Received: 23 August 2017 | Accepted: 4 December 2017
DOI: 10.1002/jcb.26603

RESEARCH ARTICLE (FASTTRACK)

WILEY Journal of Cellular Biochemistry

Database of transcription factors in lung cancer (DBTFLC): A novel resource for exploring transcription factors associated with lung cancer

Thabitha Amalraj | **Ameya A. Dravid** | **Riya Tripathi** | **Sajitha Lulu S**

Department of Biotechnology, School of Bio Sciences and Technology, VIT University, Vellore, Tamilnadu, India

Correspondence
Dr. S. Sajitha Lulu, Department of Biotechnology, School of Bio Sciences and Technology, VIT University, Vellore, Tamil Nadu, India.
Email: ssajithalulu@vit.ac.in

Abstract

Lung cancer is considered as the most prevalent form of cancer and it is found to be frequent cause of cancer related death. Even though, approved molecular targeted therapies other than chemotherapy are currently unavailable, the mechanism of pathogenesis in lung cancer remains still unclear. Transcription factors (TFs) play a critical role in cancer cell processes, such as cell proliferation, apoptosis, migration, and regulate gene expression. Thus, the identification and characterization of transcription factors involved in lung cancer would provide valuable information for further elucidation of the mechanism(s) underlying pathogenesis and the identification of potential therapeutic target types, which are critical for the development of therapeutic strategies. Through an extensive literature survey, we have identified 349 transcription factors noted for their strong involvement in lung cancer. Database of Transcription Factors in Lung Cancer (DBTFLC) was constructed as a data repository and analytical platform for systematic collection, curation of TFs and their interacting partners. The database includes all pertinent information such as lung cancer related TFs, chromosomal location, family, lung cancer type, references, TF-TF interaction-(s), and TF-target gene interaction(s); thus, it could serve as a valuable resource for therapeutic studies in lung cancer. The database is freely available at http://www.vit.ac.in/files/database/Home.php

KEYWORDS

cytoscape, lung cancer, TF-TF interaction, transcription factors, transcriptome

1 | INTRODUCTION

Lung cancer is the second most common deadly disease, leading to approximately 155 870 deaths per year globally.[1] Lung cancer can be classified into two types: non-small cell lung cancer (85% of cancerous cells) and small cell lung cancer (15% of cancerous cells).[2] Over time, cancer cells appear to have developed resistance to chemotherapeutic drugs that target cancer genes, which has led to a dire need for developing new approaches to overcome these issues. Approximately 7% of proteins have been described as transcription factors (TFs), many of which are associated with cancer progression in the human proteome.[3,4] One transcription factor can affect many gene functions and thus TFs can be proactive in inhibiting the activity of malignant cells.[5,6] The critical role played by TFs in developmental, sensory, and signaling regulatory processes seems to be highly intriguing. For example, signal-transducing

transcription factors play a major role in cell cycle progression, differentiation, and apoptosis.[7,8] TFs are promiscuous molecules for therapeutics and are often found to be overactive in diseases.[9] These proteins also affect the formation of gene products thought to be involved in metastasis and tumor growth. Importantly, TFs contain specific residues and binding sites important for protein-protein and protein-DNA interactions, which may be amenable to therapeutic targeting.[10,11] This paves the way for developing delivery systems that specifically target tumor cells by targeting transcription factors.

Here, we developed a DBTFLC database consisting of TFs involved in lung cancer with subtype information identified via the manual curation of PubMed abstracts. The database also details chromosomal location, family, and TF-target and TF-TF interactions (Figure 1). For efficient curation, we used a keyword-based text-mining approach in which keywords that might pertain to transcriptional regulation were first extracted and then subjected to manual curation. We believe that the DBTFLC data could serve as a useful benchmark for computational reconstruction of human lung cancer regulatory networks.

2 | MATERIALS AND METHODS

2.1 | Dataset collection

Lung cancer microarray datasets were downloaded from the Gene Expression Omnibus (GEO) database, accession number GSE2514, GSE43346, and GSE3268. These datasets were pre-processed and analyzed using the LIMMA package in R program. GCRMA (Gene Chip Robust Multiarray Averaging) method normalization and a moderated *t*-statistic were used to identify the differentially expressed genes (DEGs) between normal and Lung cancer samples. The log2-fold change (log2FC)>2 and adjusted *P*-value <0.05 were considered as cut-off value for up regulated and down regulated genes. The differentially expressed transcription factors (DETFs) were obtained manually from DEGs in each dataset (Supplementary Table SII). In addition to that, TFs were identified using the keyword search, "transcription factors associated with lung cancer, lung adenocarcinoma, non-small cell lung cancer, small cell lung cancer, and lung neoplasms" from the PubMed database (Figure 2).

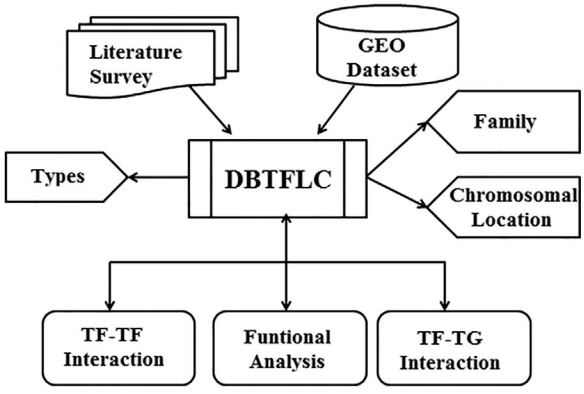


FIGURE 1 Schematic representation of DBTFLC

2.2 | Data resources

The Database of Transcription Factors in Lung Cancer (DBTFLC) integrates the following transcription factor related resources: TRRUST,[12] TcoF-DB v2,[13] PAZAR,[14] OregAnno,[15] ITFP,[16] TRED,[17] HTRI,[18] Tfacts,[19] STRING v9.1,[20] and RegNetwork.[21] Transcription factors, families, and chromosomal location are annotated according to Uniprot (http://www.uniprot.org/), HGNC (http://www.genenames.org/cgi-bin/genefamilies/), and NCBI (https://www.ncbi.nlm.nih.gov/genome/tools), respectively.

2.3 | Database construction

The data resource in the DBTFLC has been designed as a relational model and stored in the MySQL database system. PHP language scripts were used to generate the data on a web interface; briefly, the database is queried via PHP and the Apache web server is used as the server software. Website design was conducted using CSS and basic HTML. The graphic network between genes displayed with each result was executed using Cytoscape-embedded in JavaScript. All procedures were performed with the Windows Operating System. Gene symbols were used as the main keys to organize and link all of the tables.

2.4 | Functional analysis

To study the biological functions of selected transcription factors, we used Target explorer tool and Gene Ontology online tool with a threshold *P*-value <0.05.[22] The biological pathways of the identified transcription factors were retrieved using the pathway mapping tool of KEGG pathway database.[23]

3 | RESULTS AND DISCUSSION

3.1 | Database description

This database allows the user to retrieve transcription factors involved in lung cancer and pertinent factors such as, chromosome location, family, and type of lung cancer, and the network representation shows TF-TARGET GENE and TF-TF interactions (Supplementary Table SI). For a more user-friendly interface of this database, Cytoscape embedded in JavaScript were used to visualize the interactions.

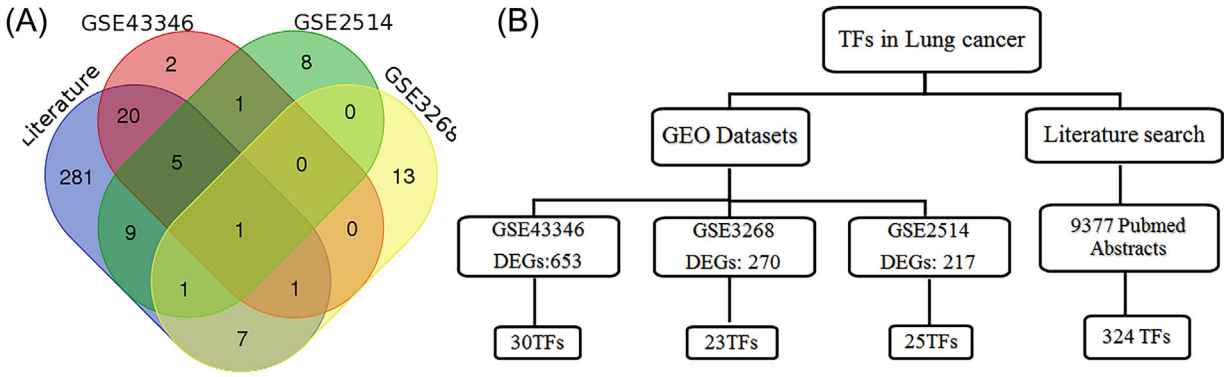


FIGURE 2 Flowchart of lung cancer involving TFs dataset construction

First, in the DBTFLC database, we have identified 68 differentially expressed TFs in lung cancer dataset and 324 lung cancer involving TFs were manually curated from the literatures till the year of 2016 (Figure 2). Overall 349 TFs were obtained, in which 249 TFs having minimum of three publications in lung cancer and 24 TFs were not having any literature evidence. Further, we recognized the target gene and interacting transcription factors for identified TFs from available TF related public databases. In this study, we assembled a human TF-TF and TF-TARGET GENE network. The overall statistics of the data are listed in Table 1.

3.2 | Database interface

An overview of available search and browse options is shown on the home page screen (Figure 3). The search option of the page includes Transcription Factors (TFs), TF-TF interaction, TF-TARGET GENE (TF-TG) interaction and functional analysis, which facilitates the user to browse the various features of TFs.

3.2.1 | Browse by transcription factor

The DBTFLC web interface allows searching by either alphabetical order of TF list or TF name, which facilitates the user to search particular transcription factors. After entering to the respective TF page, the database will provide complete information about the TF which includes TF family, chromosomal location, lung cancer type, and citation for role in lung cancer (Figure 4).

3.2.2 | Browse by TF interaction categories

This option facilitates the user can browse transcription factor interaction data in DBTFLC according to their query. The TF-TF interaction web page comprises of details about the identified 349 TFs and their interacting TFs present in human. After entering the TF of interest and click on the enter button, the same page will show the network representation of TF-TF interaction. Similarly, user can also browse target genes in the TF-TARGET GENE interaction option by entering the TF name. The page will be showing network of user query TF with TF targeting genes. For example, the interactions of transcription factor ATF1 are shown in Figures 5 and 6.

3.2.3 | Browse by function categories

This page comprises of hyperlinks for gene ontology and pathway search tool. By clicking on the tool, link redirects to the home page of Gene ontology consortium and the KEGG pathway database (Figure 7). In this option, search TF of interest to the hyperlinked tools, which enables users to do further analysis.

DBTFLC represents a valuable resource for the study of TF in lung cancer regulation and play a key controlling

TABLE 1 The available data in DBTFLC database

S.No	Data type	Statistics
1.	Transcription factors from Literature	324
2.	Transcription factors from Microarray dataset	68
3.	Lung cancer involving Transcription factors	349
4.	Number of other transcription factors in TF-TF Network	603
5.	Number of target genes in TF-TG Network	18 445

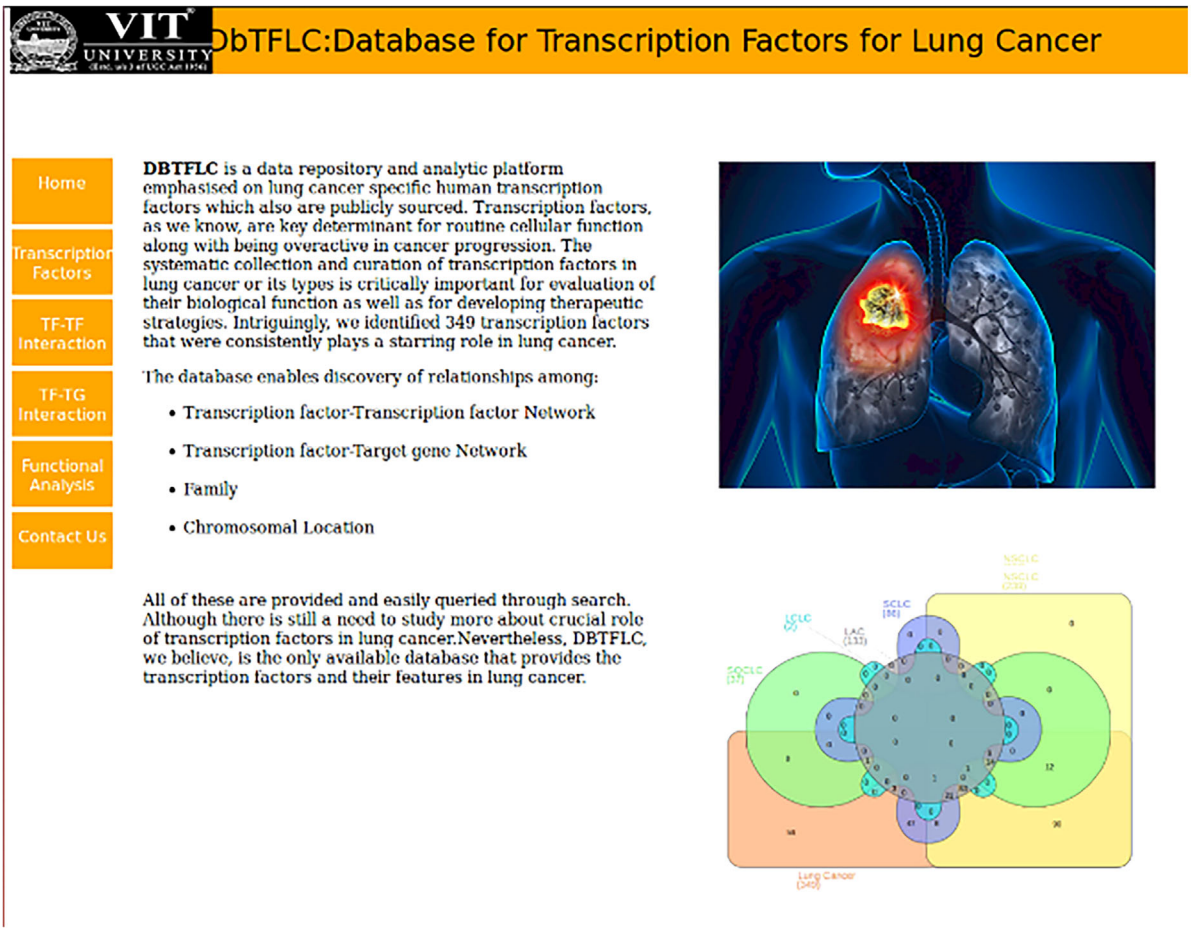
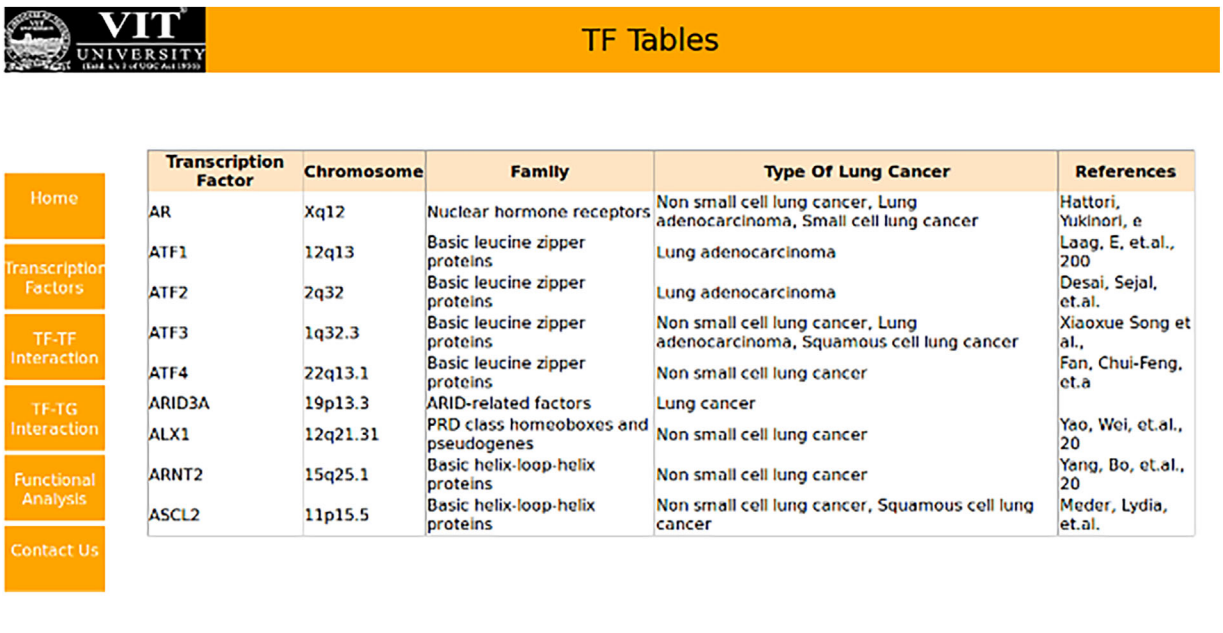


FIGURE 3 Home page of DBTFLC



Transcription Factor	Chromosome	Family	Type Of Lung Cancer	References
AR	Xq12	Nuclear hormone receptors	Non small cell lung cancer, Lung adenocarcinoma, Small cell lung cancer	Hattori, Yukinori, e
ATF1	12q13	Basic leucine zipper proteins	Lung adenocarcinoma	Laag, E, et.al., 200
ATF2	2q32	Basic leucine zipper proteins	Lung adenocarcinoma	Desai, Sejal, et.al.
ATF3	1q32.3	Basic leucine zipper proteins	Non small cell lung cancer, Lung adenocarcinoma, Squamous cell lung cancer	Xiaoxue Song et al.,
ATF4	22q13.1	Basic leucine zipper proteins	Non small cell lung cancer	Fan, Chui-Feng, et.a
ARID3A	19p13.3	ARID-related factors	Lung cancer	
ALX1	12q21.31	PRD class homeoboxes and pseudogenes	Non small cell lung cancer	Yao, Wei, et.al., 20
ARNT2	15q25.1	Basic helix-loop-helix proteins	Non small cell lung cancer	Yang, Bo, et.al., 20
ASCL2	11p15.5	Basic helix-loop-helix proteins	Non small cell lung cancer, Squamous cell lung cancer	Meder, Lydia, et.al.

FIGURE 4 Transcription factors search module

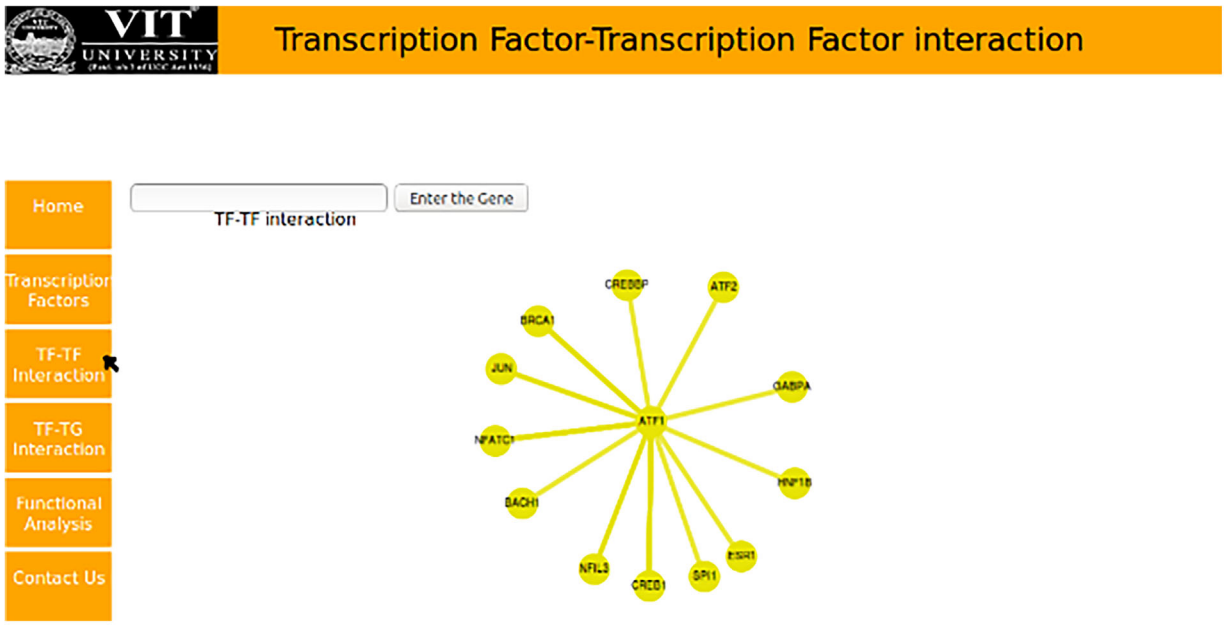


FIGURE 5 TF-TF interaction network module

role in various functional enrichments like pathway and biological processes. Here, we have done gene ontology and pathway analysis for 349 TFs identified in this study. Among them, 323TFs were involved in cellular metabolism, cellular physiological process, and biological regulation. There are 73 TFs involved in organogenesis, 10-50 TFs were involved in wounding and cell differentiation, cell cycle, stress, defense response, cell proliferation and below 10 TFs were involved in angiogenesis, cell growth, immune response, apoptosis, and cell death (Figure 8). Out of 349 TFs, 38 TFs were involved in pathways related cancer and signaling pathway which showed in Figure 9. However, only 12 TFs were reported in lung cancer pathway of KEGG database.

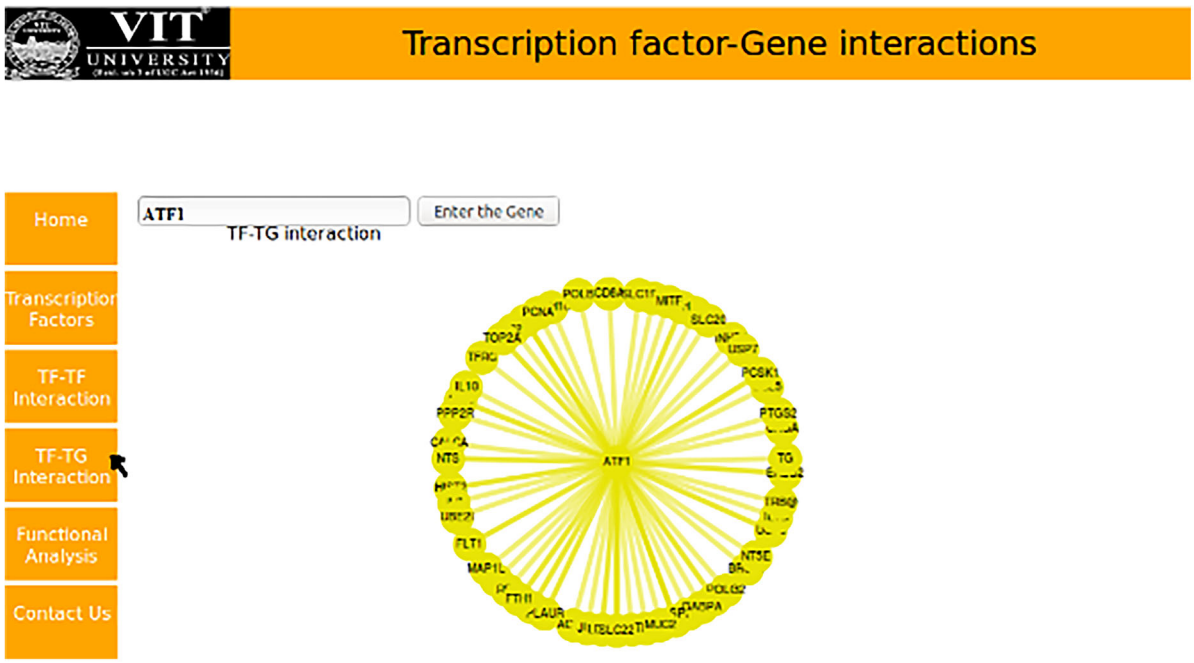


FIGURE 6 TF-TARGET GENE interaction network module

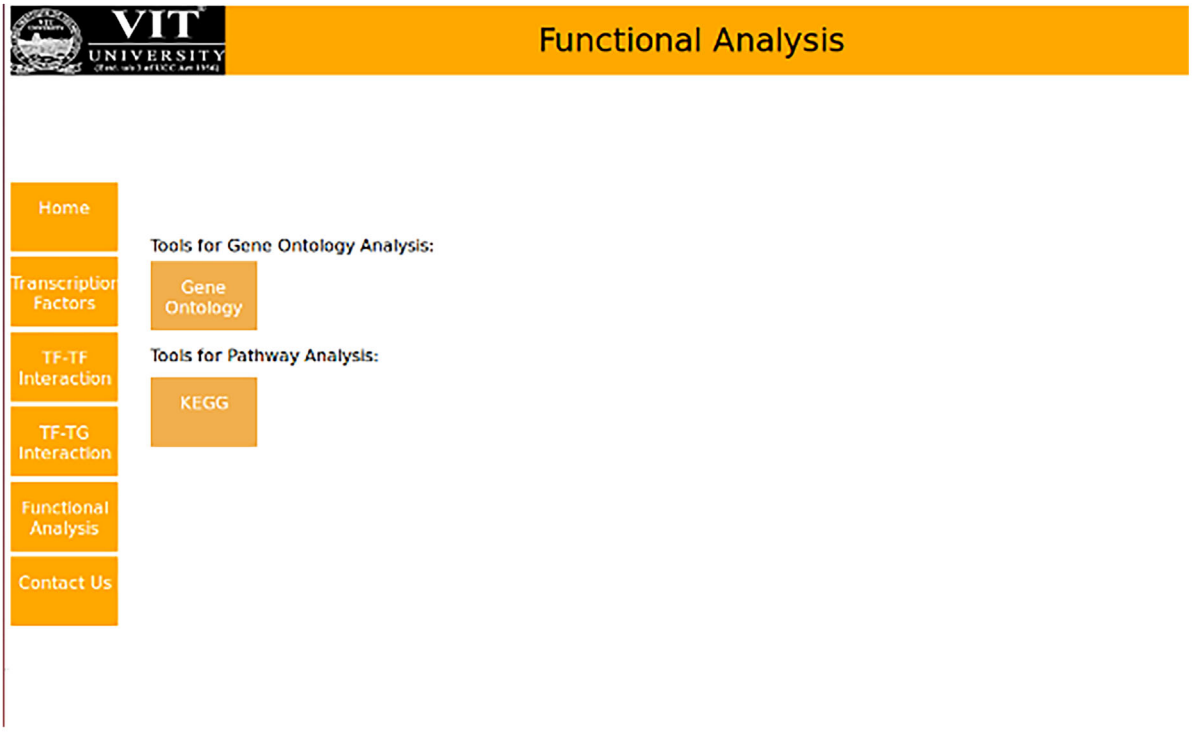


FIGURE 7 Functional analysis search module

In addition to micro-array data, we have compared with RNA-seq data in normal versus lung adenocarcinoma and normal versus lung squamous cell carcinoma from cancer nexus RNA-seq database to get substantially deeper information on differentially expressed coding transcripts relating with identified transcription factor in lung cancer. Among 349 TFs, 314 TFs (251-Lung adenocarcinoma and 282-Squamous cell lung cancer) shows significant transcript expression and tabulated in supplementary Table S3. From 64 differentially expressed TFs identified from microarray data, 59 of them showing statistically significant expression in RNA-seq data of lung cancer excluding NONAG, PAX4,

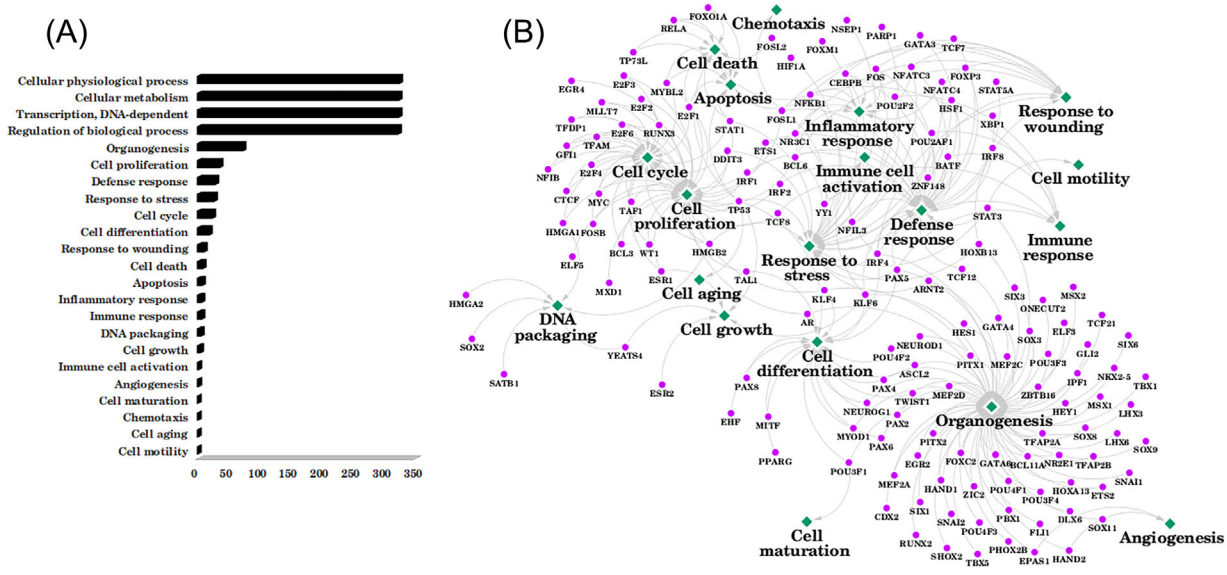


FIGURE 8 Gene Ontology analysis of transcription factors

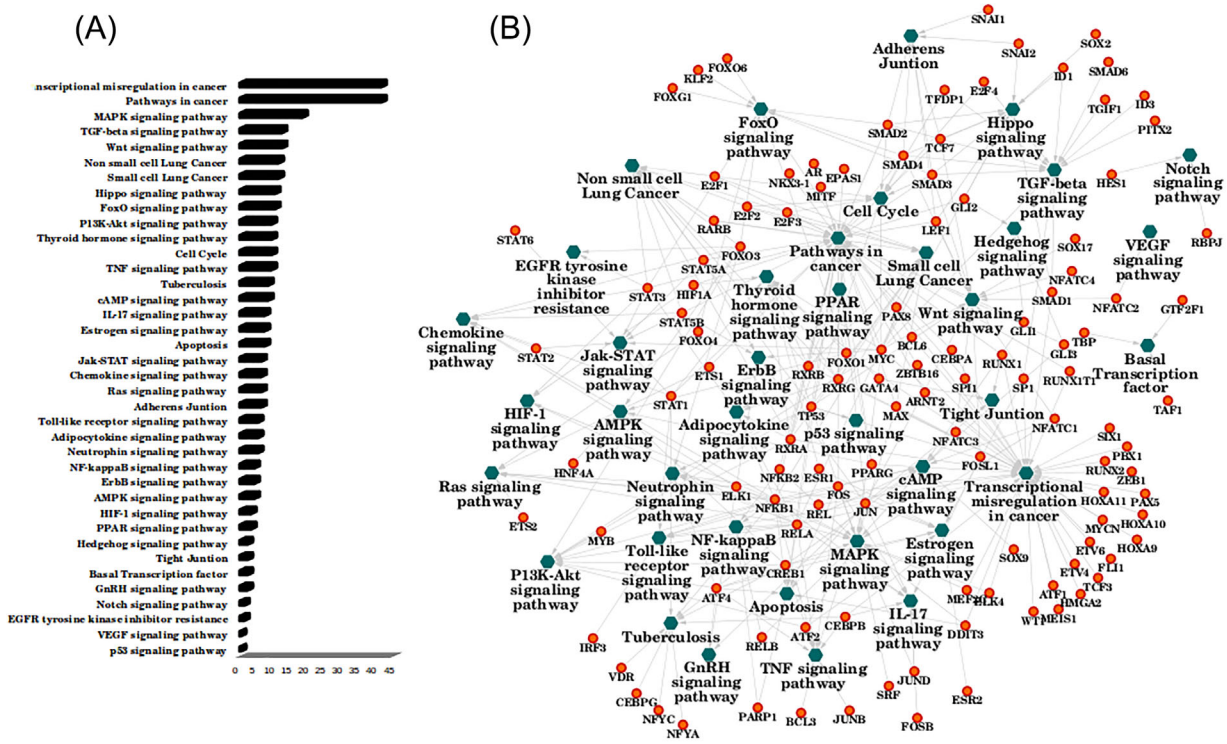


FIGURE 9 Pathway analysis of transcription factors

SOX10, SRY, and TCF7L2. To cover extensive cancer—TF associations, three well known disease databases were used such as DisgeNet, Disease Gene Annotation, Comparative Toxicogenomics Database. After removal of overlap between these databases, identified TFs were associated with 37 different types of cancer which is depicted in Figure 10 (Supplementary Table S4). Also from our analysis, we have identified 95 TFs are oncogenes and 92 TFs tumor suppressor

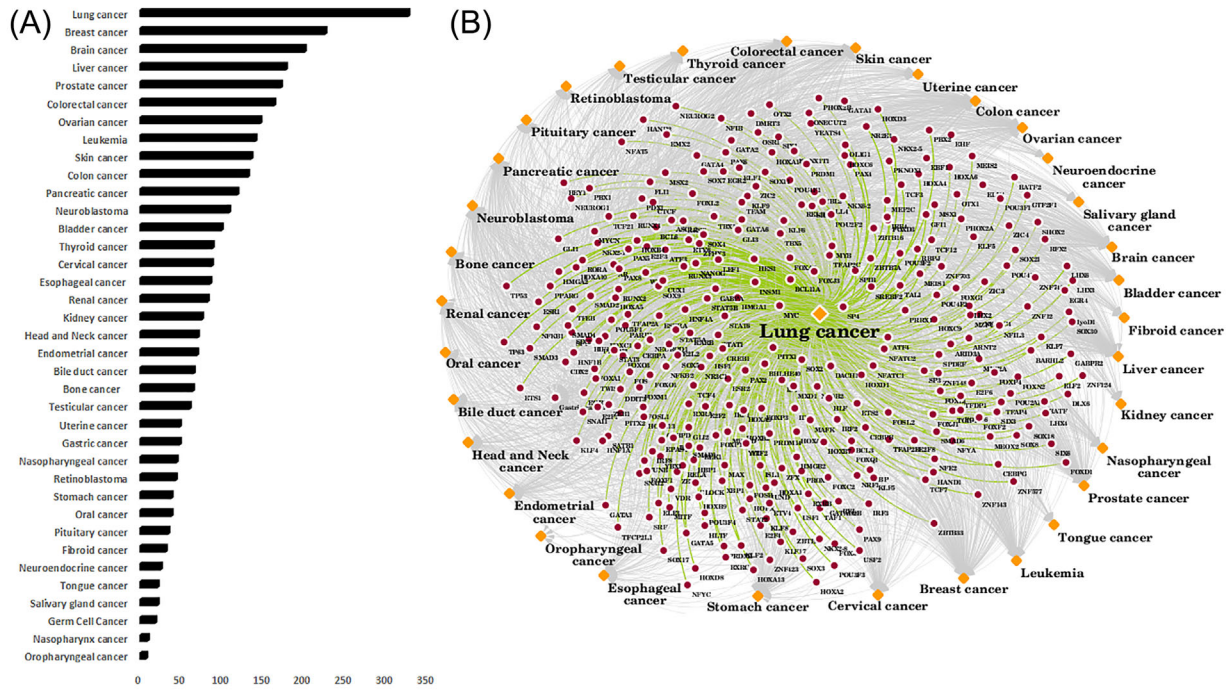


FIGURE 10 List of identified TFs associated with different cancer types

genes (Supplementary Table S5). These findings clearly indicated that the regulatory relationships of TFs in lung cancer progression are thought-provoking in therapeutics. We believe that the DBTFLC will be beneficial for the clinical researchers and assists to make an effort for identifying druggable targets in lung cancer.

4 | CONCLUSIONS

DBTFLC is a web-accessible database of transcription factors involved in lung cancer. This database includes integrated data resources related to transcription factors associated with lung cancer and their interacting partners as an easy-to-use web-based module. We believe that, DBTFLC will serve as a significant resource for transcriptional factor research in lung cancer. This database includes molecular targets and their interacting genes along with the link for Gene Ontology and Pathway analysis. Moreover, the database envisage to obtain and display the interaction of TF-TF and TF-TARGET GENE networks and to integrate their functional attributes. Thus, the database could serve as a potential resource for lung cancer research through identifying putative biomarkers and to discern enriched Gene Ontology annotations via the network biology approach. As the development in intensive oncological research, the understanding of TFs involved in lung cancer which enhance to focus as a druggable targets for the treatment in the coming years, we will continuously collect the information and the database will be updated periodically.

5 | AUTHOR SUMMARY

Lung cancer is the leading cause of cancer-related mortality worldwide for both women and men, with 1.38 million deaths annually. The two major types of lung cancer include small-cell lung cancer and non-small-cell lung cancer. As the early stage of the disease is asymptomatic, disease diagnosis at this stage can be quite challenging. Hence, understanding the mechanisms that govern lung carcinogenesis is important for developing novel therapeutic strategies. Recent studies have reported that a small subpopulation of cancer stem cells is responsible for tumor initiation, progression, metastasis, recurrence, and even resistance to treatment. The unique gene-expression patterns in different cells result in the presence of molecular markers that can be used to distinguish specific cell types. The different type of markers used to identify stem cells includes cell surface molecules, signaling pathway markers, as well as transcription factors. Transcription factors play a key role in the regulation of gene expression. The aim of developing this database was to explore and review the current data on transcription factors involved in lung carcinogenesis and their interacting molecules, gene ontology, and pathways.

ACKNOWLEDGMENTS

The authors are grateful to the management of VIT University for providing the facilities to carry out this research investigation.

ORCID

Sajitha Lulu S http://orcid.org/0000-0002-3392-4168

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2017. *CA: Cancer J Clin*. 2017;67:7–30.
2. Oser MG, Niederst MJ, Sequist LV, Engelman JA. Transformation from non-small-cell lung cancer to small-cell lung cancer: molecular drivers and cells of origin. *Lancet Oncol*. 2015;16: e165–e172.
3. Filtz TM, Vogel WK, Leid M. Regulation of transcription factor activity by interconnected, post-translational modifications. *Trends Pharmacol Sci*. 2014;35:76–85.
4. Minna JD, Schütte J, Viallet J, et al. Transcription factors and recessive oncogenes in the pathogenesis of human lung cancer. *Int J Cancer*. 1989;44:32–34.
5. Robson T, Hirst DG. Transcriptional Targeting in Cancer Gene Therapy. *J Biomed Biotechnol*. 2003;2:110–137.
6. Frank DA. Targeting transcription factors for cancer therapy. *IDrugs*. 2009;12:29–33.
7. Viola JPB, Carvalho LDS, Fonseca BPF, Teixeira LK. NFAT transcription factors: from cell cycle to tumor development. *Braz J Med Biol Res*. 2005;38:335–344.
8. Chang F, Lee JT, Navolanic PM, et al. Involvement of PI3 K/Akt pathway in cell cycle progression, apoptosis, and neoplastic transformation: a target for cancer chemotherapy. *Leukemia*. 2003;17:590–603.
9. Lee TI, Young RA. Transcriptional regulation and its misregulation in disease. *Cell*. 2013;152:1237–1251.
10. Latchman DS. Transcription factors: an overview. *Int J Biochem Cell Biol*. 1997;29:1305–1312.
11. Huang Y-F, Huang C-C, Liu Y-C, Oyang Y-J, Huang C-K. DNA-binding residues and binding mode prediction with binding-mechanism concerned models. *BMC Genomics*. 2009;10:S23.
12. Han H, Shim H, Shin D, et al. TRRUST: a reference database of human transcriptional regulatory interactions. *Sci Rep*. 2015;5:11432.
13. Schmeier S, Alam T, Essack M, Bajic VB. TcoF-DB v2: update of the database of human and mouse transcription co-factors and transcription factor interactions. *Nucleic Acids Res*. 2017;45:D145–D150.
14. Portales-Casamar E, Arenillas D, Lim J, et al. The PAZAR database of gene regulatory information coupled to the ORCA toolkit for the study of regulatory sequences. *Nucleic Acids Res*. 2009;37:D54–D60.
15. Lesurf R, Cotto KC, Wang G, et al. ORegAnno 3.0: a community-driven resource for curated regulatory annotation. *Nucleic Acids Res*. 2016;44: D126–D132.
16. Zheng G, Tu K, Yang Q, et al. ITFP: an integrated platform of mammalian transcription factors. *Bioinformatics*. 2008;24:2416–2417.
17. Zhao F, Xuan Z, Liu L, Zhang MQ. TRED: a Transcriptional Regulatory Element Database and a platform for in silico gene regulation studies. *Nucleic Acids Res*. 2005;33:D103–D107.

18. Bovolenta LA, Acencio ML, Lemke N. HTRIdb: an open-access database for experimentally verified human transcriptional regulation interactions. *BMC Genomics*. 2012;13:405.
19. Essaghir A, Toffalini F, Knoops L, Kallin A, van Helden J, Demoulin J-B. Transcription factor regulation can be accurately predicted from the presence of target gene signatures in microarray gene expression data. *Nucleic Acids Res*. 2010;38: e120.
20. Franceschini A, Szklarczyk D, Frankild S, et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res*. 2013;41:D808–D815.
21. Liu Z-P, Wu C, Miao H, Wu H. RegNetwork: an integrated database of transcriptional and post-transcriptional regulatory networks in human and mouse. *Database (Oxford) [Internet]*. 2015;30:2015.
22. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology. *Nat Genet*. 2000;25:25–29.
23. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28:27–30.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Amalraj T, Dravid AA, Tripathi R, S SL. Database of transcription factors in lung cancer (DBTFLC): A novel resource for exploring transcription factors associated with lung cancer. *J Cell Biochem*. 2018;119:5253–5261. https://doi.org/10.1002/jcb.26603