Identification of LGR4 as a prognostic biomarker in KRAS-mutant lung adenocarcinoma
Evidence from integrated bioinformatics analysis

Yasmeen Dodin, MSC[a,*] (ID)

Abstract

Globally, lung cancer is the leading cause of cancer-related deaths, primarily non-small cell lung cancer. Kirsten Rat Sarcoma Oncogene Homolog (KRAS) mutations are common in non-small cell lung cancer and linked to a poor prognosis. Covalent inhibitors targeting KRAS-G12C mutation have improved treatment for some patients, but most KRAS-mutant lung adenocarcinoma (KRAS-MT LUAD) cases lack targeted therapies. This gap in treatment options underscores a significant challenge in the field. Our study aimed to identify hub/key genes specifically associated with KRAS-MT LUAD. These hub genes hold the potential to serve as therapeutic targets or biomarkers, providing insights into the pathogenesis and prognosis of lung cancer. We performed a comprehensive analysis on KRAS-MT LUAD samples using diverse data sources. This included TCGA project data for RNA-seq, clinical information, and somatic mutations, along with RNA-seq data for adjacent normal tissues. DESeq2 identified differentially expressed genes (DEGs), while weighted gene co-expression network analysis revealed co-expression modules. Overlapping genes between DEGs and co-expression module with the highest significance were analyzed using gene set enrichment analysis and protein-protein interaction network analysis. Hub genes were identified with the Maximal Clique Centrality algorithm in Cytoscape. Prognostic significance was assessed through survival analysis and validated using the GSE72094 dataset from Gene Expression Omnibus (GEO) database. In KRAS-MT LUAD, 3122 DEGs were found (2131 up-regulated, 985 down-regulated). The blue module, among 25 co-expression modules from weighted gene co-expression network analysis, had the strongest correlation. 804 genes overlapped between DEGs and the blue module. Among 20 hub genes in the blue module, leucine-rich repeats containing G protein-coupled receptor 4 (LGR4) overexpression correlated with worse overall survival. The prognostic significance of LGR4 was confirmed using GSE72094, but surprisingly, the direction of the association was opposite to what was expected. LGR4 stands as a promising biomarker in KRAS-MT LUAD prognosis. Contrasting associations in TCGA and GSE72094 datasets reveal the intricate nature of KRAS-MT LUAD. Additional explorations are imperative to grasp the precise involvement of LGR4 in lung adenocarcinoma prognosis, particularly concerning KRAS mutations. These insights could potentially pave the way for targeted therapeutic interventions, addressing the existing unmet demands in this specific subgroup.

Abbreviations: BP = biological process, CC = cellular component, DEGs = differentially expressed genes, EGFR = epidermal growth factor receptor, GO = gene ontology, KEGG = Kyoto Encyclopedia of Genes and Genomes, KRAS = Kirsten Rat Sarcoma Oncogene Homolog, KRAS-MT LUAD = KRAS-Mutant Lung Adenocarcinoma, LGR4 = leucine-rich Repeats Containing G Protein-coupled Receptor 4, NSCLC = non-small cell lung cancer, OS = overall survival, PPI = protein-protein interactions, TCGA-LUAD = The Cancer Genome Atlas Lung Adenocarcinoma, WGCNA = weighted gene co-expression analysis.

Keywords: bioinformatics, GEO, KRAS–mutant lung adenocarcinoma, LGR4, prognosis, TCGA, WGCNA

1. Introduction

The death rate from lung cancer is the highest among cancer related deaths worldwide, accounting for 1.8 million deaths in 2020.[1] Non-small cell lung cancers (NSCLCs) account for approximately 85% of all cases of lung cancer. Specifically, adenocarcinomas comprise around 40% of NSCLC cases.[2]

Within NSCLC, the most prevalent genetic alterations occur in the Kristen Rat Sarcoma Oncogene Homolog (KRAS) gene, which is found in ~ 20% to 30% of NSCLC cases.[3] Among these mutations, the substitution of Gly12 with Ala, Cys, Asp, or Val is most commonly observed. These alterations are strongly associated with tobacco smoking, resistance to

chemotherapy and radiation treatments, and overall poor prognosis.[3]

Ethnicity significantly influences the prevalence of KRAS mutations in NSCLC.[4] Notably, there is a lower prevalence of KRAS mutations in adenocarcinoma among Asian populations compared to non-Asian groups. For example, the prevalence is 5.7% in Chinese,[5] 9.3% in Japanese,[6] 17% in African-Americans,[7] 32% in Lebanese,[8] 9% in Korean,[5] 9% in Moroccan,[9] 5.2% in Taiwanese,[10] 36.9% in Dutch,[11] and 14.0% in Latin Americans.[12] Notably, an Iranian study assessing KRAS mutation distribution in 55 patients, 13 with adenocarcinoma as the histological subtype, reported a 0% prevalence.[13]

The KRAS protein, belonging to the small guanosine triphosphatase (GTPase) family, operates by transitioning between 2 states: the active form bound to GTP and the inactive form associated with GDP. When in the active GTP-bound state, KRAS activates several crucial cellular signaling pathways, including RAF-MEK-ERK, PI3K-AKTmTOR, and RALGDS-RA.[14,15] Oncogenic KRAS mutations hinder the GTPase ability to hydrolyze GTP, leaving KRAS in a constantly active state. This disruption results in uncontrolled cell division and enhanced cell survival, contributing to the development and progression of cancer.[16]

Previously, it was widely believed that KRAS, due to its high affinity for GTP and the absence of clearly defined binding pockets for allosteric inhibitors, was considered untargetable and undruggable. However, this perspective underwent a significant shift with the advent of 2 breakthrough covalent inhibitors, adagrasib and sotorasib, specifically designed to target the KRAS-G12C mutation. These inhibitors proved to be a game-changer in the field. Before the era of KRAS-G12C inhibitors, researchers explored various therapeutic strategies aimed at improving lung cancer outcomes. These approaches involved targeting KRAS upstream regulators and downstream effectors, hoping to find alternative ways to disrupt the aberrant signaling pathways associated with KRAS-driven cancers. However, the advent of adagrasib and sotorasib marked a significant milestone, offering a direct and specific approach to inhibit KRAS-G12C and potentially revolutionize the treatment landscape for patients with lung cancer.

Adagrasib and sotorasib have demonstrated their selective and irreversible binding to the cysteine at codon 12, specifically within the KRAS mutation. This binding effectively traps KRAS in an inactive state, bound to GDP.[17,18] It is important to note that, like other kinase inhibitors, the response to both adagrasib and sotorasib is not long-lasting due to the development of resistance.[19,20] It is worth mentioning that KRAS-G12C mutations are only found in approximately 40% of patients with KRAS mutations, leaving more than half of patients with KRAS mutations without a targeted treatment option.[21]

The rapid progress in genomics and high-throughput technologies has significantly expedited the advancement of targeted therapies, including KRAS-G12C inhibitors. These inhibitors have brought about a revolutionary shift in the treatment approach and overall outcomes for certain patients with lung cancer. However, it is crucial to note that the majority of lung cancer cases are diagnosed at advanced stages when surgical interventions are no longer viable, and the absence of approved targeted therapies as first-line treatments for KRAS-mutant metastatic NSCLC[22,23] further limits the available treatment options. To address these challenges and improve patient care, further genomic research is required to gain a deeper understanding of KRAS genetic alterations and identify key prognostic genes in KRAS-mutant lung adenocarcinoma (KRAS-MT LUAD) that can serve as therapeutic targets or assist in facilitating the routine diagnosis and treatment of lung cancer (predictive biomarkers).

Weighted gene co-expression analysis (WGCNA) is a valuable scientific technique used to identify co-expressed gene modules and important biomarkers.[24] This method allows for the identification of modules containing genes that exhibit high correlations. Furthermore, it enables the exploration of the relationship between these modules and various clinical traits. In our analysis, we aimed to identify hub genes within key co-expression gene network modules specifically associated with KRAS-MT LUAD. These hub genes hold the potential to serve as therapeutic targets or biomarkers, providing insights into the pathogenesis and prognosis of lung cancer.

2. Materials and methods

Figure 1 presents a broad outline of the current analysis workflow. Further elaborations on the specific analyses can be found in the subsequent sections.

2.1. RNA-seq data, clinical, and somatic mutation information

The KRAS-MT LUAD RNA-seq data, along with corresponding clinicopathological features and somatic mutation information, were retrieved from the Cancer Genome Atlas Lung Adenocarcinoma (TCGA-LUAD) project.[25] The retrieval process utilized the TCGAbiolinks R/Bioconductor package,[26] as well as the Maftools Bioconductor package.[27] For the TCGAbiolinks package, the retrieval conditions were similar to what specified in our previous work,[28] with the exception of *sample.type* (Primary Tumor and Solid Tissue Normal were used instead of Primary Tumor) and ajcc_pathologic_stage (all stages were included instead of advanced stages). A total of 134 KRAS-MT LUAD samples and 55 normal samples were retrieved for subsequent analyses. KRAS mutations were present in 24.4% of cases within the TCGA-LUAD project, comprising 134 mutated samples out of 549 primary tumor samples.[25]

2.2. Differentially expressed genes (DEGs) screening

The DESeq2 package in R was employed to identify the DEGs between KRAS-MT LUAD and normal tissues, using raw read counts.[29] To mitigate potential bias arising from genes with low expression counts, genes with values below ten were excluded from the analysis. The criteria for identifying DEGs included an absolute log2 fold change (|log2 FC|) of ≥ 2 and an adjusted *P* value of ≤ .05. The findings were visualized through an MA plot (created using the plotMA function in DESeq2), a volcano plot, and a heatmap (generated using the ggplot2 package in R).[30]

2.3. Identification of KRAS-MT LUAD related modules with WGCNA

In this study, the TCGA-LUAD gene expression data, consisting of KRAS-MT samples and normal tissue samples, were utilized to construct a weighted co-expression network. The construction of the network was performed using the "WGCNA" package in R.[24] To ensure the reliability of the network structure, 7 outlier samples from the KRAS-MT LUAD dataset were excluded from the analysis (Fig. 2). To achieve a scale-free topology and enhance the correlation adjacency matrix, a soft-thresholding power of β = 8 (R^2 = 0.89) was chosen. The expression matrix was transformed into an adjacency matrix, employing Pearson correlation coefficients to estimate the coexpression relationships between gene pairs. Subsequently, the adjacency matrix was converted into a
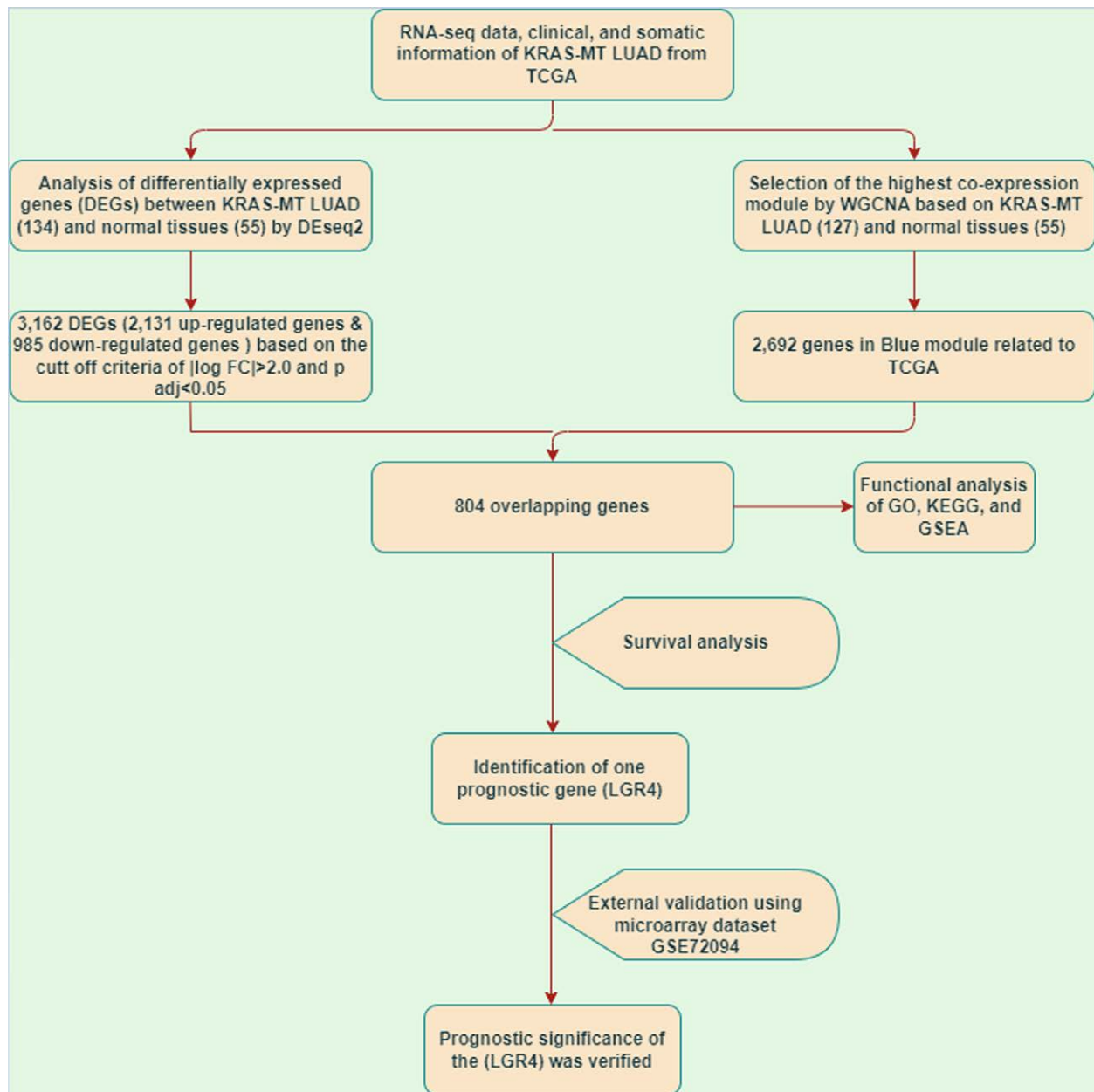
**Figure 1.** Broad outline of the current analysis workflow.

topological overlap matrix. The DynamicTreeCut algorithm was applied, using topological overlap matrix dissimilarity, to obtain the initial set of modules. Correlated modules (with a correlation coefficient of $R \geq 0.75$) were merged to form similar modules. Module eigengenes, representing the first principal components, were calculated and considered representative of the respective module genes. The Spearman correlation test was employed to assess the correlation between each eigengene and the clinical trait of interest (KRAS-MT LUAD). Further analysis focused on the genes within the most relevant KRAS-MT LUAD module.

2.4. Intersection of DEGs and WGCNA

In the subsequent steps, we focused on utilizing the most pertinent module and genes associated with KRAS-MT LUAD. These selections were made by employing WGCNA and identifying the genes that overlapped with the DEGs.

2.5. Application of gene set enrichment analysis, and construction of protein-protein interactions (PPI)

The overlapping/intersecting genes obtained from the analysis were subjected to gene ontology (GO) enrichment analysis, which included molecular function, cellular component (CC), and biological process (BP), as well as Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis. This analysis was conducted using the "clusterProfiler" package in R[31] using a minimum gene set of 20 and a P value of .05. Graphical representations of the significantly enriched BP, CC, molecular function, and KEGG pathways were generated using the ggplot2 package in R, focusing on the top 5 results.[30] The intersecting genes obtained from the analysis were subjected to further analysis using the Search Tool for the Retrieval of Interacting Genes online tool.[32] This analysis aimed to construct a PPI network with a cutoff criterion of a combined score ≥ 0.9, indicating high confidence in the interactions. The resulting PPI network was visualized using Cytoscape software
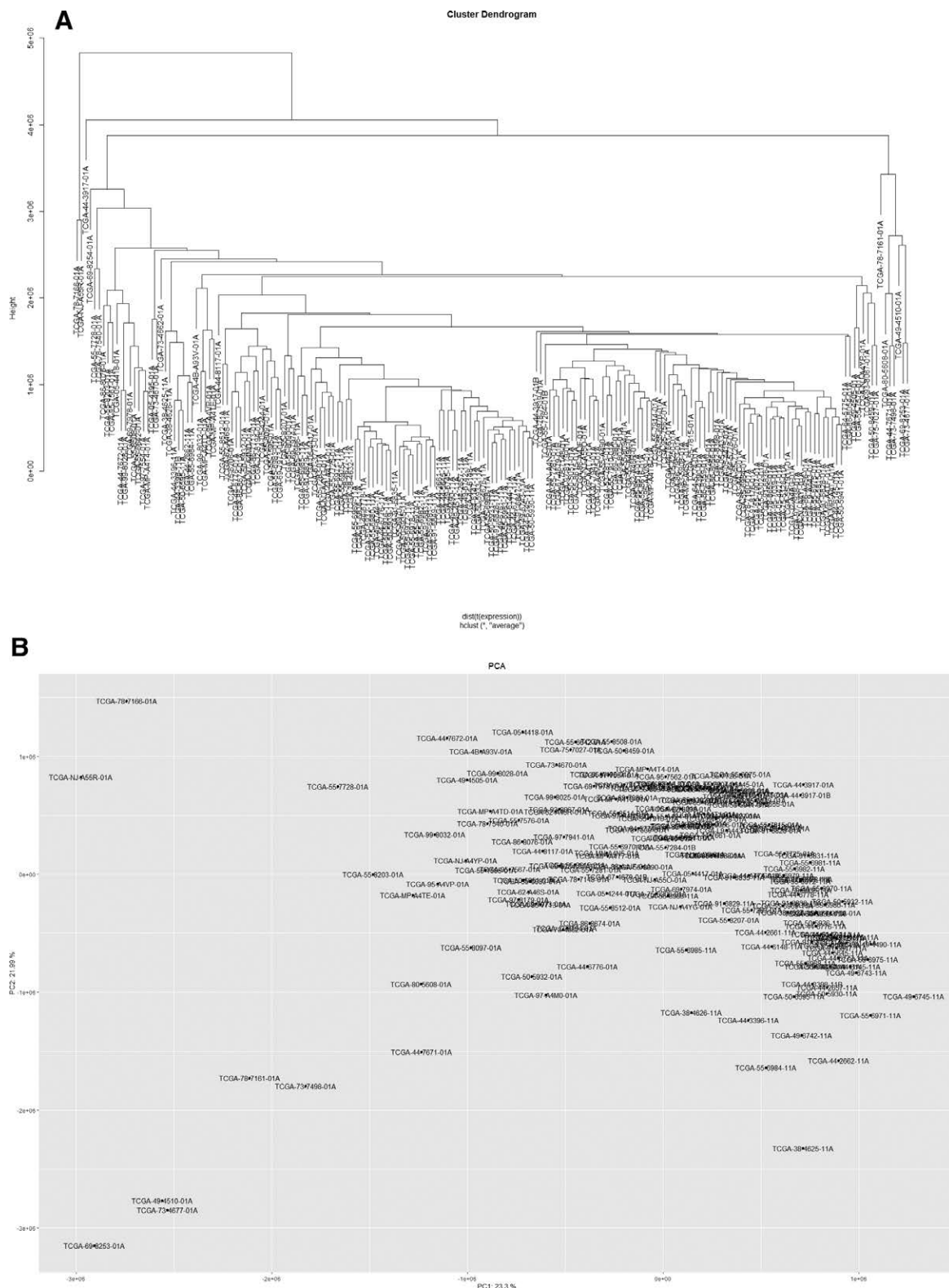
Figure 2. Hierarchical clustering method (A) and principal component analysis method (B) for detecting outlier samples.

(v3.9.1).[33]To identify the most important genes within the co-expression network, the Maximal Clique Centrality algorithm provided by the cytoHubba plugin in Cytoscape was applied.[34] This analysis allowed for the identification of the top 20 hub genes in the co-expression PPI network, which play significant roles in the network based on their connectivity and centrality.

2.6. Hub genes survival analysis

To investigate the potential association of these hub genes with overall survival (OS) in KRAS-MT LUAD patients, the expression levels of each hub gene were categorized into 2 groups: patients with expression levels greater than or equal to the median, and patients with expression levels smaller than the median. To assess the survival differences between the high
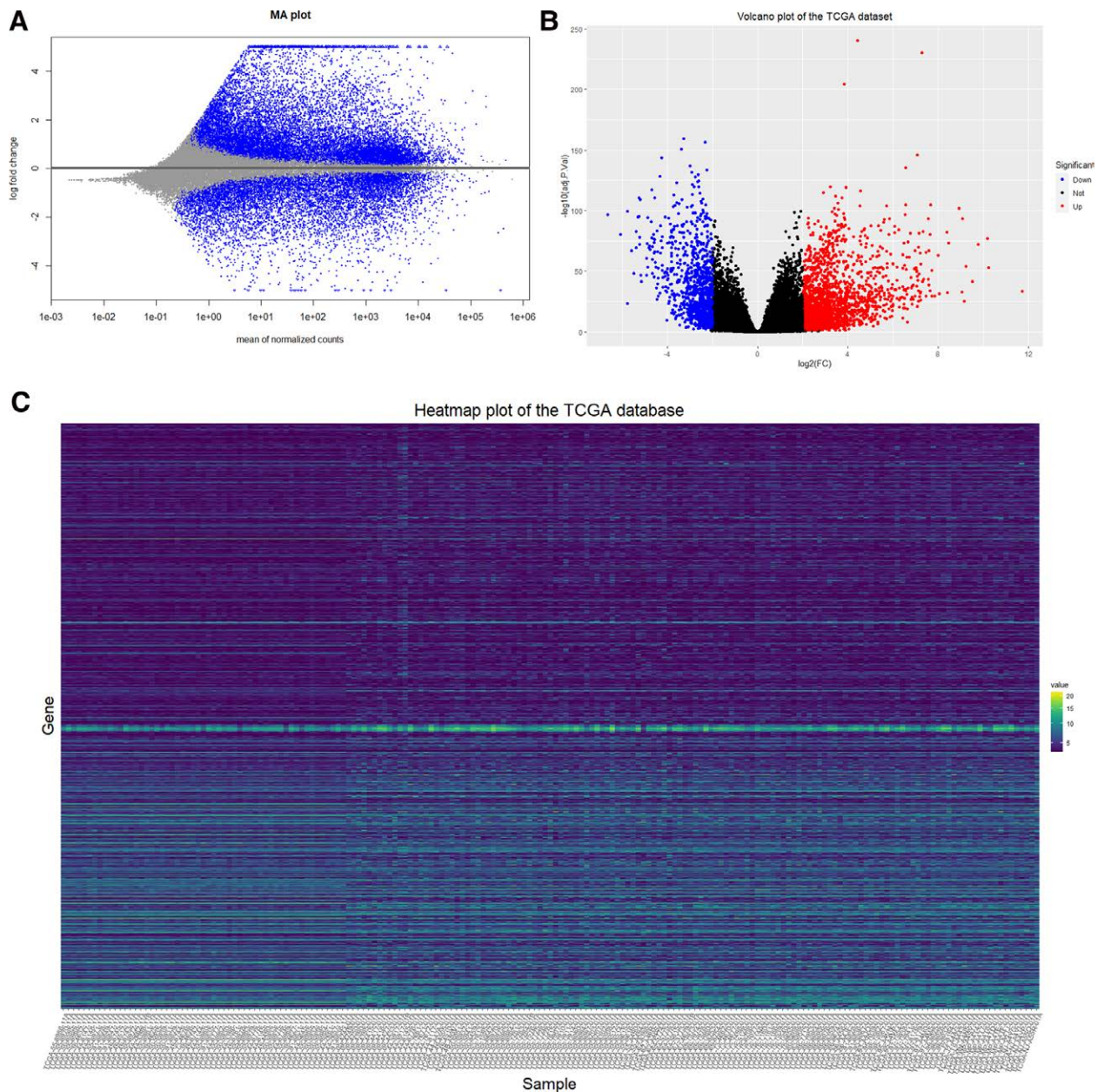
Figure 3. Differentially expressed genes (DEGs) in KRAS-MT LUAD. MA plot (A), volcano plot (B), heatmap (C). KRAS-MT LUAD = Kirsten Rat Sarcoma Oncogene Homolog-mutant lung adenocarcinoma.

and low expression groups, the R packages "survival" (v3.5–5) and "survminer" (v0.4.9) were utilized. Kaplan–Meier estimates were used to analyze the survival probabilities, and the log-rank test was employed to determine the statistical significance. A *P* value <.05 was considered indicative of a significant difference in survival between the high and low expression groups.

2.7. External validation of survival-related hub genes

For external validation, we used a microarray dataset (GSE72094) to verify survival significance of the survival related hub gene. The GSE72094 dataset enrolled a total of 442 patients, all of whom had comprehensive mRNA expression data and underwent Sanger sequencing analysis to investigate the presence of epidermal growth factor receptor (EGFR), KRAS, TP53, and STK11 genes.[35] For the survival analysis we only included KRAS positive samples (N = 139).

3. Results

3.1. TCGA gene expression data

A comprehensive analysis was conducted on a total of 182 samples, consisting of 127 KRAS-MT LUAD samples and 55 adjacent normal tissues, to identify the DEGs. Among the screened genes, a total of 3162 DEGs met the criteria for selection, with 2131 genes showing up-regulation and 985 genes showing downregulation. The DEGs were visually represented through the MA plot (Fig. 3A), volcano plot (Fig. 3B), and heatmap (Fig. 3C).

3.2. Weighted co-expression network and their key modules

By employing a soft-thresholding power β value of 8, the fit index curve for the scale-free topology flattened out at 0.89 (Fig. 4). Subsequently, a total of 25 co-expression modules were
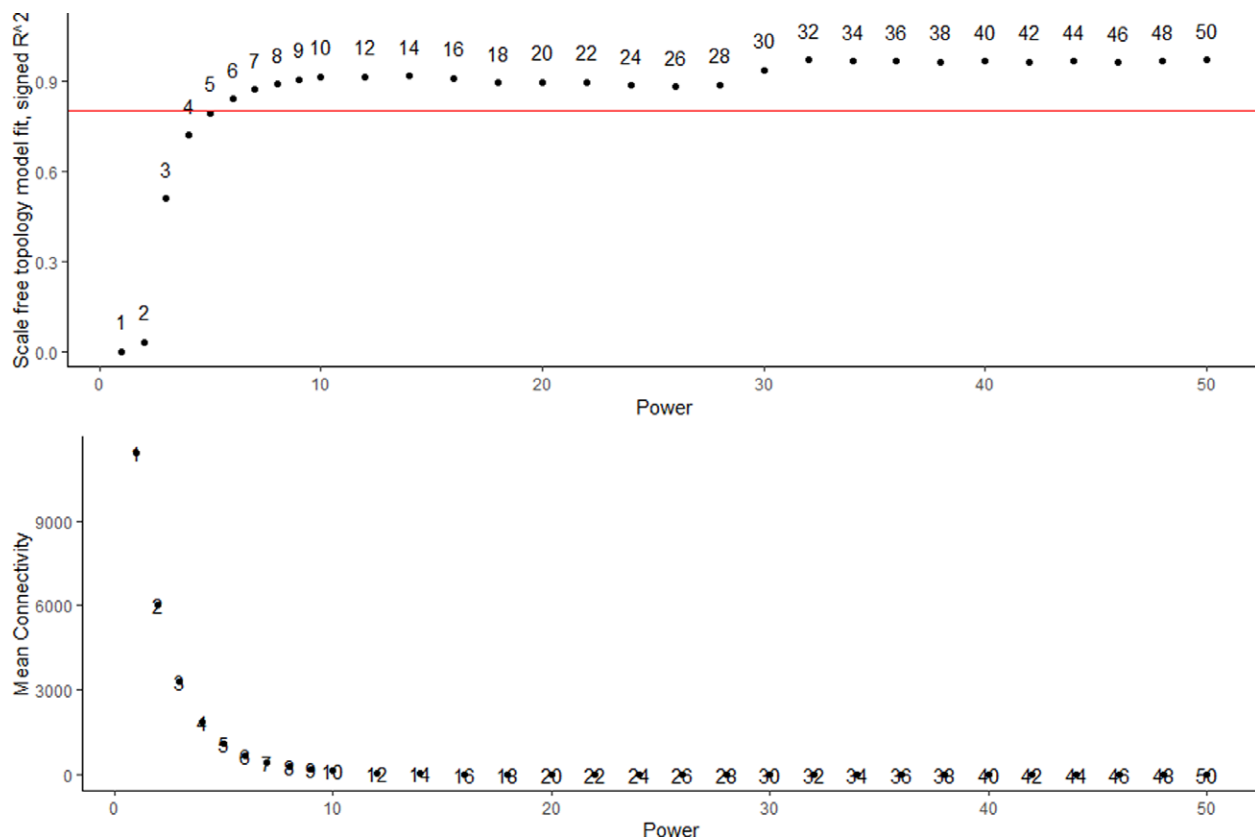
Figure 4. Analysis of the scale-free fit index and the mean connectivity of various soft-thresholding powers.

identified within the constructed weighted gene co-expression network of KRAS-MT LUAD. These modules were obtained through the process of merging correlated modules using average linkage clustering (Fig. 5).

The constructed weighted gene co-expression network in KRAS-MT LUAD comprised various modules, each represented by a distinct color such as black, light yellow, royal blue, yellow, blue, midnight blue, salmon, tan, magenta, green yellow, gray 60, dark green, pink, light green, brown, purple, light cyan, dark turquoise, green, red, turquoise, dark red, cyan, dark gray, and gray, including 21–11,221 genes. The gray module specifically consisted of genes that did not belong to any other module. Through correlation analysis between module eigengenes and KRAS-MT LUAD, it was observed that 16 modules exhibited an association with KRAS-MT LUAD (Fig. 6).

Among the identified modules in KRAS-MT LUAD, the blue module stood out as the most significant and representative. It consisted of 2692 genes, with a correlation ($R^2 = 0.40$) and statistical significance ($P = 1.227674e-85$). Due to its prominence, the blue module was chosen for further analysis.

3.3. Intersection of DEGs and WGCNA

A total of 804 genes within the blue module overlapped with the set of DEGs.

3.4. Application of gene set enrichment analysis

Functional enrichment analyses using GO and KEGG were performed on the set of 804 genes. GO enrichment analysis revealed significant enrichment in 1 BP and 5 CCs. KEGG enrichment analysis identified 1 significantly enriched pathway.

The top 5 enriched GO terms and KEGG pathways are displayed in Figure 7.

3.5. Hub gene identification

With the assistance of the Search Tool for the Retrieval of Interacting Genes database, a PPI network was constructed from the DEGs within the blue module. The resulting PPI network contained 804 nodes and 184 edges. To identify hub genes within the blue module, we employed the Maximal Clique Centrality algorithm through the cytoHubba plugin. The top 20 genes identified as hub genes within the blue module were ADAM Metallopeptidase With Thrombospondin Type 1 Motif 14, ADAMTS like 3, Somatomedin B And Thrombospondin Type 1 Domain Containing, Semaphorin 5A, ADAM Metallopeptidase With Thrombospondin Type 1 Motif 8, Endothelin 1, Adrenoceptor Beta 2, Thrombospondin Type 1 Domain Containing 1, Angiotensin II Receptor Type 2, Cholinergic Receptor Muscarinic 1, Wnt Family Member 3A, R-Spondin 1, R-Spondin 2, R-Spondin 4, Leucine-rich Repeats Containing G protein-coupled receptor 4 (LGR4), ADAM Metallopeptidase With Thrombospondin Type 1 Motif 1, Fibroblast Growth Factor Receptor 2, Receptor Activity Modifying Protein 2, Receptor Activity Modifying Protein 3 (RAMP3), and Calcitonin Receptor Like Receptor. These hub genes are depicted in Figure 8.

3.6. Hub genes survival analysis

The analysis of the 20 hub genes within the blue module revealed a significant correlation between the increased expression of 1 gene, LGR4, and a lower OS rate in KRAS-MT LUAD patients ($P = .012$). This finding is illustrated in Figure 9.
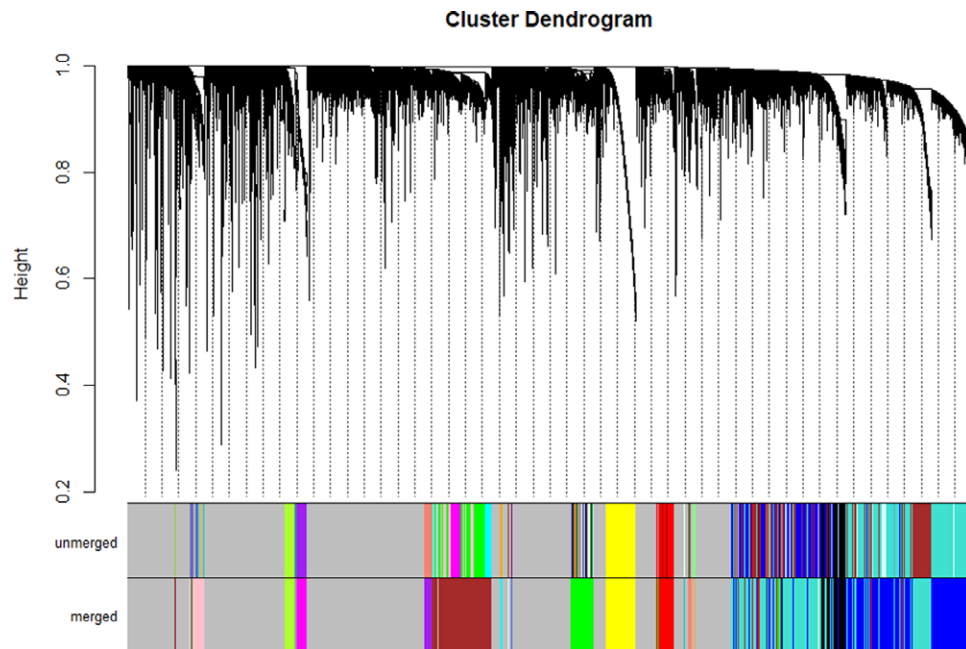
Figure 5. Hierarchical clustering dendrograms of identified co-expressed genes in KRAS-MT LUAD. KRAS-MT LUAD = Kirsten Rat Sarcoma Oncogene Homolog-mutant lung adenocarcinoma.
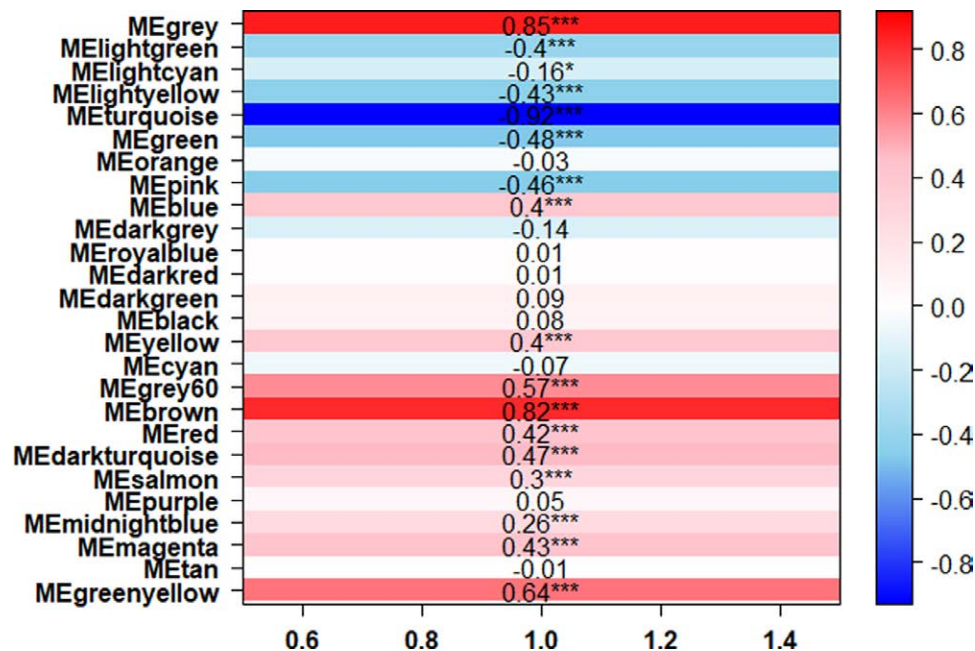
Figure 6. Heatmap of the correlation between Eigengene and the clinical treat of KRAS-MT LUAD. KRAS-MT LUAD = Kirsten Rat Sarcoma Oncogene Homolog-mutant lung adenocarcinoma.

3.7. External validation of the survival-related hub genes

In our study, we conducted survival analysis using RNA-seq data from KRAS-MT LUAD tissues. Our analysis identified a specific up-regulated DEG that overlapped with a tumor-related module known as the blue module. Remarkably, this gene exhibited a strong correlation with poor survival outcomes in LUAD patients. To validate the prognostic significance of this survival-related hub gene (LGR4), we sought external validation using a separate microarray dataset, GSE72094. The validation process involved performing survival analysis in a manner consistent with our previous approach using TCGA data. The results obtained from this external dataset are presented in Figure 10, further supporting the survival significance

of LGR4 as observed in our initial analysis, but with the opposite direction (low expression levels of LGR4 were associated with poor survival).

4. Discussion

Over recent decades, the molecular analysis of lung cancer patients has unveiled a multitude of mutations that exhibit correlations with NSCLC.[25,36] Approximately 60% of adenocarcinoma cases feature a driver mutation that serves as the catalyst for the oncogenic process. The most common driver mutations encompass KRAS (25%), EGFR at 23%, anaplastic lymphoma kinase at 6%, Phosphoinositide-3-Kinase, Catalytic, Alpha Polypeptide at 3%, v-Raf Murine Sarcoma Viral Oncogene

A

B

C
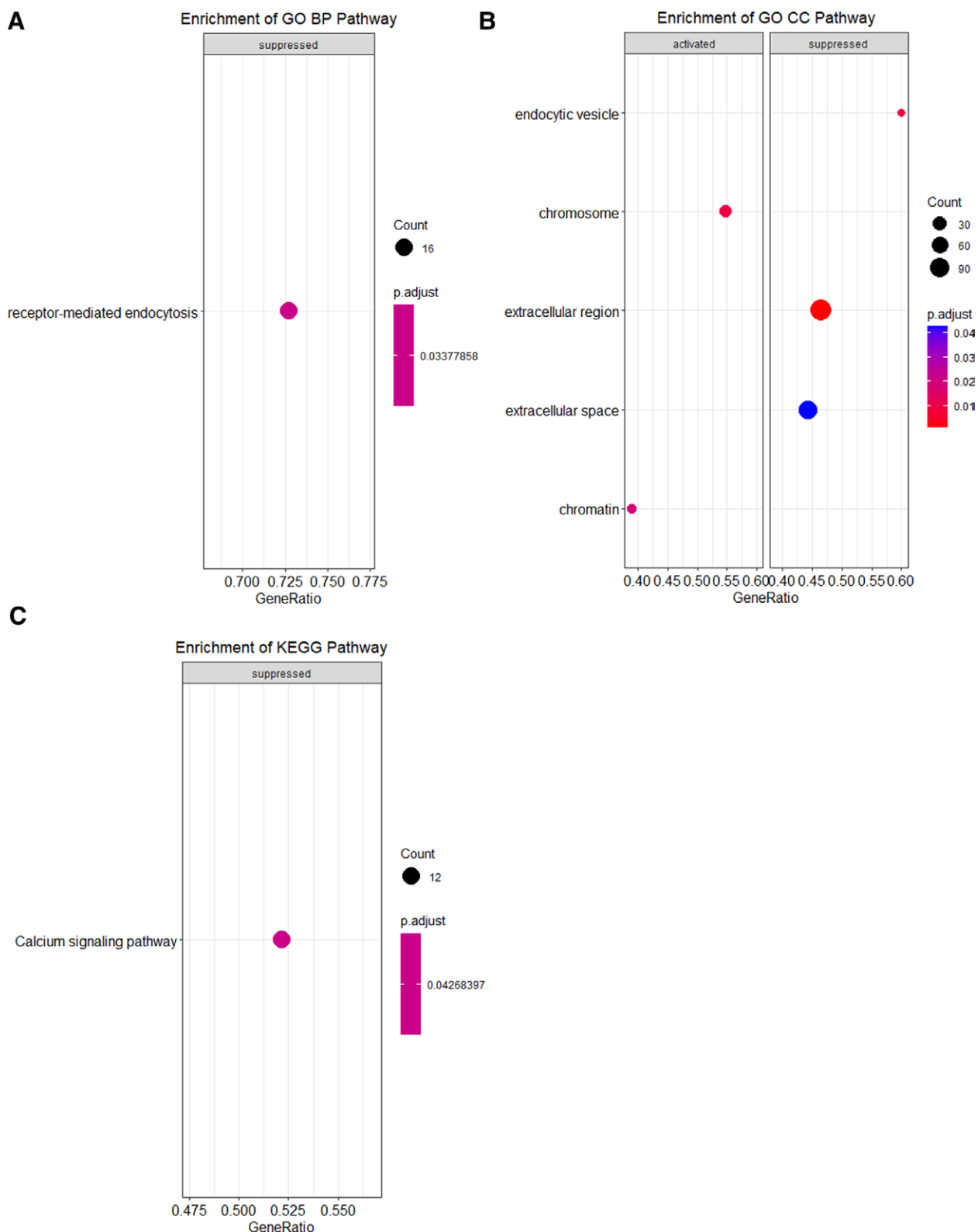
Figure 7. Gene set enrichment analysis (GSEA) for the DEGs intersecting with the blue module (*N* = 804). (A) GO biological process (BP); (B) GO cellular component and (C) KEGG pathway analysis. No molecular function (MF) was significantly enriched under the specified conditions. Spot sizes represent the number of genes, while color indicates adjusted *P* values. DEGs = differentially expressed genes, GO = gene ontology, KEGG = Kyoto Encyclopedia of Genes and Genomes.

Homolog B1 at 3%, and Human Epidermal Growth Factor Receptor 2 at 1%.[37,38] Targeted therapies have demonstrated success against EGFR mutations, anaplastic lymphoma kinase rearrangements, and KRAS-G12C mutations. Nevertheless, persistent challenges in this field revolve around disease progression, the emergence of drug resistance, and late-stage lung cancer diagnosis. Identifying novel biomarkers becomes imperative for early disease detection, prognostic assessment, and
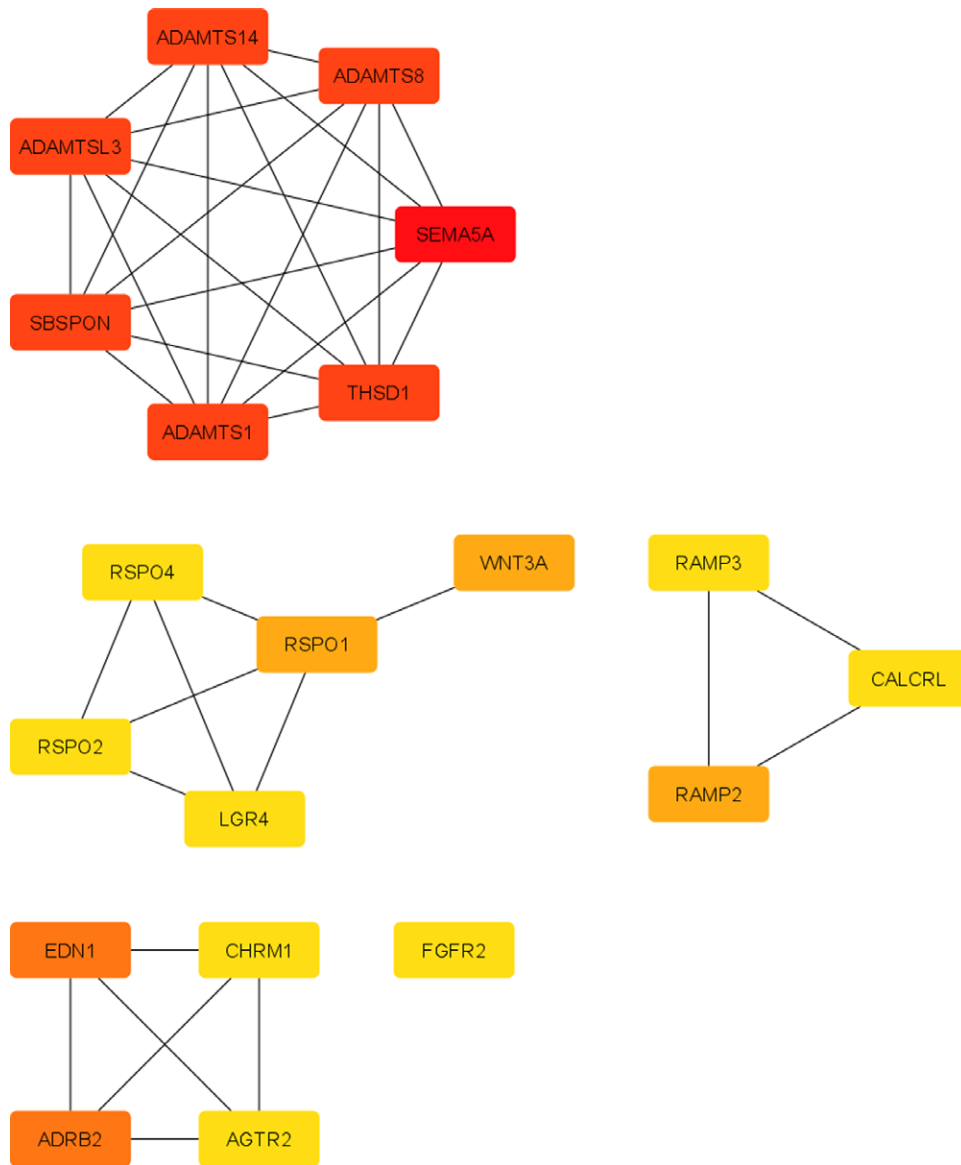
Figure 8. PPI analysis and identification of hub genes involved in the DEGs intersecting with the blue module using. STRING database and cytoHubba plug-in Cytoscape. DEGs = differentially expressed genes, PPI = protein-protein interaction, STRING = Search Tool for the Retrieval of Interacting Genes.

stratifying patients who would best respond to diverse targeted therapies.

Several research studies have employed computational methodologies to detect genetic alterations in LUAD patients, aiming to uncover predictive, diagnostic, or prognostic biomarkers. These endeavors led to the identifications of different hub genes that can serve as potential biomarkers including Angiopoietin 1[39]; Secreted Phosphoprotein 1[40]; Aurora Kinase A, DNA Topoisomerase II Alpha and Maternal Embryonic Leucine Zipper Kinase[41]; Ubiquitin Conjugating Enzyme E2 C[42]; HLF Transcription Factor, Chordin Like 1, Selenium Binding Protein 1, and Transmembrane Protein 163[43]; GINS Complex Subunit 2.[44] These investigations primarily focused on gene-level alterations and often overlooked the comprehensive molecular profiles of lung cancer patients. While they did offer valuable insights into the potential roles of biomarkers within LUAD broadly, we assert that comprehensively assessing the roles of these biomarkers necessitates an in-depth understanding of the specific mutational landscape.

Limited research efforts have been dedicated to the identification of novel biomarkers in LUAD that are specifically linked

to driver mutations, potentially serving as therapeutic targets or offering diagnostic and prognostic value. For instance, Zhang and colleagues discovered a significant correlation between the elevated expression of Beta-1,3-N-acetylglucosaminyltransferase 3 (B3GNT3) and poorer OS in EGFR-mutant LUAD patients.[45] Regarding KRAS-mutant LUAD patients, Dai et al identified 7 hub genes from an analysis of 184 stage IIB to IV LUAD samples.[46]

In this study, we aimed to identify potential prognostic biomarkers in KRAS-MT LUAD using WGCNA. Our analysis revealed the blue module as the most significant and representative module associated with in KRAS-MT LUAD, consisting of 2692 genes. We further explored the intersection of DEGs caused by KRAS mutation and the blue module, leading to the identification of 804 overlapping genes. Functional enrichment analysis using GO and KEGG pathways provided insights into the potential biological functions of these intersecting genes. In our study, we focused on identifying hub genes within the DEGs located in the blue module of in KRAS-MT LUAD. Hub genes are highly connected genes within a network and are often critical in maintaining network integrity and functionality. The top
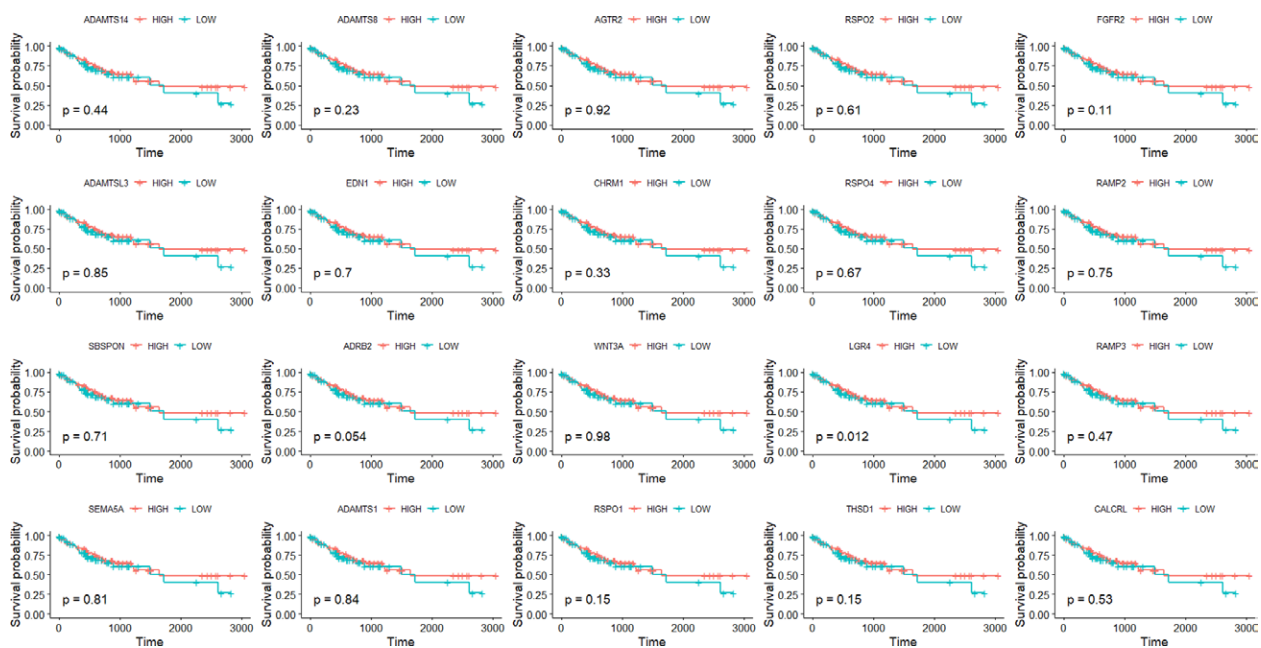
Figure 9. Kaplan–Meier survival of hub genes. The patients were categorized into high-level (red) and low-level (green) groups based on the gene median expression.
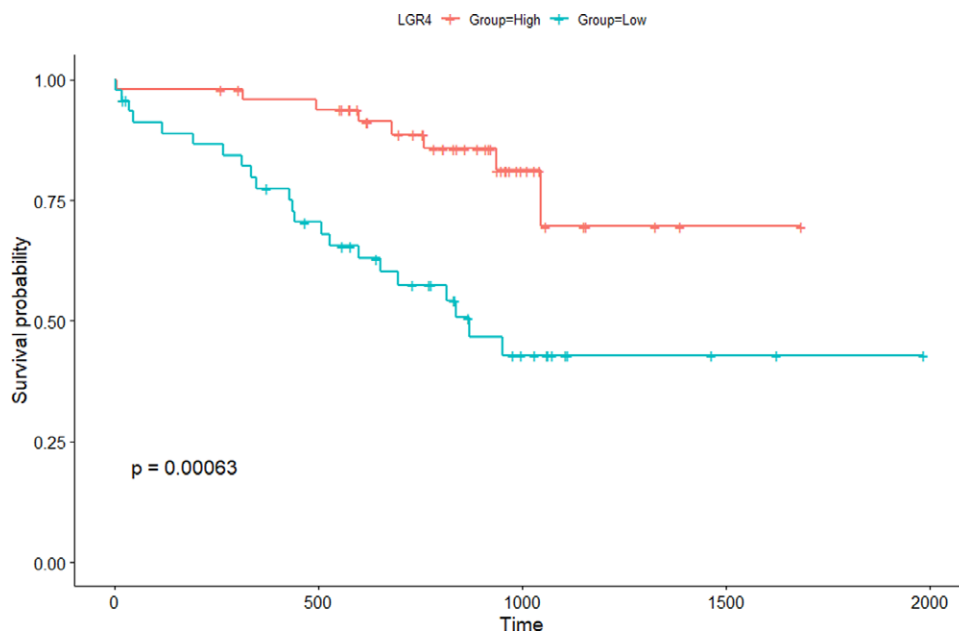
Figure 10. Kaplan–Meier survival of survival-related hub genes. The Patients were categorized into high-level (red) and low-level (green) groups based on the gene median expression.

20 hub genes identified in our analysis may hold potential as therapeutic targets or biomarkers in in KRAS-MT LUAD. To evaluate the prognostic significance of the identified hub genes, we performed survival analysis and categorized patients into high and low expression groups based on the median expression levels of each hub gene. The survival analysis indicated a potential correlation between the expression of LGR4 and the OS of patients with in KRAS-MT LUAD, where elevated levels of LGR4 were associated with poor survival outcomes. LGR4, also referred to as protein-coupled receptor 87, belongs to the group B of the LGR family, which includes LGR4, LGR5, and LGR6 receptors. These transmembrane receptors are members

of the superfamily of G protein-coupled receptors, play significant roles in developmental processes, and have implications in various cancer types.[47]

To validate our findings, we performed an external validation using a microarray dataset (GSE72094) of KRAS-MT LUAD. This validation aimed to assess the survival significance of the identified survival-related hub gene. Our analysis of the microarray dataset GSE72094 has uncovered a perplexing discrepancy that challenges the prevailing understanding of LGR4 role in LUAD prognosis. Few studies have focused on the prognostic value of LGR4, while the majority of these studies and analyses,[48–51] including our own TCGA analysis, have consistently

associated higher expression levels of LGR4 with worse OS in LUAD patients, our validation results based on the microarray dataset revealed a strikingly contrasting pattern. Surprisingly, we observed that lower expression levels of LGR were correlated with poor OS, in direct opposition to the established consensus. Notably, an earlier study by Rao et al[52] also presented similar findings, providing further support for our validation results. They demonstrated that in KRAS lung cancer, the gene expression patterns associated with an active receptor activator of nuclear factor-kB/receptor activator of nuclear factor-κB ligand system are linked to worse OS. LGR4, which acts as a membrane-bound negative regulator by countering receptor activator of nuclear factor-kB activation, is a receptor for receptor activator of nuclear factor-κB ligand.[53] This observation could provide an explanation for Rao et al's discovery that lower mRNA expression of LGR4 is linked to an unfavorable outcome.

The existence of such divergent findings raises intriguing questions about the underlying mechanisms driving LGR4 impact on LUAD prognosis. One possible explanation could be the inherent heterogeneity within LUAD, as different subtypes or molecular characteristics might respond differently to LGR4 expression. Moreover, KRAS mutations occurring within the same codon may exhibit non-identical biological effects, leading to variations in progression-free survival rates, changes in tumor gene expression, and differences in drug responsiveness.[54] Additionally, variations in the sample size, patient demographics, treatment regimens, and follow-up protocols across different studies may contribute to the observed disparities. Furthermore, it is crucial to consider the complexities of LGR4 signaling pathways and its interactions with other molecular players within the tumor microenvironment.

These contrasting outcomes highlight the need for further comprehensive investigations to elucidate the factors that determine LGR4 role in the prognosis of LUAD as a whole, and specifically in KRAS-mutant LUAD. Future functional studies should be conducted to elucidate the biological mechanisms underlying the association between LGR4 expression and survival in KRAS-MT LUAD. This could involve in vitro and in vivo experiments, such as cell line models, animal models, and molecular assays, to investigate the impact of LGR4 on tumor growth, metastasis, and response to therapy. Moreover, the association of LGR4 with clinical variables, such as stage, tumor size, lymph node involvement, and treatment response, should be assessed. This will help determine whether LGR4 expression is an independent prognostic factor or if its association with survival is confounded by other clinical factors. By gaining a deeper understanding of these complexities, we can pave the way for personalized therapeutic approaches and refine prognostic models in the context of KRAS-MT LUAD.

Like any research endeavor, our study does come with inherent limitations that warrant consideration. While we diligently conducted an extensive bioinformatics analysis to uncover potential prognostic genes in KRAS-MT LUAD in comparison to adjacent samples, it crucial to acknowledge that our approach was retrospective and built upon publicly available datasets. Consequently, it prudent to approach our study findings with caution, recognizing the need for validation through prospective investigations and experimental validations. It also important to underscore that a more comprehensive set of experimental data is necessary to comprehensively explore the underlying mechanisms and interactions of the identified genes. This underscores the necessity for future research endeavors to delve deeper into the intricate molecular aspects and functional significance of these genes in KRAS-MT LUAD. Additionally, it is vital to recognize that our initial assessment relies on data from the TCGA-LUAD project, which encompasses a comprehensive collection of genomic studies conducted through cross-sectional analysis, utilizing NGS and whole-exome sequencing techniques.

Within this dataset, approximately 70% of the patients were of Caucasian ethnicity, while only 2% were of Asian descent.[15] Consequently, it is of paramount importance to consider the variances in ethnic backgrounds when interpreting our results and when seeking to apply these findings to diverse demographic groups.

5. Conclusion

In conclusion, our study identified LGR4 as a potential prognostic biomarker in KRAS-MT LUAD using WGCNA. The differences in how LGR4 expression and survival are related in the TCGA and GSE72094 datasets emphasize the intricate nature of KRAS-MT LUAD. Further investigations are warranted to elucidate the precise role of LGR4 in lung adenocarcinoma prognosis, particularly in the context of KRAS mutations. These findings contribute to our understanding of KRAS-MT LUAD and provide insights into potential biomarkers for patient prognosis and personalized treatment strategies.

Author contributions

Conceptualization: Yasmeen Dodin.
Data curation: Yasmeen Dodin.
Formal analysis: Yasmeen Dodin.
Validation: Yasmeen Dodin.
Visualization: Yasmeen Dodin.
Writing – original draft: Yasmeen Dodin.
Writing – review & editing: Yasmeen Dodin.

References

[1] Ferlay J, Colombet M, Soerjomataram I, et al. Cancer statistics for the year 2020: an overview. Int J Cancer. 2021;149:778–89.

[2] Travis WD, Brambilla E, Burke AP, et al. Introduction to the 2015 World Health Organization classification of tumors of the lung, pleura, thymus, and heart. J Thorac Oncol. 2015;10:1240–2.

[3] Aviel-Ronen S, Blackhall FH, Shepherd FA, et al. K-ras mutations in non-small-cell lung carcinoma: a review. Clin Lung Cancer. 2006;8:30–8.

[4] Mehdi I, Abdulmonem E, Al Bahrani BJ. KRAS mutations: does ethnicity play a role? J Clin Oncol. 2014;32(15_suppl):e14628.

[5] Dearden S, Stevens J, Wu YL, et al. Mutation incidence and coincidence in non small-cell lung cancer: meta-analyses by ethnicity and histology (mutMap). Ann Oncol. 2013;24:2371–6.

[6] Izumi M, Suzumura T, Ogawa K, et al. Differences in molecular epidemiology of lung cancer among ethnicities (Asian vs Caucasian). J Thorac Dis. 2020;12:3776–84.

[7] Reinersman JM, Johnson ML, Riely GJ, et al. Frequency of EGFR and KRAS mutations in lung adenocarcinomas in African Americans. J Thorac Oncol. 2011;6:28–31.

[8] Fakhruddin N, Mahfouz R, Farhat F, et al. Epidermal growth factor receptor and KRAS mutations in lung adenocarcinoma: a retrospective study of the Lebanese population. Oncol Rep. 2014;32:2223–9.

[9] Elghissassi I, Inrhaoun H, Boukir A, et al. Frequency and spectrum of KRAS mutations in Moroccan patients with lung adenocarcinoma. ISRN Oncol. 2014;2014:192493.

[10] Luo YH, Liang KH, Huang HC, et al. State-of-the-art molecular oncology of lung cancer in Taiwan. Int J Mol Sci. 2022;23:7037.

[11] Smits AJ, Kummer JA, Hinrichs JW, et al. EGFR and KRAS mutations in lung carcinomas in the Dutch population: increased EGFR mutation frequency in malignant pleural effusion of lung adenocarcinoma. Cell Oncol (Dordr). 2012;35:189–96.

[12] Arrieta O, Cardona AF, Martin C, et al. Updated frequency of EGFR and KRAS mutations in nonsmall-cell lung cancer in Latin

America: the Latin-American Consortium for the Investigation of Lung Cancer (CLICaP). J Thorac Oncol. 2015;10:838–43.

[13] Fathi Z, Mousavi SAJ, Roudi R, et al. Distribution of KRAS, DDR2, and TP53 gene mutations in lung cancer: an analysis of Iranian patients. PLoS One. 2018;13:e0200633.

[14] McCormick F. KRAS as a therapeutic target. Clin Cancer Res. 2015;21:1797–801.

[15] Bodemann BO, White MA. Ral GTPases and cancer: linchpin support of the tumorigenic platform. Nat Rev Cancer. 2008;8:133–40.

[16] Yang H, Liang SQ, Schmid RA, et al. New horizons in KRAS-mutant lung cancer: dawn after darkness. Front Oncol. 2019;9:953.

[17] Fell JB, Fischer JP, Baer BR, et al. Identification of the clinical development candidate MRTX849, a covalent KRAS (G12C) inhibitor for the treatment of cancer. J Med Chem. 2020;63:6679–93.

[18] Skoulidis F, Li BT, Dy GK, et al. Sotorasib for lung cancers with KRAS pG12C mutation. N Engl J Med. 2021;384:2371–81.

[19] Hong DS, Fakih MG, Strickler JH, et al. KRAS(G12C) Inhibition with sotorasib in advanced solid tumors. N Engl J Med. 2020;383:1207–17.

[20] Akhave NS, Biter AB, Hong DS. Mechanisms of resistance to KRAS (G12C)-targeted therapy. Cancer Discov. 2021;11:1345–52.

[21] Biernacka A, Tsongalis PD, Peterson JD, et al. The potential utility of re-mining results of somatic mutation testing: KRAS status in lung adenocarcinoma. Cancer Genet. 2016;209:195–8.

[22] NCCN. NCCN guidelines for Patients® non-small cell lung cancer-metastatic. 2021. Available at: http://www.nccn.org/patients.

[23] Planchard D, Popat S, Kerr K, et al. Metastatic non-small cell lung cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2018;29(Suppl 4):iv192–237.

[24] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinf. 2008;9:559.

[25] Cancer Genome Atlas Research N. Comprehensive molecular profiling of lung adenocarcinoma. Nature. 2014;511:543–50.

[26] Colaprico A, Silva TC, Olsen C, et al. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Res. 2016;44:e71.

[27] Mayakonda A, Lin DC, Assenov Y, et al. Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res. 2018;28:1747–56.

[28] Dodin Y. Integrated bioinformatics approach for disclosing autophagy pathway as a therapeutic target in advanced KRAS mutated/positive lung adenocarcinoma. Open Bioinformatics J. 2023;16:e187503622304100.

[29] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.

[30] Wickham H. Ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag. ISBN 978-3-319-24277-42016.

[31] Yu G, Wang LG, Han Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16:284–7.

[32] Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019;47(D1):D607–13.

[33] Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13:2498–504.

[34] Chin C-H, Chen S-H, Wu H-H, et al. cytoHubba: identifying hub objects and sub-networks from complex interactome. BMC Syst Biol. 2014;8(S4):S11.

[35] Schabath MB, Welsh EA, Fulp WJ, et al. Differential association of STK11 and TP53 with KRAS mutation-associated gene expression,

proliferation and immune surveillance in lung adenocarcinoma. Oncogene. 2016;35:3209–16.

[36] Wigle DA, Jurisica I, Radulovich N, et al. Molecular profiling of non-small cell lung cancer and correlation with disease-free survival. Cancer Res. 2002;62:3005–8.

[37] Sequist LV, Heist RS, Shaw AT, et al. Implementing multiplexed genotyping of non-small-cell lung cancers into routine clinical practice. Ann Oncol. 2011;22:2616–24.

[38] Zappa C, Mousa SA. Non-small cell lung cancer: current treatment and future advances. Transl Lung Cancer Res. 2016;5:288–300.

[39] Roudi R, Beikzadeh B, Roviello G, et al. Identification of hub genes, modules and biological pathways associated with lung adenocarcinoma: a system biology approach. Gene Rep. 2022;27:101638.

[40] Luo X, Feng L, Xu W, et al. Weighted gene co-expression network analysis of hub genes in lung adenocarcinoma. Evol Bioinform Online. 2021;17:11769343211009898.

[41] Xu Y, Wang S, Xu B, et al. AURKA, TOP2A and MELK are the key genes identified by WGCNA for the pathogenesis of lung adenocarcinoma. Oncol Lett. 2023;25:238.

[42] Dai JJ, Zhou WB, Wang B. Identification of crucial genes associated with lung adenocarcinoma by bioinformatic analysis. Medicine (Baltim). 2020;99:e23052.

[43] Wu Y, Yang L, Zhang L, et al. Identification of a four-gene signature associated with the prognosis prediction of Lung adenocarcinoma based on integrated bioinformatics analysis. Genes (Basel). 2022;13:238.

[44] Tian W, Yang X, Yang H, et al. GINS2 functions as a key gene in lung adenocarcinoma by WGCNA co-expression network analysis. Onco Targets Ther. 2020;13:6735–46.

[45] Zhang H, Lu D, Li Q, et al. Identification of six prognostic genes in EGFR-mutant lung adenocarcinoma using structure network algorithms. Front Genet. 2021;12:755245.

[46] Dai D, Shi R, Han S, et al. Weighted gene coexpression network analysis identifies hub genes related to KRAS mutant lung adenocarcinoma. Medicine (Baltim). 2020;99:e21478.

[47] Van Loy T, Vandersmissen HP, Van Hiel MB, et al. Comparative genomics of leucine-rich repeats containing G protein-coupled receptors and their ligands. Gen Comp Endocrinol. 2008;155:14–21.

[48] Yang D, Li JS, Xu QY, et al. Inhibitory effect of MiR-449b on Cancer cell growth and invasion through LGR4 in non-small-cell lung carcinoma. Curr Med Sci. 2018;38:582–9.

[49] Bai R, Zhang J, He F, et al. GPR87 promotes tumor cell invasion and mediates the immunogenomic landscape of lung adenocarcinoma. Commun Biol. 2022;5:663.

[50] Nii K, Tokunaga Y, Liu D, et al. Overexpression of G protein-coupled receptor 87 correlates with poorer tumor differentiation and higher tumor proliferation in non-small-cell lung cancer. Mol Clin Oncol. 2014;2:539–44.

[51] Kita Y, Go T, Nakashima N, et al. Inhibition of cell-surface molecular GPR87 with GPR87-suppressing adenoviral vector disturb tumor proliferation in lung cancer cells. Anticancer Res. 2020;40:733–41.

[52] Rao S, Sigl V, Wimmer RA, et al. RANK rewires energy homeostasis in lung cancer cells and drives primary lung cancer. Genes Dev. 2017;31:2099–112.

[53] Luo J, Yang Z, Ma Y, et al. LGR4 is a receptor for RANKL and negatively regulates osteoclast differentiation and bone resorption. Nat Med. 2016;22:539–46.

[54] Ihle NT, Byers LA, Kim ES, et al. Effect of KRAS oncogene substitutions on protein behavior: implications for signaling and clinical outcome. J Natl Cancer Inst. 2012;104:228–39.