In silico characterization of transcription factors implicated in lung cancer

Akshita Chaudhary[1], Hari Singh[2,3], and Tiratha Raj Singh[1,3]*

[1] Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology, Waknaghat, Solan, Himachal Pradesh 173234, India

[2] Department of Computer Science Engineering, Jaypee University of Information Technology, Waknaghat, Solan, Himachal Pradesh 173234, India

[3] Centre of Healthcare Technologies and Informatics (CEHTI), Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology, Waknaghat, Solan, Himachal Pradesh 173234, India
*tiratharaj.singh@juitsolan.in

Abstract
Lung cancer is a leading cause of cancer-related deaths worldwide, primarily due to late-stage diagnosis and molecular heterogeneity. Transcription factors have a crucial role in tumor formation and progression. TFs that have not been adequately studied in lung cancer are numerous. The paper provides an in-silico analysis to identify unexplored TFs that may be implicated in the pathogenesis of lung cancer. A systematic curation of 13 TFs identified in databases and their analysis showed that TFs form complex regulatory interactions. The target gene network analysis revealed that the projected targets are enriched in cell death, differentiation, and proliferation. Motif analysis revealed a high level of conservation of cis-regulatory elements, suggesting high specificity of DNA-binding TFs such as ZNF266 and TSH2. The analysis of the PPI network revealed an enriched interaction network (23 nodes, 57 edges; PPI enrichment p-value < 1.0e-16), with the core cluster of proteins involved in mitotic regulation and cell-cycle control. PCGF2 is a sparsely connected group that mediates Polycomb-mediated epigenetic control and has significant interconnections with cell cycle control, p53 pathways, and cancer-related pathways. Profiling of genomic alterations in lung adenocarcinoma and squamous cell carcinoma pathways identified a common alteration in PCGF2, occurring through mutations and copy number increases, suggesting that the protein may be involved in transcriptional and epigenetic dysregulation in lung cancer. This research paper identifies new TFs linked to lung cancer and proposes a systematic bioinformatics analysis of key regulatory drivers, with the aim of developing TF-based biomarkers and therapeutic targets.

Keywords: Lung cancer, Transcription factors, In silico analysis, PCGF2

1. Introduction

Lung cancer is the most commonly diagnosed cancer type and the primary cause of cancer-related deaths globally [1]. Eighty five percent of diagnosed cases are non-small cell lung cancer (NSCLC), which is usually diagnosed at an advanced stage with a poor 5-year survival. In the past few years, transcription factors (TFs) have become the main regulators of lung tumour biology. Transcription factors are at the top of the cell signalling cascade and regulate the expression of genes involved in cell growth, survival, invasion, and immune evasion [2]. Alteration of TFs or their co-regulators is now regarded as a crucial step for the development and progression of lung cancer. Not only do these master regulators activate oncogenic programs, but they also define the different subtypes of lung cancer [3]. Thus, TFs are important targets for therapy and biomarkers, as the inhibition of a single TF can affect many downstream oncogenic genes. Moreover, TF expression can be used to stratify patients for therapy [4].

1.1 Key Transcription Factors in Lung Cancer

Various TFs have long been established to play a role in lung cancer pathogenesis: NF-κB (nuclear factor kappa B): A pro inflammatory TF that is often hyperactive in pulmonary cancer. NF-κB signalling, which leads to the expression of COX-2, cyclin D1, MMP9, and other genes that promote cell proliferation and survival, is driven in smoking-induced lung tumours. NF κB is also associated with tumour promotivity and therefore high NF-κB levels are associated with poor prognosis in NSCLC [5]. AP-1 (Activator Protein 1, JUN/FOS family): bZIP TF complex, which is frequently abnormally activated in lung tumors. Multi-omic studies demonstrate that AP-1 interacts with oncogene enhancers and promoters in NSCLC and induces oncogenic transcription programs in such cells [6]. STAT3 (Signal transducer and activator of transcription 3). This TF is also cytokine activated and is frequently constitutively active in NSCLC.
Angiogenesis, survival, stemness, and immune escape are being constitutively activated by STAT3. It controls anti-apoptotic

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

and proliferative genes, such as Bcl-xL, Mcl-1, and survivin, and induces drug resistance; its activation is usually associated with a poor prognosis [1]. FoxM1 (Forkhead Box M1): A Forkhead family TF highly expressed in lung cancer. FoxM1 controls cell cycle progression genes that participate in the G1/S and G2/M cell cycle transitions (cyclins, CDKs, PLK1). The FoxM1 overexpression is associated with proliferation and metastasis of NSCLC, whereas the suppression may suppress tumor growth [7].Other TFs are also major players in lung cancer development. For example, developmental regulators such as NKX2-1/TTF-1 (the "lung lineage TF" that plays the role of a "lineage survival oncogene" in adenocarcinomas) or EMT regulators such as Snail/Slug (responsible for the repression of epithelial cell identity for the acquisition of metastatic ability) are also involved [8].

Transcription factors (TFs), key regulators of gene expression, are critical for the initiation, progression, and therapeutic outcome of lung cancer [1]. Due to their central position in the regulatory network, TFs not only offer opportunities for therapeutic targeting but also serve as important biomarkers for clinical outcomes. In line with this, new therapeutic strategies targeting TFs, either through direct action or indirect interference, are being developed to combat disease progression [9]. For example, clinical trials targeting the STAT3 pathway in non-small cell lung cancer (NSCLC), as well as indirect interference with TF action by targeting transcriptional co-activators such as YAP1 and BET proteins, are being pursued. In addition to experimental approaches, in silico transcriptional regulatory network mapping has emerged as an important tool for studying lung cancer [10]. Bioinformatic approaches, which integrate genomics, transcriptomics, and proteomics, offer opportunities to reconstruct transcription factor centred gene regulatory networks in cancer, providing new avenues of investigation into the biology of the disease [11].

Notwithstanding these advances, many TFs with putative oncogenic or tumor suppressive activities in lung cancer development still require further exploration. A literature-based bioinformatics screening was performed to address this information gap, leading to the identification of 13 TFs that have not been the focus of research on lung cancer development. To evaluate their possible regulatory actions, a full set of in silico methods, such as motif enrichment analysis, target gene prediction, pathway enrichment analysis, and PPI network modeling, was used to have a complete analysis of these poorly studied TFs. The primary goal was to identify the most promising TF candidate to pursue in greater detail, thereby helping explain the complexity of TF regulatory processes and leading to new opportunities in the treatment of lung cancer.

2. Materials and Methods

2.1 Data Collection

We retrieved potential TFs correlated with lung cancer from TCGA datasets and transcription factor (TF) databases, as well as through the published literature. Candidate transcription factors (TFs) were first obtained from the literature and databases. Inclusion criteria included prior evidence of lung cancer involvement, significant differential expression in TCGA lung cancer datasets, and availability of mutation and clinical data in TCGA. TFs that either had inadequate data, showed non-significant expression patterns or operated redundantly were excluded. This led to the identification of thirteen TFs for further investigation.Gene expression, mutation, and clinical profiles were subsequently obtained from TCGA data, further validating downstream regulatory network, functional enrichment, and survival analyses using MEME Suite, STRING, ShinyGO, and cBioPortal.

2.2 Target Gene and Motif Analysis

TFlink and MEME Suite were used to analyze key TFs and their target genes. The regulatory interactions between TFs and target genes were predicted with TFlink, and MEME Suite was used to identify conserved DNA motifs in promoter regions. This helped clarify TF-DNA binding patterns and regulatory processes in lung cancer.

2.3 Network Building and Functional Enrichment

The protein-protein interaction (PPI) network, constructed using STRING, was used to investigate gene functional interactions. To assess gene functional interactions, a protein-protein interaction network (PPI) was constructed using the STRING tool. To ensure network reliability, multiple interaction sources, such as experimental data, co-expression, and database annotations, have been considered, and a confidence score has been applied. ShinyGO also conducted functional enrichment analysis to identify significantly enriched biological processes and cancer-related pathways using GO and KEGG terms.

2.4 Analysis of Genomic Alteration

Genomic alterations in the chosen TFs, including somatic mutations, copy number alterations, and mRNA expression changes, were analysed using cBioPortal and TCGA. These changes and their prevalence across different types of lung cancer were studied to determine their impact on tumour development.

2.5 Survival Analysis

The tools used were cBioPortal. Lastly, survival analysis is used to assess the clinical value of the candidate genes. Kaplan Meier plots are used to compare overall survival between groups of patients based on the presence or absence of gene expression or an alteration. Statistical significance between survival groups was assessed using the log-rank test; $p < 0.05$ was considered significant. This analysis enabled the investigation of how variations in the expression of selected genes correlate with patient prognosis, thereby supporting their potential utility as diagnostic and prognostic biomarkers. It is a step in which changes in the expression of the chosen genes are related to a patient's prognosis, making them more useful as diagnostic or prognostic biomarkers.

3. Results

3.1 Identification and curation of transcription factors

Thirteen transcription factors (TFs) were identified and curated, then selected based on their potential and necessary regulatory roles. The literature review and database confirmed that the mentioned TFs are novel and have not been previously described. The TFs are identified and sorted by regulatory type (activator or repressor). This classification provides their possible uses in gene regulation and a platform on which downstream research, including the identification of target genes, the establishment of interaction networks etc. can be performed.

3.2 Target Gene Analysis

Target gene networks were constructed to explore the regulatory relationships and downstream genes of 13 transcription factors (TFs) using the TFLink database. The total networks illustrate the complexity of TF regulation. Table 1 shows the regulatory complexity of 13 TFs in lung cancer, with AEBP1 exhibiting the most extensive scale, encompassing approximately 17 targets: proliferation via EGFR/PDGFB/NRAS/BIRC5; inflammation/apoptosis via TNFAIP3/TNFRSF10D; lipid metabolism via PPARγ1/LXRα; and EMT via POSTN. Contrarily, RFX2 has ~5 targets cilia/differentiation/signalling through Foxj and axoneme genes, and ZNF266 has ~3 targets-hypoxia/pHi regulation via CA9-mTOR axis, which indicates more specialised functions, which may restrict general oncogenesis but allow niche adaptation of tumours (e.g. metastasis under stress). Network centrality aligns with tumour pathways: high-degree hubs (BTNL9 cell cycle cluster, AEBP1) are the drivers, and low-degree specialists (RFX2 and ZNF266) could regulate differentiation.

Table 1: Transcription Factors and Their Target Genes

Transcription Factor (TF)	Target Genes
MYBL2	CCNA1, CCND1, COL1A1, MYBL2, MYC, NCL, PSG1, SP1
HLF	BRD2, CDX2, CYP21A2, EHMT2, EYA1, HLA-E, HOXD9, RARB, TNF, TNXB, AAA1, ABCD3, ACAT1, ACO1, ACTG2
PCGF2	ZEB2, HOXB genes, cell cycle regulators, HOXA7 and ZEB1
MAZ	CD4, CLCNKA, HRAS, HTR1A, MMP1, MMP14, MMP9, TSG101
RFX2	FGF1, IL5RA, Foxj1-co-bound cilia genes, axoneme/flagellar genes ADGB
PBX3	p53, p21, G6PD, SOX2, SALL2, MMP9, IL6, HMGA2
FOXS1	CXCL8, TGFB1, ADORA2A, IL6, MMP2
AEBP1	IRS1, EGFR, IL4R, PDGFB, NRAS, TNFAIP3, TNFAIP8, TNFRSF10D, TNFSF14, BIRC5, PPARγ1, LXRα, ABCA1, ABCG1, ApoE, AQP1, POSTN
TSHZ2	PRC1, GLI1, AEBP1, CXCR4, EGFR, Notch1, Hes1
TSHZ3	CASP4, N-cadherin (CDH2), Vimentin, E-cadherin (CDH1), Twist, Snail
SNAI3	MYOD, MYOG, CASP4, CDH1, CDH2
ZNF266	CA9, HZF1, mTOR pathway genes
BTNL9	E2F1, CDK1, CDK2, CDC25C, FOXM1, CCNA2, CCNB1, MCM2, MCM3, MCM7, ORC6, CDKN1A, BBC3, GADD45A, Cyclin B2

3.3 Motif Analysis

To analyze the promoter region as a conserved regulatory element, the motif enrichment analysis of the chosen transcription factors (TFs) was done with the MEME Suite. The MEME output Figure 1 showed that the motifs were significantly enriched on a large scale with the 13 transcription factors with each having specific binding motifs throughout the sequence of interest. AEBP1 had the highest number of motifs, with p-values of 4.81e-10 (MAZ), 0.00e+0 (TSH3), and 3.50e-74 (ZN266), indicating strong sequence conservation and possible regulatory activity. The plot of motif positions indicates that some transcription factors, including ZN266, TSH3, and TSH2, are densely located in the promoter region, suggesting they play a significant regulatory role. Conversely, the transcription factors PCGF2 and BTNL9 possessed well-located motifs and showed higher selective binding affinity. The fact that the motifs are separated in the sequence implies that the TFs cooperatively/combinatorially regulate the sequence.

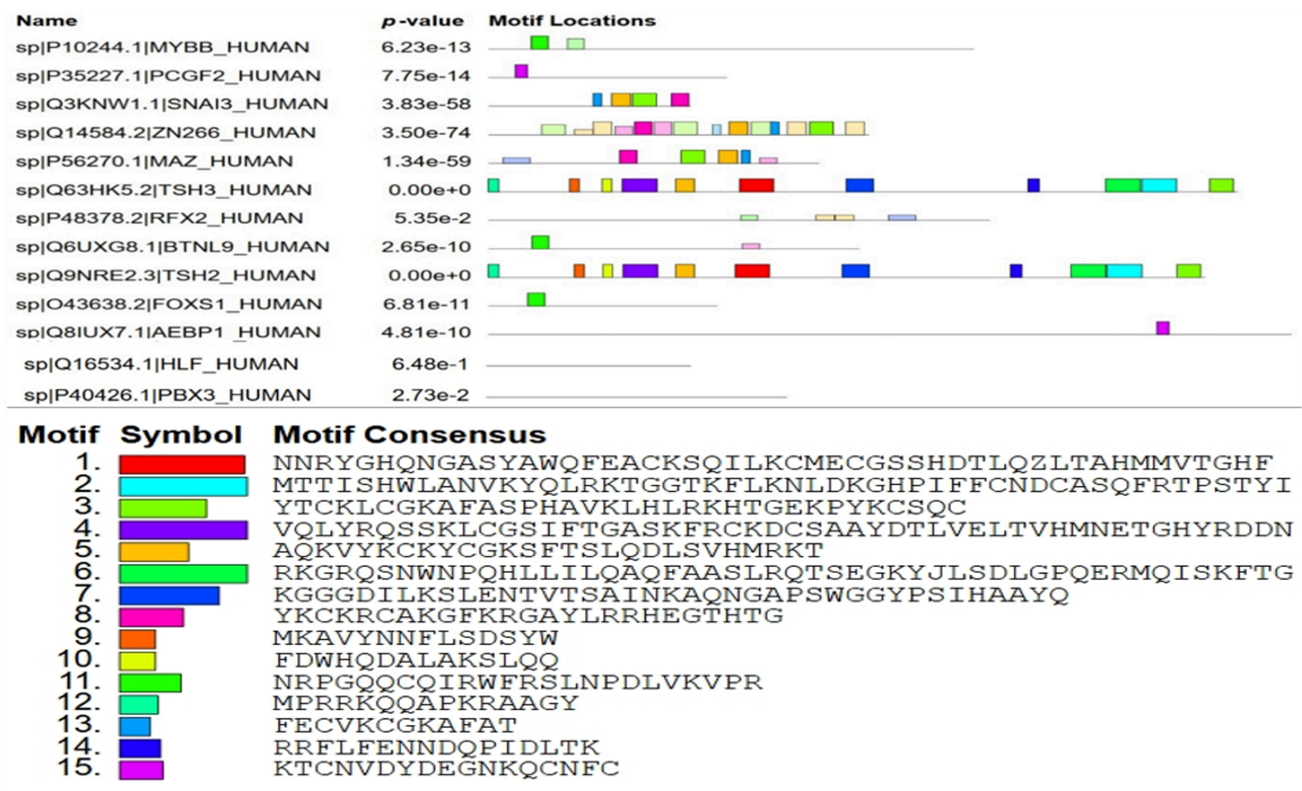


Figure 1. Motif enrichment analysis shows that transcription factor binding motifs are strongly enriched across the examined genomic regions.

3.4 Protein-protein interaction analysis

To analyse the functional relationships among the identified genes, a PPI network was constructed using the STRING database. Table 2 showed a highly interconnected network among your proteins, compared to what would be expected for a random set of proteins with a similar number and degree distribution. This suggests that the proteins are at least partially biologically related to one another. The evidence of interaction among the edges has different colors, such as experimental data, maintained databases, co-expression, and text mining. PCGF2, a Polycomb group protein (PRC1 component) known to be a transcription factor, as indicated in this network figure 2(a), is at the periphery of the interaction network and therefore has limited direct protein-protein interactions, suggesting that it regulates the cell cycle complex rather than being a part of it.

Table 2. Parameters for network construction

Network Parameter	**Value**
Number of nodes	23
Number of edges	57
Average node degree	4.96
Average local clustering coefficient	0.508
Expected number of edges	10
PPI enrichment p-value	$< 1.0 \times 10^{-16}$

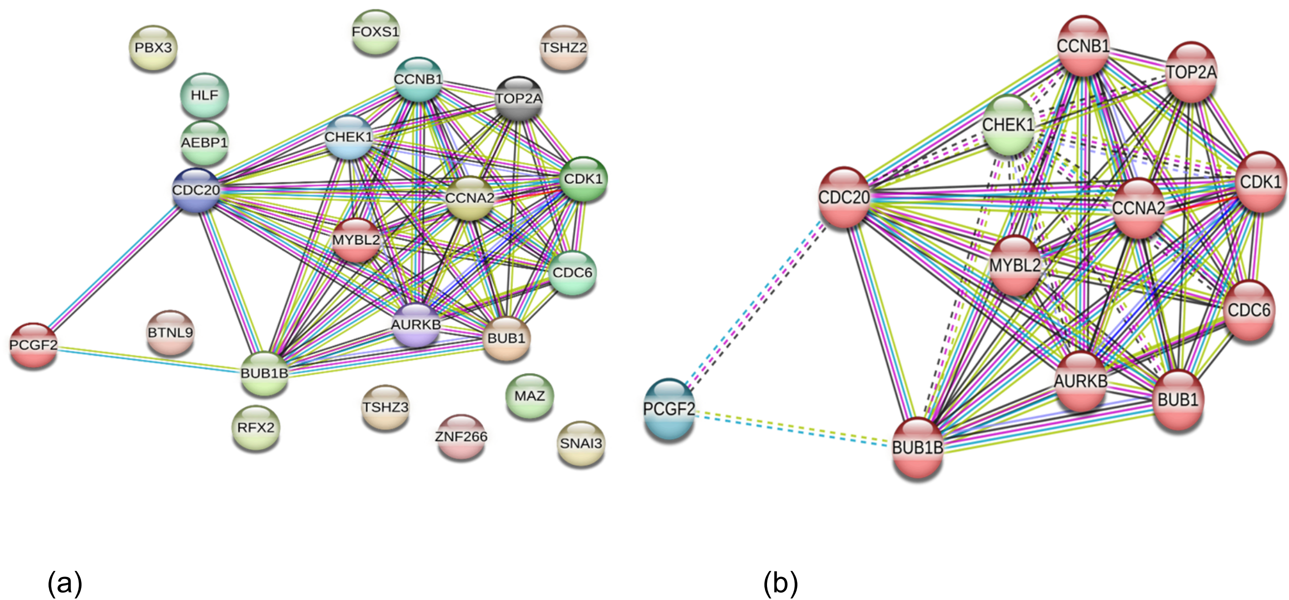


Figure 2. (a) Protein–protein interaction (PPI) network showing interactions among the selected genes; (b) STRING PPI network showing three functional gene clusters. The network is divided into distinct colour-coded clusters based on interaction density and functional similarity.

In Figure 2(b), the network is clustered using the k-means method, with colour-coded clusters based on interaction density and the functional similarity of different nodes. Cluster 1 (in red) has over 10 genes that encode the fundamental cell cycle functions and Cluster 2 (in green) CHEK1 and Cluster 3 (in blue) PCGF2 are represented as a single-gene cluster due to their functions as regulatory proteins failing to be part and parcel of the main mitotic process e.g. CHEK1 in the DNA damage checkpoint and PCGF2 in epigenetics. The list of carefully selected genes was analyzed with the help of ShinyGO to determine highly enriched GO terms and pathways.

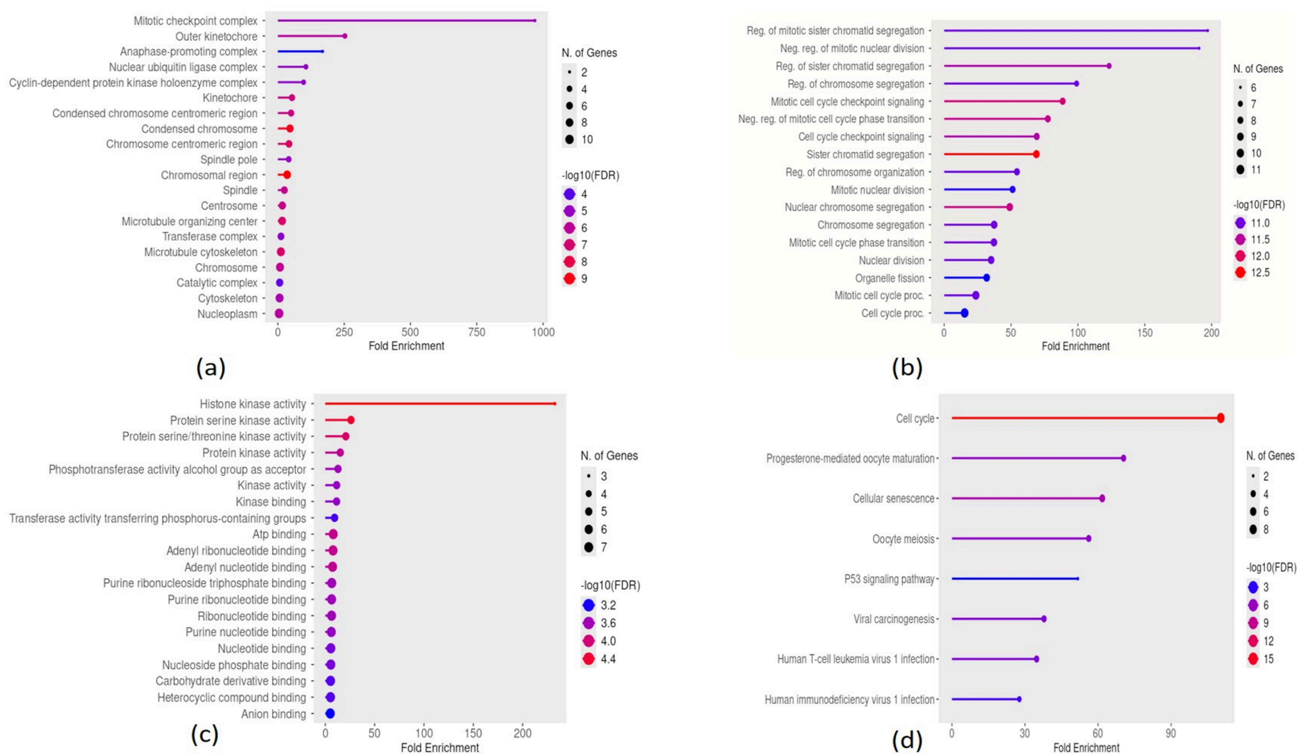


Figure 3. Function Enrichment Analysis (a) Gene Ontology (Cellular Component), (b) Biological Process, (c) Molecular Function, (d) KEGG pathway

As illustrated in this analysis, the Gene Ontology (Cellular Component) Figure 3(a) shows high enrichment for the activities of histone kinase, serine/threonine kinase, protein kinase, and other activities involving nucleotide binding. Likewise, Gene Ontology (Biological Process) (Figure 3(b) shows highly significant enrichment for the regulation of mitotic segregation of sister chromatids and for negative regulation of mitotic nuclear division. The Gene Ontology (Molecular Function) Figure 3(c) also illustrates significant enrichment for histone kinase activity, serine/threonine kinase activity, protein kinase activity, and other nucleotide-binding activities. In addition, Figure 3(d) shows KEGG pathway analysis illustrates high enrichment for cell cycle, oocyte maturation, p53 signalling, and pathways of viral carcinogenesis.

3.5 Genomic Alteration Analysis

Genomic alteration analysis of PCGF2 in lung cancer datasets is done by using cBioPortal. The figure 4 bar chart shows the frequency and nature of genomic alterations of PCGF2 in various lung cancer datasets, including lung adenocarcinoma, lung squamous cell carcinoma, and small cell lung cancer. The alterations are shown in green (mutations), red (amplifications), blue (deep deletions), and gray (multiple alterations). PCGF2 had the highest and most consistent mutation frequency in Lung Adenocarcinoma (LUAD), with approximately 3.2% in the LUAD dataset (Figure 5). Gene amplification (red) was identified as the most noticeable alteration in the majority of LUAD datasets, indicating that the mutation state of PCGF2 in adenocarcinoma is amplification-dominated rather than mutation-dominated. This amplification-dominant trend demonstrates an oncogenic dose effect, and thus it could be suggested that an elevated expression of PCGF2 could play a role in tumor growth. Lung Squamous Cell Carcinoma (LUSC), on the other hand, had lower alteration frequencies (about 2.4 per cent) and a more complex interplay between mutation and amplification events. The amplification of the dominant and repetitive PCGF2 structure, particularly in LUAD, suggests that this protein may act as a functional epigenetic factor in the Polycomb group pathway in NSCLC and warrants further mechanistic research.

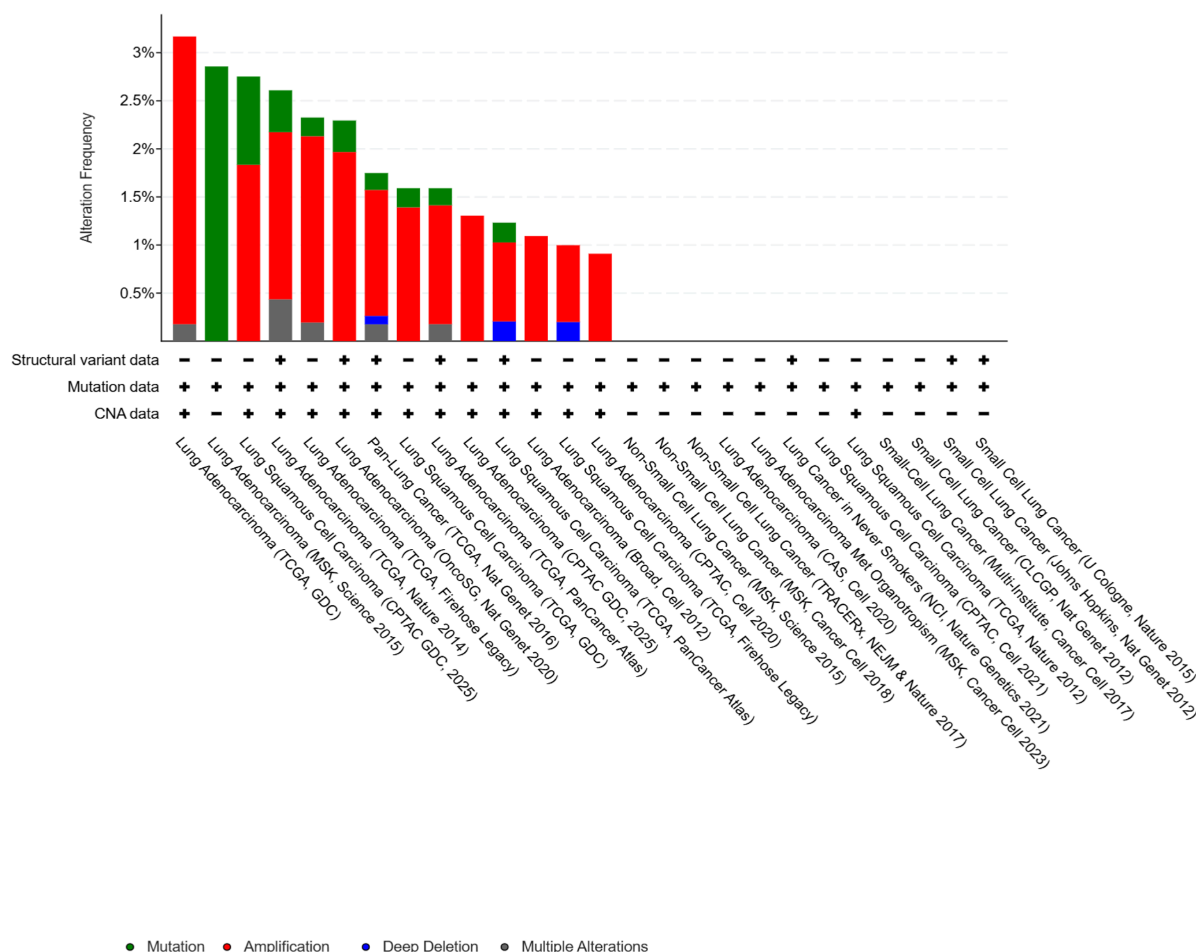


Figure 4. Alteration frequency of PCGF2 across lung cancer cohort

The Oncoprint visualisation of PCGF2 genomic alteration across multiple lung cancer cohorts in cBioPortal, showing a low overall alteration frequency of 1.5%. Most of the patient bars were grey, indicating no change, but only a few columns had colourful markers indicating missense, nonsense, truncating mutations, amplifications, and deep deletions. This means that PCGF2 was altered, but infrequently, in lung cancer studies.

In the lollipop plot in Figure 5, PCGF2 is shown with distributions and frequencies of somatic mutations across its 344-amino acid protein sequence; each dot indicates a mutation, and the height represents the number of patients with that mutation. The repetitive E116K mutation is underlined and of specific interest. E116K is a missense mutation in which glutamic acid is replaced by lysine at the 116 position, which causes a charge change that affects the RING finger domain. The green-colored zf-C3HC4 (RING finger) domain, situated at the N-terminus, has ubiquitin ligase activity itself.

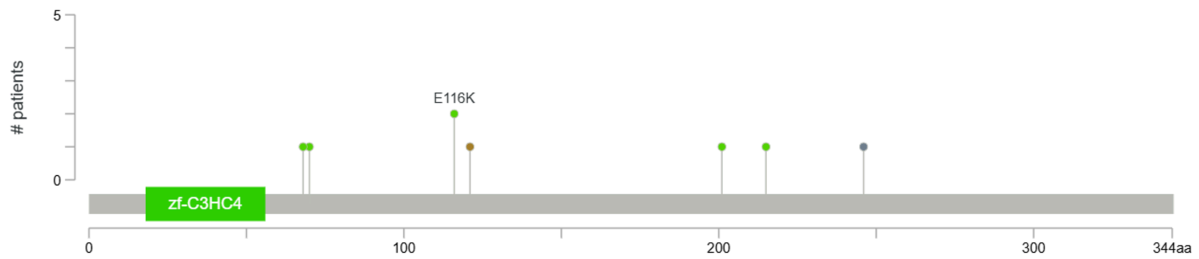
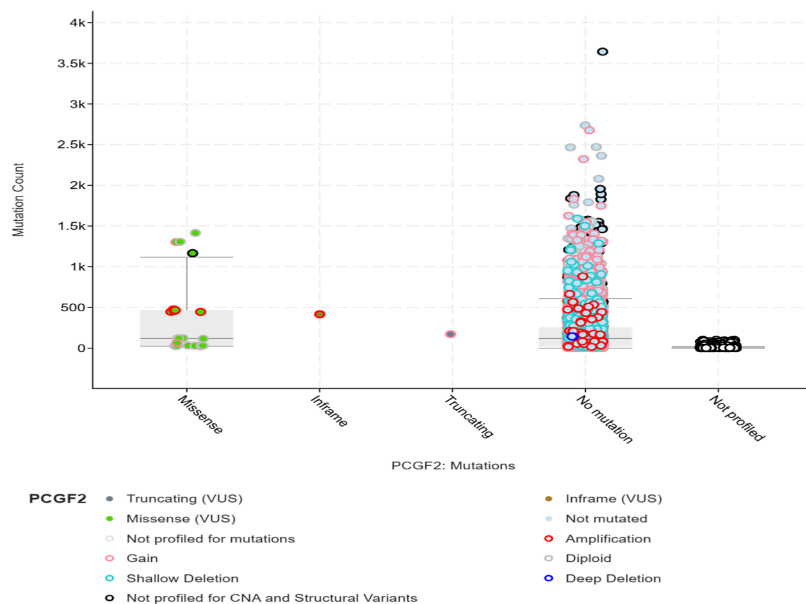


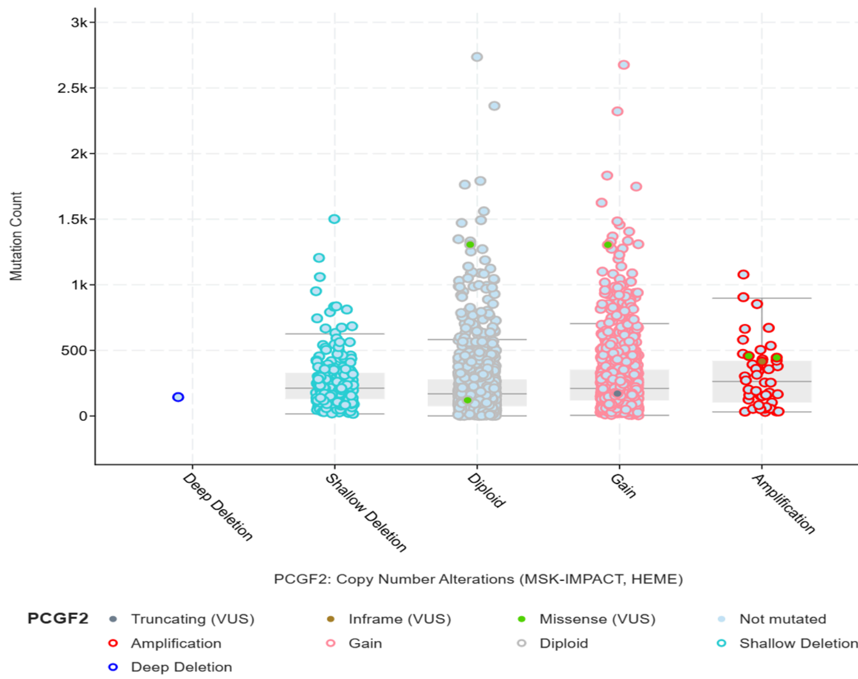
Figure 5. Lollipop plot of PCGF2 mutations, showing their positions and frequencies across functional protein domains in lung cancer.

3.5.1 Correlation of PCGF2 mRNA expression with genetic alterations

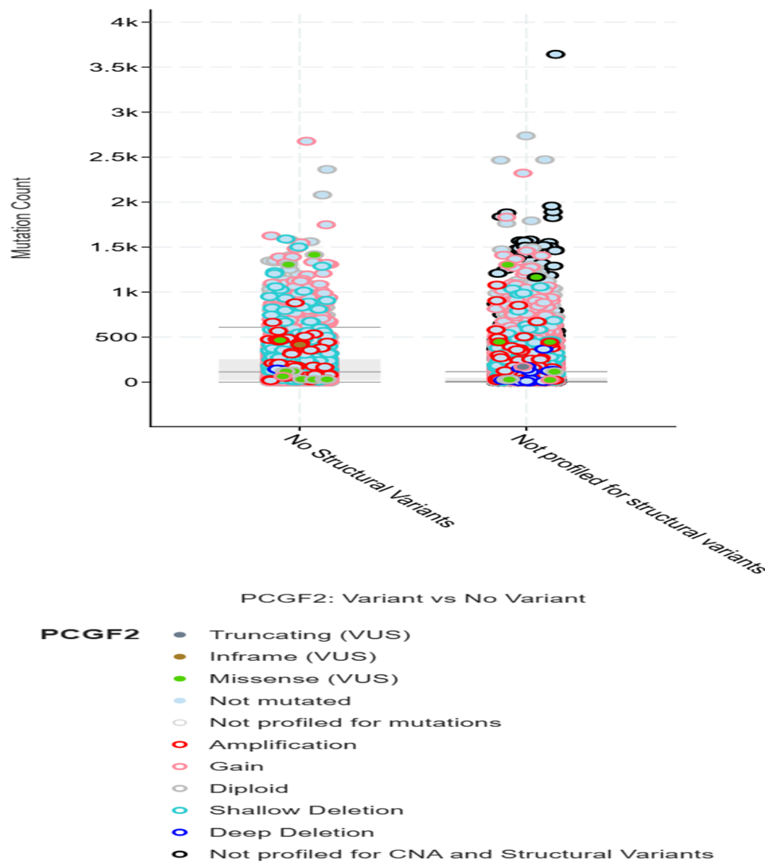
Figure 6(a)indicates how mutation burden is distributed across different categories of mutation in PCGF2. It indicates that most samples fall into the non-mutated category, with a variety of mutation counts, including many high-mutation-burden outliers. There aren't many that fall under "missense," "in-frame," or "truncating mutations," and those that do tend to have lower mutation counts. The "not profiled" group also has very few mutations. This indicates that PCGF2 mutations are quite rare and that, most of the time, tumours with a high mutation rate don't have a direct PCGF2 mutation. Figure 6(b) illustrates the distribution of mutation count in various PCGF2 copy number variation states, including deep deletion, shallow deletion, diploid, gain, and amplification. From this figure, it is clear that the majority of samples are in the diploid and gain states, with high mutation counts. In addition, samples in the amplification state have intermediate mutation loads, while those in the deep deletion state have low mutation loads, with few samples in this state. In general, this figure shows a relationship between increased PCGF2 copy number and high genomic instability, as illustrated by high mutation loads.



(a)



(b)



(c)

Figure 6. Correlation of mRNA expression with (a) mutations; (b) putative copy number alterations; (c) structural variants.

Figure 6(c) compares the mutation load in samples that have not been screened for structural variations in comparison to those

that have not been screened for structural variants in relation to PCGF2. From this figure, it is clear that in both samples, the majority have low to moderate mutation loads, although a particular group is associated with high mutation loads. In addition, there is no significant difference in mutation loads between these samples, as illustrated by the extensive range of points, including various copy number variation states with high mutation loads, indicating that there is no difference in mutation loads between samples that have been screened for structural variants in comparison to those that have not been screened for structural variants in relation to PCGF2.

3.6 Survival Analysis

Kaplan Meier analysis for overall survival. Figure 7 shows a state of stability in the difference between the altered (red) and non altered (blue) groups, suggesting poorer survival of altered genes compared to non-altered genes. It also shows that the altered genes tend to have a more rapid decline in survival state than the non-altered genes, which maintain a better survival state for a longer time, even beyond 200 months. From the summary of survival plot shown in Table 3, the altered demographic has 80 cases, 34 events, with a median overall survival of 31.73 months (CI: 23.03-NA), while the unaltered demographic has 4,939 cases, 1,840 events, with a median overall survival of 53.33 months (CI: 49.10-55.73). However, the p-value of the log-rank test, 0.0604, does not fall below the traditional level of statistical significance but is close enough to suggest a trend of biological interest, with the gene modification correlating with decreased overall survival. Overall, the study's results suggest that the change may be negative, although confirmation in larger or different populations is necessary.

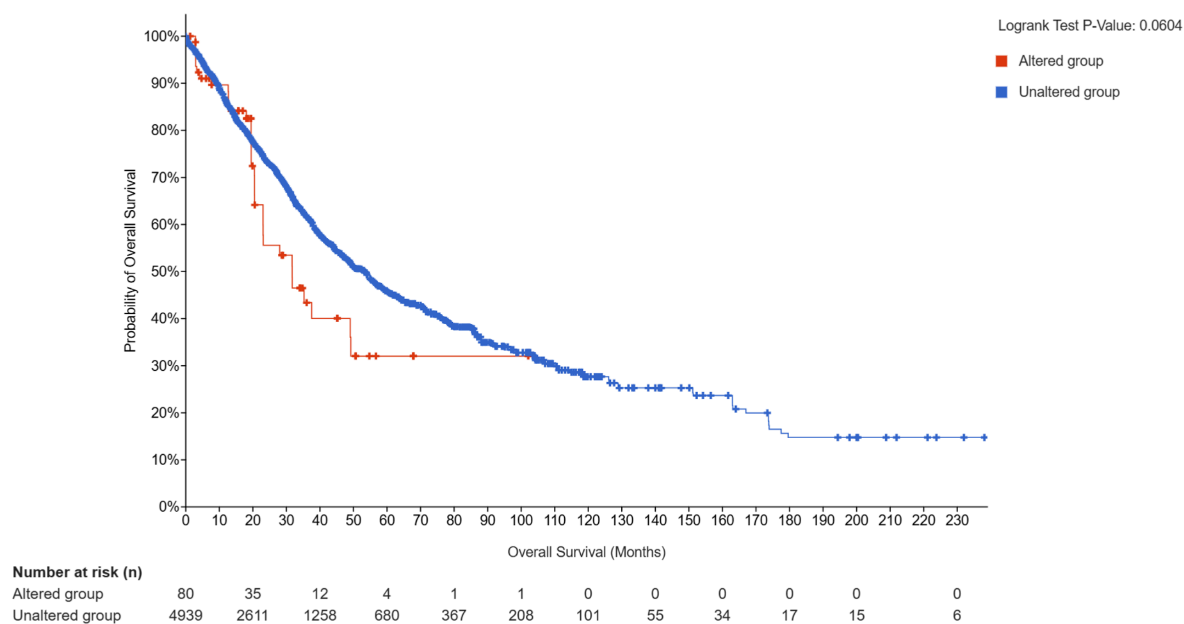


Figure 7. Kaplan–Meier overall survival study comparing patients with and without gene alterations, showing no significant difference in survival between the altered and unaffected groups (log-rank p = 0.0604).

Table 3. Survival plot summary

Group	Number of Cases, Total	Number of Events	Median Months Overall Survival (95% CI)
Altered group	80	34	31.73 (23.03 – NA)
Unaltered group	4939	1840	53.33 (49.10 – 55.73)

4. Discussions

Lung cancer is a significant cause of cancer-related deaths because it is heterogeneous in terms of its molecular heterogeneity and has no prognostic biomarkers. Transcription factors (TFs) are essential regulators of oncogenic transcriptional regulation. In this work, three understudied transcription factors were profiled using integrative in silico methods, each with distinct patterns

regarding complex regulation, functional redundancy, and clinical relevance. The TFs are heterogeneous in terms of regulatory connectivity, as revealed by target gene network analysis. Oncogenic cell proliferation, apoptosis, inflammation, and cell cycle regulation are among the genes involved that exhibit widespread downstream regulatory interactions with AEBP1 and BTNL9, suggesting they are transcriptional hubs that may increase oncogenic signalling pathways. Instead, their transcription factors, such as ZNF266 and RFX2, had fewer, yet more differentiated, target genes linked to hypoxia adaptation, differentiation, and ciliary activity, suggesting they are more context-dependent. Furthermore, motif enrichment revealed functional heterogeneity in relation to highly conserved transcription factors such as ZNF266, MAZ, and TSHZ3, suggesting that they may be highly specific in binding to DNA sequences. On the contrary, PCGF2 and BTNL9 exhibited fewer, but more specifically located, motifs that can be linked to more focused or epigenetic forms of regulation. The analysis of protein-protein interactions revealed a highly connected network, with the centre composed of mitosis control, chromosome separation, and checkpoint signalling pathways, underscoring the importance of cell cycle misregulation in the onset of lung cancer. PCGF2 lies at the edge of this cluster, but is still relevant, suggesting that it is also involved in upstream epigenetic activities. Based on this, genomic analysis has shown that PCGF2 is an amplification candidate that is highly recurring in lung adenocarcinoma and that all these transcription factors collaborate to regulate the stability of the genome and pathways of tumor suppressor activity in lung cancer initiation. In addition, comparing the alteration patterns of the PCGF2 gene between lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) is of great significance. It is evident that PCGF2 amplification is much higher in LUAD than in LUSC. It is believed that the PCGF2 gene plays a role in LUAD tumorigenesis through epigenetic regulation. Furthermore, alterations in the PCGF2 gene in LUSC might be related to other mechanisms, such as transcriptional modulation. It is evident that the molecular characteristics of the two subtypes differ, with the tumorigenesis of LUAD being related to chromatin remodeling. From a clinical perspective, the enrichment of PCGF2 amplification in LUAD emphasizes its value as a subtype specific biomarker for prognosis and therapeutic stratification, particularly for epigenetic-targeted therapies. In contrast, the lack of PCGF2 amplification in LUSC suggests that alternative mechanisms might be more important in this subtype. Based on its known function in Polycomb-mediated transcriptional repression and its ubiquitin ligase activity, PCGF2 amplification might contribute to tumor suppressor silencing and genomic instability in LUAD. Further studies will be necessary to validate its potential as a target for precision oncology. Even though mutations are not extremely repeatable in these transcription factors, mutations in the RING finger area may likely interfere with the ubiquitin ligase pathway and Polycomb-regulated transcriptional suppression. The survival study showed a poor survival rate in total control patients with changes in these transcription factors.

Collectively, our findings indicate that the effects of these transcription factors on lung cancer initiation are important and involve the regulation of coarse transcriptional activation and fine-tuning of epigenetics. PCGF2 is one of the most promising transcription factors and should be investigated further. Though our work is premised on predictions and publicly available databases, it offers a clear framework for utilising transcription factors as a priority for lung cancer biomarkers and therapeutic targets.

5. References

[1] Parakh S, Ernst M, Poh AR. Multicellular effects of STAT3 in non-small cell lung cancer: Mechanistic insights and therapeutic opportunities. Cancers (Basel) 2021;13:6228. https://doi.org/10.3390/cancers13246228.

[2] Papavassiliou KA, Anagnostopoulos N, Papavassiliou AG. Lung cancer through transcription factors. Int J Mol Sci 2023;24:9461. https://doi.org/10.3390/ijms24119461.

[3] Ozen M, Lopez CF. Data-driven structural analysis of Small Cell Lung Cancer transcription factor network suggests potential subtype regulators and transition pathways. bioRxiv.org 2023. https://doi.org/10.1101/2023.04.01.535226.

[4] Mitra P. Targeting transcription factors in cancer drug discovery. Explor Target Antitumor Ther 2020;1:401–12. https://doi.org/10.37349/etat.2020.00025.

[5] Zhang L, Ludden CM, Cullen AJ, Tew KD, Branco de Barros AL, Townsend DM. Nuclear factor kappa B expression in non-small cell lung cancer. Biomed Pharmacother 2023;167:115459. https://doi.org/10.1016/j.biopha.2023.115459.

[6] Tan X, Kroneberg M, Sun F, Diaz KA, Flora A, Carey MF. AP-1 promotes oncogenic transcription in lung cancer cells by bridging promoter-enhancer interactions. Cancer Gene Ther 2025:1–13. https://doi.org/10.1038/s41417-025- 00974-w.

[7] Wang Z, Ahmad A, Li Y, Banerjee S, Kong D, Sarkar FH. Forkhead box M1 transcription factor: a novel target for cancer therapy. Cancer Treat Rev 2010;36:151–6. https://doi.org/10.1016/j.ctrv.2009.11.006.

[8] Mollaoglu G, Jones A, Wait SJ, Mukhopadhyay A, Jeong S, Arya R, et al. The lineage defining transcription factors SOX2 and NKX2-1 determine lung cancer cell fate and shape the tumor immune microenvironment. Immunity 2018;49:764-779.e9. https://doi.org/10.1016/j.immuni.2018.09.020.

[9] Zhuang JJ, Liu Q, Wu DL, Tie L. Current strategies and progress for targeting the "undruggable" transcription factors. Acta Pharmacol Sin. 2022 Oct;43(10):2474-2481. doi: 10.1038/s41401-021-00852-9. Epub 2022 Feb 7. PMID: 35132191; PMCID: PMC9525275.

[10] Tang C, Hartley GP, Couillault C, Yuan Y, Lin H, Nicholas C, et al. Preclinical study and parallel phase II trial evaluating antisense STAT3 oligonucleotide and checkpoint blockade for advanced pancreatic, non-small cell lung cancer and mismatch repair-deficient colorectal cancer. BMJ Oncol 2024;3:e000133. https://doi.org/10.1136/bmjonc-2023-000133.

[11] Li D, Yang W, Zhang J, Yang JY, Guan R, Yang MQ. Transcription Factor and lncRNA Regulatory Networks Identify Key Elements in Lung Adenocarcinoma. Genes 2018;9(5):251. https://doi.org/10.3390/genes9050251.