England Journal of Medicine, 363(18), 1693-1703.

[4] McCusker, M. G., Russo, A., Scilla, K. A., Mehra, R., & Rolfo, C. (2019). How I treat ALK-positive non-small cell lung cancer. Esmo Open, 4, e000524.

[5] Khan, M., Lin, J., Liao, G., Tian, Y., Liang, Y., Li, R., ... & Yuan, Y. (2019). ALK inhibitors in the treatment of ALK positive NSCLC. Frontiers in oncology, 8, 557.

[6] Bang, Y. J. (2012). Treatment of ALK-positive non–small cell lung cancer. Archives of pathology & laboratory medicine, 136(10), 1201-1204.

[7] Okayama H, Kohno T, Ishii Y, Shimada Y et al. Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas. Cancer Res 2012 Jan 1;72(1):100-11. PMID: 22080568.

[8] Yamauchi M, Yamaguchi R, Nakata A, Kohno T et al. Epidermal growth factor receptor tyrosine kinase defines critical prognostic genes of stage I lung adenocarcinoma. PLoS One 2012;7(9):e43923. PMID: 23028479.

[9] Gene Expression Omnibus. (n.d.). GEO accession: GSE31210. National Center for Biotechnology Information. Retrieved July 16, 2024, from https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE31210

[10] Gene Expression Omnibus. (n.d.). GEO2R: Differential expression analysis tool. National Center for Biotechnology Information. Retrieved July 16, 2024, from https://www.ncbi.nlm.nih.gov/geo/geo2r/

[11] Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, Gable AL, Fang T, Doncheva NT, Pyysalo S, Bork P, Jensen LJ, von Mering C. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res. 2023 Jan 6;51(D1):D638-D646. doi: 10.1093/nar/gkac1000. PMID: 36370105; PMCID: PMC9825434.

[12] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003 Nov;13(11):2498-504. doi: 10.1101/gr.1239303. PMID: 14597658; PMCID: PMC403769.

[13] Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: identifying hub objects and sub-networks from complex interactome. BMC Syst Biol. 2014;8 Suppl 4(Suppl 4):S11. doi: 10.1186/1752-0509-8-S4-S11. Epub 2014 Dec 8. PMID: 25521941; PMCID: PMC4290687.

[14] Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, McDermott MG, Monteiro CD, Gundersen GW, Ma'ayan A. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. 2016 Jul 8;44(W1):W90-7. doi: 10.1093/nar/gkw377. Epub 2016 May 3. PMID: 27141961; PMCID: PMC4987924.

[15] Fan Y, Siklenka K, Arora SK, Ribeiro P, Kimmins S, Xia J. miRNet - dissecting miRNA-target interactions and functional associations through network-based visual analysis. Nucleic Acids Res. 2016 Jul 8;44(W1):W135-41. doi: 10.1093/nar/gkw288. Epub 2016 Apr 21. PMID: 27105848; PMCID: PMC4987881.

[16] Gasparini, P., Cascione, L., Landi, L., Carasi, S., Lovat, F., Tibaldi, C., ... & Croce, C. M. (2015). microRNA classifiers are powerful diagnostic/prognostic tools in ALK-, EGFR-, and KRAS-driven lung cancers. *Proceedings of the National Academy of Sciences*, *112*(48), 14924-14929.

[17] Li LL, Qu LL, Fu HJ, Zheng XF, Tang CH, Li XY, Chen J, Wang WX, Yang SX, Wang L, Zhao GH, Lv PP, Zhang M, Lei YY, Qin HF, Wang H, Gao HJ, Liu XQ. Circulating microRNAs as novel biomarkers of ALK-positive nonsmall cell lung cancer and predictors of response to crizotinib therapy. Oncotarget. 2017 Jul 11;8(28):45399-45414. doi: 10.18632/oncotarget.17535. PMID: 28514730; PMCID: PMC5542196.

[18] Murakami, Y., Mitsudomi, T., & Yatabe, Y. (2012). A screening method for the ALK fusion gene in NSCLC. *Frontiers in oncology*, *2*, 24.

[19] Du, X., Shao, Y., Qin, H. F., Tai, Y. H., & Gao, H. J. (2018). ALK-rearrangement in non-small-cell lung cancer (NSCLC). *Thoracic cancer*, *9*(4), 423-430.

[20] He, R., & Zuo, S. (2019). A robust 8-gene prognostic signature for early-stage non-small cell lung cancer. *Frontiers in Oncology*, *9*, 693.

[21] Zhang, X., Zhang, S., Yang, X., Yang, J., Zhou, Q., Yin, L., ... & Wu, Y. L. (2010). Fusion of EML4 and ALK is associated with development of lung adenocarcinomas lacking EGFR and KRAS mutations and is correlated with ALK expression. *Molecular cancer*, *9*, 1-12.

[22] Li, Z., Sang, M., Tian, Z., Liu, Z., Lv, J., Zhang, F., & Shan, B. (2019). Identification of key biomarkers and potential molecular mechanisms in lung cancer by bioinformatics analysis. *Oncology letters*, *18*(5), 4429-4440.

[23] Chen, Y., Wang, X. K., Wang, Y., Zong, J. W., Wang, S. Y., & Wan, X. Y. (2022). Bioinformatics Analysis Screened and Identified Key Genes as Potential Biomarkers for Progression of Lung Cancer. *J Coll Physicians Surg Pak*, *32*(6), 712-721.