



Article

Integrated Bioinformatics and RT-qPCR Validation Identify Candidate Biomarkers Associated with EGFR-TKI Resistance in *EGFR*-Mutant Lung Adenocarcinoma

Muhammad Alfin Hanif ^{1,2}, Fadilah ^{1,3,4,*}, Noorwati Sutandyo ⁵, Rafika Indah Paramita ^{1,3,4},
Ajeng Megawati Fajrin ³, and Septelia Inawati Wanandi ^{1,6}

- ¹ Doctoral Programme in Biomedical Sciences, Faculty of Medicine, Universitas Indonesia, Jakarta 10430, Indonesia; malfinhanif2@gmail.com (M.A.H.); rafikaindah@ui.ac.id (R.I.P.); septelia.inawati@ui.ac.id (S.I.W.)
- ² Department of Thoracic, Dharmas Cancer Hospital, National Cancer Center, Jakarta 11420, Indonesia
- ³ Department of Medical Chemistry, Faculty of Medicine, Universitas Indonesia, Jakarta 10430, Indonesia; ajengmegawatif@gmail.com
- ⁴ Bioinformatics Core Facilities—IMERI, Faculty of Medicine, Universitas Indonesia, Jakarta 10430, Indonesia
- ⁵ Department of Medical-Haematology, Dharmas Cancer Hospital, National Cancer Center, Jakarta 11420, Indonesia; noorwatis3@yahoo.com
- ⁶ Department of Biochemistry and Molecular Biology, Faculty of Medicine, Universitas Indonesia, Jakarta 10430, Indonesia
- * Correspondence: fadilah.msi@ui.ac.id

Abstract

Background: Acquired resistance to epidermal growth factor receptor tyrosine kinase inhibitors (EGFR-TKIs) remains a major challenge in the treatment of EGFR-mutant lung adenocarcinoma. Identifying biomarkers associated with resistance may improve understanding of the underlying molecular mechanisms. **Objective:** This study aimed to identify candidate biomarkers associated with EGFR-TKI resistance using integrated bioinformatics analysis and RT-qPCR validation. **Methods:** Two Gene Expression Omnibus (GEO) datasets (GSE123066 and GSE178755) were analysed to identify differentially expressed genes (DEGs). Overlapping DEGs were subjected to protein–protein interaction (PPI) network construction, Gene Ontology (GO), KEGG pathway enrichment, and Kaplan–Meier survival analyses. Selected candidate genes were subsequently validated by RT-qPCR in an independent clinical cohort. **Results:** A total of 76 overlapping DEGs were identified, including 52 upregulated genes enriched in cell proliferation, hypoxia response, EGFR signalling, and metabolic processes. KEGG analysis identified the PI3K–Akt signalling pathway as significantly enriched. Kaplan–Meier analysis showed that higher expression of TGFA, DDIT4, and SOX9 was associated with shorter progression-free survival. RT-qPCR validation demonstrated significantly increased SOX9 expression in the resistant group, whereas TGFA and DDIT4 showed no statistically significant differences. **Conclusions:** Integrated bioinformatics analysis identified TGFA, DDIT4, and SOX9 as candidate genes associated with EGFR-TKI resistance in EGFR-mutant lung adenocarcinoma. These findings provide a basis for further investigation of resistance-associated biomarkers, although larger independent cohorts and functional studies are required before clinical application.

Keywords: EGFR-TKI resistance; lung adenocarcinoma; PI3K–Akt pathway; multigene biomarker; overall survival

Academic Editor: Hans Binder

Received: 6 May 2026

Revised: 16 July 2026

Accepted: 23 July 2026

Published: 28 July 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Lung cancer is the second most common cancer in Indonesia, following breast cancer, yet it remains the leading cause of cancer-related mortality nationwide [1]. The five-year survival rate is markedly low at approximately 22%, underscoring its poor prognosis [2]. Non-small cell lung cancer (NSCLC) accounts for 80–85% of cases, with adenocarcinoma representing the predominant histological subtype [3]. In Indonesia, adenocarcinoma accounts for more than 70% of NSCLC cases, emphasizing its major clinical burden [4].

The discovery of activating mutations in the epidermal growth factor receptor (EGFR) has transformed the therapeutic landscape of advanced lung adenocarcinoma. Since the early 2000s, *EGFR* mutations have become a key molecular target due to their association with improved clinical outcomes when treated with EGFR tyrosine kinase inhibitors (EGFR-TKIs), compared with conventional platinum-based chemotherapy [5]. *EGFR* mutations occur in approximately 14–32% of lung cancer cases globally, with higher prevalence in East Asian populations [6]. Notably, Indonesian data report mutation frequencies reaching up to 44% among newly diagnosed NSCLC patients, underscoring the clinical relevance of EGFR-targeted therapy in this population [7].

Activating *EGFR* mutations promote persistent downstream signalling through pathways involved in cellular proliferation, survival, angiogenesis, migration, and inhibition of apoptosis [8]. EGFR-TKIs effectively block this aberrant signalling, resulting in higher response rates and a median progression-free survival (PFS) of approximately 10–12 months [9]. However, therapeutic resistance remains a major clinical challenge. Both primary resistance (failure to respond from treatment initiation) and acquired resistance (disease progression after an initial response) significantly limit the durability of EGFR-TKI efficacy [10]. Primary resistance alone occurs in approximately 20% of patients and contributes substantially to early treatment failure, highlighting the need for predictive biomarkers capable of identifying high-risk individuals before therapy initiation [11].

Despite considerable advances in understanding the molecular mechanisms underlying EGFR-TKI resistance, clinically reliable biomarkers that can accurately predict resistance and disease progression before treatment failure remain lacking. Numerous resistance-associated genes and signalling pathways have been reported; however, few have demonstrated sufficient reproducibility and biological relevance for clinical application. Transcriptomic analysis of differentially expressed genes (DEGs) provides a systematic approach to identifying genes and molecular pathways that are consistently associated with therapeutic resistance and tumour progression. Integrating DEG analysis with protein–protein interaction networks, functional enrichment, and survival analyses may facilitate the identification of biologically meaningful candidate biomarkers involved in resistance-related signalling pathways.

Therefore, the present study aimed to identify DEGs associated with EGFR-TKI resistance in EGFR-mutant lung adenocarcinoma through integrated bioinformatic analyses of publicly available transcriptomic datasets. Protein–protein interaction analysis, Gene Ontology enrichment, KEGG pathway mapping, Kaplan–Meier survival analysis, and experimental RT-qPCR validation were subsequently performed to prioritize and validate candidate biomarkers with potential associations with resistance mechanisms and disease progression.

2. Materials and Methods

2.1. Study Design and Data Collection

This study employed an analytical observational approach with a cross-sectional design, integrating bioinformatics and laboratory-based validation. The study commenced after obtaining ethical approval in February 2025. The research was conducted in four

sequential phases. First, an *in silico* bioinformatics analysis was performed at the Bioinformatics Laboratory, Department of Medical Chemistry, Faculty of Medicine, Universitas Indonesia. Publicly available transcriptomic datasets were analysed to identify differentially expressed genes (DEGs) associated with lung adenocarcinoma and treatment response. Upon completion of the bioinformatics analysis, formalin-fixed paraffin-embedded (FFPE) tissue specimens were collected from the archives of the Department of Anatomical Pathology at Dharmas Cancer Hospital, Jakarta, Indonesia. The archived specimens were obtained from patients with histologically confirmed lung adenocarcinoma. Next, experimental validation was conducted by measuring the relative expression levels of the selected DEGs using quantitative reverse transcription polymerase chain reaction (RT-qPCR). Patients included in the validation cohort were diagnosed with EGFR-mutant lung adenocarcinoma and had received EGFR tyrosine kinase inhibitor (EGFR-TKI) therapy. The validation cohort comprised EGFR-TKI-sensitive and EGFR-TKI-resistant patients in a 1:2 ratio. Because this study was designed as a pilot validation of candidate biomarkers identified through bioinformatics analyses, a small independent cohort (3 sensitive and 6 resistant cases) was used for validation. The EGFR-TKI-sensitive group consisted of patients who remained progression-free during EGFR-TKI treatment, whereas the EGFR-TKI-resistant group included patients who developed progressive disease according to the Response Evaluation Criteria in Solid Tumors (RECIST) following EGFR-TKI therapy. Patients who had received systemic anticancer treatments other than EGFR-TKIs before disease progression were excluded to minimize potential confounding effects on gene expression. Finally, the RT-qPCR results were integrated with the bioinformatics findings, followed by statistical analysis and interpretation of the data to evaluate the association between gene expression and lung adenocarcinoma.

2.2. Identification and Analysis of Differentially Expressed Genes (DEGs)

Transcriptomic datasets were retrieved from the Gene Expression Omnibus (GEO) database using the following search string: “EGFR-mutant lung adenocarcinoma gefitinib/erlotinib/afatinib”. Two datasets were included in this study. GSE123066 is a microarray dataset generated from the EGFR-mutant NSCLC HCC4006 cell line, comprising three parental gefitinib-sensitive and three acquired gefitinib-resistant samples. Although the original dataset contains nine samples, three samples treated with the TGF- β receptor inhibitor SB431542 were excluded to ensure a direct comparison between sensitive and resistant phenotypes. GSE178755 is an RNA-sequencing dataset generated from the EGFR-mutant NSCLC HCC827 cell line, consisting of three DMSO-treated controls and three erlotinib-treated samples. This dataset was included to investigate transcriptomic changes induced by EGFR-TKI exposure but does not represent an acquired resistance model. Detailed characteristics of the datasets, including sample size, experimental platform, and treatment conditions, are summarized in Table 1.

Raw expression data were analysed using the GEO2R web-based platform (<http://www.ncbi.nlm.nih.gov/geo/geo2r>, accessed on 16 March 2026). GEO2R utilizes the R-based limma package (version 3.64.0) to compare gene expression between control and treatment groups. The limma background Correct function was used to remove non-specific signals. Statistical thresholds were defined as an adjusted *p*-value (*p*_{adj}) < 0.05 and an absolute log₂ fold change ($|\log_2FC|$) ≥ 1 . A volcano plot was generated to visualize the distribution of significantly upregulated and downregulated genes. To identify genes consistently dysregulated across datasets, an intersection analysis was conducted using a Venn diagram. Venny 2.1.0 (<https://bioinfogp.cnb.csic.es/tools/venny/>, accessed on 16 March 2026) was utilized to visualize overlapping gene sets and to identify candidate biomarkers potentially associated with disease progression. The identified overlapping DEGs were subsequently selected for downstream analyses.

Table 1. Characteristics of GEO datasets included in this study.

GEO Dataset	Platform	Sample Type	EGFR-TKI-Sensitive (<i>n</i>)	EGFR-TKI-Resistant/Comparison (<i>n</i>)	Total (<i>n</i>)
GSE123066	Microarray (Illumina HumanHT-12 V4.0 Expression BeadChip, GPL10904)	Human EGFR-mutant NSCLC HCC4006 cell line (parental vs. gefitinib-resistant)	3	3	6 *
GSE178755	RNA-seq (Illumina NovaSeq 6000, GPL24676)	Human EGFR-mutant NSCLC HCC827 cell line (DMSO-treated vs. erlotinib-treated)	3	3 †	6

Abbreviations: NSCLC, non-small cell lung cancer; EGFR-TKI, epidermal growth factor receptor tyrosine kinase inhibitor. Notes: * GSE123066 contains nine samples in GEO, including three gefitinib-resistant cells treated with the TGF- β receptor inhibitor SB431542. Only the parental HCC4006 cells ($n = 3$) and gefitinib-resistant HCC4006 cells ($n = 3$) were included in this study to compare EGFR-TKI-sensitive and EGFR-TKI-resistant phenotypes. † GSE178755 does not represent an acquired EGFR-TKI-resistant model. The comparison consists of untreated (DMSO) HCC827 cells and HCC827 cells exposed to erlotinib for 24 h, reflecting the acute transcriptional response to EGFR inhibition rather than acquired resistance.

2.3. PPI Network Construction and Functional Enrichment Analysis

The list of overlapping differentially expressed genes (DEGs) identified from GSE123066 and GSE178755 following GEO2R analysis ($p < 0.05$ and $|\log_2FC| \geq 1$) was uploaded to the STRING database (version 12.0) [12] to construct the protein–protein interaction (PPI) network. A minimum required interaction confidence score of 0.400 (medium confidence) was applied. The PPI network generated by STRING was subsequently interpreted in conjunction with Gene Ontology (GO) enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses to identify biologically relevant signalling pathways associated with EGFR-TKI resistance. Functional enrichment analysis was subsequently performed, and the PI3K–Akt signalling pathway was selected under the “Functional Enrichment Visualization” option to highlight genes associated with this pathway. In the resulting network, highlighted (red) nodes represent genes annotated to the PI3K–Akt signalling pathway and do not indicate the direction or magnitude of differential gene expression. No formal hub-gene analysis based on network topology metrics (e.g., degree centrality or maximal clique centrality) was performed.

Gene Ontology (GO) enrichment analysis was performed using the Enrichr platform [13,14], whereas KEGG pathway enrichment analysis was conducted using ShinyGO version 0.82 [15] with the Kyoto Encyclopedia of Genes and Genomes (KEGG) database [16] as the reference. Genes significantly enriched in the PI3K–Akt signalling pathway were subsequently mapped onto the canonical KEGG pathway using KEGG Mapper [17] and the Pathview R/Bioconductor package (version 1.44.0) [18] to determine their functional positions within the signalling cascade. Based on their reported biological functions and pathway locations, genes were classified into five functional categories: (1) receptor activation, (2) signal transduction, (3) transcription and cell cycle regulation, (4) metabolism and stress response, and (5) cell survival and anti-apoptotic processes. This functional categorization was developed by the authors to facilitate biological interpretation of the identified DEGs and to identify candidate genes potentially involved in bypass signalling associated with EGFR-TKI resistance. Functional annotation and pathway integration were performed manually using Microsoft Excel.

2.4. Kaplan–Meier Survival Analysis

The prognostic significance of the prioritized differentially expressed genes (DEGs) was evaluated using the Kaplan–Meier Plotter web server (<https://kmplot.com/analysis/>, accessed on 14 July 2026), an online survival analysis platform integrating gene expression and clinical data from publicly available cancer datasets [19]. Survival analysis was performed in patients with adenocarcinoma. Patients were stratified into high- and low-expression groups using the median gene expression value as the default cutoff. Kaplan–Meier survival curves were generated automatically by the platform, and differences between groups were evaluated using the log-rank test. Hazard ratios (HRs) with 95% confidence intervals (CIs) were calculated by the Kaplan–Meier Plotter platform, and a two-sided p -value < 0.05 was considered statistically significant [12].

2.5. RNA Isolation and Quality Assessment

Total RNA was extracted from FFPE tumour tissue of lung adenocarcinoma patients using the Fastpure cell/tissue total RNA extraction (Nanjing Vazyme Biotech Co., Ltd., Nanjing, China) according to the manufacturer's protocol. RNA purity was assessed using the A260/A280 ratio via NanoDrop™ spectrophotometry (Thermo Fisher Scientific, Waltham, MA, USA). RNA concentration was quantified using a Quantus™ Fluorometer (Promega Corporation, Madison, WI, USA) in combination with the QuantiFluor® RNA System (Promega Corporation, Madison, WI, USA). Extracted RNA was aliquoted to minimize repeated freeze–thaw cycles and stored at -80°C until further use. The isolated RNA served as the template for downstream RT-qPCR validation experiments. RNA purity (A260/A280) and concentration are summarized in Supplementary Table S1, with purity values ranging from 1.46 to 2.20 and concentrations from 11.84 to 33.92 ng/ μL , indicating that the RNA quality was adequate for downstream qRT-PCR analysis.

2.6. Validation of DEG Biomarkers by qRT-PCR

Sequences of *TGFA*, *DDIT4*, and *SOX9* gene primers as DEG biomarkers were downloaded from NCBI. The primer used for amplification was designed using Primer-BLAST [20]. Primer specificity was evaluated using the Integrated DNA Technologies (IDT) OligoAnalyzer tool to meet standard qRT-PCR requirements, including optimal primer length, melting temperature (T_m), GC content, amplicon size, and absence of secondary structures, and in silico PCR was performed using the UCSC In Silico PCR platform [21] to verify the expected amplification products, and the most suitable primer pairs were selected for each target gene [22–24]. The results of the primary sequence design and the results of the primary annealing temperature optimization of the gene were successfully obtained, as seen in Table 2.

Table 2. Designed primers for biomarkers.

Primer Name	Primer Sequences	
	Forward	Reverse
SOX9	AGGAAGCTCGCGGACCAGTAC	GGTGGTCCTTCTTGCTGCAC
TGFA	GGTCCGAAAACACTGTGAGTGG	CAAACCTCCTCTGGGCTCTT
DDIT4	GTTTGACCGCTCCACGAGCCT	GCACACAAGTGTTTCATCCTCAGG

Gene expression levels were quantified by real-time quantitative PCR (RT-qPCR). Relative gene expression was calculated using the comparative C_t ($2^{-\Delta\Delta C_t}$) method described by Livak and Schmittgen (2001) [25], with ACTB serving as the endogenous reference gene. Complementary DNA (cDNA) synthesis was performed using the Vazyme HiScript III RT SuperMix for qPCR (+gDNA Wiper) (Vazyme Biotech Co., Ltd., Nanjing,

China) to remove genomic DNA contamination prior to reverse transcription. Quantitative PCR amplification was carried out under the following thermal cycling conditions: initial denaturation at 95 °C for 15 min; 38 amplification cycles consisting of denaturation at 95 °C for 50 s, annealing at 59 °C for 50 s, and extension at 72 °C for 50 s; followed by a final extension at 72 °C for 10 min. The reaction mixture (total volume 20 µL) consisted of 1 µL cDNA (15 ng/µL), 10 µL of 2× Vazyme SYBR Green PCR Master Mix (Vazyme Biotech Co., Ltd., Nanjing, China), 0.8 µL each of forward and reverse primers (100 µM), and nuclease-free water to volume. All reactions were performed in duplicate to ensure reproducibility. Differences in relative gene expression between EGFR-TKI-sensitive and EGFR-TKI-resistant groups were evaluated using an independent-sample *t*-test. Statistical significance was defined as a two-sided *p*-value < 0.05.

2.7. Statistical Analysis

Statistical analyses were performed according to the requirements of each analytical platform. Differential gene expression analysis of the GEO datasets was conducted using the GEO2R web tool, which is based on the **limma** package in R (version 3.64.0). Genes were considered differentially expressed if they met the predefined criteria of an adjusted *p* < 0.05 and an absolute log₂ fold change ($|\log_2FC| \geq 1$). Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were performed using the Enrichr and ShinyGO platforms, respectively. Enriched terms with an adjusted *p* < 0.05 were considered statistically significant.

Survival analysis was performed using the Kaplan–Meier Plotter database. Patients were stratified into high- and low-expression groups according to the optimal expression cutoff determined by the platform. Progression-free survival (PFS) was compared using the log-rank test, and hazard ratios (HRs) with 95% confidence intervals (CIs) were obtained directly from the Kaplan–Meier Plotter output.

For RT-qPCR validation, relative gene expression levels were calculated using the comparative Ct ($2^{-\Delta\Delta Ct}$) method with GAPDH as the internal reference gene. The normality of the data distribution was assessed using the Shapiro–Wilk test. Normally distributed data were compared using an unpaired Student's *t*-test, whereas non-normally distributed data were analysed using the Mann–Whitney *U* test. Data are presented as mean ± standard deviation (SD). All statistical tests were two-sided, and a *p*-value < 0.05 was considered statistically significant. Statistical analyses of the RT-qPCR data were performed using Microsoft Excel.

3. Results

3.1. Differentially Expressed Genes (DEGs) Biomarker Analysis

Gene expression profiles were obtained from the GSE123066 and GSE178755 datasets. The discovery and validation phases of this study were conducted using independent cohorts, which is a common strategy in biomarker discovery studies. Candidate genes were initially identified through integrated analysis of publicly available transcriptomic datasets and subsequently validated by RT-qPCR in an independent cohort of patients with EGFR-mutant lung adenocarcinoma. Furthermore, the discovery datasets were derived from NSCLC cell lines, whereas the validation cohort comprised FFPE tumour tissue specimens from patients. Although these sample types differ in their biological complexity, both are tumour-derived and reflect molecular alterations associated with EGFR-TKI response. Cell-line models provide a controlled system for identifying resistance-associated genes, while validation in clinical FFPE tumour tissues enhances the translational relevance of the findings. Nevertheless, these differences represent a limitation of the present study and should be considered when interpreting the findings. Future validation in

larger, prospectively collected cohorts using standardized sample types will be necessary to confirm the clinical utility of the proposed biomarker panel.

Differential expression analysis was performed using the GEO2R web-based platform to identify significantly differentially expressed genes (DEGs) in lung adenocarcinoma based on predefined threshold criteria. In the GSE123066 dataset, a total of 479 genes were identified as significantly upregulated and 326 genes as significantly downregulated. In the GSE178755 dataset, 2512 upregulated genes and 2726 downregulated genes were detected (Table 3). The distribution and magnitude of these differentially expressed genes in each dataset are visualized in Figure 1A.

Table 3. Summary of differentially expressed genes (DEGs) identified using GEO2R (limma) with selection criteria of $p < 0.05$ and $|\log_2FC| \geq 1$.

GEO Dataset	Platform	Sample Type	Sample Size (n)	Total Gene	Upregulated DEGs	Downregulated DEGs	Total DEGs
GSE123066	Microarray	NSCLC cell line	6	25,906	479	326	805
GSE178755	RNA-seq	NSCLC cell line	6	17,043	2512	2726	5238

Abbreviations: DEG, differentially expressed gene; EGFR-TKI, epidermal growth factor receptor tyrosine kinase inhibitor. Selection criteria: Differential expression analysis was performed using the GEO2R platform based on the limma package. Genes were considered differentially expressed if they satisfied both the statistical significance criterion ($p < 0.05$) and the fold-change threshold ($|\log_2FC| \geq 1$).

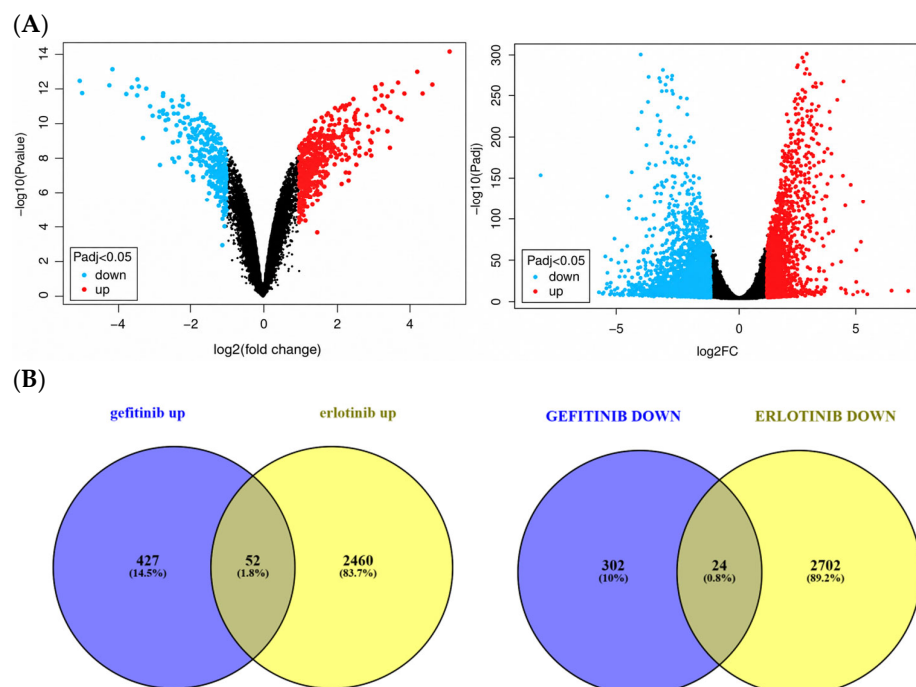


Figure 1. Identification of differentially expressed genes (DEGs) from the GEO datasets: (A) Volcano plots illustrating the distribution of differentially expressed genes in the GSE123066 and GSE178755 datasets. Coloured dots indicate genes meeting the predefined differential expression criteria. Red dots represent significantly upregulated genes, whereas blue dots represent significantly downregulated genes. Black dots represent genes that did not meet these criteria. (B) Venn diagram showing

the overlapping DEGs identified between GSE123066 and GSE178755, which were subsequently used for downstream protein–protein interaction, functional enrichment, and pathway analyses.

To identify genes showing consistent dysregulation across both datasets, an intersection analysis was conducted using a Venn diagram, which revealed a total of 76 overlapping DEGs. Among these, 52 genes were consistently upregulated, while 24 genes were consistently downregulated (Figure 1B). These common DEGs were then selected for subsequent PPI network construction and functional enrichment analyses.

3.2. PPI Network and Enrichment Analysis

3.2.1. PPI Network

To explore the functional interactions of proteins encoded by the subtype-specific DEGs, a PPI network analysis was conducted using the STRING database. This analysis aimed to identify key genes and elucidate their roles at the molecular level. Interactions among the DEGs were mapped, and the resulting PPI networks are presented in Figure 2. The PPI analysis demonstrated that several upregulated genes were highly interconnected within key signalling pathways associated with cell proliferation, survival, and drug resistance, particularly the PI3K–Akt signalling pathway. This pathway is well-established for its critical role in lung adenocarcinoma tumorigenesis and in mediating resistance to EGFR tyrosine kinase inhibitors (EGFR-TKIs). Therefore, subsequent analyses focused primarily on the upregulated genes, as their involvement in critical signalling cascades highlights their potential as biologically significant markers and as candidates for further investigation, including target identification and structure-based drug discovery.

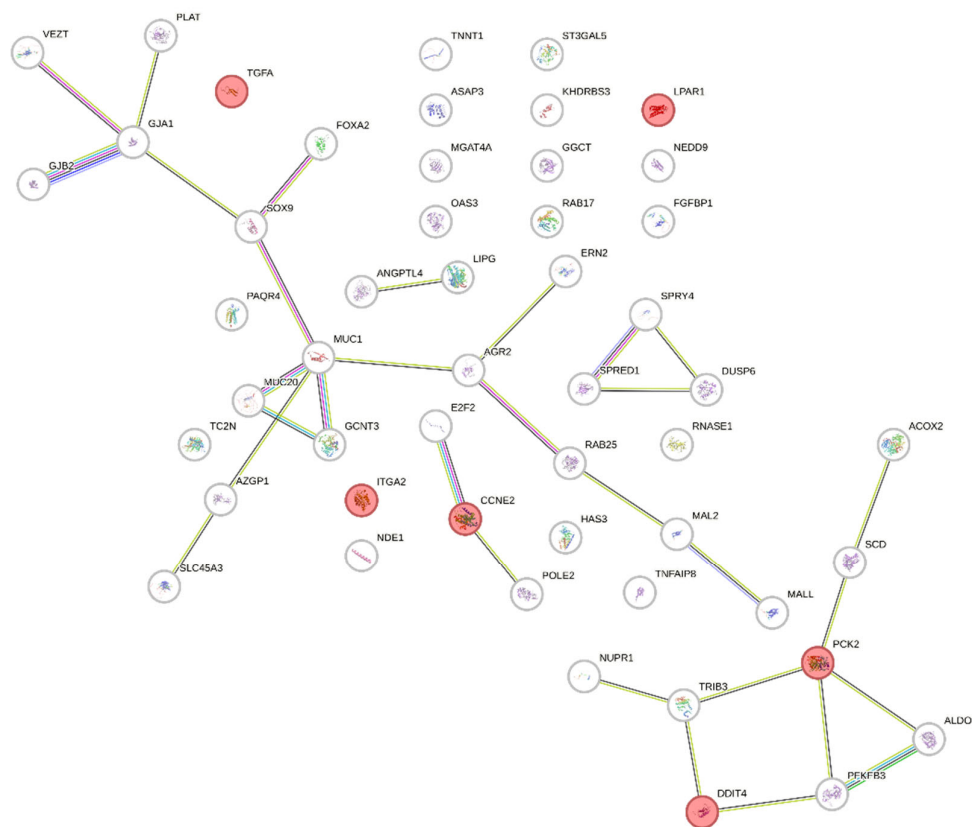


Figure 2. Protein–protein interaction (PPI) network of overlapping differentially expressed genes (DEGs) associated with the PI3K–Akt signalling pathway. The PPI network was generated using

the overlapping DEGs identified from the GSE123066 and GSE178755 datasets. Genes annotated to the KEGG PI3K–Akt signalling pathway are highlighted in red using the STRING Functional Enrichment Visualization feature. Red nodes indicate pathway membership rather than the direction or magnitude of differential gene expression. Coloured edges represent different evidence channels supporting protein–protein associations in STRING, including curated databases, experimentally determined interactions, gene neighbourhood, gene fusion, gene co-occurrence, text mining, co-expression, and protein homology.

3.2.2. Functional Enrichment Analysis

Fifty-two consistently upregulated genes identified from GEO2R analysis were subjected to Gene Ontology (GO) enrichment analysis and classified into three primary GO categories: Biological Process (BP), Molecular Function (MF), and Cellular Component (CC). Table 4 presents the significantly enriched GO terms, the corresponding contributing genes and their p-values ($p < 0.05$). The enrichment analysis indicates that these upregulated genes are predominantly involved in biological processes related to cell proliferation, cell cycle regulation, hypoxia response, and modulation of EGFR signalling, suggesting their potential roles in tumour progression and response to EGFR-targeted therapies. In terms of molecular function, the enriched terms are associated with activities involved in signalling regulation and protein interactions, while the cellular component analysis indicates that many of these genes are localized to cellular compartments that facilitate signal transduction and adaptive responses in cancer cells. Collectively, the GO enrichment results summarized in Table 4 provide functional evidence that the identified upregulated genes may contribute to tumour growth, EGFR signalling, and mechanisms underlying resistance to EGFR-TKI therapy.

Table 4. Functional enrichment analysis of consistently upregulated genes.

Category	GO Term	Genes Involved	p-Value
Biological Process (GO-BP)	Regulation of cell proliferation	E2F2, CCNE2, TGFA, DUSP6, SOX9	1.2×10^{-6}
	Cell cycle progression	E2F2, CCNE2, POLE2	3.8×10^{-6}
	Response to hypoxia	DDIT4, ANGPTL4, PFKFB3	2.1×10^{-5}
	Regulation of EGFR signalling	TGFA, SPRY4, DUSP6, SPRED1	4.5×10^{-5}
	Epithelial cell differentiation	MUC1, MUC20, AGR2, FOXA2, SOX9	6.9×10^{-5}
	Cell migration and adhesion	ITGA2, NEDD9, VEZT, GJA1	1.3×10^{-4}
	Metabolic process regulation	PFKFB3, ALDOC, SCD, ACOX2, PCK2	2.0×10^{-4}
	ER stress and unfolded protein response	AGR2, ERN2, NUPR1	3.6×10^{-4}
Molecular Function (GO-MF)	Growth factor activity	TGFA, ANGPTL4	2.4×10^{-6}
	Protein kinase regulator activity	DUSP6, SPRY4, SPRED1	5.1×10^{-6}
	Transcription factor activity	FOXA2, SOX9, E2F2	8.7×10^{-6}
	Receptor binding	TGFA, ANGPTL4, LPAR1	1.9×10^{-5}
	Cell adhesion molecule binding	ITGA2, VEZT, GJA1	3.4×10^{-5}
	Metabolic enzyme activity	PFKFB3, SCD, ACOX2, ALDOC	6.2×10^{-5}
	RNA polymerase activity regulation	POLE2, NUPR1	1.1×10^{-4}
	Plasma membrane	MUC1, ITGA2, GJA1, LPAR1	9.8×10^{-7}

Cellular Component (GO-CC)	Cell surface	MUC1, MUC20, TGFA	1.6×10^{-6}
	Cytoplasm	DUSP6, PFKFB3, ALDOC, TRIB3	2.7×10^{-5}
	Nucleus	E2F2, FOXA2, SOX9, POLE2	4.1×10^{-5}
	Endoplasmic reticulum	AGR2, ERN2, NUPR1	7.9×10^{-5}
	Cell junction	GJA1, VEZT	1.3×10^{-4}

The KEGG pathway enrichment analysis identified the PI3K–Akt signalling pathway as one of the most significantly enriched pathways among the overlapping differentially expressed genes (DEGs). The functional relationships among these DEGs were first explored through the protein–protein interaction (PPI) network generated using the STRING database, which highlighted genes associated with the PI3K–Akt signalling pathway (Figure 3). These pathway-associated genes were subsequently integrated with Gene Ontology (GO) annotations to facilitate functional interpretation of their biological roles within the signalling cascade.

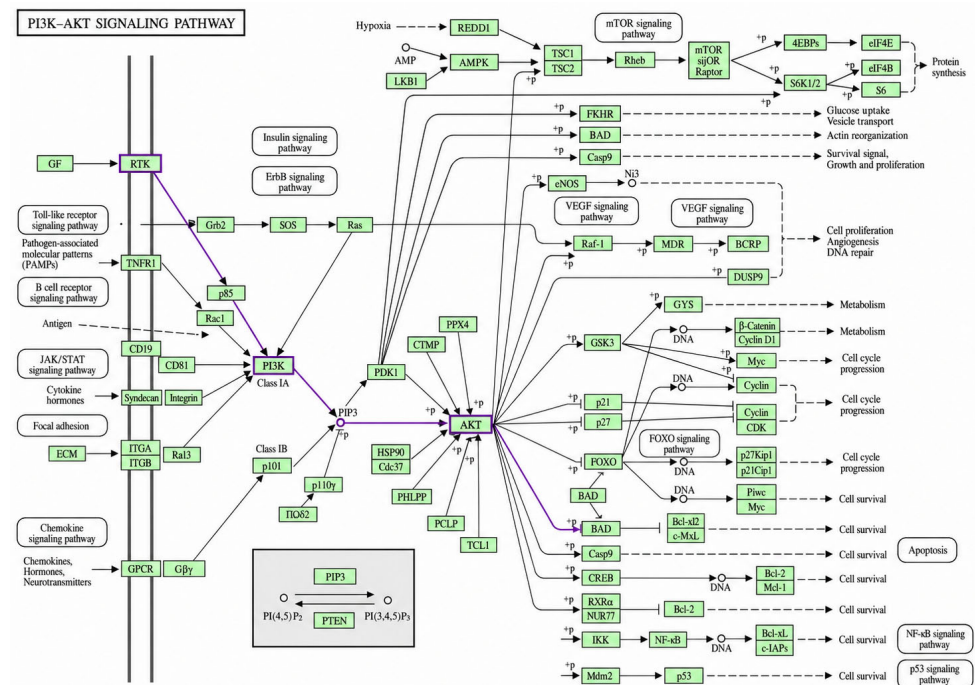


Figure 3. Mapping of overlapping differentially expressed genes (DEGs) onto the KEGG PI3K–Akt signalling pathway. Overlapping DEGs identified from the integrated bioinformatics analysis were mapped onto the canonical KEGG PI3K–Akt signalling pathway using KEGG Mapper. Green boxes indicate genes from the uploaded DEG list that were successfully mapped to the pathway, whereas white boxes represent pathway components that were not identified in the present study. Arrows denote the canonical molecular interactions and signalling relationships defined by the KEGG database. Purple lines highlight the interactions involving the candidate genes identified in this study within the PI3K–Akt signalling pathway.

To further investigate their biological relevance, the significantly upregulated DEGs were mapped onto the canonical PI3K–Akt signalling pathway using Pathview. As shown in Figure 4, the mapped genes illustrate the distribution of the identified DEGs within the signalling cascade and highlight potential molecular mechanisms underlying EGFR-TKI resistance in EGFR-mutant lung adenocarcinoma.

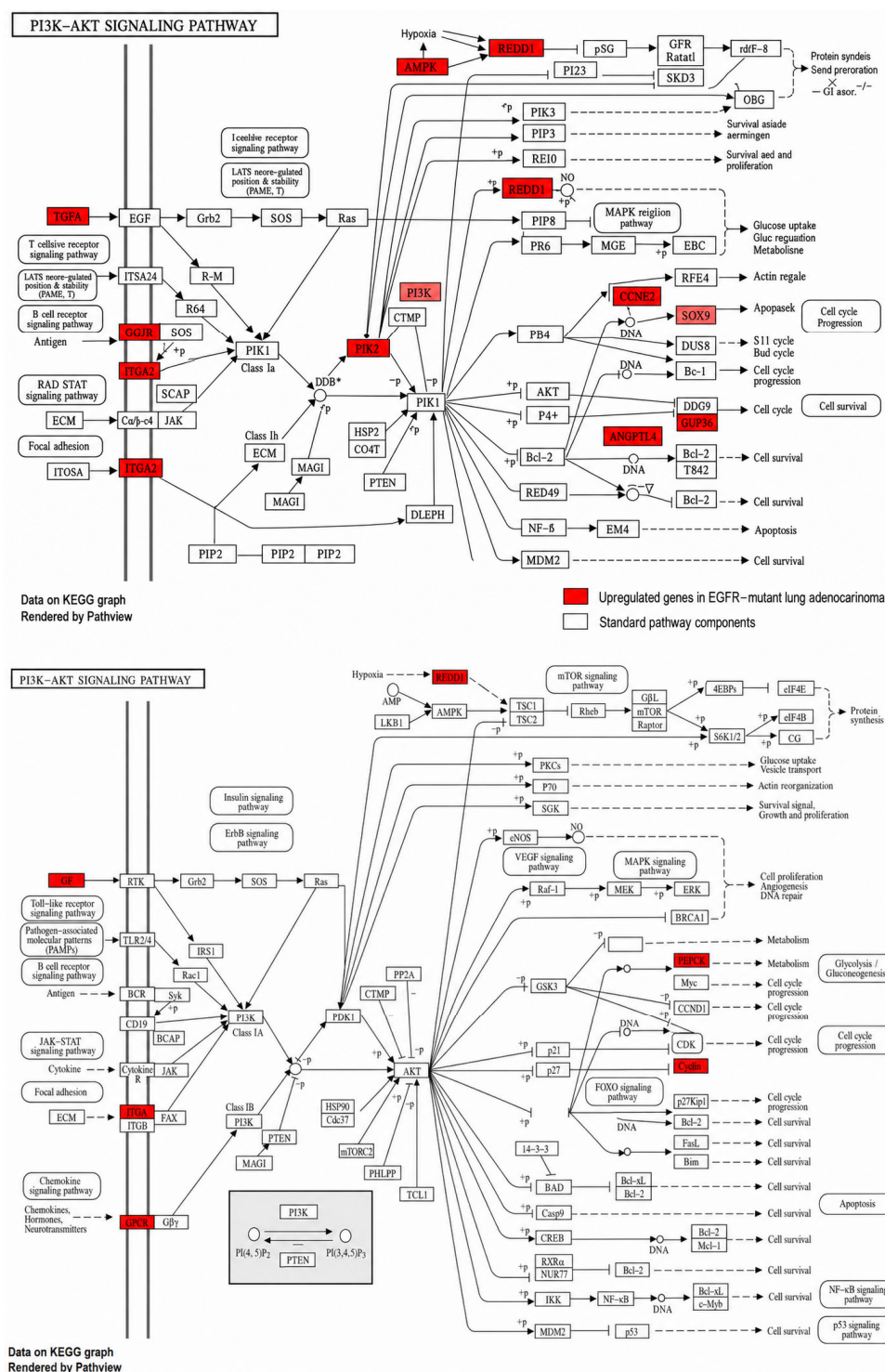


Figure 4. Mapping of significantly upregulated genes onto the canonical EGFR–PI3K–Akt signalling pathway. Genes highlighted in red represent significantly upregulated DEGs identified from the integrated transcriptomic analysis and mapped to the canonical KEGG PI3K–Akt pathway. Arrows and connecting lines represent curated molecular interactions, including activation, inhibition, and phosphorylation events, as defined by the KEGG database. The figure illustrates the distribution of candidate genes within the pathway and should not be interpreted as direct evidence of pathway activation.

The pathway mapping demonstrated that the significantly upregulated genes were distributed across multiple functional levels of the PI3K–Akt signalling cascade (Figure 4). At the receptor level, *TGFA* and *ITGA2* were localized to components involved in receptor activation and cell adhesion, suggesting potential involvement of EGFR- and integrin-mediated signalling. At the intermediate level, *DDIT4* (*REDD1*) and other metabolism-related genes were associated with signalling and metabolic regulatory processes. Downstream, genes involved in cell cycle regulation and transcriptional control, including *CCNE2* and *SOX9*, were mapped to components associated with cell proliferation and survival. Overall, the identified DEGs were distributed throughout the PI3K–Akt signalling pathway rather than being confined to a single functional module, indicating that multiple components of the pathway may contribute to the molecular mechanisms associated with EGFR-TKI resistance in EGFR-mutant lung adenocarcinoma.

3.3. Prognostic Significance of Candidate Genes

The results of GO enrichment and PI3K–Akt pathway mapping were subsequently integrated into a conceptual GO–OS model by linking gene function and pathway position to the clinical outcome of overall survival (OS) (Table 5). Genes involved in cell proliferation, metabolic regulation, and cell survival were considered potential contributors to tumour cell adaptation during EGFR-TKI therapy. Functionally, the activation of these processes may support sustained tumour growth and therapeutic resistance, which could clinically influence the duration of OS in patients with EGFR-mutant lung adenocarcinoma.

Table 5. Identification of dominant bypass pathways to EGFR-TKI therapy and their clinical implications for overall survival (OS).

Dominant Bypass Pathway	Key Upregulated Genes	Molecular Evidence (GO & KEGG)	Bypass Mechanism Against EGFR-TKI	Clinical Implication (OS)
PI3K–Akt signalling pathway	<i>TGFA</i> , <i>ITGA2</i> , <i>PFKFB3</i> , <i>DDIT4</i> , <i>E2F2</i> , <i>CCNE2</i> , <i>SOX9</i>	Most significant KEGG enrichment ($p < 1 \times 10^{-6}$); genes distributed from upstream to downstream levels	Sustained activation of survival and proliferative signalling despite EGFR inhibition	Shorter OS; primary resistance
MAPK/ERK signalling pathway	<i>DUSP6</i> , <i>SPRY4</i> , <i>SPRED1</i>	Significant enrichment: genes function as feedback regulators	Activation of alternative signalling via PI3K/MAPK crosstalk	Adaptive resistance to EGFR-TKI
HIF-1/metabolic pathway	<i>DDIT4</i> , <i>PFKFB3</i> , <i>ANGPTL4</i>	Enrichment in HIF-1 pathway; GO-BP hypoxia response	Metabolic adaptation and cellular stress response	Early progression; reduced OS
Cell cycle pathway	<i>E2F2</i> , <i>CCNE2</i> , <i>POLE2</i>	Enrichment in cell cycle pathway; GO-BP cell proliferation	EGFR-independent cell proliferation	Accelerated disease progression

Candidate genes for biomarkers were selected based on a combination of predefined criteria. First, genes were required to demonstrate consistent upregulation across at least two independent GEO datasets. Second, biological significance was considered, with priority given to genes enriched in Gene Ontology Biological Processes (GO-BP) related to proliferation, metabolism, and cell survival. Third, pathway relevance was assessed, focusing on genes involved in the PI3K–Akt signalling pathway or other identified bypass pathways. Fourth, strategic pathway position was evaluated, prioritizing genes located at critical upstream or downstream nodes within the signalling cascade. Finally, clinical relevance was considered, emphasizing genes with potential influence on response to EGFR-TKI therapy. Based on these integrated criteria, the following genes were prioritized for

Kaplan–Meier survival analysis: *TGFA*, *PFKFB3*, *DDIT4*, *E2F2*, *CCNE2*, *SOX9*, and *DUSP6* (Table 6).

Table 6. Prioritized candidate biomarkers, functional pathways, and rationale.

Gene	Pathway/Function	Rationale as Biomarker
<i>TGFA</i>	EGFR activation (up-stream)	Indicates autocrine EGFR activation
<i>PFKFB3</i>	Glycolytic metabolism	Marker of metabolic adaptation
<i>DDIT4</i>	HIF-1/mTOR pathway	Indicator of cellular stress and hypoxia response
<i>E2F2</i>	Cell cycle regulation	Marker of EGFR-independent proliferation
<i>CCNE2</i>	G1/S phase transition	Associated with aggressive cell cycle progression
<i>SOX9</i>	Transcription factor	Linked to stemness and therapeutic resistance
<i>DUSP6</i>	MAPK feedback regulation	Reflects activation of bypass signalling pathways

To evaluate the clinical relevance of the candidate genes identified through the integrated bioinformatics analysis, Kaplan–Meier survival analysis was performed using the Kaplan–Meier Plotter database. Patients with EGFR-mutant lung adenocarcinoma were stratified into high- and low-expression groups according to the optimal expression cutoff determined by the platform, and overall survival (OS) was compared using the log-rank test.

Among the seven candidate genes evaluated, high expression of *TGFA*, *DDIT4*, *E2F2*, *CCNE2*, and *SOX9* was significantly associated with shorter OS, whereas elevated *PFKFB3* and *DUSP6* expression was associated with longer OS (Table 3). Specifically, patients with high *TGFA* expression exhibited a shorter median OS than those with low expression (59.53 vs. 79 months, $p = 9.0 \times 10^{-4}$). Similarly, elevated *DDIT4* expression was associated with reduced median OS (62 vs. 80 months, $p = 0.017$), while high *SOX9* expression was also associated with poorer prognosis (61.3 vs. 74 months, $p = 0.022$) (Figure 5).

Although *E2F2* and *CCNE2* demonstrated stronger statistical significance in the survival analysis, these genes primarily function as downstream regulators of cell cycle progression and proliferation. In contrast, *TGFA*, *DDIT4*, and *SOX9* were selected for subsequent experimental validation because they represent distinct biological mechanisms implicated in EGFR-TKI resistance. *TGFA* functions as an EGFR ligand capable of activating bypass signalling through the EGFR–PI3K–Akt pathway; *DDIT4* regulates cellular stress responses and mTOR signalling under therapeutic pressure; and *SOX9* is a transcription factor associated with epithelial–mesenchymal transition, cancer stemness, and acquired drug resistance. Together, these genes provide complementary biological insights into receptor activation, stress adaptation, and transcriptional regulation underlying EGFR-TKI resistance. Based on their significant association with poorer OS and their established biological relevance to EGFR-TKI resistance, *TGFA*, *DDIT4*, and *SOX9* were prioritized for subsequent validation by RT-qPCR.

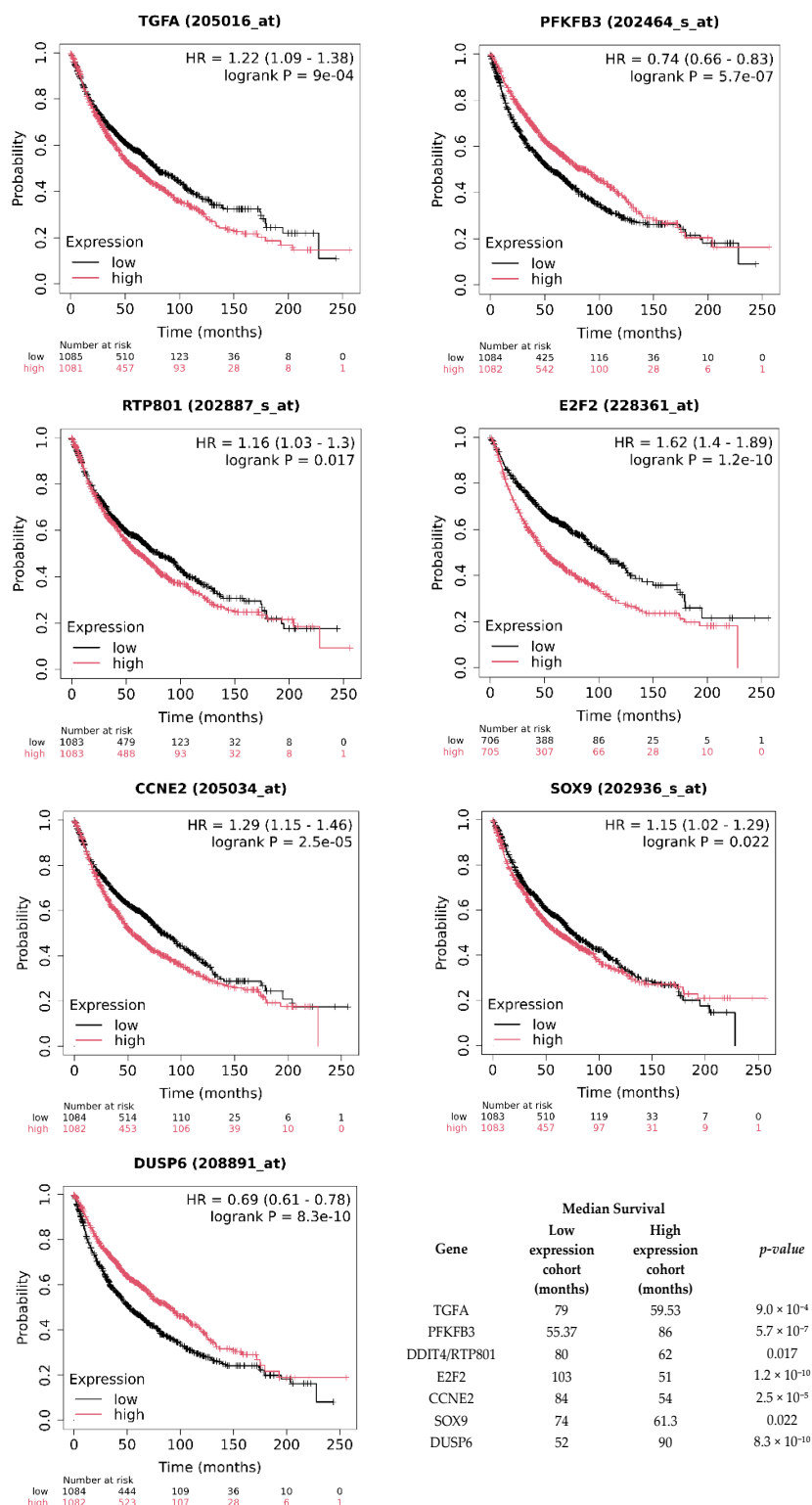


Figure 5. Kaplan–Meier survival analysis of candidate genes in EGFR-mutant lung adenocarcinoma. Kaplan–Meier overall survival (OS) curves generated using the Kaplan–Meier Plotter database for seven genes. Patients were stratified into high- and low-expression groups according to the optimal expression cutoff determined by the Kaplan–Meier Plotter platform. The corresponding hazard ratio (HR), 95% confidence interval (CI), and log-rank *p*-value are displayed in each panel.

3.4. DEG Biomarker Validation Using RT-qPCR

RT-qPCR analysis was performed using RNA isolated from FFPE tumour tissue samples obtained from patients with EGFR-mutant lung adenocarcinoma who were classified as either EGFR-TKI-sensitive or EGFR-TKI-resistant according to the study inclusion criteria. Relative gene expression was calculated using the comparative Ct ($2^{-\Delta\Delta Ct}$) method with GAPDH as the internal reference gene.

Experimental validation using RT-qPCR further confirmed the upregulated expression pattern of the three-gene panel (TGFA, DDIT4, and SOX9) in the resistant group compared with the sensitive group, as detailed in Supplementary Tables S2 and S3. As shown in Figure 6, the expression levels of TGFA, DDIT4, and SOX9 were consistently higher in the EGFR-TKI-resistant group than in the EGFR-TKI-sensitive group. RT-qPCR analysis demonstrated significantly higher expression of SOX9 in the resistant group compared with the sensitive group ($p = 0.044$). Although TGFA exhibited a similar direction of expression change as observed in the transcriptomic datasets, the difference did not reach statistical significance ($p = 0.122$). In contrast, DDIT4 showed no significant difference between the two groups ($p = 0.951$). Despite these findings, the RT-qPCR results provide preliminary experimental support for SOX9 while indicating that further validation of TGFA and DDIT4 in larger independent cohorts is required.

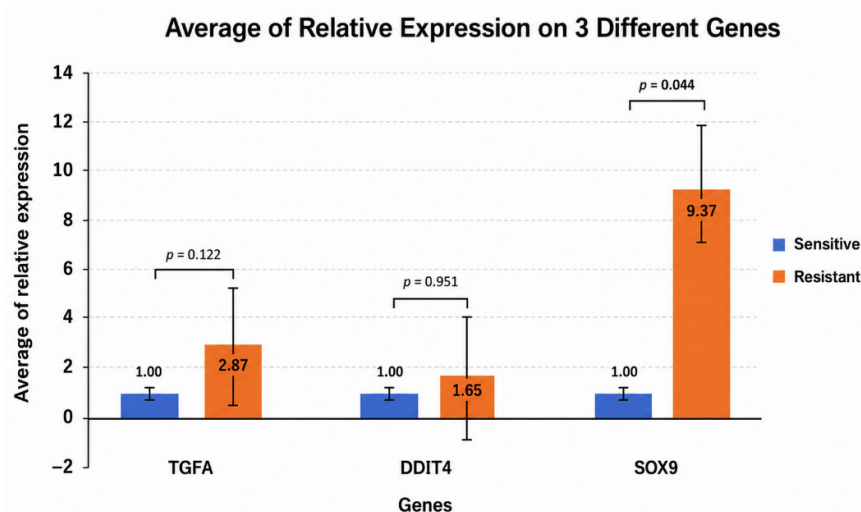


Figure 6. Relative expression of TGFA, DDIT4, and SOX9 in EGFR-TKI-sensitive and EGFR-TKI-resistant patients. Relative mRNA expression levels of TGFA, DDIT4, and SOX9 were quantified by RT-qPCR and normalized to GAPDH using the comparative Ct ($2^{-\Delta\Delta Ct}$) method. Gene expression levels are presented relative to the EGFR-TKI-sensitive group (set to 1.0). Bars represent the mean $\pm$ standard deviation (SD). Statistical significance between the EGFR-TKI-sensitive and EGFR-TKI-resistant groups was determined using the unpaired *t*-test, with $p < 0.05$ considered statistically significant.

4. Discussion

The development of acquired resistance remains one of the major challenges in the management of EGFR-mutant lung adenocarcinoma. Although EGFR tyrosine kinase inhibitors (EGFR-TKIs) have substantially improved clinical outcomes, most patients eventually develop resistance through mechanisms extending beyond secondary EGFR mutations. Increasing evidence suggests that activation of bypass signalling pathways, metabolic adaptation, tumour plasticity, and changes in the tumour microenvironment collectively contribute to therapeutic resistance [10,26]. Accordingly, identifying biomarkers

that capture these complementary biological processes may improve patient stratification and facilitate precision oncology.

In the present study, integrated bioinformatics analyses identified TGFA, DDIT4, and SOX9 as candidate genes consistently associated with EGFR-TKI resistance. Differential expression analysis, protein–protein interaction network analysis, GO and KEGG enrichment, and Kaplan–Meier survival analysis collectively indicated that these genes represent complementary biological processes rather than isolated molecular events. KEGG enrichment identified the PI3K–Akt signalling pathway as one of the most significantly enriched pathways. Because this pathway regulates proliferation, survival, metabolism, and apoptosis, its dysregulation has frequently been implicated in acquired resistance to EGFR-targeted therapy [27,28]. However, our findings are based on transcriptomic enrichment analysis and therefore suggest association with PI3K–Akt signalling, rather than direct biochemical activation. Confirmation of pathway activation would require functional studies evaluating phosphorylation of AKT, mTOR, or other downstream signalling proteins.

The three selected genes represent complementary mechanisms that collectively support tumour adaptation during EGFR-TKI treatment. TGFA encodes an EGFR ligand that promotes ligand-dependent receptor activation and sustained downstream signalling through pathways including PI3K–Akt and MAPK [23,29]. DDIT4 (REDD1) functions as a stress-responsive regulator of mTOR signalling and is induced by hypoxia, oxidative stress, and nutrient deprivation, enabling tumour cells to adapt to therapeutic stress [24,30]. SOX9 regulates epithelial–mesenchymal transition (EMT), stemness, and tumour plasticity, all of which have been implicated in EGFR-TKI resistance [31,32]. Rather than acting within a single linear signalling cascade, these genes represent complementary biological processes involving persistent receptor signalling, stress adaptation, and tumour plasticity, thereby providing a mechanistic rationale for proposing them as a combined biomarker panel.

Kaplan–Meier survival analysis further supported the potential clinical relevance of these candidate genes. Increased expression of TGFA, DDIT4, and SOX9 was associated with shorter progression-free survival, although PFKFB3, E2F2, and CCNE2 also demonstrated significant survival associations. We selected TGFA, DDIT4, and SOX9 for experimental validation because together they capture three distinct yet complementary mechanisms underlying EGFR-TKI resistance. In contrast, PFKFB3, E2F2, and CCNE2 primarily reflect metabolic regulation and cell cycle progression, processes that overlap biologically and were therefore not prioritized for validation. From a translational perspective, a three-gene panel may provide a practical balance between biological representation and clinical feasibility, although future studies should evaluate whether incorporating additional candidate genes improves predictive performance.

RT-qPCR provided partial experimental support for the bioinformatics findings. Among the three selected genes, SOX9 showed significantly higher expression in the resistant group, whereas TGFA and DDIT4 did not demonstrate statistically significant differences in the present validation cohort. These discrepancies may reflect the relatively small sample size, biological heterogeneity among patients, differences between tumour tissue and peripheral blood gene expression, or the complexity of acquired resistance mechanisms. Consequently, TGFA and DDIT4 should be regarded as promising candidate biomarkers requiring validation in larger independent cohorts. Nevertheless, integrating transcriptomic profiling, pathway enrichment, survival analysis, and experimental validation provide a stronger biological framework than relying on any single analytical approach and supports further investigation of this mechanism-based biomarker panel.

The present study has several limitations. First, the RT-qPCR validation cohort was relatively small, which limited statistical power and may explain why TGFA and DDIT4 did not reach statistical significance despite their identification through transcriptomic analyses. Second, the validation cohort was recruited from a single institution using a cross-sectional case–control design comparing patients with established EGFR-TKI sensitivity and resistance. Consequently, the study identified candidate biomarkers associated with resistance status rather than biomarkers that prospectively predict disease progression. Third, the discovery phase relied on publicly available GEO datasets with differences in patient characteristics, experimental platforms, and sample processing. To minimize dataset-specific bias, each dataset was analysed independently, and only overlapping differentially expressed genes were retained for downstream analyses; however, residual heterogeneity cannot be excluded. Finally, pathway activation was inferred from transcriptomic enrichment analysis without protein-level or functional validation. Furthermore, the enrichment of the PI3K–Akt pathway was inferred from transcriptomic analysis rather than direct biochemical evidence. Future multicentre prospective studies with larger cohorts and mechanistic experiments are required to confirm the biological and clinical utility of these candidate biomarkers.

5. Conclusions

This study identified TGFA, DDIT4, and SOX9 as candidate genes associated with EGFR-TKI resistance in EGFR-mutant lung adenocarcinoma through an integrated bioinformatics approach combining differential expression analysis, protein–protein interaction network analysis, functional enrichment, and survival analysis. These genes represent complementary biological processes involving receptor activation, cellular adaptation to stress, and tumour plasticity, supporting their potential relevance to mechanisms of acquired resistance.

Experimental validation using RT-qPCR provided preliminary support for SOX9, whereas TGFA and DDIT4 did not demonstrate statistically significant differential expression in the present validation cohort. Therefore, these findings should be interpreted as exploratory and hypothesis-generating rather than definitive evidence of clinical biomarkers.

Overall, the study provides a rationale for further investigating a mechanism-based biomarker panel for EGFR-TKI resistance. However, larger independent patient cohorts, multicentre validation studies, and functional experiments are required to confirm the biological significance and clinical utility of these candidate genes before they can be considered for routine clinical application.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/biomedinformatics6040052/s1>, Table S1: Concentration and purity of RNA isolates from FFPE lung adenocarcinoma cancer samples; Table S2: Gene's relative expression results in a resistant and sensitive group; Table S3: T-test for differences in TGFA, DDIT, and SOX9 expression.

Author Contributions: Conceptualization, M.A.H. and F.; methodology, F., S.I.W. and N.S.; software, F., R.I.P. and A.M.F.; validation, F.; formal analysis, F., R.I.P., A.M.F. and M.A.H.; investigation, M.A.H.; resources, F., S.I.W. and N.S.; data curation, F., S.I.W. and N.S.; writing—original draft preparation, M.A.H.; writing—review and editing, M.A.H., F., S.I.W. and N.S.; visualization, M.A.H. and F.; supervision, F., S.I.W. and N.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Dharmais Cancer Hospital, which provided funding for publication.

Institutional Review Board Statement: This study was conducted in accordance with the principles of the Declaration of Helsinki and was approved by the Ethics Committee of Dharmais Cancer Hospital (Approval No. DP.04.03/11.10/056/2025) on 17 February 2025.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments: The authors would like to acknowledge Dharmais Cancer Hospital, Jakarta, Indonesia, and the Faculty of Medicine, Universitas Indonesia, Jakarta, Indonesia, for their support.

Conflicts of Interest: The authors declare that there are no conflicts of interest related to this study.

References

1. Ervik, M.; Laversanne, M.; Lam, F.; Colombet, M.; Ferlay, J.; Filho, A.M.; Bray, F. Global Cancer Observatory: Cancer Today (2024 estimates, version 1.0). Lyon, France: International Agency for Research on Cancer. Available from: <https://gco.iarc.who.int/today> (accessed on 16 March 2026).
2. Siegel, R.L.; Miller, K.D.; Fuchs, H.E.; Jemal, A. Cancer statistics, 2022. *CA Cancer J. Clin.* **2022**, *72*, 7–33.
3. Dancey, J.; Forde, P.M.; Mak, R.H.; Rajan, A.; Szabo, E. *Non-Small Cell Lung Cancer Treatment (PDQ®)—Health Professional Version*; National Cancer Institute: Bethesda, MD, USA, 2016.
4. Hanafi, A.R.; Hanif, M.A.; Pangaribuan, M.T.G.; Ariawan, W.P.; Sutandyo, N.; Kurniawati, S.A.; Setiawan, L.; Cahyanti, D.; Rayhani, F.; Imelda, P. Genomic features of lung cancer patients in Indonesia’s national cancer center. *BMC Pulm. Med.* **2024**, *24*, 43.
5. Maemondo, M.; Inoue, A.; Kobayashi, K.; Sugawara, S.; Oizumi, S.; Isobe, H.; Gemma, A.; Harada, M.; Yoshizawa, H.; Kinoshita, I.; et al. Gefitinib or chemotherapy for non-small-cell lung cancer with mutated EGFR. *N. Engl. J. Med.* **2010**, *362*, 2380–2388.
6. Zhang, Y.L.; Yuan, J.Q.; Wang, K.F.; Fu, X.H.; Han, X.R.; Threapleton, D.; Yang, Z.Y.; Mao, C.; Tang, J.L. The prevalence of EGFR mutation in patients with non-small cell lung cancer: A systematic review and meta-analysis. *Oncotarget* **2016**, *7*, 78985–78993.
7. Syahrudin, E.; Wulandari, L.; Muktiati, N.S.; Rima, A.; Soeroso, N.; Ermayanti, S.; Levi, M.; Hidajat, H.; Widjajahakim, G.; Utomo, A.R.H. Uncommon EGFR mutations in cytological specimens of 1,874 newly diagnosed Indonesian lung cancer patients. *Lung Cancer* **2018**, *9*, 25–34.
8. Thomas, R.; Weihua, Z. Rethink of EGFR in cancer with its kinase-independent function on board. *Front. Oncol.* **2019**, *9*, 800.
9. Sun, R.; Hou, Z.; Zhang, Y.; Jiang, B. Drug resistance mechanisms and progress in the treatment of EGFR-mutated lung adenocarcinoma. *Oncol. Lett.* **2022**, *24*, 382.
10. Chhour, H.; Alexandre, D.; Grumolato, L. Mechanisms of acquired resistance and tolerance to EGFR targeted therapy in non-small cell lung cancer. *Cancers* **2023**, *15*, 504.
11. Wang, R.; Pan, S.; Song, X. Research advances of EGFR–TP53 co-mutation in advanced non-small cell lung cancer. *Chin. J. Lung Cancer* **2022**, *25*, 174–182.
12. Szklarczyk, D.; Kirsch, R.; Koutrouli, M.; Nastou, K.; Mehryary, F.; Hachilif, R.; Gable, A.L.; Fang, T.; Doncheva, N.T.; Pyysalo, S.; et al. The STRING database in 2023: Protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* **2023**, *51*, D638–D646.
13. Chen, E.Y.; Tan, C.M.; Kou, Y.; Duan, Q.; Wang, Z.; Meirelles, G.V.; Clark, N.R.; Ma’Ayan, A. Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinform.* **2013**, *14*, 128.
14. Kuleshov, M.V.; Jones, M.R.; Rouillard, A.D.; Fernandez, N.F.; Duan, Q.; Wang, Z.; Koplev, S.; Jenkins, S.L.; Jagodnik, K.M.; Lachmann, A.; et al. Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **2016**, *44*, W90–W97.
15. Ge, S.X.; Jung, D.; Yao, R. ShinyGO: A graphical gene-set enrichment tool for animals and plants. *Bioinformatics* **2020**, *36*, 2628–2629.
16. Kanehisa, M.; Furumichi, M.; Sato, Y.; Kawashima, M.; Ishiguro-Watanabe, M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.* **2023**, *51*, D587–D592.
17. Kanehisa, M.; Sato, Y. KEGG Mapper for inferring cellular functions from protein sequences. *Protein Sci.* **2019**, *29*, 28–35. <https://doi.org/10.1002/pro.3711>.

18. Luo, W.; Brouwer, C. Pathview: An R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics* **2013**, *29*, 1830–1831.
19. Györfy, B.; Surowiak, P.; Budczies, J.; Lánczky, A. Online survival analysis software to assess the prognostic value of biomarkers using transcriptomic data in non-small-cell lung cancer. *PLoS ONE* **2013**, *8*, e82241.
20. Ye, J.; Coulouris, G.; Zaretskaya, I.; Cutcutache, I.; Rozen, S.; Madden, T.L. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinform.* **2012**, *13*, 134.
21. Kent, W.J.; Sugnet, C.W.; Furey, T.S.; Roskin, K.M.; Pringle, T.H.; Zahler, A.M.; Haussler, D. The human genome browser at UCSC. *Genome Res.* **2002**, *12*, 996–1006.
22. Kalendar, R.; Shevtsov, A.; Otarbay, Z.; Ismailova, A. In silico PCR analysis: A comprehensive bioinformatics tool for enhancing nucleic acid amplification assays. *Methods* **2024**, *4*, 1464197.
23. Tomoshige, K.; Guo, M.; Tsuchiya, T.; Fukazawa, T.; Fink-Balduf, I.M.; Stuart, W.D.; Naomoto, Y.; Nagayasu, T.; Maeda, Y. An EGFR ligand promotes EGFR-mutant but not KRAS-mutant lung cancer in vivo. *Oncogene* **2018**, *37*, 3894–3908.
24. Song, L.; Chen, Z.; Zhang, M.; Zhang, M.; Lu, X.; Li, C.; Miao, L. DDIT4 overexpression associates with poor prognosis in lung adenocarcinoma. *J. Cancer* **2021**, *12*, 6422–6428.
25. Livak, K.J.; Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* **2001**, *25*, 402–408.
26. Leonetti, A.; Sharma, S.; Minari, R.; Perego, P.; Giovannetti, E.; Tiseo, M. Resistance mechanisms to osimertinib in EGFR-mutated non-small cell lung cancer. *Br. J. Cancer* **2019**, *121*, 725–737.
27. Fruman, D.A.; Chiu, H.; Hopkins, B.D.; Bagrodia, S.; Cantley, L.C.; Abraham, R.T. The PI3K pathway in human disease. *Cell* **2017**, *170*, 605–635.
28. Vivanco, I.; Sawyers, C.L. The phosphatidylinositol 3-kinase–AKT pathway in human cancer. *Nat. Rev. Cancer* **2002**, *2*, 489–501.
29. Normanno, N.; De Luca, A.; Bianco, C.; Strizzi, L.; Mancino, M.; Maiello, M.R.; Carotenuto, A.; De Feo, G.; Caponigro, F.; Salomon, D.S. Epidermal growth factor receptor (EGFR) signaling in cancer. *Gene* **2006**, *366*, 2–16.
30. Ellisen, L.W. Growth control under stress: mTOR regulation through the REDD1–TSC pathway. *Cell Cycle* **2005**, *4*, 1500–1502.
31. Tripathi, S.K.; Biswal, B.K. SOX9 promotes epidermal growth factor receptor-tyrosine kinase inhibitor resistance via targeting β -catenin and epithelial-to-mesenchymal transition in lung cancer. *Life Sci.* **2021**, *277*, 119608.
32. Dongre, A.; Weinberg, R.A. New insights into the mechanisms of epithelial–mesenchymal transition and implications for cancer. *Nat. Rev. Mol. Cell Biol.* **2019**, *20*, 69–84.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.