

RESEARCH ARTICLE

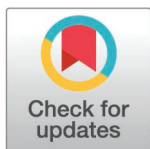
Identifying regulatory driver motifs in non-small cell lung carcinoma via a systematic approach

Rahul Kumar¹, Sheersh Massey², Sarah Albogami³, Abdulaziz A. Aloliqi⁴, Abdulaziz Asiri⁵, Maher M. Aljohani⁶, Hashim M. Aljohani⁷, Atul Kumar¹, Kapil Dev^{1*}

1 Medical Biotechnology Lab, Department of Biotechnology, Jamia Millia Islamia, New Delhi, India, **2** Human Genetics Lab, Department of Biosciences, Jamia Millia Islamia, New Delhi, India, **3** Department of Biotechnology, College of Science, Taif University, Taif, Saudi Arabia, **4** Department of Basic Health Sciences, College of Applied Medical Sciences, Qassim University, Buraydah, Al-Qassim, Saudi Arabia, **5** Department of Medical Laboratory Sciences, College of Applied Medical Sciences, University of Bisha, Bisha, Saudi Arabia, **6** Department of Basic Medical Sciences, College of Medicine, Taibah University, Medina, Saudi Arabia, **7** Department of Clinical Laboratory Sciences, College of Applied Medical Sciences, Taibah University, Medina, Saudi Arabia

✉ These authors contributed equally to this work.

* kdev@jmi.ac.in



OPEN ACCESS

Citation: Kumar R, Massey S, Albogami S, Aloliqi AA, Asiri A, Aljohani MM, et al. (2026) Identifying regulatory driver motifs in non-small cell lung carcinoma via a systematic approach. PLoS One 21(2): e0340798. <https://doi.org/10.1371/journal.pone.0340798>

Editor: Mohammad Zeeshan Najm, Apeejay Stya University, INDIA

Received: July 30, 2025

Accepted: December 28, 2025

Published: February 11, 2026

Copyright: © 2026 Kumar et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: All relevant data are within the manuscript and its [Supporting information](#) files.

Funding: The author(s) received no specific funding for this work.

Abstract

Background

Lung cancer exhibits highest incidence among all cancer types worldwide and even after rigorous research and advanced treatment strategies, it constitutes a primary cause of cancer-related mortality. Non-small cell lung cancer is the predominant subtype, constituting the majority of lung cancer cases. Therefore, exploring novel biomarkers is crucial for betterment of diagnostic and therapeutic approaches.

Methods

The meta-analysis was performed using GEO datasets, to explore the differentially expressed genes (DEGs) and miRNAs (DEMs) in the non-small cell lung cancer (NSCLC) cases. We explored the ChEA database to extract the relevant transcription factors regulating the expression of our hub genes. Further, based on the highest degree of centrality, the feed-forward loop was identified with highest sub-network motif comprising of gene-TF-miRNA. We used pathway and GO term enrichment analysis to determine the importance of these DEGs in different biological processes.

Results

In NSCLC, we found 950 differentially expressed miRNAs and 1761 genes were recognized exhibiting the significant change in expression ($p < 0.05$). Further, we investigated the role of sub-network motif in patient survival, hsa-miR-5010 was found to be significantly linked with patient outcome in Lung Adenocarcinoma (LUAD) ($p = 0.033$)

Competing interests: The authors have declared that no competing interests exist.

Abbreviations: NSCLC, Non-small cell lung cancer; DEGs, Differentially expressed genes; DEMs, Differentially expressed miRNA; TF, Transcription factor; WHO, World Health Organization; lncRNA, long non-coding RNAs; GEO database, Genome Expression Omnibus database; GO, Gene Ontology; FFL, Feed-Forward Loop; NRG1, Neuregulin 1; TCGA, The Cancer Genome Atlas; LUAD, Lung Adenocarcinoma; LUSC, Lung Squamous Cell Carcinoma; RB1, Retinoblastoma 1.

and Lung Squamous Cell Carcinoma (LUSC) ($p=0.013$) while SMAD4 ($p<0.001$) and NRG1 ($p<0.001$) expression exhibited prognostic significance in LUAD cohort only.

Conclusion

Our data indicated that NRG1-SMAD4-miR-5010-5p was the most prominent sub-network motif engaged in NSCLC patients based on the degree of centrality. In vitro mechanistic studies will provide better understanding on the role of NRG1-SMAD4-miR-5010-5p motif in NSCLC cases.

1. Introduction

For numerous decades, lung cancer is ranked amongst the most prevalent type of cancer globally [1]. Although it stands as the third most common cancer, following breast and prostate cancer, it holds the highest proportion of cancer-related fatalities, accounting for 22% of such deaths [2]. Approximately 80% of lung cancer associated deaths can be linked to smoking. Other factors known to increase the risk of developing lung cancer include exposure to radon, prolonged and cumulative exposure to air pollution—especially emissions comprising polycyclic aromatic hydrocarbons (PAH), asbestos, as well as personal or family history of lung cancer [3,4]. According to the World Health Organization (WHO), lung tumors are categorized into two main types: non-small cell lung cancer (NSCLC), which makes up 80–85% of all cases, and small cell lung cancer (SCLC), accounting for the remaining 15% [1,5].

Compelling evidence supports the significant connections of long non-coding RNAs (lncRNAs) in several diseases, with a particular emphasis on cancer, including NSCLCs. Recent research has explored the expression patterns of lncRNAs in NSCLCs. MicroRNAs (miRNAs), a class of small non-coding RNAs, play vital roles in post-transcriptional regulation of gene. Dysregulation of miRNA expression has been implicated in several cancer types, including NSCLC, where they can function both as tumor suppressors and oncogenes [6,7]. Xu et al. found 2420 lncRNAs with differential expression (fold change ≥ 2) between lung adenocarcinoma (LAD) and normal tissue (NT) samples using high-throughput microarrays. Out of these, 1213 lncRNAs showed upregulation, while the remaining 1207 lncRNAs demonstrated downregulation [8,9]. Gaining insights into the complex interactions involving miRNAs, messenger RNAs (mRNAs), and transcription factors (TFs) is crucial for unravelling the molecular mechanisms of NSCLC, providing potential avenues for diagnosis and therapy.

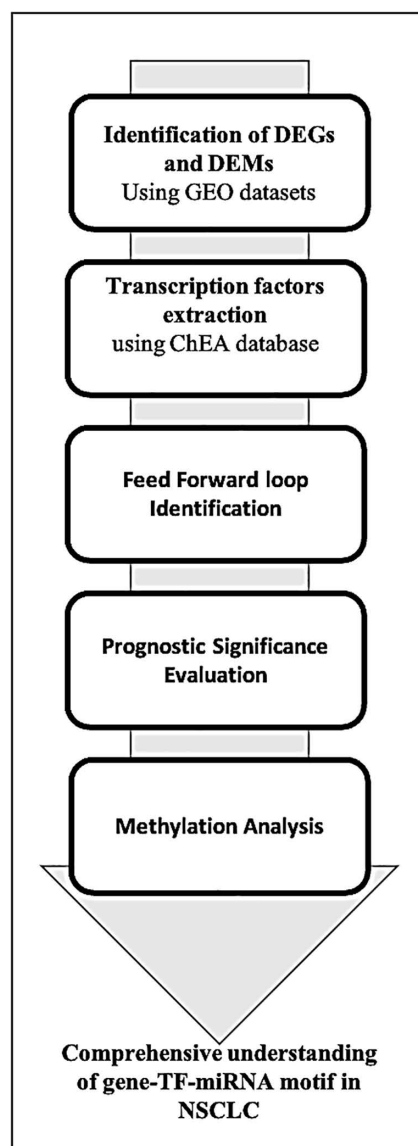
In the last few years, the domain of bioinformatics has witnessed the emergence of potent tools for unravelling intricate regulatory networks within cancer cells. The integrated examination of high-throughput data encompassing miRNA and mRNA expression, alongside the prediction of regulatory motifs involving miRNAs, target mRNAs, and related transcription factors (TFs), has yielded fresh perspectives on the regulatory setting of non-small cell lung carcinoma (NSCLC) [10]. Harnessing

computational approaches enables the identification of pivotal regulatory modules governing the progression of NSCLC, thereby aiding in the design of targeted interventions [11].

In this investigation, we conducted a thorough bioinformatics examination of miRNA and mRNA expression patterns in NSCLC with the goal of unveiling regulatory motifs that govern gene networks associated with cancer. Our systematic computational analysis includes gene and miRNA expression curation followed by their validation using publicly available datasets. Also, evaluating their significance in patient outcome via survival analysis and finally identifying novel regulatory motif involving miRNA, genes, and transcription factors (TFs), presenting itself as a promising candidate for potential therapeutic targeting in NSCLC. The insights derived from our study illuminate the complex regulatory mechanisms that underlie NSCLC and establish a basis for subsequent experimental validation and exploration of therapeutic avenues.

2. Materials and methods

Work Flow Chart



2.1. Procurement of Microarray data and preliminary processing

We obtained miRNA and mRNA expression profiles from publicly available datasets related to Non-Small Cell Lung Cancer (NSCLC) from the Genome Expression Omnibus (GEO) database [12]. We refined our search by applying specific inclusion criteria: (1) we considered only samples from the 'Homo sapiens' species and of the 'expression profiling by array' type; (2) we selected datasets that included both processed and raw microarray data; (3) we focused on datasets with sample sizes ranging from 20 to 1000. The miRNA and mRNA expression datasets were acquired using the accession numbers GSE53882 and GSE31799, respectively. We downloaded the series matrix (processed) files for both datasets. Additionally, as the data was pre-normalized, preprocessing steps involved background correction and removal of duplicates.

2.2. Identification of DEGs and DEMs

We employed the R statistical software (<https://www.r-project.org/>), specifically employing the Limma package [13], to conduct an analysis of the expression profiles obtained from the selected datasets. This package facilitated the implementation of the t-test for assessing significance, resulting in corresponding p-values. Subsequently, miRNAs and genes exhibiting p-values less than 0.05 and fold changes within the range of 0.5 to 1.0 were identified and classified as differentially expressed miRNAs (DEMs) and differentially expressed genes (DEGs), respectively [14,15].

2.3. Enrichment analysis of pathways and Gene Ontology (GO) terms

We gathered information on the top 10 significantly enriched pathways and Gene Ontology (GO) terms (p-value < 0.05) associated with the identified DEGs using the Enrichr web-based tool [16]. Our selection included the Reactome pathway database, as well as the GO Biological Process (GO-BP), GO Molecular Function (GO-MF), and GO Cellular Component (GO-CC) libraries for compiling the pathway and GO term data.

2.4. Analysis and construction of 3-node miRNA FFL

Relevant transcription factors (TFs) with a significance score (p-value) less than 0.05, which regulate our central mRNAs, were obtained from the ChEA v3.0 database as it offers experimentally validated TF-mRNA interactions [17]. Subsequently, the corresponding miRNAs (score greater than 0.95 and binding exclusively to the 3' UTR region) that suppress our central mRNAs, along with the significant TFs obtained from ChEA, were retrieved from cross-validated comprehensive list of miRNAs in miRWalk v3.0 database [18]. To design a 3-node miRNA Feed-Forward Loop (FFL), only the common miRNAs that repress both our mRNAs and transcription factors were maintained. Ultimately, all three interaction pairs "miRNA-mRNA, TF-mRNA, and mRNA-TF"—were calibrated according to shared miRNAs and definitive TFs. The pairings were subsequently combined to create a 3-node miRNA feedforward loop, and the network was analysed using Cytoscape (<http://www.cytoscape.org/>) [19].

2.5. Survival analysis

We performed survival analysis utilising the Kaplan-Meier Plotter to assess the prognostic importance of hsa-miR-5010, SMAD4, and NRG1 (<https://kmplot.com/analysis/>) [20], a publicly accessible database enables the gene and miRNA expression on clinical outcomes across diverse cancer types. Following the selection of the specific miRNA or gene of interest, our focus was directed towards lung adenocarcinoma and lung squamous cell carcinoma. To ensure data accuracy, we applied two essential filters: 'Invert HR values below 1' to highlight adverse survival associations and 'Compute median over the entire dataset' to obtain robust estimates.

2.6. MethSurv analysis

The predictive relevance of CpG methylation in LUAD and LUSC cohorts was assessed using the web-based program MethSurv (<https://biit.cs.ut.ee/methsurv>). This tool is useful for survival analysis based on individual CpG sites using TCGA data. CpG sites of SMAD4 and NRG1 were analyzed to determine their association with patient overall survival. The results provided insights into potential methylation-based prognostic biomarkers in NSCLC.

3. Results

3.1. Acquisition of microarray data and pre-processing

Under the defined search conditions in the methods, we acquired 36 miRNA and 66 mRNA expression profiles. We specifically chose datasets with accession numbers GSE53882 (miRNA, including tumor and control samples) and GSE31799 (mRNA, including Adenocarcinoma and Squamous cell carcinoma samples). The miRNA dataset comprised 397 samples from patients diagnosed with tumors and 151 corresponding adjacent non-cancerous tissues. Conversely, the mRNA expression data consisted of 31 samples with adenocarcinoma and 20 samples with Squamous cell carcinoma.

We enhanced dataset reliability through vital pre-processing steps, including normalization and probe ID mapping using GEO2R. This yielded accurate quantile normalized and RMA normalized expression values for GSE53882 and GSE31799, respectively, tied to probes. The resulting tab-delimited text file after batch correction using PCA-based method formed the basis for robust analyses.

3.2. Identification of DEGs and DEMs

Upon implementing the specified thresholds, we identified a total of 950 differentially expressed miRNAs (DEMs) and 1761 genes (DEGs) in relation to their respective varying conditions. By incorporating criteria accommodating low-fold changes, we ensured a comprehensive assessment of tumor vs. control group variations, maximizing DEMs and DEGs detection.

[Table 1](#) details the compiled upregulated and downregulated DEMs and DEGs, while [Fig 1A, 1B](#) showcases volcano plots, illustrating significant vs. nonsignificant miRNAs and genes. Additionally, [Fig 2A, 2B](#) presents heatmap plots for the top 10 down and upregulated DEMs and DEGs, providing insights into their expression patterns.

3.3. Pathway and GO term enrichment analyses

We conducted a comprehensive Reactome pathway analysis using the Enrichr tool, which unveiled the participation of 88 differentially expressed genes (DEGs) in the top 10 Reactome pathways ($p < 0.05$). Among these, the pathway with the highest gene count was identified as “Keratinization.” Particularly noteworthy was the identification of the “Formation of Cornified Envelope” pathway as the most significant, indicating its pronounced impact in the biological context. The graphical representation in [Fig 3A](#) visually presents the top 10 Reactome pathways.

Our analysis further extended to the exploration of gene expression differences, with a combined total of 96, 66, and 85 DEGs exhibiting distinct expression patterns and engaging in the foremost 10 significant terms across the “Biological Process (BP) ($p < 0.05$), Cellular Compartment (CC) ($p < 0.05$), and Molecular Function (MF) ($p < 0.05$)”. In [Fig 3B](#), we provide a visual depiction of the distribution of noteworthy Gene Ontology (GO) Biological Process terms, demonstrating the

Table 1. Upregulated and downregulated DEMs and DEGs.

GEO accession	Number of upregulated	Number of Downregulated
GSE53882(miRNAs)	516	434
GSE31799(mRNAs)	1012	749

<https://doi.org/10.1371/journal.pone.0340798.t001>

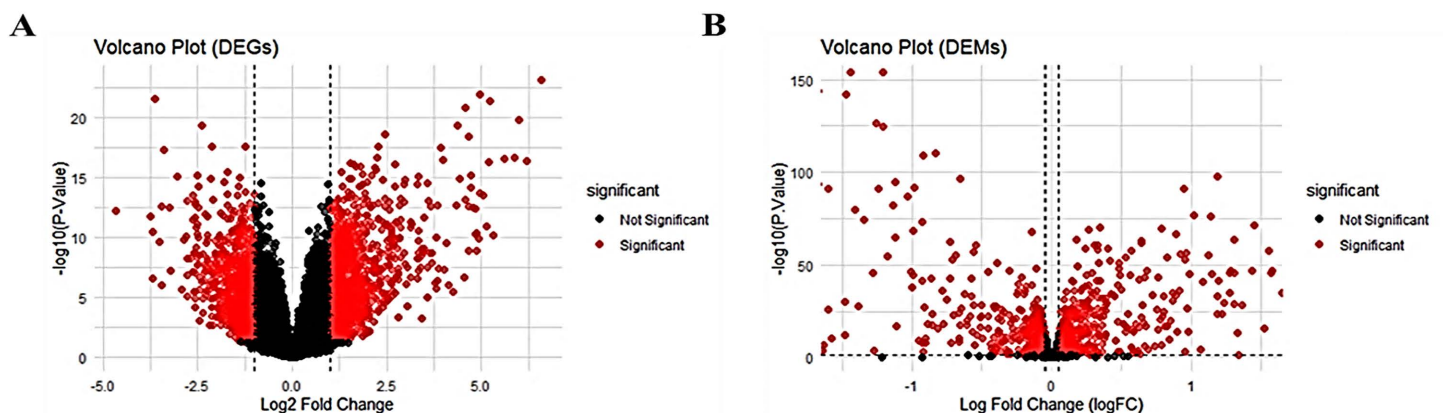


Fig 1. Volcano plot representing DEGs and DEMs: Log2 Fold Change, Y-axis: -Log10 (P-Value). Red dots significant and black dots are insignificant. (A) DEGs. (B) DEMs.

<https://doi.org/10.1371/journal.pone.0340798.g001>

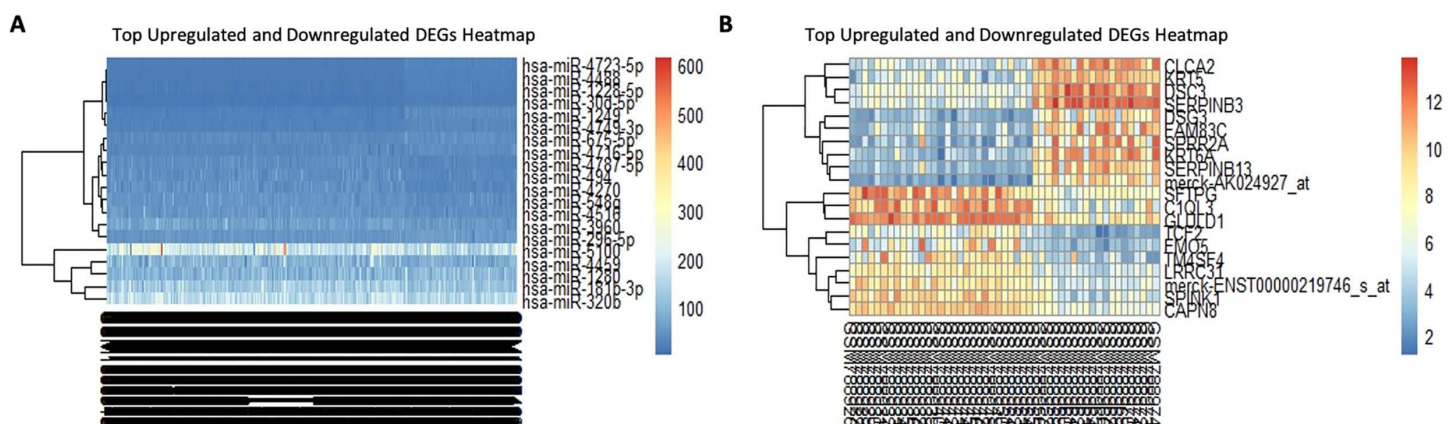


Fig 2. Heatmap illustrating. (A) Top 10 up and downregulated DEMs. Gene's positions are indicated by vertical multicolour bands in the right panel. (B) Top 10 up and downregulated DEGs. Gene's positions are highlighted by vertical multicolour bands in the right panel.

<https://doi.org/10.1371/journal.pone.0340798.g002>

gene count (x-axis) for each term (y-axis). Particularly notable was the prominence of “Epithelium Development,” with a comprehensive gene count of 30.

3.4. 3-node miRNA FFL construction and analysis

The 3-node miRNA feedforward loop, seen in Fig 4A, consisted of 1871 nodes and 116,395 edges. The TF-mRNA pair comprised 9,051 edges, while the miRNA-TF and miRNA-mRNA couples accounted for 16,824 and 90,520 edges, respectively. Of the total nodes, 142 were transcription factors (TFs), 921 were messenger RNAs (mRNAs), and 808 were microRNAs (miRNAs). Utilising the Degree centrality metric, we identified the predominant subnetwork motif consisting of one transcription factor (SMAD4), one mRNA (NRG1), and one microRNA (miR-5010-5p), as illustrated in Fig 4B.

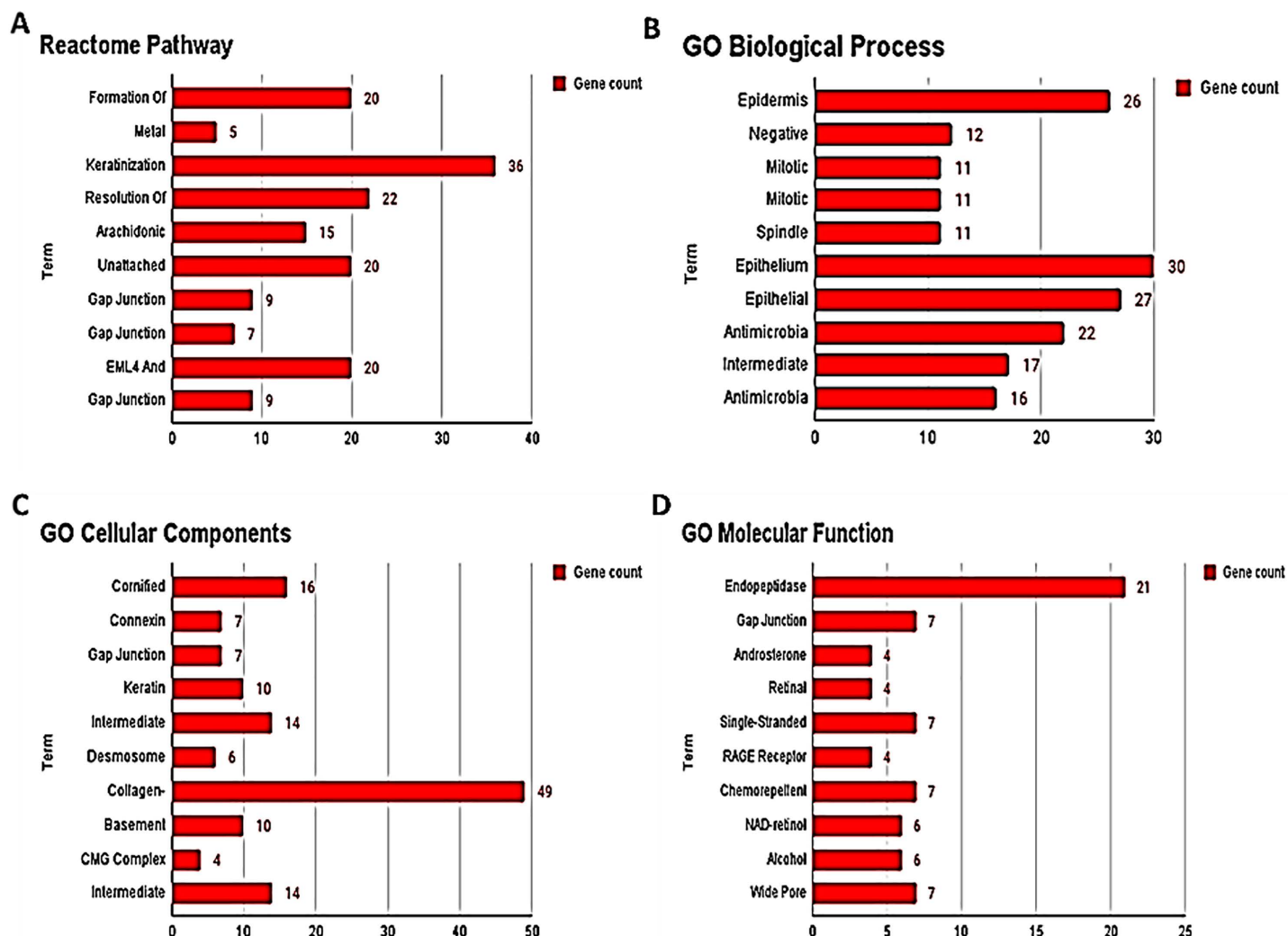


Fig 3. Pathway and GO term enrichment analyses. (A) Reactome Pathway Analysis (B) GO Biological Process Analysis (C): GO Cellular Compartment Analysis (D) GO Molecular Function Analysis.

<https://doi.org/10.1371/journal.pone.0340798.g003>

3.5. Survival analysis

Our survival analysis revealed the prognostic insights regarding SMAD4, NRG1 and hsa-miR-5010 in lung adenocarcinoma and lung squamous cell carcinoma. Patients with NRG1 overexpression ($p < 0.001$) and downregulation of tumor suppressor SMAD4 ($p < 0.001$) as expected correlate significantly with poor patient survival in LUAD cohort. While lower hsa-miR-5010 levels were linked to significantly shorter median survival ($p < 0.05$) in both LUAD ($p = 0.033$) as well as LUSC ($p = 0.013$) cohorts. These results suggest NRG1 and SMAD4 as potential prognostic markers in lung adenocarcinoma cases whereas, hsa-miR-5010 is prognostically significant in both the subtypes (Fig 5).

3.6. Promoter methylation analysis of SMAD4 and NRG1

The promoter methylation status of various CpG islands of NRG1 and SMAD4 were analyzed (Fig 6). Heatmaps demonstrating the promoter methylation pattern across the various CpG islands throughout the LUAD and LUSC cohorts were

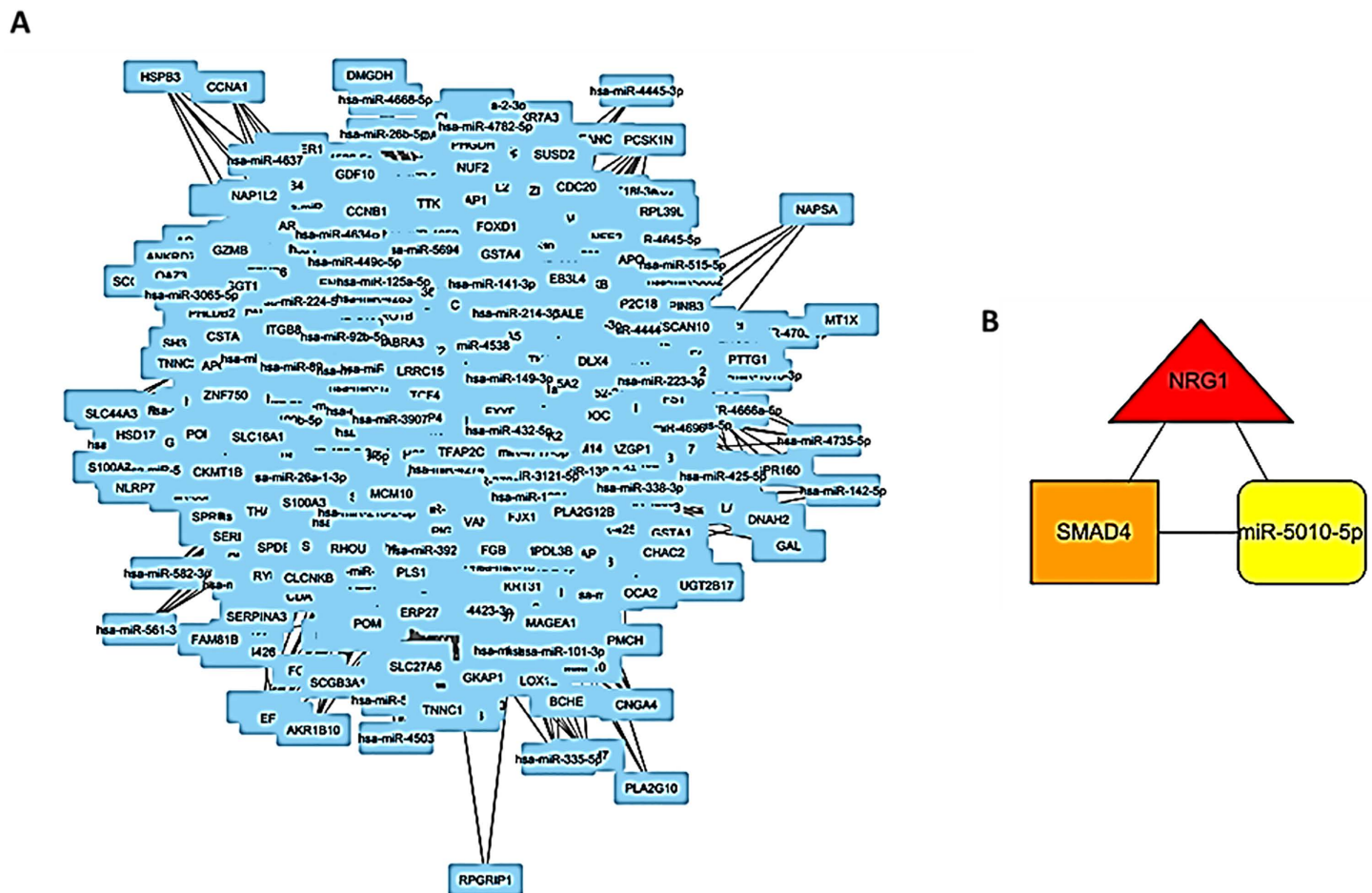


Fig 4. Interaction network of a miRNA feed-forward loop. (A) 3-node miRNA feed-forward loop (FFL) network Fig (B) The visual representation of the regulatory motif NRG1-SMAD4-miR-5010-5p.

<https://doi.org/10.1371/journal.pone.0340798.g004>

produced. Further, these CpG islands were then investigated for prognostic significance to analyze their role in carcinogenesis and patient survival. SMAD4 CpG islands cg26909431 and cg6329143 while NRG1 CpG islands cg17457560 and cg23637605 were identified to be prognostically most significant in LUAD and LUSC respectively.

4. Discussion

Over the decades, the treatment strategies for non-small cell lung carcinoma have become more inclined towards a biomarker-driven personalized approach in addition to the existing therapies [21]. Molecular analysis is currently a potent tool for more efficient diagnosis and management of disease. In the current era, bioinformatics has emerged to be a promising platform for computer-aided biomarker discovery and provides the advantage of processing big data for biomarker screening. Identification and analysis of the biological molecules at the different omics levels can be efficient tools for early diagnosis and better prognosis of disease [22,23]. In this study, by using suitable inclusion and exclusion criteria a dataset from GEO having microarray profiles from two varying samples was selected. We identified DEGs from the dataset and overall, 1761 genes were discovered to exhibit differential expression, of which 1012 were upregulated and 749 were downregulated.

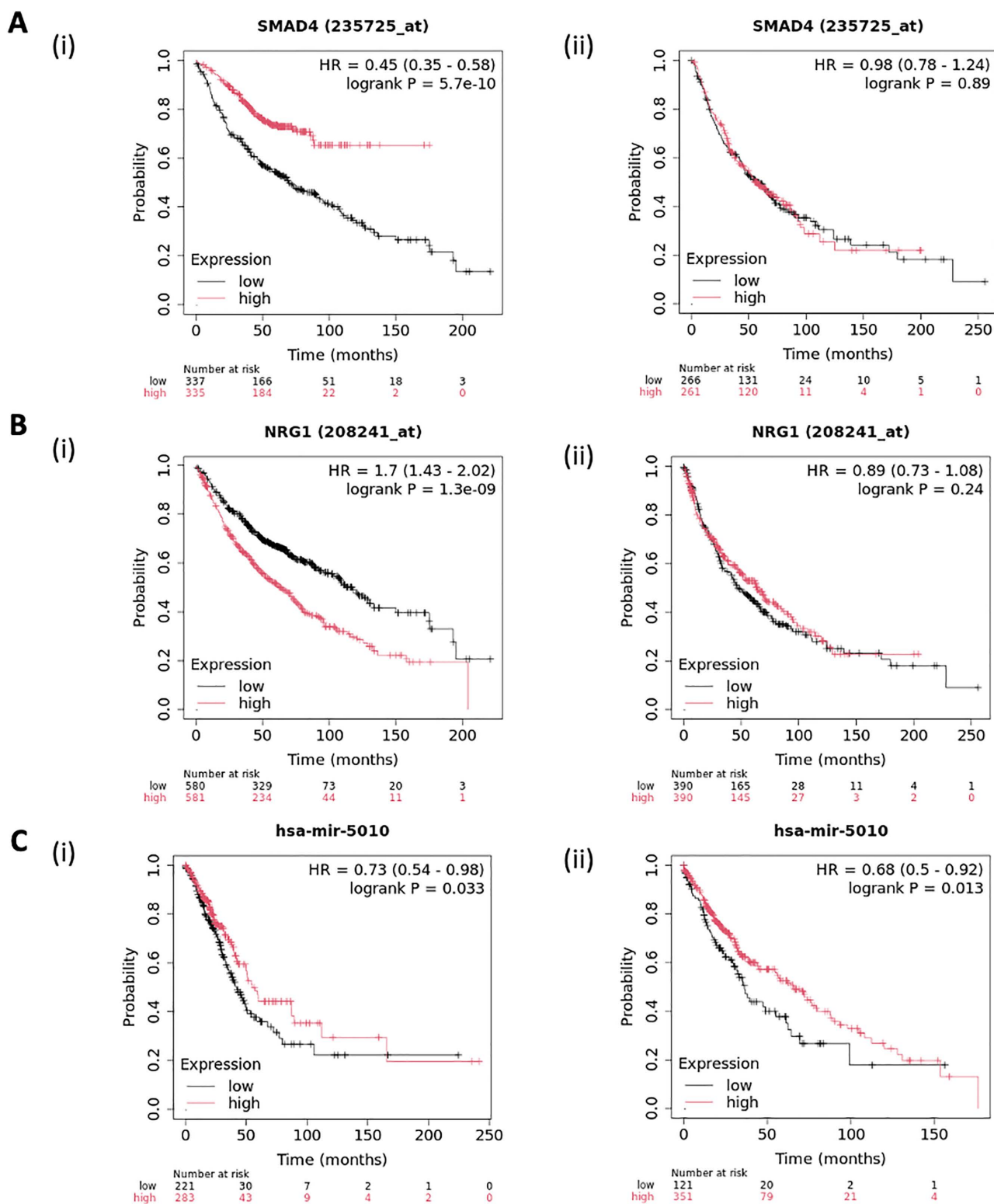


Fig 5. Kaplan-Meier (KM) plots showing high and low expression groups for survival analysis. (A) (i) SMAD4 in LUAD (ii) SMAD4 in LUSC (B) (i) NRG1 in LUAD (ii) NRG1 in LUSC (C) (i) hsa-miR-5010 in LUAD (ii) hsa-miR-5010 in LUSC.

<https://doi.org/10.1371/journal.pone.0340798.g005>

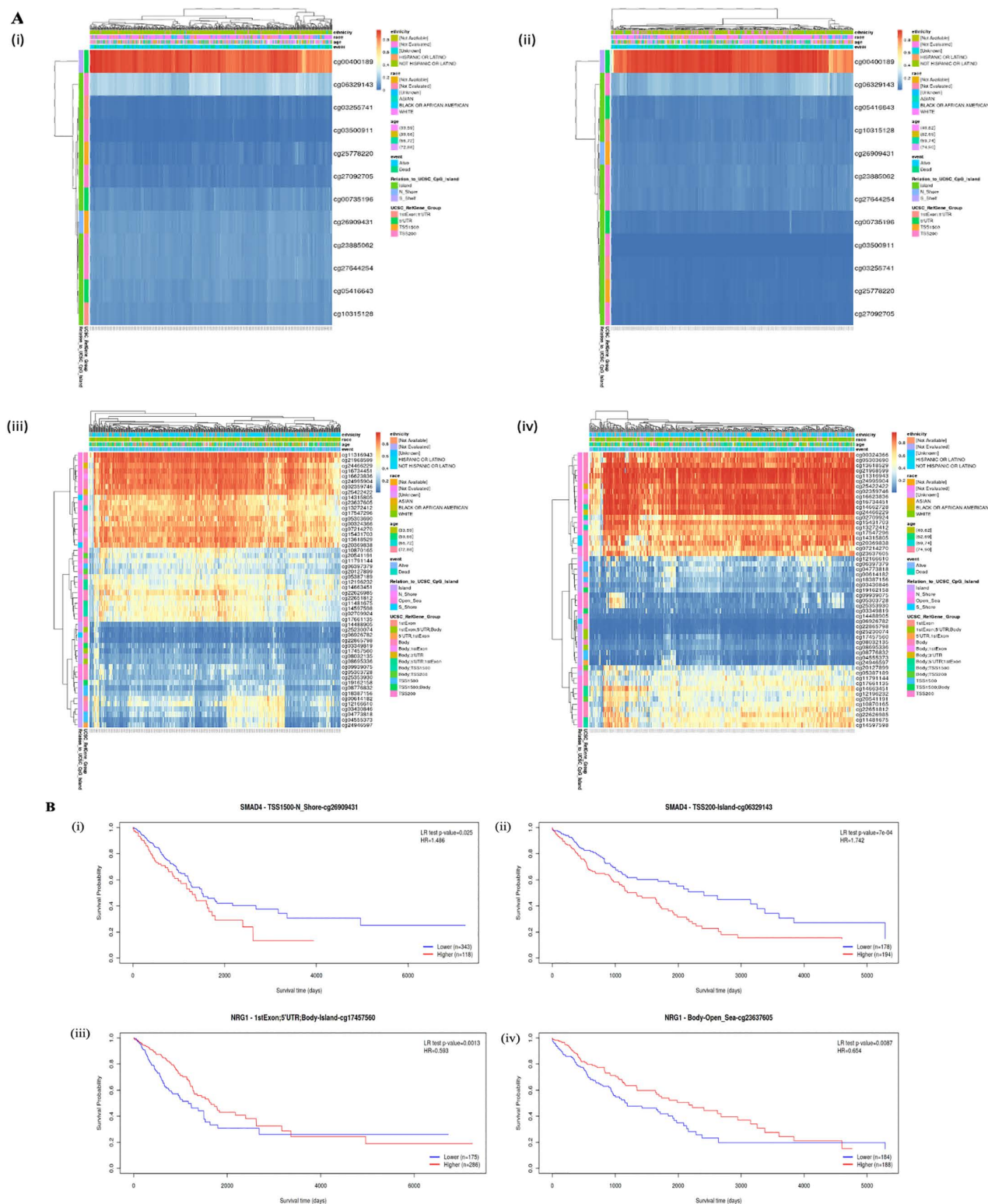


Fig 6. Promoter methylation analysis. (A) Heat map showing methylation across CpG of (i) SMAD4 in LUAD (ii) SMAD4 in LUSC (iii) NRG1 in LUAD (iv) NRG1 in LUSC. (B) Kaplan Meier plot for prognostically most significant CpG island (i) SMAD4 in LUAD (ii) SMAD4 in LUSC (iii) NRG1 in LUAD (iv) NRG1 in LUSC.

<https://doi.org/10.1371/journal.pone.0340798.g006>

The Enrichr tool was used and the identified DEGs were further subjected to Reactome pathway analysis considering p value <0.05 to be significant. We identified a strong association between 88 DEGs in the top 10 Reactome pathways, with “Keratinization” exhibiting the greatest gene count. Furthermore, it was discovered that the “Formation of Cornified Envelope” was the most important pathway, demonstrating its substantial effect in the biological context. The functional significance of the identified DEGs was then evaluated using the GO enrichment tool and 96 DEGs displayed significant association across Biological Processes (BP), of which the highest gene count was observed in epithelium development. Similarly, 66 and 88 DEGs were found to have a strong connection with Cellular Compartment (CC) and Molecular Function (MF) respectively.

Epigenetic regulation plays a crucial role in controlling the gene expression [24]. Among various epigenetic regulators, miRNAs are known for their vital regulatory role in various malignancies [25,26]. In this context, we identified the differentially expressed miRNA with p -values less than 0.05 and categorized them as DEMs. Overall, 950 DEMs were identified in our study of which 516 were found to be upregulated whereas 434 were significantly downregulated. Moreover, to explore the role of transcription factors (TFs) in controlling the gene expression along with the DEMs, the ChEA database was used to extract the TFs having regulatory functions and significant association ($p < 0.05$) with our hub mRNA. To explore the connection among our hub mRNA, DEMs, and TFs, a feed-forward loop was constructed with the mRNA, miRNA, and TFs that were common in all three interactions i.e. miRNAs-mRNA, TFs-mRNAs, and miRNAs-TFs.

NRG1 (mRNA), SMAD4 (TF), and miR-5010-5p (miRNA) were identified to be the highest sub network motif with the highest degree of centrality measure. Neuregulin 1 (NRG1) belongs to the family of epidermal growth factors. The NRG1 expression has been reported in various malignancies like breast, lung, and prostate cancer. It is also detected to have a critical role in the prognosis of the disease and resistance to therapeutic drugs like trastuzumab [27–29]. Also, NRG1 fusion with other oncogenic genes due to chromosomal rearrangements can consequently form chimeric proteins resulting in activation of downstream pathways like PI3K-AKT. Moreover, oncogenic variants of NRG1 promote proliferation and metastasis via activating MAPK or NF- κ B/MMP9 pathway and can be responsible for resistance to targeted therapies like EGFR and HER2 [30,31]. In NSCLC patients, NRG1 was found to be upregulated and its elevated expression is associated with poor survival of the patients [32]. The SMAD4 and NRG1 fusion have been reported in lung cancer cases [33]. SMAD4 is a known transcription factor located on chromosome 18q21 region and is well known to be associated with TGF β signalling [34,35]. SMAD4 mediated TGF β signalling plays vital role in maintenance of tissue homeostasis while its downregulation is responsible for cell proliferation and tumorigenesis. Moreover, the downregulation of SMAD4 can lead to activation of oncogenic pathways like ErbB2 and Akt leading to uncontrolled cellular growth and invasion [36,37]. Previously published literature revealed the tumor suppressive role of SMAD4 in various cancers including NSCLC, and is reported to be downregulated in cancer [38,39]. Our findings corroborated with the earlier researches and survival analysis revealed significant association of SMAD4 downregulation with good overall survival in LUAD cohort [40]. Further, miR-5010 was found to be overexpressed in the selected dataset for our study and its expression is associated with the poor survival in the non-small cell lung cancer patient. Similar role of miR-5010 is reported in gastric cancer where its expression is reported to be elevated [41]. It is important to clarify that miR-5010-5p and miR-5010 refer to the same miRNA entity; miR-5010-5p corresponds to the mature 5' arm product derived from the miR-5010 precursor. Our survival analysis revealed. In our FFL analysis, NRG1-SMAD4-miR-5010-5p is the novel axis to be associated with the NSCLC [30,36]. Further methylation analysis was done using MethSurv tool to evaluate the promoter methylation landscape of NRG1 and SMAD4. Promoter methylation status of SMAD4 CpG islands cg26909431 and cg6329143 and of NRG1 CpG islands cg17457560 and cg23637605 was observed to have most significant role in patients' survival in LUAD and LUSC cohorts. Promoter methylation plays critical role in tumor initiation, maintenance as well as progression [42]. Promoter hypermethylation of SMAD4 gene is observed in prostate and gastric carcinoma [43,44]. Similarly, epigenetic dysregulation of NRG1 is frequently reported in cervical and breast cancer tissue [45,46]. The studies suggest the association of

NRG1 expression and SMAD 4 downregulation as key players in activation of oncogenic pathways like PI3K-AKT. Also, their dysregulation is linked with disruption of cellular homeostasis and cancer progression [30,36].

The gene expression regulatory motifs plays decisive roles in carcinogenesis and extensive studies are being done to identify such motifs. For instance MYC/RB1/miR-106a exhibits oncogenic role by suppressing tumor suppressor RB1 gene at transcriptional; as well as translational levels in solid tumors. Similarly, the association of p53 and miR-34a with CDK4 have been reported in regulation of cell cycle arrest and proliferation [47–49]. Earlier studies have revealed the association of NRG1 in NSCLC; however, its regulatory interactions have not been explored. Hereby, our findings suggest the significant role of NRG1 and its regulatory partners SMAD4 and miR-5010-5p in NSCLC which could be pioneer for further studies exploring the role of respective motif in lung carcinogenesis. This study considers the publicly available GEO datasets separately for mRNA and miRNA therefore can further be validated using experimental and clinical studies. This will provide better insight into our understanding and will highlight the diagnostic and therapeutic significance of the NRG1-SMAD4-miR-5010-5p axis.

Supporting information

S1 File. Downregulated-miR.

(TXT)

S2 File. Upregulated-miR.

(TXT)

S3 File. Downregulated-mRNA.

(TXT)

S4 File. Upregulated-mRNA.

(TXT)

S5 File. miRNA_TF.

(CSV)

S6 File. TF_Genes.

(CSV)

Author contributions

Conceptualization: Rahul Kumar, Sheersh Massey, Kapil Dev.

Data curation: Abdulaziz Asiri.

Formal analysis: Sheersh Massey.

Funding acquisition: Sarah Albogami, Abdulaziz A. Aloliqi, Maher M. Aljohani, Hashim M. Aljohani.

Investigation: Rahul Kumar, Atul Kumar.

Methodology: Rahul Kumar, Sheersh Massey, Atul Kumar.

Project administration: Kapil Dev.

Resources: Abdulaziz Asiri, Kapil Dev.

Software: Rahul Kumar, Sheersh Massey, Atul Kumar.

Supervision: Kapil Dev.

Validation: Sarah Albogami, Abdulaziz A. Aloliqi, Kapil Dev.

Visualization: Sarah Albogami.

Writing – original draft: Sheers Massey.

Writing – review & editing: Maher M. Aljohani, Hashim M. Aljohani, Kapil Dev.

References

1. World Health Organization [Internet]. 2022 [cited 2022 Feb 24]. Available from: https://www.who.int/health-topics/cancer#tab=tab_1
2. IARC [Internet]. 2023 [cited 2023 Jul 6]. Available from: <https://www.iarc.who.int/cancer-topics/>
3. Thandra KC, Barsouk A, Saginala K, Aluru JS, Barsouk A. Epidemiology of lung cancer. *Contemp Oncol (Pozn)*. 2021;25(1):45–52. <https://doi.org/10.5114/wo.2021.103829> PMID: 33911981
4. Alexander M, Kim SY, Cheng H. Update 2020: Management of Non-Small Cell Lung Cancer. *Lung*. 2020;198(6):897–907. <https://doi.org/10.1007/s00408-020-00407-5> PMID: 33175991
5. Travis WD, Brambilla E, Nicholson AG, Yatabe Y, Austin JHM, Beasley MB, et al. The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification. *J Thoracic Oncol*. 2015;10(9):1243–60.
6. Iorio MV, Croce CM. Causes and consequences of microRNA dysregulation. *Cancer J*. 2012;18(3):215–22. <https://doi.org/10.1097/PPO.0b013e318250c001> PMID: 22647357
7. McManus MT. MicroRNAs and cancer. *Semin Cancer Biol*. 2003;13(4):253–8. [https://doi.org/10.1016/s1044-579x\(03\)00038-5](https://doi.org/10.1016/s1044-579x(03)00038-5) PMID: 14563119
8. Xu G, Chen J, Pan Q, Huang K, Pan J, Zhang W, et al. Long noncoding RNA expression profiles of lung adenocarcinoma ascertained by microarray analysis. *PLoS One*. 2014;9(8):e104044. <https://doi.org/10.1371/journal.pone.0104044> PMID: 25089627
9. Zhou X, Ao X, Jia Z, Li Y, Kuang S, Du C, et al. Non-coding RNA in cancer drug resistance: underlying mechanisms and clinical applications. *Front Oncol*. 2022;12:951864. <https://doi.org/10.3389/fonc.2022.951864>
10. Zhang J, Nie R, Liu M, Zhang X. A novel strategy for identifying NSCLC microRNA biomarkers and their mechanism analysis based on a brand-new ceRNA-hub-FFL network. *Int J Mol Sci*. 2022;23(19):11303.
11. Chan BA, Hughes BGM. Targeted therapy for non-small cell lung cancer: current standards and the promise of the future. *Transl Lung Cancer Res*. 2015;4(1):36–54. <https://doi.org/10.3978/j.issn.2218-6751.2014.05.01> PMID: 25806345
12. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res*. 2013;41(Database issue):D991–5. <https://doi.org/10.1093/nar/gks1193> PMID: 23193258
13. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47. <https://doi.org/10.1093/nar/gkv007> PMID: 25605792
14. Qu H, Zhou P, Yang Z, Wang H, Deng K, Wang N, et al. Comprehensive analysis of prognosis and drug sensitivity of programmed cell death pattern-related genes in gastric cancer patients. *Sci Rep*. 2025;15(1):22535. <https://doi.org/10.1038/s41598-025-06424-9> PMID: 40596319
15. Wang J, Ren M, Yu J, Hu M, Wang X, Ma W, et al. Single-cell RNA sequencing highlights the functional role of human endogenous retroviruses in gallbladder cancer. *EBioMedicine* [Internet]. 2022 [cited 2025 Nov 7];85. Available from: <https://www.thelancet.com/action/showFullText?pii=S2352396422005011>
16. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research*. 2016;44(W1):W90–7. <https://doi.org/10.1093/nar/gkw377>
17. Keenan AB, Torre D, Lachmann A, Leong AK, Wojciechowicz ML, Utti V. ChEA3: transcription factor enrichment analysis by orthogonal omics integration. *Nucleic Acids Research*. 2019;47(W1):W212–24.
18. Sticht C, De La Torre C, Parveen A, Gretz N. miRWalk: An online resource for prediction of microRNA binding sites. *PLoS One*. 2018;13(10):e0206239. <https://doi.org/10.1371/journal.pone.0206239> PMID: 30335862
19. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498–504. <https://doi.org/10.1101/gr.1239303> PMID: 14597658
20. Lanczky A, Gyorffy B. Web-Based Survival Analysis Tool Tailored for Medical Research (KMplot): Development and Implementation. *J Med Internet Res*. 2021;23(7):e27633. <https://doi.org/10.2196/27633> PMID: 34309564
21. Pennell NA, Arcila ME, Gandara DR, West H. Biomarker Testing for Patients With Advanced Non-Small Cell Lung Cancer: Real-World Issues and Tough Choices. *Am Soc Clin Oncol Educ Book*. 2019;39:531–42. https://doi.org/10.1200/EDBK_237863 PMID: 31099633
22. Lin Y, Qian F, Shen L, Chen F, Chen J, Shen B. Computer-aided biomarker discovery for precision medicine: data resources, models and applications. *Brief Bioinform*. 2019;20(3):952–75. <https://doi.org/10.1093/bib/bbx158> PMID: 29194464
23. Yang SR, Schultheis AM, Yu H, Mandelker D, Ladanyi M, Buttner R. Precision medicine in non-small cell lung cancer: Current applications and future directions. *Semin Cancer Biol*. 2022;84:184–98.
24. Biswas S, Rao CM. Epigenetics in cancer: fundamentals and beyond. *Pharmacol Therap*. 2017;173:118–34.

25. Inamura K, Ishikawa Y. MicroRNA In Lung Cancer: Novel Biomarkers and Potential Tools for Treatment. *J Clin Med*. 2016;5(3):36. <https://doi.org/10.3390/jcm5030036> PMID: [27005669](#)
26. Petrek H, Yu A-M. MicroRNAs in non-small cell lung cancer: Gene regulation, impact on cancer cellular processes, and therapeutic potential. *Pharmacol Res Perspect*. 2019;7(6):e00528. <https://doi.org/10.1002/prp2.528> PMID: [31859460](#)
27. De Iuliis F, Salerno G, Taglieri L, Lanza R, Cardelli P, Scarpa S. Circulating Neuregulin-1 and Galectin-3 Can Be Prognostic Markers in Breast Cancer. *Int J Biomed*. 2017;32(3):e333-6. <https://doi.org/10.5301/ijbm5000262>
28. Tal-Or P, Di-Segni A, Lupowitz Z, Pinkas-Kramarski R. Neuregulin promotes autophagic cell death of prostate cancer cells. *Prostate*. 2003;55(2):147–57.
29. Montero JC, Rodríguez-Barrueco R, Ocaña A, Díaz-Rodríguez E, Esparís-Ogando A, Pandiella A. Neuregulins and cancer. *Clin Cancer Res*. 2008;14(11):3237–41. <https://doi.org/10.1158/1078-0432.CCR-07-5133> PMID: [18519747](#)
30. Wang Z, Wang H, Zhao B, Luo X, Yang Y, Guo R, et al. NRG1 fusion glycoproteins in cancer progression - Structural mechanisms and therapeutic potential: A review. *Int J Biol Macromol*. 2025;323(Pt 1):147057. <https://doi.org/10.1016/j.ijbiomac.2025.147057> PMID: [40850415](#)
31. Hou G, Niu T, Jia A, Zhang Y, Chen X, Wei H, et al. NRG1 promotes tumorigenesis and metastasis and afatinib treatment efficiency is enhanced by NRG1 inhibition in esophageal squamous cell carcinoma. *Biochem Pharmacol*. 2023;218:115920. <https://doi.org/10.1016/j.bcp.2023.115920> PMID: [37989416](#)
32. Fang C, Kang B, Zhao P, Ran J, Wang L, Zhao L, et al. Targeting Neuregulin 1 (NRG1): A Novel Biomarker for Non-Small-Cell Lung Cancer. *J Environ Pathol Toxicol Oncol*. 2021;40(4):61–72.
33. Laskin J, Liu SV, Tolba K, Heining C, Schlenk RF, Cheema P, et al. NRG1 fusion-driven tumors: biology, detection, and the therapeutic role of afatinib and other ErbB-targeting agents. *Ann Oncol*. 2020;31(12):1693–703. <https://doi.org/10.1016/j.annonc.2020.08.2335> PMID: [32916265](#)
34. Xia X, Wu W, Huang C, Cen G, Jiang T, Cao J, et al. SMAD4 and its role in pancreatic cancer. *Tumour Biol*. 2015;36(1):111–9. <https://doi.org/10.1007/s13277-014-2883-z> PMID: [25464861](#)
35. Alazzouzi H, Alhopuro P, Salovaara R, Sammalkorpi H, Järvinen H, Mecklin J-P, et al. SMAD4 as a prognostic marker in colorectal cancer. *Clin Cancer Res*. 2005;11(7):2606–11. <https://doi.org/10.1158/1078-0432.CCR-04-1458> PMID: [15814640](#)
36. Wang Q, Xiong F, Wu G, Wang D, Liu W, Chen J, et al. SMAD Proteins in TGF- β Signalling Pathway in Cancer: Regulatory Mechanisms and Clinical Applications. *Diagnostics (Basel)*. 2023;13(17):2769. <https://doi.org/10.3390/diagnostics13172769> PMID: [37685308](#)
37. Liu J, Cho SN, Akkanti B, Jin N, Mao J, Long W, et al. ErbB2 pathway activation upon smad4 loss promotes lung tumor growth and metastasis. *Cell Reports*. 2015;10(9):1599–613. Available from: <https://www.cell.com/action/showFullText?pii=S2211124715001424>
38. Guo X, Li M, Wang X, Pan Y, Li J. Correlation between loss of Smad4 and clinical parameters of non-small cell lung cancer: an observational cohort study. *BMC Pulm Med*. 2021;21(1):111. <https://doi.org/10.1186/s12890-021-01480-z> PMID: [33794845](#)
39. Ziemke M, Patil T, Nolan K, Tippimanchai D, Malkoski SP. Reduced Smad4 expression and DNA topoisomerase inhibitor chemosensitivity in non-small cell lung cancer. *Lung Cancer*. 2017;109:28–35. <https://doi.org/10.1016/j.lungcan.2017.04.017> PMID: [28577946](#)
40. Li Z, Huang Y, Zhou R, Li Z, Yan Q. Clinicopathological and prognostic significance of SMAD4 in non-small cell lung cancer: A meta-analysis and database validation. *Medicine (Baltimore)*. 2023;102(29):e34312. <https://doi.org/10.1097/MD.00000000000034312> PMID: [37478236](#)
41. Ma F-C, Zhang G-L, Chi B-T, Tang Y-L, Peng W, Liu A-Q, et al. Blood-based machine learning classifiers for early diagnosis of gastric cancer via multiple miRNAs. *World J Gastrointest Oncol*. 2025;17(4):103679. <https://doi.org/10.4251/wjgo.v17.i4.103679> PMID: [40235889](#)
42. Lakshminarasimhan R, Liang G. The Role of DNA Methylation in Cancer. *Adv Exp Med Biol*. 2016;945:151–72. https://doi.org/10.1007/978-3-319-43624-1_7 PMID: [27826838](#)
43. Aitchison AA, Veerakumarasivam A, Vias M, Kumar R, Hamdy FC, Neal DE, et al. Promoter methylation correlates with reduced Smad4 expression in advanced prostate cancer. *Prostate*. 2008;68(6):661–74. <https://doi.org/10.1002/pros.20730> PMID: [18213629](#)
44. Wang LH, Kim SH, Lee JH, Choi YL, Kim YC, Park TS, et al. Inactivation of SMAD4 Tumor Suppressor Gene During Gastric Carcinoma Progression. *Clin Cancer Res [Internet]*. 2007 [cited 2025 May 1];13(1):102–10. Available from: <https://clincancerres/article/13/1/102/193320/Inactivation-of-SMAD4-Tumor-Suppressor-Gene-During>
45. Yan X, An F, Ding Z, Ma D, Yang X. Study on the relationship between Methylation of Neuregulin (NRG) Gene and Cervical Carcinoma. *Pak J Med Sci [Internet]*. 2024 [cited 2025 May 1];40(6):1207. Available from: <https://pmc.ncbi.nlm.nih.gov/articles/PMC11190413/>
46. Chua YL, Ito Y, Pole JCM, Newman S, Chin SF, Stein RC, et al. The NRG1 gene is frequently silenced by methylation in breast cancers and is a strong candidate for the 8p tumour suppressor gene. *Oncogene [Internet]*. 2009 [cited 2025 May 1];28(46):4041–52. Available from: <https://www.nature.com/articles/onc2009259>
47. Wu Q, Qin H, Zhao Q, He X-X. Emerging role of transcription factor-microRNA-target gene feed-forward loops in cancer. *Biomed Rep*. 2015;3(5):611–6. <https://doi.org/10.3892/br.2015.477> PMID: [26405533](#)
48. El Baroudi M, Corà D, Bosia C, Osella M, Caselle M. A curated database of miRNA mediated feed-forward loops involving MYC as master regulator. *PLoS One*. 2011;6(3):e14742. <https://doi.org/10.1371/journal.pone.0014742> PMID: [21390222](#)
49. He L, He X, Lim LP, de Stanchina E, Xuan Z, Liang Y, et al. A microRNA component of the p53 tumour suppressor network. *Nature*. 2007;447(7148):1130–4. <https://doi.org/10.1038/nature05939> PMID: [17554337](#)