

**INTEGRATIVE OMICS ANALYSIS FOR THE  
IDENTIFICATION OF KEY GENOMIC  
BIOMARKERS DRIVING TUMOREGENESIS IN  
LUNG ADENOCARCINOMA PATIENTS**

BY

**HASAN SHUJA**

A Thesis Presented to the  
**COLLEGE OF GRADUATE AND INTERDISCIPLINARY STUDIES**  
**KING FAHD UNIVERSITY OF PETROLEUM & MINERALS**  
DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the  
Requirements for the Degree of

**MASTER OF SCIENCE**

**In**

**LIFE SCIENCES**

**MAY 2025**

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN- 31261, SAUDI ARABIA

DEANSHIP OF GRADUATE STUDIES

This thesis, written by Hasan Shuja under the direction of his thesis advisor and approved by his thesis committee, has been presented and accepted by the Dean of Graduate Studies, in partial fulfillment of the requirements for the degree of **MASTER OF SCIENCE IN LIFE SCIENCES**.



Dr. Yaqub Mahanshi  
Department Chairman



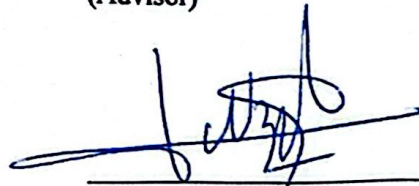
Dr. Suliman Al-Homidan  
Dean of Graduate Studies

21-05-2025

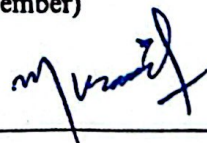
Date



Dr. Irshad Ahmad  
(Advisor)



Dr. Alexis Nzila  
(Member)



Dr. Muzammil Behzad  
(Member)



© Hasan Shuja

2025

To my all-family members,  
whose continuous support, endless prayers, and unconditional love have been the pillars  
of my strength throughout this journey

## **ACKNOWLEDGMENTS**

First and foremost, I would like to express my deepest gratitude to my advisor, Dr. Irshad Ahmad, for his invaluable guidance, patience, and encouragement throughout the course of this research. His insight and support throughout helped me in shaping both my thesis and my academic growth.

I am truly thankful to my thesis committee members, Dr. Alexis Nzila and Dr. Muzammil Behzad, for their support, time, and suggestions, which greatly enhanced the quality of this work.

A heartfelt thanks goes to my family for being my constant source of motivation and strength. Your endless sacrifices, love, and prayers have made this achievement possible.

I would also like to thank my friends and colleagues especially Mr. Uzair for helping me with the write up. And other friends for believing in me, and for providing the much-needed moments of laughter and encouragement.

Finally, I am deeply grateful to King Fahd University of Petroleum & Minerals (KFUPM) for providing me with the resources, environment, and opportunities to pursue my academic aspirations.

This journey would not have been the same without the vibrant academic community I was lucky to be a part of.

Thank you, all, from the bottom of my heart.

# TABLE OF CONTENTS

|                                                             |      |
|-------------------------------------------------------------|------|
| ACKNOWLEDGMENTS .....                                       | V    |
| TABLE OF CONTENTS .....                                     | VI   |
| LIST OF TABLES.....                                         | IX   |
| LIST OF FIGURES.....                                        | X    |
| LIST OF ABBREVIATIONS.....                                  | XII  |
| ABSTRACT .....                                              | XV   |
| ملخص الرسالة .....                                          | XVII |
| CHAPTER 1 INTRODUCTION.....                                 | 1    |
| 1.1 Classification and Staging of Lung Adenocarcinoma ..... | 1    |
| 1.2 Key Genomic Alterations in LUAD .....                   | 2    |
| 1.3 Targeted Therapies in LUAD .....                        | 2    |
| 1.4 Advancements in Omics Technologies .....                | 3    |
| 1.5 Research Gaps.....                                      | 4    |
| 1.6 Our Study .....                                         | 4    |
| CHAPTER 2 LITERATURE REVIEW .....                           | 6    |
| 2.1 Epidemiology .....                                      | 6    |
| 2.1.1 Global Incidence and Mortality .....                  | 7    |
| 2.1.2 Demographic Factors.....                              | 7    |
| 2.1.3 Risk Factors .....                                    | 7    |
| 2.1.4 Regional Variations .....                             | 8    |
| 2.2 Key Genetic Aberrations in Lung Adenocarcinoma .....    | 8    |
| 2.2.1 EGFR Mutations.....                                   | 8    |
| 2.2.2 KRAS Mutations.....                                   | 8    |
| 2.2.3 TP53 Mutations .....                                  | 9    |

|                                   |                                                                 |           |
|-----------------------------------|-----------------------------------------------------------------|-----------|
| 2.2.4                             | ALK Fusions .....                                               | 9         |
| 2.2.5                             | Tumor Suppressor Genes .....                                    | 9         |
| 2.3                               | Key Mechanisms of Resistance in LUAD .....                      | 10        |
| 2.3.1                             | Immune Evasion .....                                            | 10        |
| 2.3.2                             | Tumor Heterogeneity.....                                        | 10        |
| 2.3.3                             | Chronic Inflammation .....                                      | 11        |
| 2.4                               | Resistance to Targeted Therapies .....                          | 11        |
| 2.5                               | Omics Data availability and its usage.....                      | 13        |
| <b>CHAPTER 3 METHODOLOGY.....</b> |                                                                 | <b>14</b> |
| 3.1                               | Data Collection and Patient Stratification.....                 | 14        |
| 3.2                               | Differential Gene Expression Analysis .....                     | 14        |
| 3.3                               | Gene Set Enrichment Analysis (GSEA).....                        | 15        |
| 3.4                               | Identification of Mutation Hotspots and Driver Genes .....      | 16        |
| 3.5                               | Data Integration of Mutation and Expression .....               | 17        |
| 3.6                               | Variant Effect Predictor .....                                  | 18        |
| 3.7                               | Transcription Factor Binding Affinity Analysis.....             | 18        |
| 3.8                               | Final Filtering and Biomarker Identification .....              | 18        |
| <b>CHAPTER 4 RESULTS.....</b>     |                                                                 | <b>20</b> |
| 4.1                               | Patient Cohort Overview .....                                   | 20        |
| 4.2                               | Differential Gene Expression Analysis .....                     | 23        |
| 4.3                               | Enriched Pathways via Gene Set Enrichment Analysis (GSEA) ..... | 27        |
| 4.4                               | Mutation Hotspot and Driver Gene Detection .....                | 31        |
| 4.5                               | Integration of DEGs and Mutation Hotspots Data .....            | 33        |
| 4.6                               | Transcription Factor Binding Affinity Prediction (TRAP).....    | 35        |
| 4.7                               | Final Biomarker Shortlisting .....                              | 37        |
| <b>CHAPTER 5 DISCUSSION .....</b> |                                                                 | <b>39</b> |

|                                                         |           |
|---------------------------------------------------------|-----------|
| <b>CHAPTER 6 CONCLUSION AND FUTURE DIRECTIONS .....</b> | <b>43</b> |
| <b>REFERENCES.....</b>                                  | <b>46</b> |
| <b>VITAE.....</b>                                       | <b>54</b> |

## LIST OF TABLES

|                                                                                            |    |
|--------------------------------------------------------------------------------------------|----|
| <b>Table 1.</b> Showing Key mechanisms of resistance .....                                 | 12 |
| <b>Table 2.</b> Showing mechanisms and pathways leading to resistance against Therapy..... | 30 |

## LIST OF FIGURES

|                                                                                                                                                                                                                                                                                                                                                  |    |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| <b>Figure 1.</b> Distribution of LUAD patients based on TNM staging .....                                                                                                                                                                                                                                                                        | 20 |
| <b>Figure 2.</b> Demographic breakdown of LUAD patients by stage.....                                                                                                                                                                                                                                                                            | 22 |
| <b>Figure 3.</b> Heatmap showing expression patterns of the top 30 genes (by variance) in early stage LUAD across 40 tumour and 40 normal samples. No clear differential expression patterns were observed .....                                                                                                                                 | 23 |
| <b>Figure 4.</b> Proportion of significantly differentially expressed genes in late-stage LUAD patients. (A) Approximately 9% of genes showed statistically significant expression changes. (B) Proportion of significantly upregulated (19.5%) versus downregulated (80.5%) genes among all significant DEGs identified in late-stage LUAD..... | 24 |
| <b>Figure 5.</b> Heatmap of the top 50 differentially expressed genes in late-stage LUAD. The expression profiles distinguish tumor from normal samples based on log <sub>2</sub> fold change values .....                                                                                                                                       | 25 |
| <b>Figure 6.</b> MA plot showing the distribution of log <sub>2</sub> fold changes relative to average expression levels. Upregulated and downregulated genes are highlighted in red and blue, respectively .....                                                                                                                                | 26 |
| <b>Figure 7.</b> Volcano plot showing the distribution of log <sub>2</sub> fold changes relative to average expression levels. Upregulated and downregulated genes are highlighted in red and blue, respectively .....                                                                                                                           | 26 |
| <b>Figure 8.</b> Bar graph of top 20 Upregulated and Downregulated genes.....                                                                                                                                                                                                                                                                    | 27 |

|                                                                                                                                                                                                                  |    |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| <b>Figure 9.</b> GSEA enrichment plot for the E2F Targets pathway, showing strong positive enrichment in tumor samples. This pathway is associated with cell cycle progression and is linked to therapeutic..... | 28 |
| <b>Figure 10.</b> GSEA enrichment plot for the E2F Targets pathway, showing strong positive enrichment in tumor samples. ....                                                                                    | 29 |
| <b>Figure 11.</b> Detection of driver genes with mutation clustering in LUAD using OncodriveCLUSTL.....                                                                                                          | 32 |
| <b>Figure 12.</b> Distribution of mutation types among the selected genes based on Variant Effect Predictor (VEP) analysis. ....                                                                                 | 34 |
| <b>Figure 13.</b> Heatmap displaying transcription factor binding affinity changes computed for selected genes.....                                                                                              | 36 |
| <b>Figure 14.</b> Funnel-shaped workflow summarizing the stepwise gene filtering and prioritization pipeline used to identify LUAD biomarkers.....                                                               | 38 |

## LIST OF ABBREVIATIONS

|         |   |                                          |
|---------|---|------------------------------------------|
| LUAD    | : | Lung Adenocarcinoma                      |
| NSCLC   | : | Non-Small Cell Lung Cancer               |
| TNM     | : | Tumor-Node-Metastasis                    |
| EGFR    | : | Epidermal Growth Factor Receptor         |
| ALK     | : | Anaplastic Lymphoma Kinase               |
| ROS1    | : | c-ros Oncogene 1                         |
| KRAS    | : | Kirsten Rat Sarcoma Viral Oncogene       |
| BRAF    | : | B-Raf Proto-Oncogene                     |
| MET     | : | Mesenchymal-Epithelial Transition Factor |
| RET     | : | REarranged during Transfection           |
| HER2    | : | Human Epidermal Growth Factor Receptor 2 |
| TKI     | : | Tyrosine Kinase Inhibitor                |
| WES     | : | Whole Exome Sequencing                   |
| RNA-seq | : | RNA Sequencing                           |
| DEG     | : | Differentially Expressed Gene            |
| GSEA    | : | Gene Set Enrichment Analysis             |

|         |   |                                                |
|---------|---|------------------------------------------------|
| TCGA    | : | The Cancer Genome Atlas                        |
| VEP     | : | Variant Effect Predictor                       |
| TRAP    | : | Transcription Factor Affinity Prediction       |
| TME     | : | Tumor Microenvironment                         |
| TAMs    | : | Tumor-Associated Macrophages                   |
| MDSCs   | : | Myeloid-Derived Suppressor Cells               |
| NES     | : | Normalized Enrichment Score                    |
| FDR     | : | False Discovery Rate                           |
| MSigDB  | : | Molecular Signatures Database                  |
| UTR     | : | Untranslated Region                            |
| IL-6    | : | Interleukin-6                                  |
| IL-10   | : | Interleukin-10                                 |
| TNF     | : | Tumor Necrosis Factor                          |
| PCYT1A  | : | Phosphate Cytidyltransferase 1, Choline, Alpha |
| SPP1    | : | Secreted Phosphoprotein 1 (Osteopontin)        |
| COL11A1 | : | Collagen Type XI Alpha 1 Chain                 |
| SLC6A4  | : | Solute Carrier Family 6 Member 4               |

|                  |   |                                                        |
|------------------|---|--------------------------------------------------------|
| SPATA18          | : | Spermatogenesis Associated 18                          |
| OBSCN            | : | Obscurin Cytoskeletal Calmodulin And Titin-Interacting |
| MUC16            | : | Mucin 16, Cell Surface Associated                      |
| ATXN7L2          | : | Ataxin 7 Like 2                                        |
| LCP1             | : | Lymphocyte Cytosolic Protein 1                         |
| COL5A2           | : | Collagen Type V Alpha 2 Chain                          |
| KALRN            | : | Kalirin RhoGEF Kinase                                  |
| LRRC15           | : | Leucine Rich Repeat Containing 15                      |
| OncodriveCLUSTL: |   | Oncodrive Clustering Local Analysis Tool               |

## ABSTRACT

Full Name : Hasan Shuja

Thesis Title : Integrative Omics Analysis for the Identification of Key Genomic Biomarkers Driving Tumorigenesis in Lung Adenocarcinoma patients

Major Field : Life Sciences

Date of Degree : May 2025

Lung adenocarcinoma (LUAD), the most common subtype of non-small cell lung cancer, is driven by a complex alteration both at genomic as well as transcriptomic level that vary across disease stages. Despite advances in targeted therapies, the molecular transitions leading to LUAD progression remain incompletely understood. This study employed an integrative multi-omics approach, combining whole exome sequencing (WES) and RNA sequencing (RNA-seq) datasets from The Cancer Genome Atlas (TCGA), to systematically investigate stage-specific molecular changes.

Patients were stratified into early-stage and late-stage cohorts based on clinical TNM annotations. Mutation hotspot analysis using OncodriveCLUSTL identified both canonical drivers, such as KRAS and EGFR, and emerging candidates like OBSCN and MUC16, particularly enriched in late-stage disease. Differential gene expression analysis revealed minimal transcriptional differences in early stages but a substantial dysregulation of approximately 9.2% of genes in late-stage tumors, with a predominance of downregulation. Gene Set Enrichment Analysis (GSEA) highlighted critical oncogenic pathways—including E2F Targets, G2M Checkpoint, and Glycolysis—as significantly upregulated in late-stage tumors, while adipogenesis and inflammatory response pathways were notably suppressed.

By integrating mutational and transcriptomic data, 80 dual-affected genes were identified, subsequently filtered based on predicted functional impact and transcription factor binding affinity shifts. A final shortlist of four biomarker candidates—LCP1, COL5A2, KALRN, and LRRC15—was established, each implicated in hallmark cancer processes such as KRAS signaling, epithelial-mesenchymal transition, and angiogenesis.

These findings underscore the dynamic molecular evolution of LUAD and illustrate the power of integrative omics strategies in biomarker discovery. The study provides a foundation for future research aimed at experimental validation and therapeutic translation of the identified candidates to improve LUAD diagnosis and treatment.

## ملخص الرسالة

|                      |                                                                                                                         |
|----------------------|-------------------------------------------------------------------------------------------------------------------------|
| الاسم الكامل         | : حسن شجاع                                                                                                              |
| عنوان الرسالة        | : تحليل الأوميكس التكاملية لتحديد المؤشرات الحيوية الجينومية الرئيسية التي تسبب نشوء الأورام لدى مرضى سرطان الرئة الغدي |
| التخصص               | : علوم الحياة                                                                                                           |
| تاريخ الدرجة العلمية | : مايو ٢٠٢٥                                                                                                             |

وهو النوع الفرعي الأكثر شيوعا من سرطان الرئة ذو الخلايا غير الصغيرة، يسببه تغيرات معقدة على المستويين الجينومي والنسخي، تختلف باختلاف مراحل المرض. على الرغم من التقدم في العلاجات الموجهة، لا تزال التحولات الجزيئية المؤدية إلى تطور سرطان الرئة الغدي غير صغير الخلايا غير مفهومة تماما. اعتمدت هذه الدراسة نهجا تكامليا متعدد الأوميكس، يجمع بين مجموعات بيانات تسلسل الإكسوم الكامل (WES) وتسلسل الحمض النووي الريبوزي (RNA-seq) من أطلس جينوم السرطان (TCGA)، للتحقيق بشكل منهجي في التغيرات الجزيئية الخاصة بكل مرحلة تم تصنيف المرضى إلى مجموعات في المراحل المبكرة والمتأخرة بناء على شروح TNM السريرية. حدد تحليل نقطة الطفرة الساخنة باستخدام OncodriveCLUSTL كلا من العوامل الرئيسية، مثل KRAS وEGFR، والمرشحين الناشئين مثل OBSCN وMUC16، والتي تزداد أهميتها بشكل خاص في المراحل المتأخرة من المرض. كشف تحليل التعبير الجيني التفاضلي عن اختلافات نسخية طفيفة في المراحل المبكرة، مع خلل كبير في تنظيم حوالي 9.2% من الجينات في أورام المراحل المتأخرة، مع غلبة لانخفاض التنظيم. أبرز تحليل إثراء مجموعة الجينات (GSEA) مسارات مسرطنة مهمة - بما في ذلك أهداف E2F ونقطة تفتيش G2M وتحلل الجلوكوز - على أنها ارتفعت بشكل ملحوظ في أورام المراحل المتأخرة، بينما تم تثبيط مسارات تكوين الخلايا الدهنية والاستجابة الالتهابية بشكل ملحوظ من خلال دمج البيانات الطفرية والنسخية، تم تحديد 80 جينا مزدوج التأثير، ثم فصلت بناء على التأثير الوظيفي المتوقع وتحولات تقارب ارتباط عوامل النسخ. وضعت قائمة نهائية مختصرة لأربعة جينات مرشحة للمؤشرات الحيوية LCP1 - و COL5A2 و KALRN و - LRRRC15 كل منها متورط في عمليات السرطان المميزة مثل إشارات KRAS، والتحول الظهاري المتوسطي، وتكوين الأوعية الدموية

تؤكد هذه النتائج على التطور الجزيئي الديناميكي لمتلازمة LUAD ، وتظهر قوة استراتيجيات الجينوم التكاملية في اكتشاف المؤشرات الحيوية. توفر الدراسة أساسا لأبحاث مستقبلية تهدف إلى التحقق التجريبي والتطبيق العلاجي للمرشحين المحددين لتحسين تشخيص وعلاج متلازمة

# **CHAPTER 1**

## **INTRODUCTION**

Lung cancer remains the most fatal cancer worldwide, with non-small cell lung cancer (NSCLC) comprising the majority of cases over 85%. Among its subtypes, the most common one to be diagnosed is the lung adenocarcinoma (LUAD), representing about half of the NSCLCs. LUAD tends to arise from the distal airways with glandular histology and unique expression of markers TTF1 and KRT7 [1]. It essentially derives from progenitor clones of cells that incur early genomic mutations, which are responsible for the initiation of tumour formation. Subclonal populations arise from these clones with the passage of time, becoming adapted to microenvironment conditions and leading to changes in cell identity and regulation of lineage [2].

### **1.1 Classification and Staging of Lung Adenocarcinoma**

According to the tumor-node-metastasis (TNM) classification system, lung adenocarcinoma (LUAD) is categorized into four stages: I, II, III, and IV [3]. Stage I represents the initial phase of the disease, with no metastasis. Stages II and III are intermediate stages that involve regional lymph node metastasis, and stage III exhibits a greater extent of lymphatic spread than stage II. Stage IV is the advanced phase of the disease, that typically includes distant metastasis

LUAD is characterized by a diverse range of genomic alterations at molecular level, that drive tumor development and progression. The occurrence of LUAD has been rising annually at a global level, involving a shift toward younger patients, a faster disease progression, absence of early symptoms, and is associated with high mortality and an unfavorable prognosis [4]. The main reasons include increased industrialization and tobacco consumption, especially in developing nations. The continuous exposure of lungs to pollutants and other environmental factors significantly influences LUAD development through the activation of progenitor cells, genetic alterations, and tumorigenesis [4]

## **1.2 Key Genomic Alterations in LUAD**

Key mutations and rearrangements occur in genes such as *EGFR*, *ALK*, *ROS1*, *KRAS*, *BRAF*, *MET*, *RET*, and *HER2*, many of which serve as actionable targets for precision therapies [5]. For instance, *EGFR* mutations—especially exon 19 deletions and the L858R point mutation—are common in non-smokers and Asian populations and are effectively treated with tyrosine kinase inhibitors (TKIs). Similarly, rearrangements in *ALK* and *ROS1* are mutually exclusive with *EGFR* mutations and have led to the development of specific targeted agents. Other alterations, including *KRAS* mutations (notably G12C), *BRAF* V600E mutations, *MET* exon 14 skipping, and *HER2* insertions, contribute to the molecular heterogeneity of LUAD and are increasingly being explored for therapeutic intervention [5].

## **1.3 Targeted Therapies in LUAD**

Lung cancer, and particularly non-small cell lung cancer (NSCLC), often shows promising responses to targeted treatments like EGFR and ALK inhibitors in the early phases.

However, one of the key obstacles in long-term management is the development of drug resistance [6]. Numerous processes, such as additional mutations (EGFR T790M or C797S), activation of different signaling pathways like MET or HER2, enhance cellular flexibility, or immunological escape by modifying the tumor microenvironment. Disease relapse and reduced therapeutic efficacy are caused by these factors. As a result, current research is concentrated on multi-target drug combinations, customized treatment plans based on molecular profiling, and advanced therapies that simultaneously address both initial and resistance-inducing alterations. Furthermore, new approaches are being developed, including improving immunotherapy, focusing on epigenetic changes, utilizing synthetic lethality, and employing innovative drug delivery systems. These approaches are aimed to achieve more effective and long-lasting control over lung cancer progression [7].

#### **1.4 Advancements in Omics Technologies**

Large-scale, accurate molecular profiling of LUAD is now possible owing to the development of high-throughput omics technologies [8]. This has helped make feasible to analyze omics datasets pertinent to the disease and conduct thorough research into the intricate molecular mechanisms underlying LUAD [9]. Numerous biomarkers associated with diagnosis and prognosis have been discovered as a result of these investigations, providing opportunities for targeted treatments. However, most of the omics-driven cancer research mainly depends on survival-based expression profiling, differential gene expression analyses, or mutation significance tests. They frequently generate a large number of candidate genes or proteins, many of which need additional testing to ensure their clinical applicability [10].

## **1.5 Research Gaps**

Critical gaps remain in the interpretation of genomic alterations and their functional implications, despite tremendous advancements in our understanding of the molecular biology of LUAD. The majority of recent research tends to look at transcriptomic alterations and genomic mutations independently, which limits our understanding of how somatic mutations directly affect gene expression and subsequent biological pathways. Furthermore, despite the identification of many oncogenic mutations, There is limited knowledge about the actual mechanism by which these mutations drive transcriptional changes leading to advancement of disease [11].

The need for integrative multi-omics techniques that integrate whole exome sequencing (WES) and RNA sequencing (RNA-seq) to determine the mechanistic association between mutation hotspots and changes in gene expression and pathway dysregulation across disease stages is needed in order to overcome these gaps [12]. Particularly there is a lack of studies that investigate how driver mutations evolve from early to late-stage LUAD, and how these mutations contribute to hallmark cancer pathways by disrupting gene expression

## **1.6 Our Study**

In this study, a comprehensive multi-omics approach was employed to investigate molecular alterations in lung adenocarcinoma (LUAD) using data retrieved from The Cancer Genome Atlas (TCGA). RNA-Seq and Whole Exome Sequencing (WES) datasets were downloaded and prepared for LUAD patients. Clinical stage annotations were used to stratify patients into early-stage and late-stage cohorts. To identify driver mutations with

positional clustering, **OncodriveCLUSTL** was applied separately to each cohort, resulting in a list of genes exhibiting significant mutation hotspots. Parallel to this, differential expression analysis was performed using RNA-Seq data, applying statistical tests to identify differentially expressed genes (DEGs) between tumor and normal samples within each group. While no significant DEGs were observed in the early-stage group, nearly 4,000 DEGs were identified in the late-stage cohort, with several genes significantly enriched in hallmark cancer pathways.

To investigate pathway-level alterations, **Gene Set Enrichment Analysis (GSEA)** was performed on the late-stage DEGs, revealing twelve hallmark pathways that were significantly dysregulated. The study then focused on identifying genes that overlapped between those carrying clustered mutations (from OncodriveCLUSTL) and those that were differentially expressed. These intersecting genes were further analyzed to assess the functional impact of their mutations. Variant effect annotations were obtained using the **Ensembl Variant Effect Predictor (VEP)**, while changes in transcription factor binding affinity were evaluated using the **TSAR (Transcription factor affinity changes)** tool. Finally, pathways that included these dual-affected genes as core enriched members were filtered and prioritized, offering novel insights into the mechanistic links between somatic mutations, gene dysregulation, and cancer progression in LUAD.

## **CHAPTER 2**

### **LITERATURE REVIEW**

Lung adenocarcinoma is the most prevalent form of lung cancer and a significant subtype of non-small cell lung cancer (NSCLC) [13]. It is characterized by its origin in the glandular cells lining the lungs, specifically in the distal airways, and accounts for about 40% of all cases of lung cancer worldwide. This form of cancer is characterized by a high level of intra- and inter-tumor heterogeneity, which makes treatment more difficult and results in worse prognoses [14]. Histological patterns of LUAD include lepidic, acinar, papillary, micropapillary, and solid [15]. These patterns are categorized according to their growth characteristics and have the potential to impact clinical outcomes. Lepidic-predominant adenocarcinomas, for example, have higher survival rates than solid-predominant types, which typically have worse results [16].

Genetic mutations play a major role in the development of LUAD. EGFR gene mutations and other chromosomal abnormalities are frequent that can cause carcinogenesis [17]. Developing targeted therapies requires an understanding of these genetic factors.

#### **2.1 Epidemiology**

LUAD is a significant subtype of lung cancer that has seen increasing incidence rates globally. Here are the key epidemiological characteristics and trends associated with LUAD:

### **2.1.1 Global Incidence and Mortality**

Globally, lung cancer—including LUAD—is the primary cause of cancer-related mortality. Lung cancer accounted for 18.4% of all cancer deaths and 11.6% of all cancer incidences in 2018, with an estimated 2.09 million new cases [18].

LUAD represents about 38.5% of all lung cancers, with its incidence rising particularly among non-smokers and women in many regions [19].

### **2.1.2 Demographic Factors**

**Age:** The average age of diagnosis for lung cancer is around 70 years. A significant proportion of cases occur in individuals aged 55 to 74 years, with the highest incidence rates observed in older age groups [20].

**Sex:** Historically, lung cancer incidence has been higher in men due to smoking patterns. However, recent studies indicate a convergence in incidence rates between sexes, with some populations reporting higher rates among young women compared to men. In Taiwan, for instance, women exhibited higher age- and tumor stage-specific incidence rates than men [21].

### **2.1.3 Risk Factors**

**Smoking:** Tobacco smoking remains the primary risk factor for lung adenocarcinoma, accounting for up to 90% of cases globally. However, the proportion of LUAD cases among non-smokers is increasing, particularly in certain demographics [22].

**Environmental Factors:** Exposure to air pollution, occupational hazards (like asbestos), and second-hand smoke significantly contribute to the risk of developing LUAD [23]. Additionally, living in severely polluted cities has been linked to higher incidence rates

#### **2.1.4 Regional Variations**

In regions like North China (Hebei Province), LUAD has become the most common histological type of lung cancer, with a notable rise among non-smokers. The incidence rates vary significantly by geography; for instance, African American men in the U.S. have the highest incidence rates compared to other racial groups [24].

### **2.2 Key Genetic Aberrations in Lung Adenocarcinoma**

#### **2.2.1 EGFR Mutations**

The EGFR gene is among the top mutated genes in LUAD patients, especially among never-smokers and of Asian ethnicity [25]. Their mutation often lead to the activation of downstream signaling pathways that promotes tumor growth and survival, providing a target for specific inhibitors.

#### **2.2.2 KRAS Mutations**

KRAS mutations are also common in LUAD, almost exclusively among smokers. The common gain-of-function mutations occur at codons 12, 13, and 61 of KRAS gene and leads to activation of pathways that drive oncogenic processes [26].

### **2.2.3 TP53 Mutations**

Among the most important tumor suppressors, TP53 is mutated in almost 70% of LUAD cases, which generally cause loss of function, enabling uninhibited cell division and tumor progression [17].

### **2.2.4 ALK Fusions**

ALK rearrangements have equal importance in LUAD. The result of these fusions is the expression of an active tyrosine kinase fusion protein, driving the growth of cancer cells through continuous activity. Targeted treatments against ALK fusions have resulted in promising results [27].

Additional genetic alterations include mutations in *BRAF*, *HER2*, *MET*, and fusions involving *RET* and *ROS1*. The skipping of exon 14 of the *MET* gene has also been identified as a relevant alteration that can occur independently of other driver mutations [28].

### **2.2.5 Tumor Suppressor Genes**

Besides *TP53*, other tumor suppressor genes like *STK11* and *CDKN2A* are frequently altered in LUAD, contributing to tumorigenesis through loss of regulatory control over cell growth and division [29].

## 2.3 Key Mechanisms of Resistance in LUAD

LUAD is characterized by significant genetic and immune resistance mechanisms that complicate treatment. Understanding these mechanisms is crucial for developing effective therapies.

### 2.3.1 Immune Evasion

LUAD cells can evade the immune system through various strategies:

**Altered Antigen Presentation:** Tumor cells may downregulate major histocompatibility complex (MHC) molecules, reducing their visibility to T cells [30].

**Immunosuppressive Microenvironment:** The tumor microenvironment (TME) is often enriched with immunosuppressive cells, such as tumor-associated macrophages (TAMs) and myeloid-derived suppressor cells (MDSCs). These cells secrete factors that inhibit T cell activation and promote tumor growth [31].

**Cytokine Secretion:** Cytokines secreted by the LUAD cells may also attract some immunosuppressive cells and promote a type of TME that inhibits effective anti-tumor responses. They could upregulate the interleukins such as IL-6 and IL-10, along with tumor necrosis factor, TNF, contributing to immune suppression [32].

### 2.3.2 Tumor Heterogeneity

LUAD is highly genetically heterogeneous; it contains many different mutations within a single tumor. This kind of diversity makes treatment difficult because various subpopulations of malignant cells may respond differently to therapies. For example:

**Mutational Landscape:** Most of the mutations commonly involve genes such as EGFR, KRAS, and TP53, each of which may present different resistance profiles, especially concerning targeted therapies [33].

**Intratumoral Variability:** Even in patients with identical driver mutations, the genetic landscape of the tumors may be quite different [33].

### 2.3.3 Chronic Inflammation

Moreover, tobacco smoke, among other environmental factors, influences the development of chronic inflammation, which plays a significant role in the development and resistance of LUAD:

**Inflammatory Signaling:** This can be mediated by the inflammatory response brought about by tobacco smoke, which might facilitate tumorigenesis. Such an environment may enhance genomic instability and support cancer cell survival [34].

**Immune Modulation:** Inflammation can alter the behavior of the immune cells, polarizing them towards a more suppressive phenotype that favors tumor growth rather than destruction

## 2.4 Resistance to Targeted Therapies

Although targeted therapies have shown promise in treating LUAD, resistance remains a significant challenge:

### Primary and Secondary Resistance:

Majority of patients eventually develop resistance through specific therapies directed against mutations, such as inhibitors of EGFR, due to secondary mutations or activation of other pathways. KRAS indeed carries a particularly bad reputation in the light of its resistance mechanisms. Patients with KRAS mutations generally have few effective targeted treatment options and, as a result, have very poor outcomes [35].

**Table 1. Showing Key mechanisms of resistance**

| <b>Mechanism of Resistance</b>   | <b>Population</b>                        | <b>Causes</b>                                                                       | <b>Key Changes</b>                                    | <b>References</b> |
|----------------------------------|------------------------------------------|-------------------------------------------------------------------------------------|-------------------------------------------------------|-------------------|
| Intrinsic Resistance             | 4–10% patients                           | Existence of activating yet non-sensitizing driver mutations                        | EGFR-T790M, MET amplification and BCL2L11             | [36]              |
| Secondary or acquired resistance | 50% of patients with reported resistance | Secondary somatic mutations of the target gene which hinder the binding of the drug | EGFR-T790M<br><br>ALK-L1196M and ALK-G1269A mutations | [37]              |

|                       |                            |                                                                |                                                     |      |
|-----------------------|----------------------------|----------------------------------------------------------------|-----------------------------------------------------|------|
| Adoptive Resistance   | ~ 20%                      | Activating bypass signaling pathways                           | PI3K/AKT or RAF/MEK/ERK pathways MAPK activation    | [37] |
| Phenotypic Resistance | ~ 5–14% of resistant cases | Tumor transforms from NSCLC to SCLC to escape targeted therapy | Loss of EGFR expression, RB1 and TP53 inactivation, | [38] |

## 2.5 Omics Data availability and its usage

While high-throughput technologies and data integration methods continue to advance, access to omics data has grown exponentially—from genomics and transcriptomics to proteomics and metabolomics. Such wealth has consequently facilitated in-depth analysis of biological systems and disease mechanisms by researchers, particularly in cancer research [39]. The integration of various omics data can allow the researcher to explore the complex biological relationships and identify new biomarkers important for early detection, diagnosis, and treatment stratification in cancer patients. So far, machine learning techniques are increasingly being used to analyze multi-omics data to improve their capability in predicting disease outcomes and tailoring personalized treatments. This integrative approach is promoting further insights into tumor biology and, at the same time, enables precision medicine strategies that will finally benefit oncology patient care [40].

## CHAPTER 3

# METHODOLOGY

### 3.1 Data Collection and Patient Stratification

In this study, transcriptomic and genomic data were obtained for lung adenocarcinoma (LUAD) patients from The Cancer Genome Atlas (TCGA) database. Specifically, RNA sequencing (RNA-seq) data and whole exome sequencing (WES) data were collected for a total of 235 LUAD patients. Clinical metadata associated with each patient, including tumor staging information based on the TNM classification system [41], was used to stratify the cohort into two groups: early-stage patients (stage IA to IIB;  $n = 176$ ) and late-stage patients (stage IIIA to IV;  $n = 59$ ). Prior to analysis, both RNA-seq and WES datasets were subjected to quality control and preprocessing steps. These included the removal of low-quality reads, normalization of expression values (e.g., TPM or FPKM for RNA-seq), and alignment to the human reference genome. Duplicate entries and samples lacking complete clinical annotations were excluded to ensure dataset consistency and reliability for downstream analyses.

### 3.2 Differential Gene Expression Analysis

Differential expression analysis was conducted using the RNA-Seq data obtained from TCGA to compare tumor and normal samples in LUAD patients. The dataset, expressed in Transcripts Per Million (TPM), was first normalized using a log transformation.

$$\log_2(TPM + 1)$$

which reduces expression variance and mitigates skewness, ensuring all values remain positive. Patients were previously categorized into early-stage (IA–IIB) and late-stage (IIIA–IV) cohorts. For each gene, an independent two-sample **t-test** was applied to assess statistical differences in mean expression between tumor and normal tissues [42]. The t-test computes a t-statistic considering group variances and sizes, followed by the calculation of a **p-value**, which indicates the probability that observed differences occurred by chance. To address the issue of multiple hypothesis testing across thousands of genes, **Benjamini-Hochberg correction** was used to control the false discovery rate (FDR). Genes with an adjusted p-value < 0.05 and absolute **log<sub>2</sub> fold change (Log<sub>2</sub>FC) > 1** were considered significantly differentially expressed. The fold change was computed using the formula:

$$\log_2FC = \log_2\left(\frac{\text{mean}(TPM\_tumor)}{\text{mean}(TPM\_normal)}\right)$$

### 3.3 Gene Set Enrichment Analysis (GSEA)

Following differential gene expression analysis, **Gene Set Enrichment Analysis (GSEA)** was performed to identify biologically significant pathways associated with late-stage LUAD. A ranked gene list was constructed based on the log<sub>2</sub> fold change values obtained from the RNA-Seq data analysis, with genes ordered from highest to lowest expression change between tumor and normal samples. The **GSEA desktop application (Broad Institute)** was used with the **Hallmark gene sets** from the **Molecular Signatures Database (MSigDB v7.5.1)** as the reference pathway collection [43]. Enrichment scores (ES) were calculated using a Kolmogorov–Smirnov-like statistic that evaluates whether

genes from each set are overrepresented at the extremes (top or bottom) of the ranked list. Significance was assessed through permutation testing (1,000 permutations), and results were corrected for multiple hypothesis testing using the **False Discovery Rate (FDR)**. Pathways with FDR q-values < 0.25 and nominal p-values < 0.05 were considered significantly enriched. This analysis allowed us to identify key cancer-related pathways dysregulated in late-stage LUAD, guiding downstream mutation impact and pathway prioritization.

### 3.4 Identification of Mutation Hotspots and Driver Genes

To uncover key driver mutations and mutation hotspots in LUAD, we utilized whole exome sequencing (WES) data from TCGA, previously stratified into early-stage (IA–IIB) and late-stage (IIIA–IV) cohorts. The detection of mutation hotspots was based on the principle that a high concentration of somatic mutations within specific regions of a gene suggests potential disruption of functionally critical domains, offering a selective growth advantage to cancer cells and increasing the likelihood of oncogenic relevance [44]. For this purpose, we employed **OncodriveCLUSTL**, a position-based clustering tool that identifies genes harboring non-random aggregations of mutations.

For each gene, OncodriveCLUSTL computes a **clustering score** ( $S_g$ ) using the following formula:

$$S_g = \sum_{i=1}^{n_c} \frac{m_i}{M_g} \cdot \left(1 - \frac{l_i}{L_g}\right)$$

Where:

- $n_c$  is the number of clusters in gene  $g$

- $m_i$  is the number of mutations in cluster  $i$
- $M_g$  is the total number of mutations in gene  $g$
- $l_i$  is the length of cluster  $i$
- $L_g$  is the total coding length of gene  $g$

This score captures both the density and compactness of mutation clusters, emphasizing tightly packed mutational regions

The tool was applied independently to both early and late-stage cohorts to identify potential stage-specific drivers. For each gene, we computed the number of significant mutation clusters, the total number of mutations falling within these clusters, and the proportion of clustered mutations relative to the total mutation count in that gene. Furthermore, **p-values** and **false discovery rate (FDR)** corrections were calculated to assess the statistical significance of the clustering. This approach allowed us to highlight genes with strong positional bias in mutational patterns, which are likely to play a role in LUAD development and progression

### 3.5 Data Integration of Mutation and Expression

We identified genes that carried clustered mutations and were significantly differentially expressed by performing an intersection after differential expression and mutation hotspot analysis. Following this filtering step, 80 genes that were shared by both datasets were identified as potential candidates for additional functional impact evaluation and pathway-level analysis.

### 3.6 Variant Effect Predictor

We employed Ensembl's Variant Effect Predictor (VEP) tool to assess whether the discovered mutations might affect gene expression [45]. This tool was used to annotate the genomic location and functional impact of mutations in the 80 candidate genes, including whether they were found in untranslated regions (UTRs), coding regions, splice sites, or regulatory elements. The possible structural or regulatory effects of the observed variations were identified with the aid of this analysis.

### 3.7 Transcription Factor Binding Affinity Analysis

For every mutation in the 80 dual-affected genes, TRAP was used to calculate the change in transcription factor binding scores between the reference and mutated sequences, so assessing the regulatory impact of the identified mutations [46]. TRAP was used to compute the change in transcription factor binding scores between the reference and mutated sequences. Variants showing substantial shifts in affinity were considered to have a potential role in modulating gene expression, thereby contributing to transcriptional dysregulation in LUAD.

### 3.8 Final Filtering and Biomarker Identification

As a final step, we applied a multi-criteria filtering strategy to prioritize potential biomarkers. From the pool of 80 candidate genes, we retained only those that met all of the following criteria: (1) presence within the **core enrichment region** of significantly dysregulated cancer-related pathways identified via GSEA; (2) evidence of **mutation clustering** based on OncodriveCLUSTL results; (3) functionally significant variants located in **critical genomic regions** (e.g., coding domains, regulatory elements, splice sites) as determined by Ensembl VEP; and (4) substantial **changes in transcription factor**

**binding affinity**, as quantified by TRAP scores. Genes satisfying all these conditions were designated as high-confidence **biomarker candidates**, given their combined transcriptional dysregulation, mutational impact, and potential regulatory influence in LUAD progression.

## CHAPTER 4

# RESULTS

### 4.1 Patient Cohort Overview

A total of 235 lung adenocarcinoma (LUAD) patients were included in this study, with clinical staging information derived from the TNM classification system. This system stratifies patients based on tumour size (T), lymph node involvement (N), and the presence of metastasis (M), offering a standardized framework to assess cancer progression. Based on this classification, patients were divided into two major cohorts: an **early-stage group (Stage IA–IIB)** consisting of 176 individuals (74.9%) and a **late-stage group (Stage IIIA–IV)** comprising 59 individuals (25.1%) (Figure 1).

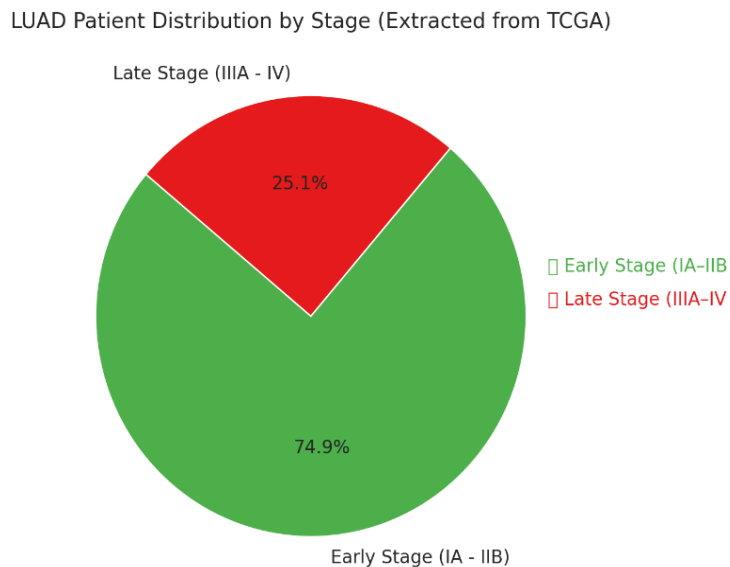


Figure 1. Distribution of LUAD patients based on TNM staging

Further demographic breakdowns revealed stage-specific patterns (Figure 2). Age distribution showed that the majority of early-stage diagnoses occurred in the 60–70 age group, whereas late-stage cases were more evenly distributed across age brackets, albeit still peaking within the 50–70 range. Gender analysis indicated a male predominance across both cohorts, with 111 males and 65 females in the early-stage group and 39 males and 20 females in the late-stage group. Race distribution analysis showed that most patients identified as **Asian** or **White**, with a notably higher representation of these groups in the early-stage cohort. These demographic patterns provide important context for interpreting downstream molecular analyses.

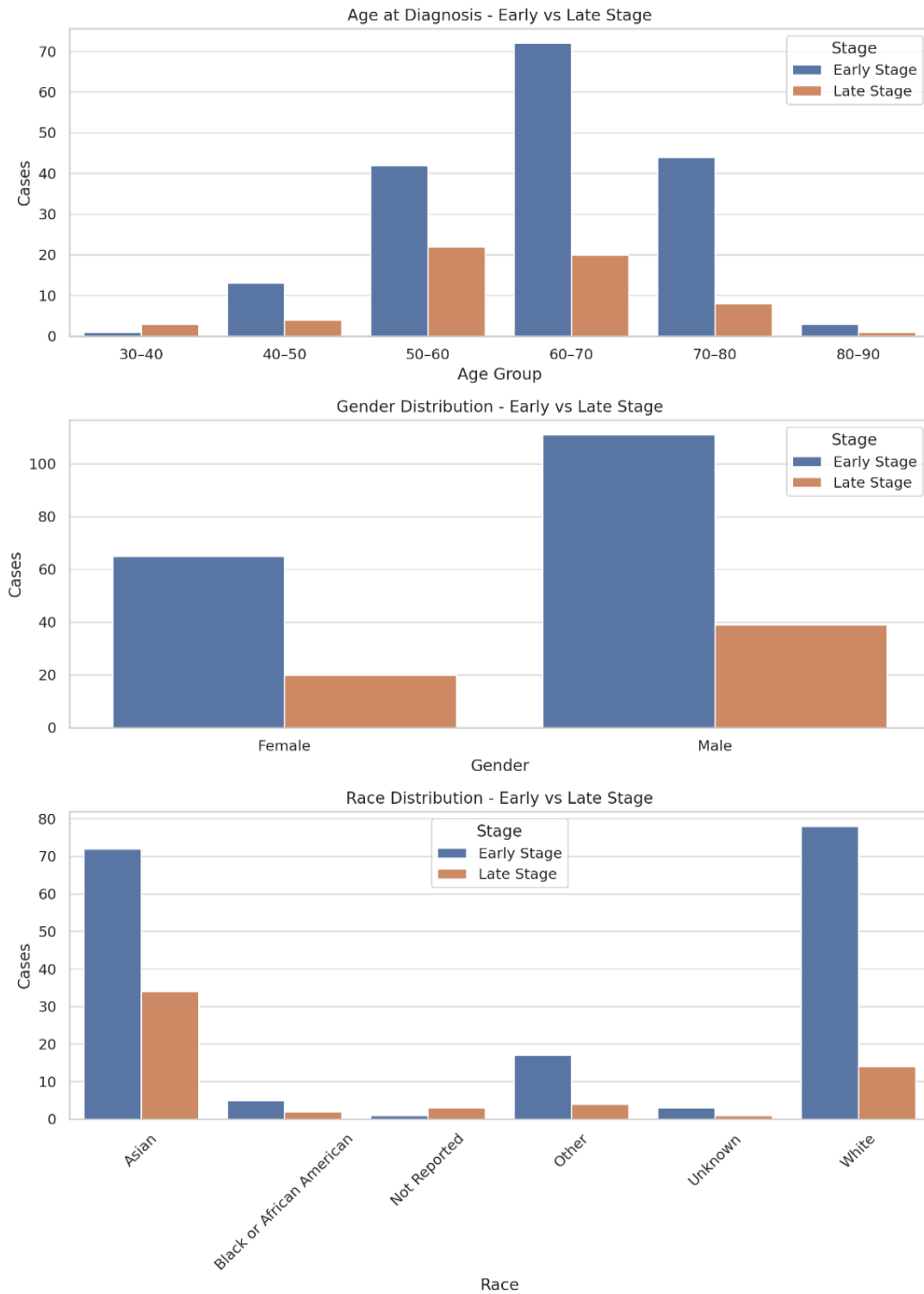


**Figure 2. Demographic breakdown of LUAD patients by stage**

## 4.2 Differential Gene Expression Analysis

Differential expression analysis was performed to investigate transcriptional changes in LUAD across different stages. For early-stage patients (IA–IIB), no genes were found to be significantly differentially expressed. A heatmap illustrating the expression profiles of the top 30 genes (by variance) across 40 tumour and 40 normal samples confirmed a lack of distinct expression patterns between conditions (Figure 3), supporting the absence of statistically significant differences.

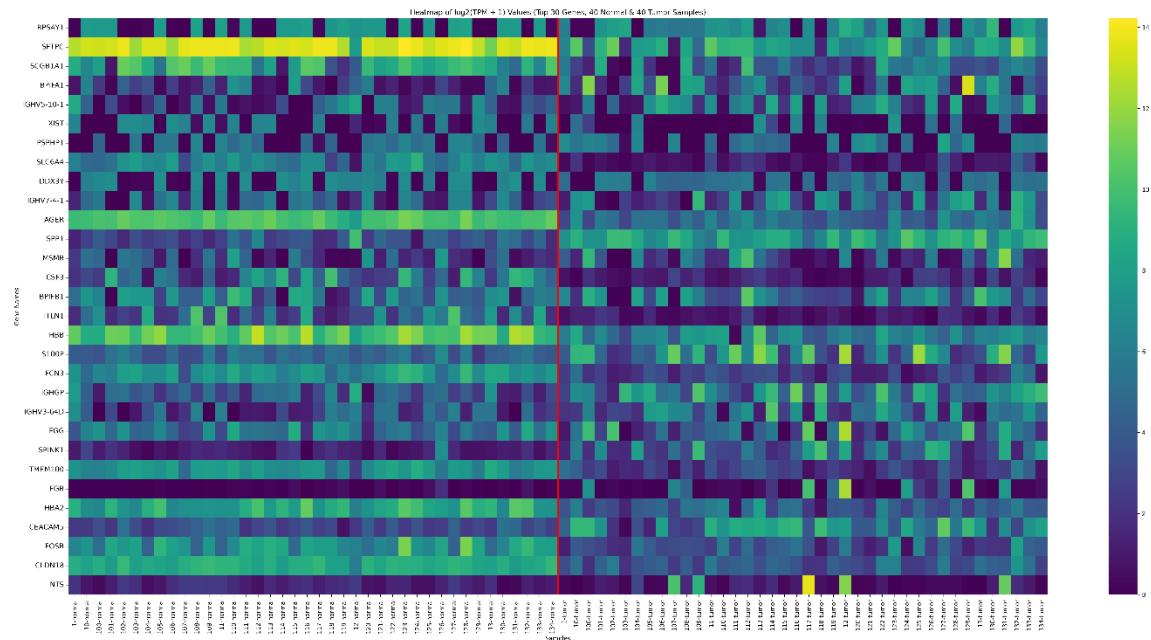


**Figure 3. Heatmap showing expression patterns of the top 30 genes (by variance) in early-stage LUAD across 40 tumour and 40 normal samples. No clear differential expression patterns were observed**

In contrast, analysis of the late-stage cohort (Stage IIIA–IV) revealed a marked shift in gene expression. Out of approximately 60,000 genes analysed, **about 9.2%** were found to be significantly differentially expressed, as visualized in the pie chart (Figure 4a). These genes were further analysed to uncover expression trends. The majority were found to be

**downregulated (80.5%)**, while a smaller fraction was **upregulated (19.5%)** in tumour samples (Figure 4b).

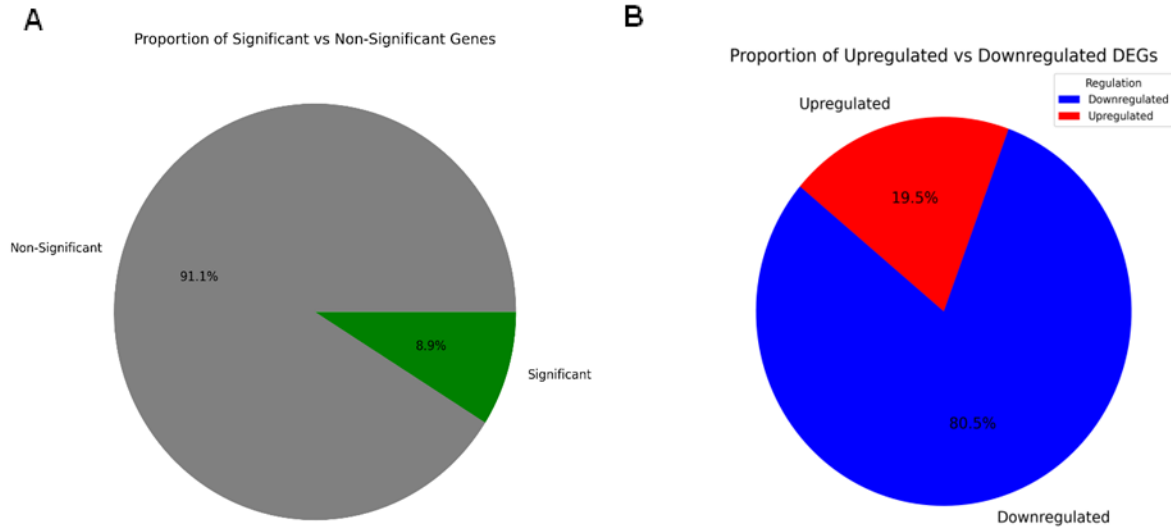


**Figure 4. Proportion of significantly differentially expressed genes in late-stage LUAD patients. (A) Approximately 9% of genes showed statistically significant expression changes. (B) Proportion of significantly upregulated (19.5%) versus downregulated (80.5%) genes among all significant DEGs identified in late-stage LUAD**

Additional visualizations provided deeper insights into the expression dynamics. A heatmap of the top 50 differentially expressed genes (DEGs) based on  $\log_2$  fold change clearly showed separation between tumour and normal profiles (Figure 5). The **MA plot** (Figure 6) and **volcano plot** (Figure 7) further highlighted the magnitude and statistical significance of these changes, with several genes such as *SPP1*, *COL11A1*, and *SLC6A4* standing out due to their strong differential expression. The top 20 upregulated and downregulated genes emphasized the sharp contrast in expression (Figure 8), particularly for genes like *SFTPC* and *SPP1*, which displayed extreme  $\log_2FC$  values.

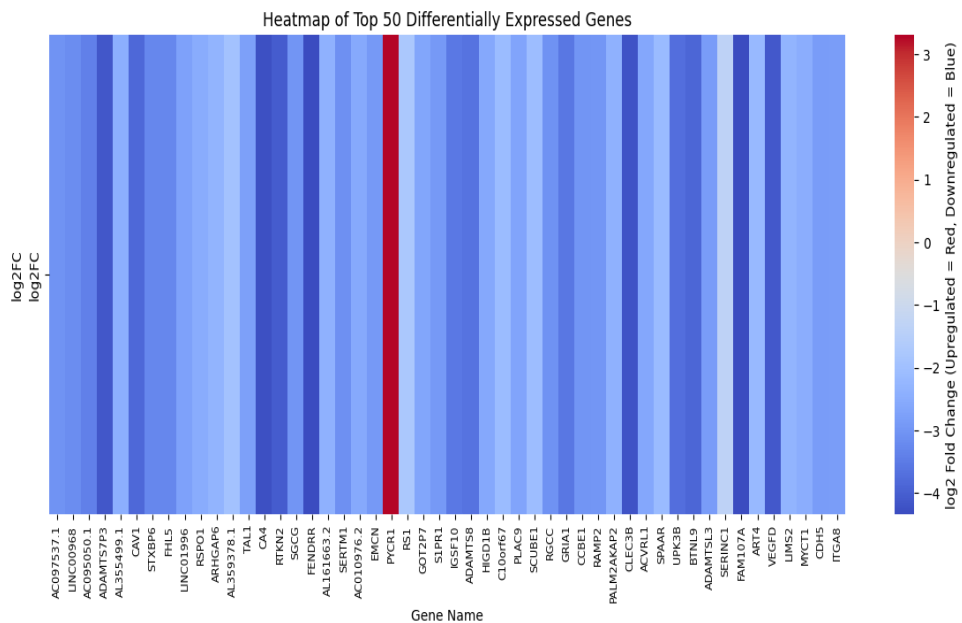


**Figure 5. Heatmap of the top 50 differentially expressed genes in late-stage LUAD. The expression profiles distinguish tumor from normal samples based on log<sub>2</sub> fold change values**

Together, these findings suggest a pronounced reprogramming of gene expression in late-stage LUAD, implicating many genes in processes potentially related to tumor progression and metastasis.

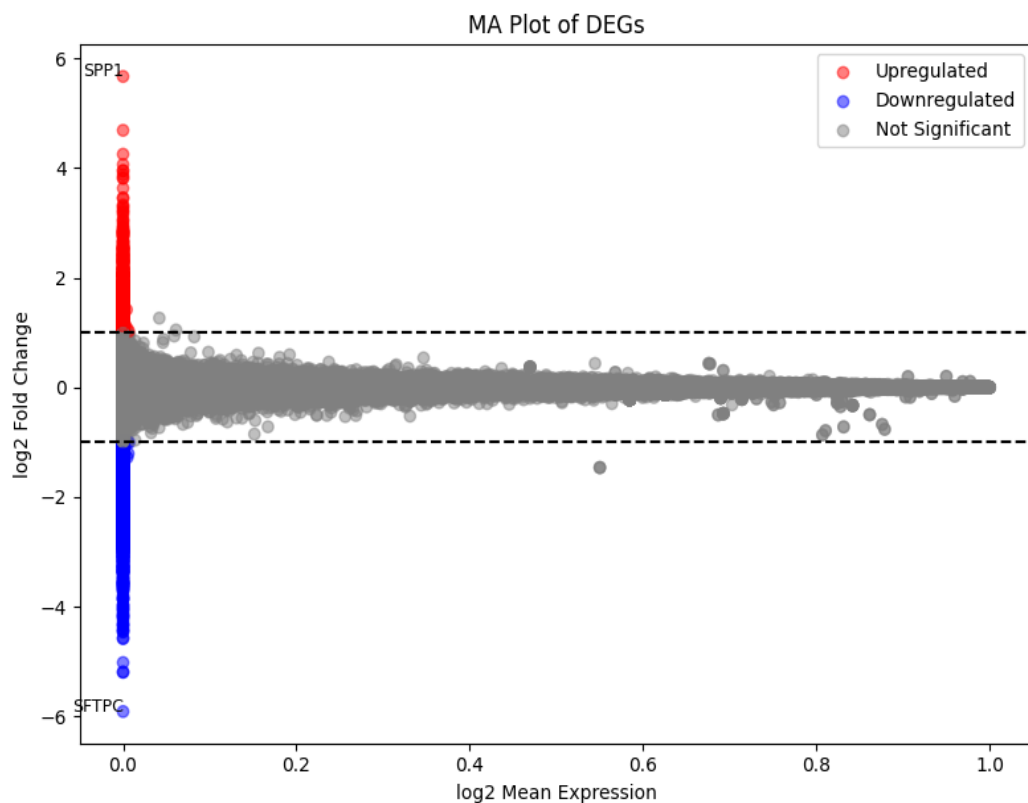


Figure 6. MA plot showing the distribution of log<sub>2</sub> fold changes relative to average expression levels. Upregulated and downregulated genes are highlighted in red and blue, respectively

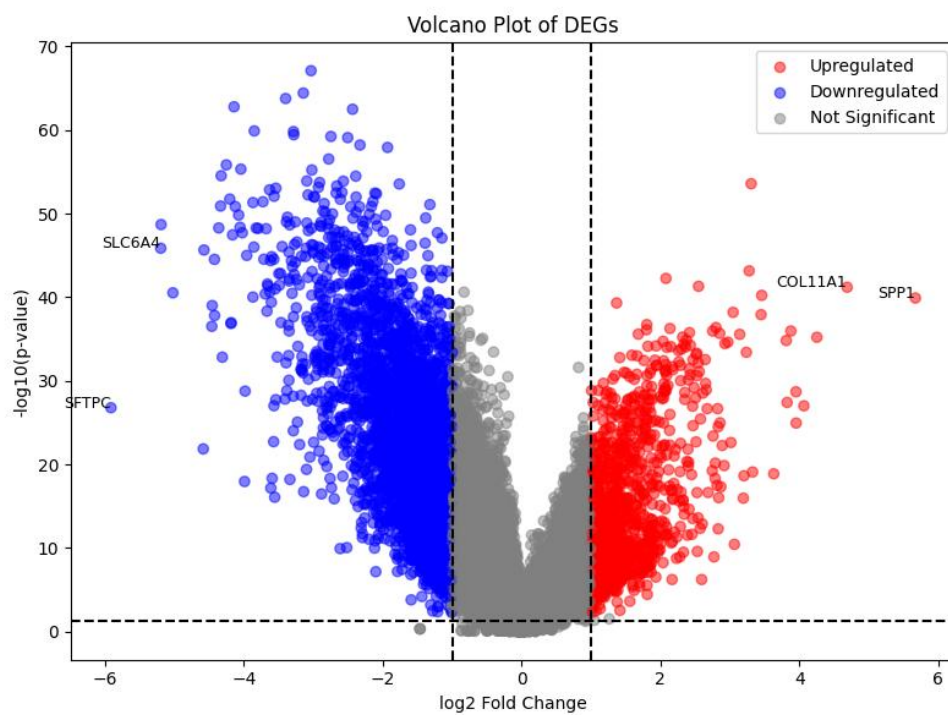


Figure 7. Volcano plot showing the distribution of log<sub>2</sub> fold changes relative to average expression levels. Upregulated and downregulated genes are highlighted in red and blue, respectively

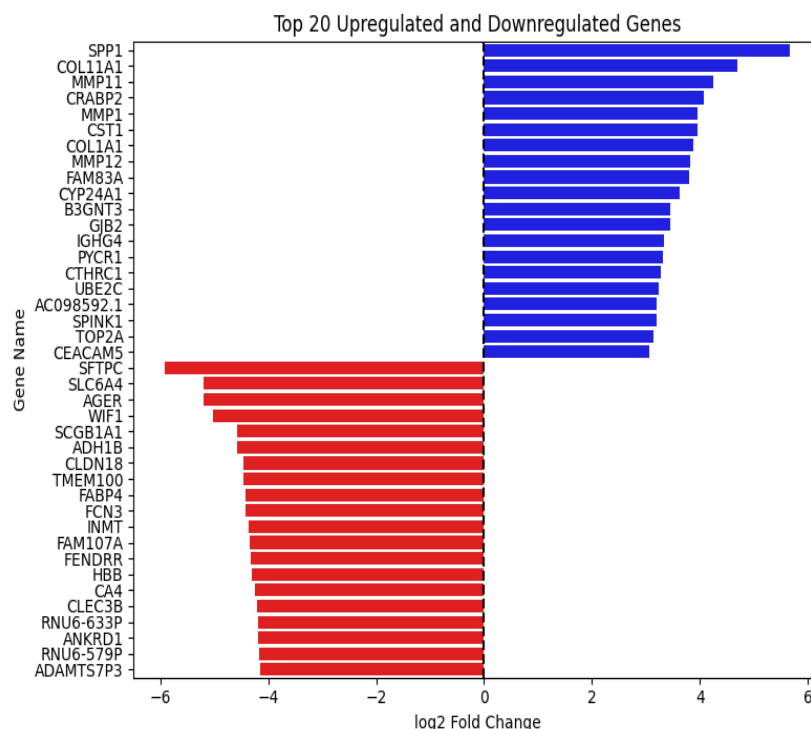


Figure 8. Bar graph of top 20 Upregulated and Downregulated genes

### 4.3 Enriched Pathways via Gene Set Enrichment Analysis (GSEA)

To gain biological insights into the transcriptional alterations in late-stage LUAD, Gene Set Enrichment Analysis (GSEA) was performed using the ranked gene list from differential expression results. The analysis focused on **Hallmark gene sets** from the MSigDB collection, providing a curated view of canonical biological processes and signaling pathways.

Out of the evaluated pathways, several were found to be significantly enriched (FDR  $q$ -value  $< 0.05$ ), with distinct patterns of **upregulation and downregulation**. Among the most strongly **upregulated** pathways were **E2F Targets**, **G2M Checkpoint**, **Glycolysis**, and **mTORC1 Signalling**. These pathways are commonly associated with **cell cycle progression**, **metabolic reprogramming**, and **proliferation**, suggesting a highly active

oncogenic environment in late-stage tumours. The pathway **E2F Targets**, for instance, showed the highest normalized enrichment score (NES = 7.08) and is visualized in the enrichment plot (Figure 9), indicating widespread upregulation of genes associated with DNA synthesis and cell cycle regulation.

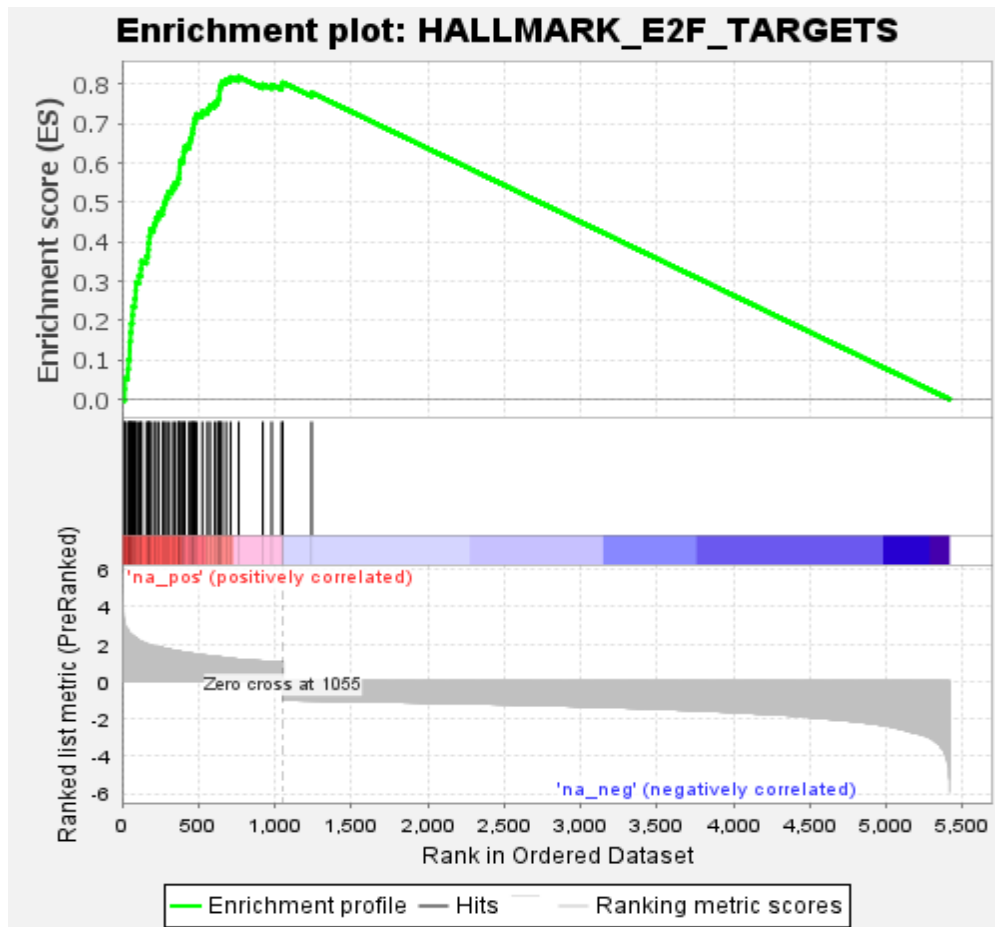


**Figure 9.** GSEA enrichment plot for the E2F Targets pathway, showing strong positive enrichment in tumor samples. This pathway is associated with cell cycle progression and is linked to therapeutic

Conversely, several pathways were significantly downregulated in tumor tissues. Notable examples include Adipogenesis, Fatty Acid Metabolism, Inflammatory Response, and UV Response Up. These negative enrichments suggest a loss of differentiation, immune suppression, and diminished tissue repair or stress responses in advanced LUAD. An example is the Adipogenesis pathway (Figure 10), which demonstrated significant negative enrichment (NES = -2.47), highlighting reduced expression of genes related to lipid

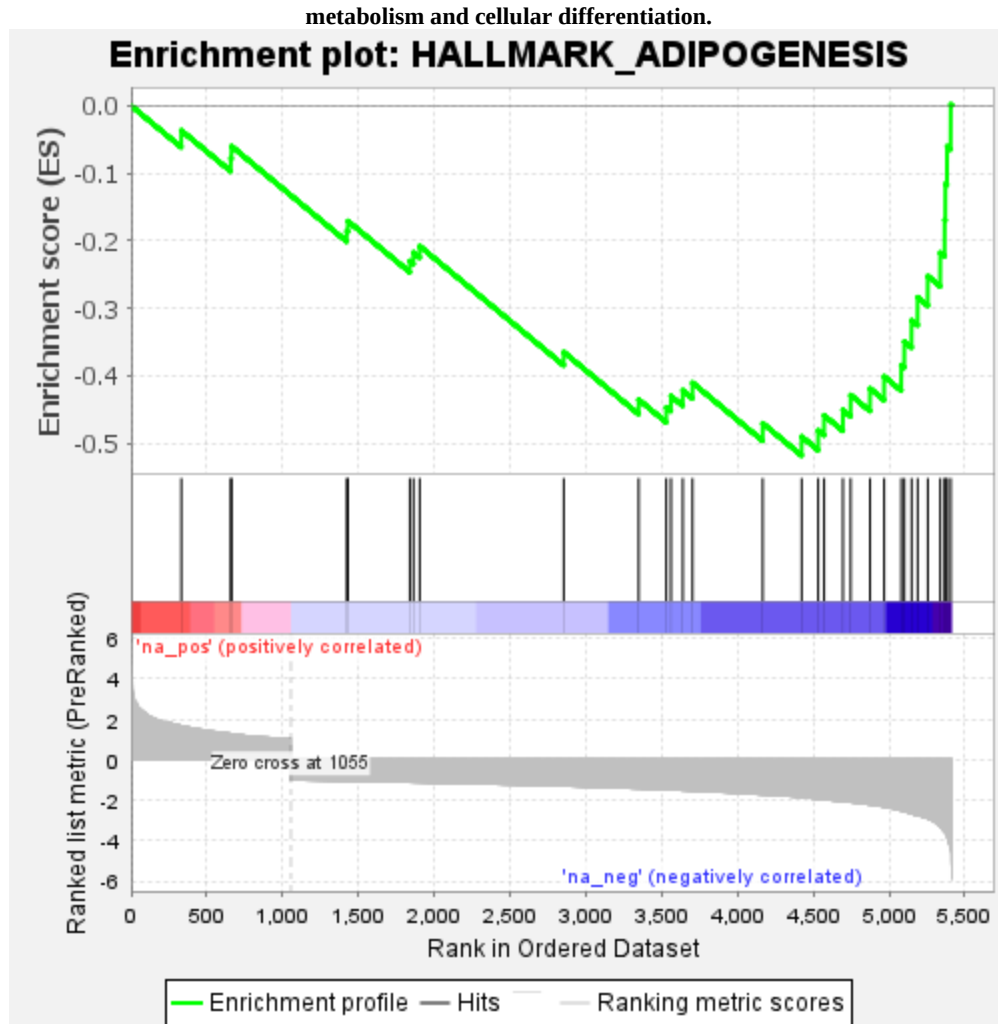


Figure 10. GSEA enrichment plot for the E2F Targets pathway, showing strong positive enrichment in tumor samples.

In total, **23 pathways** were negatively enriched, and **11 were positively enriched**, with statistical significance supported by both **normalized enrichment scores** and **false discovery rate (FDR) corrections** (FDR  $q < 0.05$  for top pathways). These enriched gene sets provide functional context to the differential gene expression patterns observed and offer potential targets for further validation and therapeutic exploration.

**Table 2. Showing mechanisms and pathways leading to resistance against Therapy**

| <b>Pathway Name</b>     | <b>NES</b> | <b>Regulation</b> | <b>Associated Function</b>                    | <b>Relevance to Resistance</b>                         | <b>Reference</b> |
|-------------------------|------------|-------------------|-----------------------------------------------|--------------------------------------------------------|------------------|
| <b>E2F Targets</b>      | 7.08       | Upregulated       | Cell cycle progression, DNA replication       | Enhances DNA repair and survival, promoting resistance | [47]             |
| <b>G2M Checkpoint</b>   | 6.71       | Upregulated       | Controls mitosis entry, DNA damage checkpoint | Allows division of damaged cells                       | [48]             |
| <b>Glycolysis</b>       | 4.87       | Upregulated       | Energy production, Warburg effect             | Supports tumor growth, undermines drug efficacy        | [49]             |
| <b>mTORC1 Signaling</b> | 4.12       | Upregulated       | Cell growth, nutrient sensing                 | Promotes survival and therapy resistance               | [50]             |
| <b>Adipogenesis</b>     | -<br>2.47  | Downregulated     | Lipid storage, metabolic differentiation      | Loss linked to dedifferentiation                       | [51]             |

|                                  |           |               |                                              |                                                       |      |
|----------------------------------|-----------|---------------|----------------------------------------------|-------------------------------------------------------|------|
|                                  |           |               |                                              | and therapy<br>resistance                             |      |
| <b>Inflammatory<br/>Response</b> | -<br>1.87 | Downregulated | Immune<br>signaling,<br>cytokine<br>response | Reduced<br>inflammation<br>impairs immune<br>response | [52] |

#### 4.4 Mutation Hotspot and Driver Gene Detection

To identify potential driver genes harboring non-random mutational clustering, OncodriveCLUSTL was applied separately to early-stage and late-stage LUAD patient cohorts using whole exome sequencing data. The analysis leverages the statistical enrichment of somatic mutations within specific regions of genes, suggesting functional disruption and selective advantage during tumor evolution.

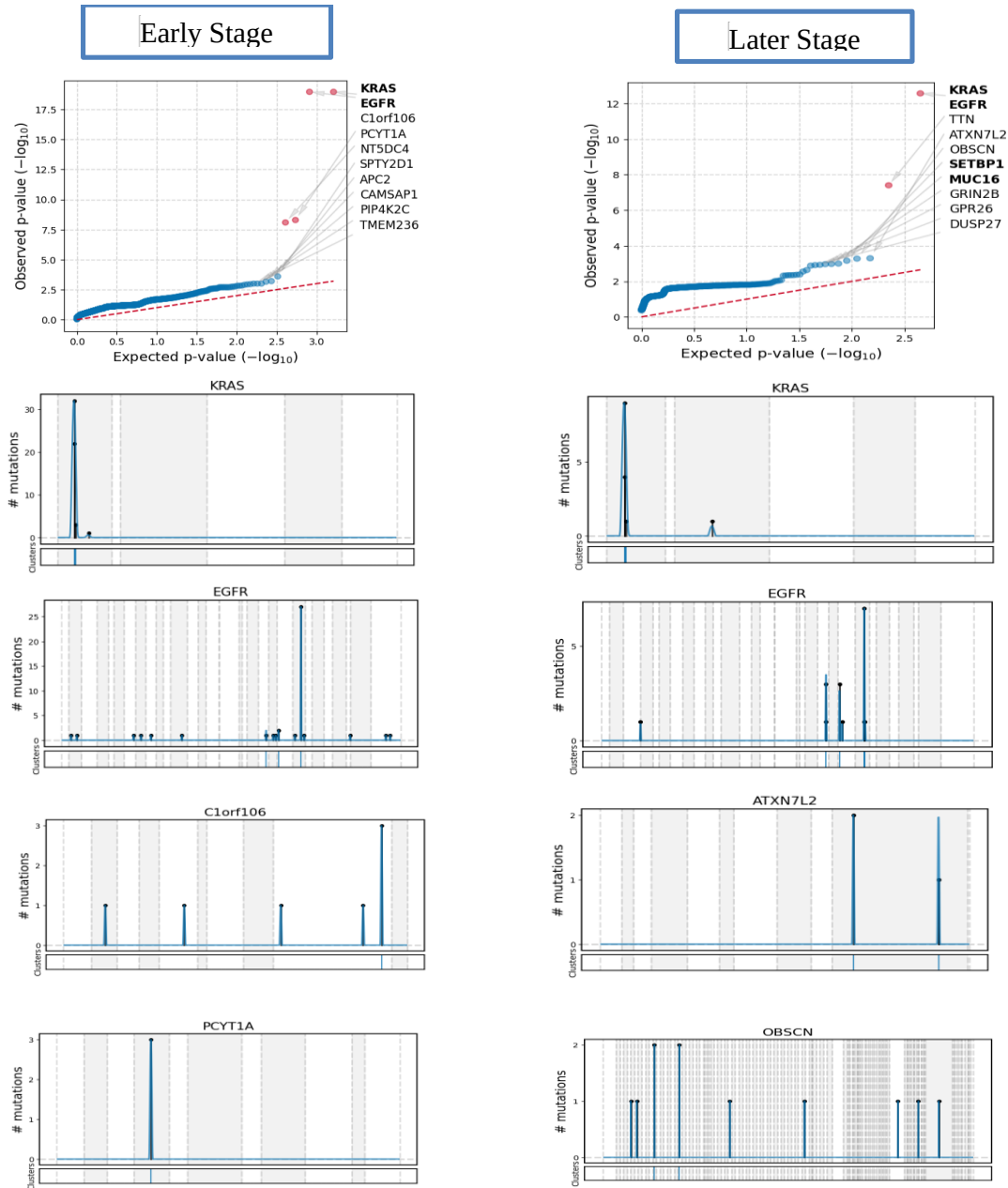


Figure 11. Detection of driver genes with mutation clustering in LUAD using OncodriveCLUSTL.

The upper panels show quantile-quantile (Q-Q) plots comparing observed versus expected  $-\log_{10}$  p-values for mutation clustering in early-stage (left) and late-stage (right) cohorts. Genes located above the red dotted line represent statistically significant clustering, suggesting potential oncogenic drivers. Highlighted genes such as KRAS and EGFR exhibit strong clustering in both stages. The lower panels display mutation distribution plots for selected genes. These plots show the spatial distribution of mutations along the protein sequence, with mutation counting on the y-axis and amino acid positions on the x-axis. Peaks represent clusters of mutations likely under positive selection. The clustering score is derived from both mutation density and compactness, capturing the tendency of driver mutations to occur within functional domain

In the early-stage group, several genes were identified with significant clustering patterns, most notably **KRAS**, **EGFR**, **C1orf106**, and **PCYT1A**, as shown in the quantile-quantile (Q–Q) plot (Figure 11, left). Both **KRAS** and **EGFR**, well-established oncogenes in LUAD, exhibited pronounced mutation aggregation. The mutation distribution plots confirm these clusters, with **KRAS** showing dense hotspot regions typically associated with G12 and G13 codons, and **EGFR** displaying multiple concentrated mutation peaks along its coding sequence.

In the late-stage cohort, **KRAS** and **EGFR** remained prominent but were accompanied by additional candidates, including **ATXN7L2**, **OBSCN**, **SETBP1**, and **MUC16**, several of which have emerging associations with cancer progression or therapeutic resistance. The corresponding Q–Q plot (Figure 12, right) illustrates stronger deviation from the null distribution, indicating a higher density of clustering in late-stage mutations. Visual inspection of gene-specific plots for **ATXN7L2** and **OBSCN** also revealed narrowly focused mutation sites, suggesting potential late-stage–specific oncogenic relevance.

Collectively, these findings reinforce the canonical driver status of **KRAS** and **EGFR** across disease stages, while highlighting novel or stage-specific candidate genes that may play a role in LUAD progression and warrant further functional validation

#### **4.5 Integration of DEGs and Mutation Hotspots Data**

To identify genes with both transcriptional dysregulation and mutational significance, an intersection analysis was performed between the differentially expressed genes (DEGs) and those carrying mutation hotspots detected by OncodriveCLUSTL. This resulted in a subset of **80 genes** shared between both datasets. To further assess the potential regulatory

impact of these mutations, we utilized the **Variant Effect Predictor (VEP)** by Ensembl. TRAP annotated mutations depending on their genomic location and expected functional consequences under categories including missense, splice region, start/stop codon modification, and 5'/3' UTR variants.

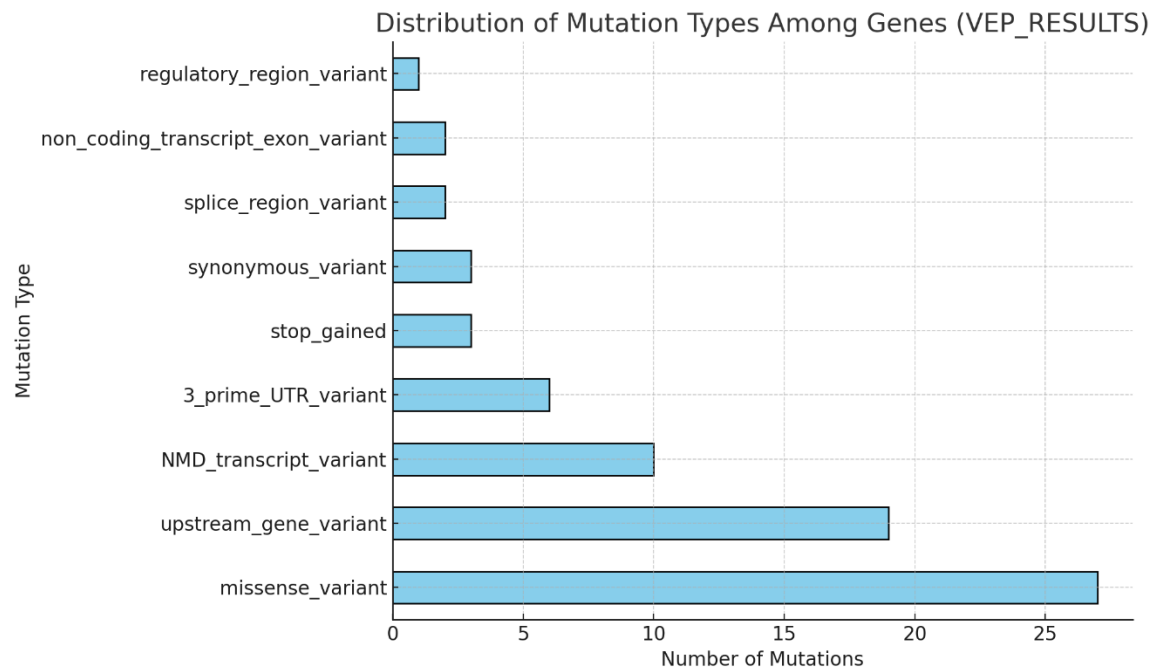


**Figure 12. Distribution of mutation types among the selected genes based on Variant Effect Predictor (VEP) analysis.**

**The bar graph shows the frequency of different mutation categories, including upstream gene variants, regulatory region variants, UTR variants, and others, reflecting the diversity of genomic alterations potentially impacting gene regulation**

Further study excluded genes including types of mutations less likely to influence gene expression—such as intronic variants or synonymous mutations. Following filtering, a refined list of 61 genes remained, each bearing mutations expected to affect expression either by regulatory disruption or protein-coding impact. Transcriptional control analysis and downstream integration with pathway-level data ranked these genes in order of their significance.

#### 4.6 Transcription Factor Binding Affinity Prediction (TRAP)

Using the TRAP (Transcription Factor Affinity Prediction) tool, we assessed the possible regulatory consequences of mutations on transcription factor (TF) binding. Using a biophysical model of TF-DNA interaction, TRAP computes binding scores that then predict how point mutations change the binding affinity of transcription factors to DNA. TRAP was applied to generate a reference for every mutation site found in the 61 shortlisted genes. **affinity**, **mutated affinity**, and the resulting change ( $\Delta$ **Affinity**). These scores reflect how strongly a transcription factor is predicted to bind to the wild-type versus mutated sequence.

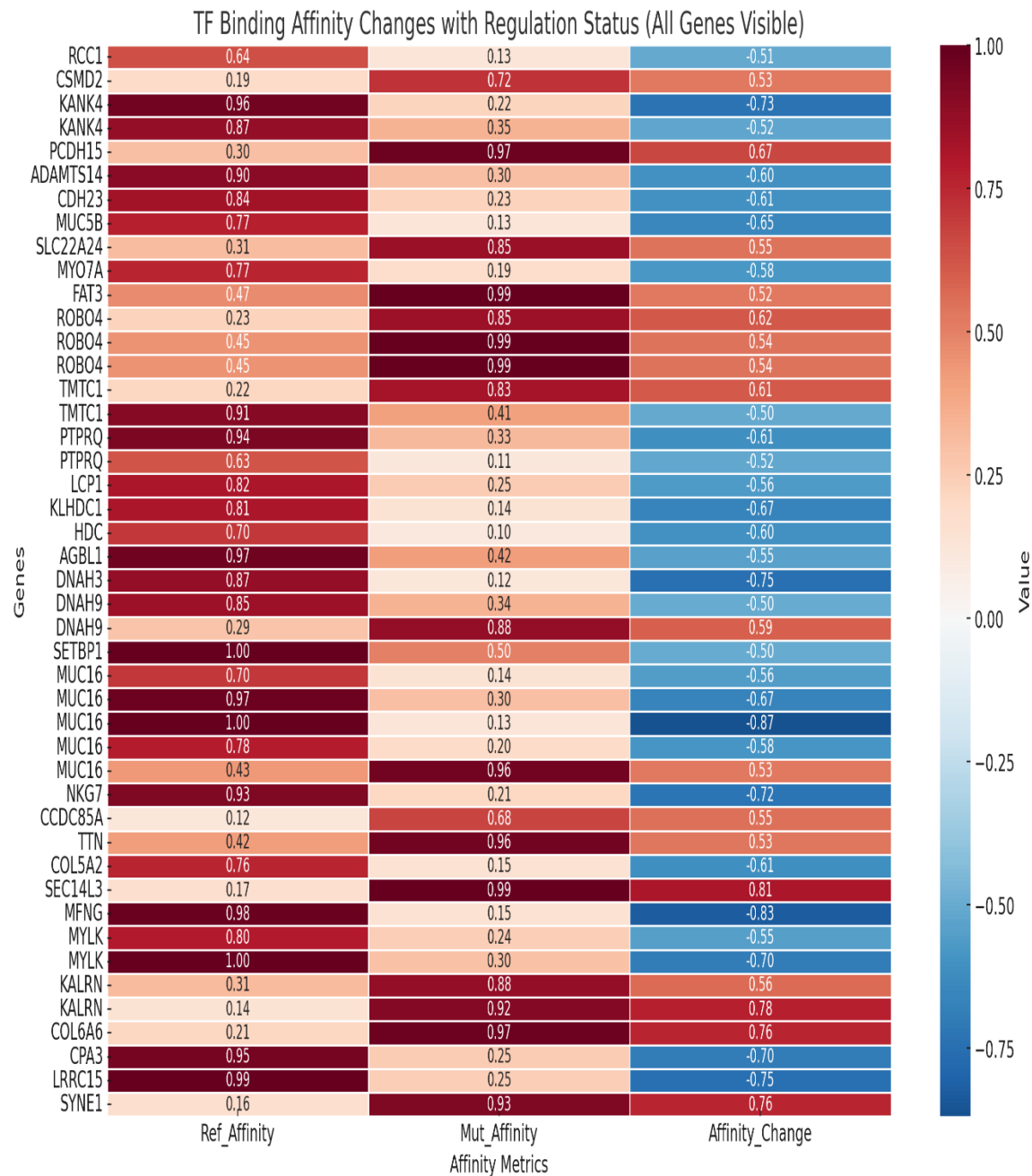


**Figure 13.** Heatmap displaying transcription factor binding affinity changes computed for selected genes.

Columns represent reference affinity (Ref\_Affinity), mutated sequence affinity (Mut\_Affinity), and the resulting change in binding potential (Affinity\_Change). Color gradients indicate the strength and direction of affinity shifts, with red denoting increased and blue indicating decreased binding in the mutated sequence. These changes suggest potential regulatory impacts of mutations on gene expression

The results, visualized in **Figure 14**, show a range of binding affinity changes, with several genes exhibiting substantial shifts. Genes such as **PCDH15**, **SEC14L3**, **ROBO4**, and **SPATA18** displayed either increased or decreased TF binding scores, suggesting that mutations in these regions could significantly affect gene regulation. Notably, some mutations led to a **gain in TF affinity** (e.g., *PCDH15*, *SPATA18*), while others caused a **reduction in affinity** (e.g., *KANK4*, *KLHDC1*), which may contribute to altered gene expression in LUAD.

By identifying mutations that likely disrupt or enhance TF-DNA binding, TRAP analysis added further evidence of the **regulatory potential** of these variants, reinforcing the candidacy of these genes in LUAD pathogenesis and their possible utility as biomarkers.

#### **4.7 Final Biomarker Shortlisting**

To identify high-confidence LUAD biomarkers, we applied a multi-layered filtering strategy combining transcriptomic, genomic, regulatory, and pathway-level evidence. Genes were shortlisted only if they met **all of the following criteria**: (1) significantly **differentially expressed**, (2) harbored **mutation hotspots** identified by OncodriveCLUSTL, (3) possessed mutations with predicted functional consequences according to **VEP** (e.g., missense, splice region, regulatory variants), (4) showed substantial **transcription factor binding affinity changes** as predicted by **TRAP**, and (5) were present within the **core enrichment region of significantly altered cancer-related pathways** based on **GSEA** analysis. Through this integrative pipeline, a refined list of **four genes** emerged as high-confidence **biomarker candidates**:

- **LCP1**, enriched in the *KRAS\_SIGNALING\_UP* pathway,

- **COL5A2**, associated with *ANGIOGENESIS*,
- **KALRN**, implicated in *UV\_RESPONSE\_DN*, and
- **LRRC15**, involved in *EPITHELIAL\_MESENCHYMAL\_TRANSITION*.

These genes demonstrated consistent evidence of functional impact across all analytical layers and are biologically linked to core cancer hallmarks, reinforcing their potential role in LUAD progression and their value as putative therapeutic targets or diagnostic markers.

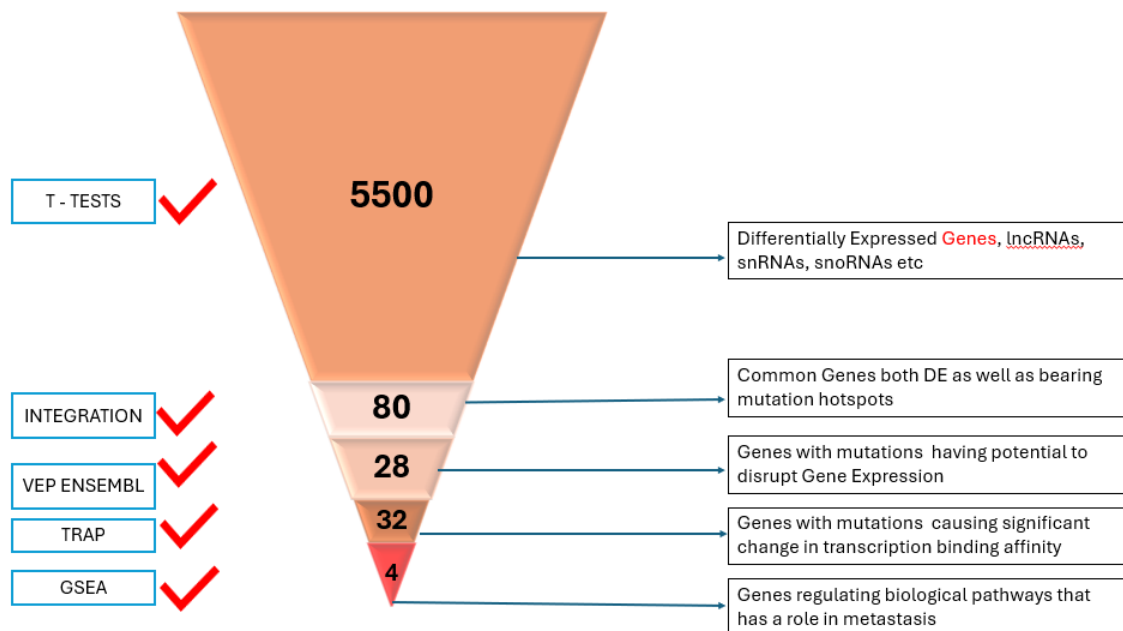


**Figure 14. Funnel-shaped workflow summarizing the stepwise gene filtering and prioritization pipeline used to identify LUAD biomarkers.**

Starting from 5,500 differentially expressed transcripts (including mRNAs, lncRNAs, snRNAs, snoRNAs), successive filters were applied using mutation clustering, functional variant analysis (VEP), transcription factor binding impact (TRAP), and pathway enrichment (GSEA). The process culminated in 4 final biomarker candidates with potential regulatory roles in metastasis

## CHAPTER 5

### DISCUSSION

The integrative multi-omics analysis conducted in this study provides a comprehensive understanding of the genomic and transcriptomic alterations underlying lung adenocarcinoma (LUAD) progression. By leveraging both whole exome sequencing (WES) and RNA sequencing (RNA-seq) datasets from The Cancer Genome Atlas (TCGA), distinct molecular patterns were uncovered between early-stage and late-stage LUAD, highlighting novel biomarker candidates and critical pathways implicated in tumorigenesis and therapeutic resistance.

One of the most striking findings of this study was the absence of significant differential gene expression in early-stage LUAD, contrasted by the extensive transcriptomic dysregulation observed in late-stage disease. This shift emphasizes the relative molecular quiescence of early-stage tumors and implies that until extra layers of gene control are disturbed, genomic changes alone might not be enough to produce aggressive phenotypes [53]. According to earlier research, late-stage LUAD is associated with increased proliferative capacity and immune evasion mechanisms. The late-stage cohort showed a marked downregulation of important homeostatic and immune pathways along with an upregulation of oncogenic processes like E2F targets and G2M checkpoint regulation [54].

The crucial role of known drivers like KRAS and EGFR across disease stages was further supported by the use of OncodriveCLUSTL to identify important mutation hotspots.

Notably, the late-stage group identified more promising candidates, including OBSCN and MUC16, whose clustered mutation patterns point to possible roles in the development of the disease and resistance to treatment [55]. These observations align with prior findings that implicate late-stage LUAD as genomically more unstable and heterogeneous compared to early stages.

An important advancement achieved in this study was the integration of mutation hotspot data with differential expression profiles. By focusing on genes that were both mutated in clustered patterns and differentially expressed, a more functionally relevant subset of candidates was prioritized. This approach addresses a critical gap in previous LUAD studies, where genomic and transcriptomic alterations were often considered in isolation, limiting mechanistic insights [56]. Variant effect prediction using Ensembl VEP and transcription factor binding affinity analysis via TRAP further refined the gene list by evaluating functional consequences at both the coding and regulatory levels.

The final shortlist of biomarkers, LCP1, COL5A2, KALRN, and LRRC15 (emerged as a novel therapeutic target), demonstrates how multi-layered filtering can converge on genes with plausible roles in LUAD pathogenesis. These genes are associated with hallmark processes such as KRAS signaling, angiogenesis, epithelial-mesenchymal transition (EMT), and UV response modulation, all of which are central to cancer metastasis and resistance mechanisms [57-60]. Notably, LCP1, enriched in the KRAS\_SIGNALING\_UP pathway, has been previously implicated in promoting cytoskeletal reorganization and metastasis in lung cancer models [61]. Similarly, COL5A2 and LRRC15 have been linked

to extracellular matrix remodeling and EMT, providing biological plausibility to their candidacy as metastasis-related biomarkers.

Moreover, the findings on pathway enrichment resonate with established knowledge about LUAD progression. Upregulated pathways such as glycolysis and mTORC1 signaling reflect a metabolic shift toward aerobic glycolysis (the Warburg effect) and growth-promoting signals, both of which have been extensively reported in aggressive lung tumors [62]. Conversely, the downregulation of pathways like adipogenesis and inflammatory responses highlights a suppression of normal differentiation and immune signaling processes, likely facilitating tumor escape from immunosurveillance [63].

Nevertheless, several limitations must be acknowledged. While the TCGA cohort offers a rich and diverse dataset, it represents primarily early to moderately advanced stages, with fewer very late-stage (e.g., extensive metastasis) cases. This limitation could bias pathway findings towards moderately progressive rather than terminal disease characteristics. Furthermore, while the computational predictions regarding transcription factor affinity shifts provide valuable insights, functional validation in wet-lab experiments remains essential to conclusively establish the regulatory consequences of these mutations.

In conclusion, this study underscores the power of integrative multi-omics approaches in revealing complex molecular landscapes of cancer. We were able to pinpoint important genomic biomarkers and pathways that are probably responsible for the late-stage progression of LUAD by connecting somatic mutational profiles with transcriptomic dysregulation. Proteomics and single-cell omics technologies should be used in future

research to further analyze tumor heterogeneity and validate the discovered biomarkers experimentally to hasten their conversion into clinical uses.

## CHAPTER 6

### CONCLUSION AND FUTURE DIRECTIONS

The molecular foundations of lung adenocarcinoma (LUAD) progression were investigated in this work using an integrated multi-omics framework. We were able to identify the different genomic and transcriptomic landscapes linked with early and late stages of the disease by aggregating whole exome sequencing data with RNA-seq datasets. Strong evidence indicates from the results that tumor evolution in LUAD is characterized by a change from minimal transcriptomic disturbance in early stages to widespread dysregulation of important biological pathways in advanced disease.

Especially, the multi-omics approach helped to identify a small subset of highly confident biomarker candidates — LCP1, COL5A2, KALRN, and LRRC15 — that show both notable mutation clustering and expression dysregulation. Closely related to several important cancer hallmarks including epithelial-mesenchymal transition, angiogenesis, KRAS signaling, and stress response were these genes. Their possible involvement as drivers of LUAD progression emphasizes the need of combining several molecular layers to find significant targets for therapy and prognosis.

Gene Set Enrichment Analysis (GSEA) pathway-level data revealed consistent enrichment of oncogenic programs including cell cycle control, metabolic reprogramming, and immune suppression in late-stage tumors. These results reflect more general observations in the LUAD research area, whereby late-stage tumors show enhanced proliferative and survival capacity at the expense of immune system interaction. Particularly the inhibition

of adipogenesis and inflammatory response pathways emphasizes how tumors reinterpret their microenvironment to enable immune evasion and metastases.

While the study offers valuable contributions to the field, it also highlights ongoing challenges. Though their depth, the reliance on TCGA datasets limits generalizability across many ethnic backgrounds and rare disease subtypes. Furthermore, computational estimates of mutation effects on transcription factor binding demand experimental validation to verify their biological significance. Translating omics-driven results into clinical practice depends on addressing these constraints in next studies.

Finally, this thesis shows how well integrated omics studies offer a potent approach to expose the intricate molecular architecture of LUAD. We have not only enlarged the list of possible biomarkers but also advanced knowledge of disease progression mechanisms by connecting somatic mutations to functional gene expression changes. Future directions should include functional validation of candidate biomarkers, integration of other omics layers such proteomics and epigenomics, and investigation of their relevance and their utility in precision oncology settings. To conclude, the main findings from this study lay a strong foundation for advancing biomarker discovery and therapeutic innovation in LUAD.

The insights gained from this integrative multi-omics analysis pave the way for several promising future research directions in the study of lung adenocarcinoma (LUAD). The identification of stage-specific molecular signatures and functionally relevant biomarkers like LCP1, COL5A2, KALRN, and LRRC15 suggests strong candidates for therapeutic targeting and prognostic modeling. Functional validation of these biomarkers *in vitro* and *in vivo* will be a critical next step to confirm their role in LUAD progression and treatment

resistance. Moreover, their integration into predictive models could enhance early diagnosis and stratify patients based on risk or likely response to therapies. Extending this research into prospective clinical trials would help evaluate their utility in real-world precision oncology applications.

Beyond biomarker validation, future work should broaden the molecular scope by incorporating additional omics layers such as proteomics, single-cell RNA sequencing, and epigenomics. These techniques could uncover further complexities of tumor heterogeneity, microenvironmental interactions, and regulatory dynamics that are not captured by bulk genomic and transcriptomic analyses alone. Importantly, such expansions will aid in resolving the limitations associated with TCGA's relatively homogenous patient population and underrepresentation of terminal-stage tumors. As multi-omics technologies become more accessible, their integration will be essential for constructing robust, clinically actionable molecular maps of LUAD that inform targeted therapy development, resistance prediction, and personalized treatment strategies.

## References

1. Chen, Z., et al., *Non-small-cell lung cancers: a heterogeneous set of diseases*. Nature Reviews Cancer, 2014. **14**(8): p. 535-546.
2. Testa, U., G. Castelli, and E. Pelosi, *Lung cancers: molecular characterization, clonal heterogeneity and evolution, and cancer stem cells*. Cancers, 2018. **10**(8): p. 248.
3. Detterbeck, F.C., D.J. Boffa, and L.T. Tanoue, *The new lung cancer staging system*. Chest, 2009. **136**(1): p. 260-271.
4. Bray, F., et al., *Erratum: Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries*. Ca Cancer J Clin, 2020. **70**(4): p. 313.
5. Foiss, S.S., et al., *Molecular epidemiology of the main druggable genetic alterations in non-small cell lung cancer*. International journal of molecular sciences, 2021. **22**(2): p. 612.
6. Yuan, M., et al., *The emerging treatment landscape of targeted therapy in non-small-cell lung cancer*. Signal transduction, 2019. **4**(1): p. 61.
7. Kumar, M. and A. Sarkar, *Current therapeutic strategies and challenges in NSCLC treatment: A comprehensive review*. Experimental oncology, 2022. **44**(1): p. 7-16.
8. Network, C.G.A.R., *Comprehensive molecular profiling of lung adenocarcinoma*. Nature Reviews Cancer, 2014. **511**(7511): p. 543.

9. Okayama, H., et al., *Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas*. Cancer research, 2012. **72**(1): p. 100-111.
10. Lv, Z. and T. Lei, *Systematical identifications of prognostic meaningful lung adenocarcinoma subtypes and the underlying mutational and expressional characters*. BMC cancer, 2020. **20**: p. 1-12.
11. Zhao, Y., et al., *Multi-omics analysis of genomics, epigenomics and transcriptomics for molecular subtypes and core genes for lung adenocarcinoma*. BMC cancer, 2021. **21**: p. 1-18.
12. Ruan, X., et al., *Multi-omics integrative analysis of lung adenocarcinoma: an in silico profiling for precise medicine*. Frontiers in Medicine, 2022. **9**: p. 894338.
13. Wang, W., H. Liu, and G. Li, *What's the difference between lung adenocarcinoma and lung squamous cell carcinoma? Evidence from a retrospective analysis in a cohort of Chinese patients*. Front Endocrinol (Lausanne), 2022. **13**: p. 947443.
14. He, D., et al., *Single-cell RNA sequencing reveals heterogeneous tumor and immune cell populations in early-stage lung adenocarcinomas harboring EGFR mutations*. Oncogene, 2021. **40**(2): p. 355-368.
15. Hutchinson, B.D., et al., *Spectrum of Lung Adenocarcinoma*. Semin Ultrasound CT MR, 2019. **40**(3): p. 255-264.
16. Zhao, Z., et al., *Specific organ metastases and prognosis in lung adenocarcinoma*. Thorac Cancer, 2023. **14**(8): p. 736-745.

17. Zhang, S., et al., *Comparison of gene mutation profile in different lung adenocarcinoma subtypes by targeted next-generation sequencing*. Med Oncol, 2023. **40**(12): p. 349.
18. Barta, J.A., C.A. Powell, and J.P.J.A.o.g.h. Wisnivesky, *Global epidemiology of lung cancer*. 2019. **85**(1).
19. Siegel, D.A., et al., *Proportion of never smokers among men and women with lung cancer in 7 US states*. 2021. **7**(2): p. 302-304.
20. Staaf, J., et al., *Molecular characteristics of lung adenocarcinoma with respect to patient age at diagnosis*. 2023. **153**(1): p. 197-209.
21. Chien, L.-H., et al., *Incidence of lung adenocarcinoma by age, sex, and smoking status in Taiwan*. 2023. **6**(11): p. e2340704-e2340704.
22. Zhang, Y., et al., *Identification of cigarette smoking-related novel biomarkers in lung adenocarcinoma*. 2022. **2022**.
23. Corrales, L., et al., *Lung cancer in never smokers: The role of different risk factors other than tobacco smoking*. 2020. **148**: p. 102895.
24. Zhang, S., *Adenocarcinoma*, in *Diagnostic Imaging of Lung Cancers*. 2024, Springer. p. 3-49.
25. Chen, Y.J., et al., *Proteogenomics of Non-smoking Lung Cancer in East Asia Delineates Molecular Signatures of Pathogenesis and Progression*. Cell, 2020. **182**(1): p. 226-244.e17.
26. Abe, J., et al., *Novel activating KRAS mutation candidates in lung adenocarcinoma*. Biochem Biophys Res Commun, 2020. **522**(3): p. 690-696.

27. Liu, Z., et al., *A novel intergenic (between REG3A and CTNNA2-AS1)-ALK fusion responds to alectinib in lung adenocarcinoma*. Lung Cancer, 2023. **186**: p. 107386.
28. Saito, M., et al., *Gene aberrations for precision medicine against lung adenocarcinoma*. Cancer Sci, 2016. **107**(6): p. 713-20.
29. Lu, M., et al., *Identification of significant genes as prognostic markers and potential tumor suppressors in lung adenocarcinoma via bioinformatical analysis*. BMC Cancer, 2021. **21**(1): p. 616.
30. Ye, F., et al., *TGF $\beta$  Antagonizes IFN $\gamma$ -Mediated Adaptive Immune Evasion via Activation of the AKT–Smad3–SHP1 Axis in Lung Adenocarcinoma*. 2023. **83**(13): p. 2262-2277.
31. Budczies, J., et al., *Deciphering the immunosuppressive tumor microenvironment in ALK-and EGFR-positive lung adenocarcinoma*. 2022. **71**(2): p. 251-265.
32. Spella, M. and G.T.J.C. Stathopoulos, *Immune resistance in lung adenocarcinoma*. 2021. **13**(3): p. 384.
33. Wu, H.-J., et al., *Spatial intra-tumor heterogeneity is associated with survival of lung adenocarcinoma patients*. 2022. **2**(8).
34. Zhang, C., et al., *Clinical significance and inflammatory landscapes of a novel recurrence-associated immune signature in early-stage lung adenocarcinoma*. 2020. **479**: p. 31-41.
35. Liu, W.-j., et al., *Drug resistance to targeted therapeutic strategies in non-small cell lung cancer*. 2020. **206**: p. 107438.

36. Bagrodia, S., T. Smeal, and R.T. Abraham, *Mechanisms of intrinsic and acquired resistance to kinase-targeted therapies*. Pigment cell & melanoma research, 2012. **25**(6): p. 819-831.
37. Lim, Z.-F. and P.C. Ma, *Emerging insights of tumor heterogeneity and drug resistance mechanisms in lung cancer targeted therapy*. Journal of hematology & oncology, 2019. **12**(1): p. 134.
38. Kemper, K., et al., *Phenotype switching: tumor cell plasticity as a resistance mechanism and target for therapy*. Cancer research, 2014. **74**(21): p. 5937-5941.
39. Das, T., et al., *Integration of online omics-data resources for cancer research*. 2020. **11**: p. 578345.
40. Wang, T.-H., et al., *Integrated Omics Analysis of Non-Small-Cell Lung Cancer Cells Harboring the EGFR C797S Mutation Reveals the Potential of AXL as a Novel Therapeutic Target in TKI-Resistant Lung Cancer*. 2020. **13**(1): p. 111.
41. Rami-Porta, R., J.J. Crowley, and P. Goldstraw, *Review the revised TNM staging system for lung cancer*. Ann Thorac Cardiovasc Surg, 2009. **15**(1): p. 5.
42. Bandyopadhyay, S., S. Mallik, and A. Mukhopadhyay, *A survey and comparative study of statistical tests for identifying differential expression from microarray data*. IEEE/ACM transactions on computational biology and bioinformatics, 2013. **11**(1): p. 95-115.
43. Subramanian, A., et al., *GSEA-P: a desktop application for Gene Set Enrichment Analysis*. Bioinformatics, 2007. **23**(23): p. 3251-3253.
44. Arnedo-Pac, C., et al., *OncodriveCLUSTL: a sequence-based clustering method to identify cancer drivers*. Bioinformatics, 2019. **35**(22): p. 4788-4790.

45. Hunt, S.E., et al., *Annotating and prioritizing genomic variants using the Ensembl Variant Effect Predictor—A tutorial*. Human mutation, 2022. **43**(8): p. 986-997.
46. Thomas-Chollier, M., et al., *Transcription factor binding predictions using TRAP for the analysis of ChIP-seq data and regulatory SNPs*. Nature protocols, 2011. **6**(12): p. 1860-1869.
47. Wei, J., et al., *The E2F family: a ray of dawn in cardiomyopathy*. Molecular and Cellular Biochemistry, 2024: p. 1-15.
48. Zhuang, Z., et al., *Enhancement of cancer chemotherapy by simultaneously altering cell cycle progression and DNA-damage defenses through global modification of the serine/threonine phospho-proteome*. Cell cycle, 2009. **8**(20): p. 3303-3306.
49. Luengo, A., et al., *Increased demand for NAD<sup>+</sup> relative to ATP drives aerobic glycolysis*. Molecular cell, 2021. **81**(4): p. 691-707.
50. Orozco, J.M., et al., *Dihydroxyacetone phosphate signals glucose availability to mTORC1*. Nature metabolism, 2020. **2**(9): p. 893-901.
51. Tencerova, M., et al., *Metabolic programming determines the lineage-differentiation fate of murine bone marrow stromal progenitor cells*. Bone Research, 2019. **7**(1): p. 35.
52. Mantovani, A., et al., *Cancer-related inflammation*. nature, 2008. **454**(7203): p. 436-444.
53. Sharma, A., et al., *Non-genetic intra-tumor heterogeneity is a major predictor of phenotypic heterogeneity and ongoing evolutionary dynamics in lung tumors*. Cell reports, 2019. **29**(8): p. 2164-2174.

54. Seguin, L., M. Durandy, and C.C. Feral, *Lung adenocarcinoma tumor origin: a guide for personalized medicine*. Cancers, 2022. **14**(7): p. 1759.
55. Zhang, L., X. Han, and Y. Shi, *Association of MUC16 mutation with response to immune checkpoint inhibitors in solid tumors*. JAMA network open, 2020. **3**(8): p. e2013201-e2013201.
56. Dasgupta, S., *Next-generation cancer phenomics: A transformative approach to unraveling lung cancer complexity and advancing precision medicine*. OMICS: A Journal of Integrative Biology, 2024. **28**(12): p. 585-595.
57. Pan, S., et al., *LCPI correlates with immune infiltration: a prognostic marker for triple-negative breast cancer*. BMC immunology, 2024. **25**(1): p. 42.
58. Cai, S., et al., *Integrative analysis and experiments to explore angiogenesis regulators correlated with poor prognosis, immune infiltration and cancer progression in lung adenocarcinoma*. Journal of Translational Medicine, 2021. **19**(1): p. 361.
59. Li, M., et al., *KALRN mutations promote antitumor immunity and immunotherapy response in cancer*. Journal for Immunotherapy of Cancer, 2020. **8**(2): p. e000293.
60. Ray, U., et al., *Exploiting LRRC15 as a novel therapeutic target in cancer*. Cancer research, 2022. **82**(9): p. 1675-1681.
61. Chen, L., et al., *Predicting immunotherapy-related adverse events in late-stage non-small cell lung cancer with KARS G12C mutation treated with PD-1 inhibitors through combined assessment of LCPI and ADPGK expression levels*. American Journal of Cancer Research, 2024. **14**(10): p. 4803.

62. Makinoshima, H., et al., *Signaling through the phosphatidylinositol 3-kinase (PI3K)/mammalian target of rapamycin (mTOR) axis is responsible for aerobic glycolysis mediated by glucose transporter in epidermal growth factor receptor (EGFR)-mutated lung adenocarcinoma*. Journal of Biological Chemistry, 2015. **290**(28): p. 17495-17504.
63. Zhao, H., et al., *Inflammation and tumor progression: signaling pathways and targeted intervention*. Signal transduction and targeted therapy, 2021. **6**(1): p. 26

## Vitae

|                     |                                                                                                                                                                                                                                                                                                                                                                                                                         |
|---------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Name                | Hasan Shuja                                                                                                                                                                                                                                                                                                                                                                                                             |
| Nationality         | Pakistani                                                                                                                                                                                                                                                                                                                                                                                                               |
| Date of Birth       | 11 <sup>th</sup> November 1996                                                                                                                                                                                                                                                                                                                                                                                          |
| Email               | hasan.shuja@outlook.com                                                                                                                                                                                                                                                                                                                                                                                                 |
| Address             | House no 834, Block D, Satellite Town Rawalpindi, 46300,<br>Punjab, Pakistan                                                                                                                                                                                                                                                                                                                                            |
| Academic Background | Bachelor of Biotechnology from International Islamic<br>University Islamabad, Pakistan<br><br>Masters of Life Sciences from King Fahd University of<br>Petroleum and Minerals Dhahran, Saudia Arabia.                                                                                                                                                                                                                   |
| Publications        | Shuja, H., Rehman, O., Ahmad, I., <i>Identification of Genomic<br/>Biomarkers in Non-Small Cell Carcinoma: Smoking Factors<br/>and Associated Biological Pathways</i> . Gene (Under Review)<br><a href="http://dx.doi.org/10.2139/ssrn.4940708">http://dx.doi.org/10.2139/ssrn.4940708</a><br><br>Ikram, M., A., Iqbal, W., Shuja, H., 2021. <i>Review on<br/>pharmacokinetics of vinblastine drug and its delivery</i> |

system. World Journal of Pharmaceutical and Life Science,  
Vol. 7, Issue 8, 28 – 35

#### Conferences

The 65th Annual Meeting of the Japan Lung Cancer Society  
Presented as an oral speaker Research Paper “*Non-Small  
Cell Carcinoma Genomic biomarkers: Smoking Factors and  
Associated Biological Pathways*”

National Cancer Center Hospital, Yokohama Japan,  
[31/10/2024 – 01/11/2024]