Determination of Biomarkers for Diagnosis of Lung Cancer Using Cytoscape-based GO and Pathway Analysis

Mona Maharjan[1], Raihanul Bari Tanvir[1], Kamal Chowdhury[2] and Ananda Mohan Mondal[1*]
[1]School of Computing and Information Sciences, Florida International University, Miami, FL, USA
[2]School of Natural Sciences and Mathematics, Claflin University, Orangeburg, SC, USA
Email: {mmaha021, rtanv003, amondal}@fiu.edu; kchowdhury@claflin.edu
[*]Corresponding author: amondal@fiu.edu

Abstract – *Lung cancer is the second most common cancer in the world. The aim of this study is to identify biomarkers for lung cancer that can aid in its diagnosis and treatment. The gene expression profiles from GEO database were analyzed by GEO2R to identify Differentially Expressed Genes (DEGs) and further analyzed using Cytoscape. The data was divided into two categories: non-treatment and treatment groups. A total of 407 DEGs (254 upregulated and 153 downregulated) and 259 DEGs (124 upregulated and 135 downregulated) were isolated for non-treatment and treatment studies respectively. The significant Gene Ontologies and pathways enriched with DEGS were identified using Cytoscape apps, BiNGO and ReactomeFIPlugIn, respectively. Hub genes based on network parameters - Degree, Closeness and Betweenness - were isolated using CytoHubba. In conclusion, DEGs identified in this study may play an important role in early diagnosis or as biomarkers of lung cancer.*

Keywords: Lung Cancer, Biomarkers, GO enrichment, Hub Genes.

1 Introduction

Lung cancer is the second most common cancer in the world. It is classified into two types - Non-Small Cell Lung Carcinoma (NSCLC) consisting of 80% of all lung cancer cases and Small Cell Lung Carcinoma (SCLC) recorded in 20% of all lung cancers [1]. The NSCLC is further divided into Adenocarcinoma (40%), Squamous Cell Carcinoma (27%) and large cell carcinoma (8%) [2]. Lung cancer has a low 5-year survival rate of 18% [3]. The low survival rate of lung cancer is due to relapse of lung cancer after treatment and late diagnosis of lung cancer [4]. The late diagnosis plays a significant role in the survival of patients. So, new methods for screening and diagnosis of lung cancer patients which would improve the prognosis need to be developed. The advancement in research and technology has been slowly shifting the focus of lung cancer diagnosis, prognosis and treatment towards underlying cause of disease progression such as protein-protein interaction (PPI) networks and molecular pathways. The networks are of special interest because the genes do not act alone. They act as a group to achieve a collective goal. The networks may correlate to specific functions. The activation or inactivation of key genes in the network may alter the function of these genes.

In this study, we aim to find genes which are common to both non-treatment and treatment studies. Both oncogenes that are upregulated and tumor suppressor genes that are downregulated could be target for cancer treatment. Any common gene upregulated in non-treatment and downregulated in treatment studies and vice versa could be potential biomarkers for lung cancer.

2 Materials and Methods

2.1 Datasets Preparation

The data for lung cancer was collected from NCBI (National Center for Biotechnology Information) GEO (Gene Expression Omnibus) database. It provides genome-wide gene expression profiles including DEGs. Figure 1 shows the cleaning and identification of top 250 DEGs that could be probable biomarkers.

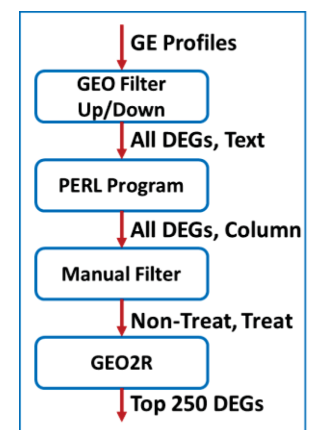

Figure 1. Data Cleaning and Identification of Top 250 DEGs.

Querying GEO database with phrase "lung cancer" retrieved 1,050,133 gene expression (GE) profiles. Using the GEO built-in filter "up/down genes" reduced the retrieved GE profiles to 16876 genes, which are all the DEGs based on GEO filter. The retrieved DEGs were downloaded as a text file, in which each GE profile consists of six lines of record. A PERL script was used to obtain four important features for each DEG - Gene Symbol, GDS number (study the gene belongs to), organism's name, and the number of samples used for the study. This intermediate data was stored in column format. From this data, we discovered that the retrieved 16,876 DEGs belong to 27 unique studies.

The scope of the present study is to consider the studies with treatment and non-treatment. Upon using manual filtering - reading the title and abstract of the main publications resulted from these 27 studies - we discovered that only 3 of these studies are non-treatment (GDS1312, GDS2499 and GDS5201) and 2 are treatment (GDA1208 and GDS4794). Table 1 shows the summary of these datasets including number of case and control.

Table 1. Summary of Datasets.

Study	#Samples	#Control	#Case
GDS1204	18	9	9
GDS1312	10	5	5
GDS2499	12	6	6
GDS4794	68	43	25
GDS5201	6	2	4

Finally, GEO2R, a LIMMA R package in NCBI GEO, was used to isolate top 250 DEGs from each of the non-treatment and treatment studies. The cutoff criteria used were P-value < 0.05 and absolute log Fold Change (FC) >1. Benjamini & Hochberg (False Discovery Rate) method was used for adjusting P-values. The duplicate DEGs and DEGs with missing symbols were removed. Finally, a list of 254 upregulated and 153 downregulated DEGs were discovered for non-treatment studies. Similarly, for treatment studies, 124 upregulated and 135 downregulated DEGs were discovered.

2.2 Methodology

Figure 2 shows the overall methodology. In order to analyze discovered DEGs in terms of i) enriched pathways, ii) enriched GO terms and iii) hub genes, a protein/gene network is required. The discovered DEGs from two study groups non-treatment and treatment were imported in ReactomeFIPlugIn [5], a Cytoscape app, to generate these networks using the functional interactions available in Reactome database. Cytoscape version 3.4.0 was used for this study.

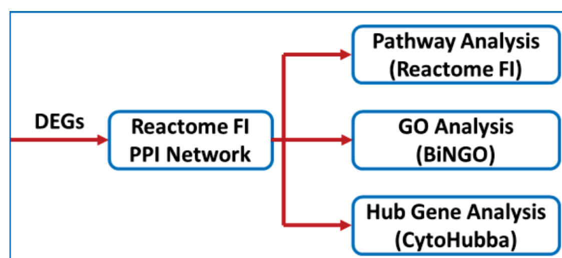

Figure 2. Overview of Data Analysis.

2.2.1 Pathway Enrichment Analysis

The pathway enrichment for both non-treatment and treatment network were performed using ReactomeFIPlugIn [5]. This plugin is designed to find pathways and network patterns related to cancer and other types of diseases. The top ten pathways for non-treatment and treatment along with DEGs based on P-values are identified.

2.2.2 GO Enrichment Analysis

GO enrichment analysis for three categories - Biological Process (BP), Cellular Component (CC) and Molecular Function (MF) - was performed using BiNGO [6] in Cytoscape 3.4.0. The statistical measures used were hypergeometric test for significance (p-value < 0.05) and Benjamini & Hochberg correction for multiple testing correction.

2.2.3 Hub Genes Identification

Hub genes are highly connected nodes in the network. Hub gene analysis was done using Cytoscape app, CytoHubba [7]. The hub genes were identified based on three parameters- degree, closeness and betweenness.

Degree represents the total number of nodes connected to its adjacent nodes [8]. It also provides the count of the number of interactions of a given node. ***Closeness*** represents the shortest path to access all the other nodes in the network. It uses the distance between vertex of interest and all other vertices as well as the sum of distance between the vertex of interest and all the other vertices[9]. ***Betweenness*** represents how the nodes are interconnected and determines the frequency with which a node is on the shortest path between the two other nodes. The nodes are not able to communicate with each other without the intermediate node [9].

3 Results

3.1 Functionally Interacting Network

Figure 3 presents the functionally interacting network for non-treatment DEGs generated using ReactomeFIPlugIn.

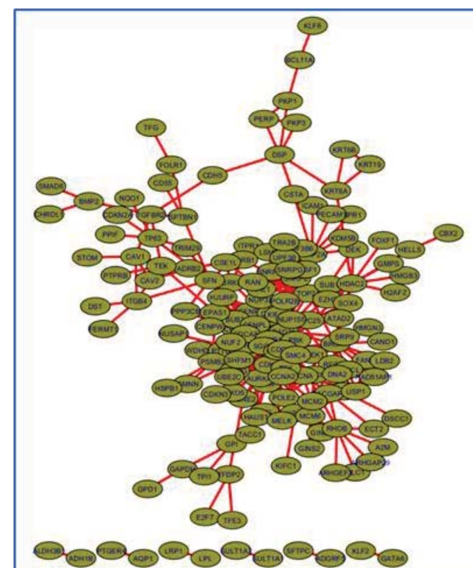

Figure 3. Functionally interacting network of non-treatment DEGs created using ReactomeFIPlugIn.

This network has 164 DEGs and 534 interactions. We originally had 407 DEGs for non-treatment studies but 243 DEGs do not have any functional interaction based on Reactome database so, they are not included in the network. The reason might be that those genes are not related to any pathways or they are yet to be determined whether they belong to any pathway or not. Further study is required to position those genes on appropriate pathways. Similarly, only 67 out of 259 DEGs are found in Reactome with 87 functional interaction for treatment studies.

3.1.1 Enriched Pathways and DEGs in Non-Treatment Studies

Table 2 shows the top ten enriched pathways along with up- and down-regulated DEGs in non-treatment studies.

Table 2. Top 10 enriched pathways in non-treatment with up- and down- regulated DEGs. ReactomeFIViz was used to identify enriched pathways. P-values ranges between 1.80E-07 and 1.11E-16.

Mitotic Prometaphase (R)	
Up	BUB3, **CCNB1**, CENPF, CENPI, CENPL, CENPU, KIF18A, KNTC1, MAD21, NCAPG, NDC80, NUF2, RAD21, SGO2, SMC4, SPC25, ZWINT
Down	**CCNB1**, CCNB2, CDC20, CDCA8, CDK1, NUP37
Cell cycle(K)	
Up	BUB3, CCNA2, **CCNB1**, CCNE2, CDKNA, CHEK1, HDAC2, MAD21, MCM6, ORC6, PCNA, RAD21, TTK
Down	**CCNB1**, CCNB2, CDC20, CDK1, MCM2, PTTG1, SFN, TFDP2
Mitotic Metaphase and Anaphase(R)	
Up	BUB3, CENPF, CENPI, CENPL, CENPU, FBXO5, KIF18A, KNTC1, MAD21, NDC80, NUF2, RAD21, SGO2, SHFM1, SPC25, UBE2C, ZWINT
Down	CDC20, CDCA8, NUP37, PSMB2, PTTG1
Signaling by Rho GTPases(R)	
Up	ARHGEF3, BUB3, CENPF, CENPI, CENPL, CENPU, DLC1, ECT2, KIF14, KIF18A, KNTC1, MAD2L1, NUF2, ARHGAP29, NDC80, RACGA1, RHOB, SGO2, SPC25, ZWINT
Down	CDC20, CDCA8, NUP37, RACGAP1SFN
Cell Cycle Checkpoints(R)	
Up	BRCA1, BRIP1, BUB3, **CCNB1**, CHEK1, DNA2, MAD21, MCM6, ORC6, RFC4, SHFM1, UBE2C
Down	**CCNB1**, CCNB2, CDC20, CDK1, MCM2, PSMB2, SFN
SUMOylation(R)	
Up	AURKA, BRCA1, CBX2, CDKNA, PCNA, RAD21, TDG, TOP2A
Down	CDCA8, NUP155, NUP37, RAE1, TFAP2A
Mitotic G1-G1/S phases(R)	
Up	**CCNB1**, CCNE2, CDKNA, FBXO5, MCM6, ORC6, PCNA, POLE2, SHFM1, TOP2A
Down	**CCNB1**, CDK1, MCM2, PSMB2, TFDP2
APC/C-mediated degradation of cell cycle proteins(R)	
Up	AURKA, BUB3, **CCNB1**, FBXO5, MAD21, SHFM1, UBE2C
Down	**CCNB1**, CDC20, CDK1, PSMB2, PTTG1
Synthesis of DNA(R)	
Up	DNA2, GINS1, GINS2, MCM6, ORC6, PCNA, POLE2, RFC4, SHFM1
Down	MCM2, PSMB2
S Phase(R)	
Up	DNA2, GINS1, GINS2, MCM6, ORC6, PCNA, POLE2, RAD21, RFC4, SHFM1
Down	MCM2, PSMB2

The letters in parenthesis corresponds to the sources of pathways: C-CellMap, R-Reactome, K-KEGG (Kyoto Encyclopedia of Gene and Genomes), N-NCI PID (National Cancer Institute Pathway Interaction Database), P-Panther and B-Biocarta. The DEGs were compared to a list of up and down regulated DEGs to identify which of the enriched DEGs are up-regulated or down-regulated. The mismatched genes that appear in both up- and down-regulated categories are highlighted in red. CCNB1 was the only DEG mismatched between up and down regulated DEGs in 5 out of 10 pathways. There were no mismatched DEGs in other 5 pathways.

3.1.2 Enriched Pathways and DEGs in Treatment Studies

Table 3 shows top ten enriched pathways along with the up- and down-regulated DEGs in each pathway of Treatment studies.

Table 3. Top 10 enriched pathways in treatment with up- and down-regulated DEGs. ReactomeFIViz was used to identify enriched pathways. P-values ranges between 1.09E-04 and 9.45E-12.

Metallothioneins bind metals(R)	
Up	**MT1E, MT1F, MT1G, MT1H, MT1X, MT2A**
Down	**MT1E, MT1F, MT1G, MT1H, MT1X, MT2A**
Mineral absorption(K)	
Up	**MT1E, MT1F, MT1G, MT1H, MT1X, MT2A**
Down	**MT1E, MT1F, MT1G, MT1H, MT1X, MT2A**
AP-1 transcription factor network(N)	
Up	EGR1, HLA-A, **JUN, MT2A**
Down	CDKN2A, ETS1, **JUN, MT2A**
Rapid glucocorticoid signaling(N)	
Up	MAPK8
Down	GNB1, MAPK1
Th1 and Th2 cell differentiation(K)	
Up	IL12, RB1, **JUN**, MAPK8
Down	**JUN**, MAPK14, NFKBIA, RBPJ
Osteopontin-mediated events(N)	
Up	**JUN**, MAPK8
Down	**JUN**, NFKBIA, ITGB3
CD40/CD40L signaling(N)	
Up	**JUN**, MAPK8
Down	**JUN**, MAPK14, NFKBIA
BCR signaling pathway(N)	
Up	**JUN**, MAPK8
Down	ETS1, **JUN**, MAPK14, NFKBIA
TGF-beta signaling pathway(P)	
Up	**JUN**, MAPK8
Down	ACVR1B, BMP4, **JUN**,
HTLV-I infection(K)	
Up	EGR1, HLA-A, **JUN**, MAPK8, TCF3
Down	CDKN2A, ETS1, **JUN**, NFKBIA

The DEGs involved in these pathways are MT1E, MT1F, MT1G, MT1H, MT1X, MT2G, EGR1, HLA-A, JUN, MAPK8, IL12RB1, TCF3, CDKN2A, ETS1, GNB1, MAPK14, NFKBIA, RBPJ, ITGB3, ACVR1B and BMP4. Similar to the non-treatment pathways, the DEGs were compared to a list of up and down regulated DEGs to identify which of the enriched DEGs are up-regulated or down-regulated. The DEGs - MT1E, MT1F, MT1G, MT1H, MT1X, MT2A and JUN - were found to be present in both up and down regulated gene lists. This could be due to the fact that our analysis consists of a group of studies and not a single study.

It is clear from Table 2 and 3 that there is no pathway in common between non-treatment and treatment studies. This could be due to the fact that there is no DEG that is common between these two study groups.

3.2 Hub Genes

The genes do not work alone, they are always present in some form of a network that works in conjugation with other gene networks. The position of the genes in the network or pathway may signify its relevance in the network. The centrally located and highly connected genes may have more important roles to play in the overall function of an organism.

Figure 4 shows the network of non-treatment DEGs with ten hub genes based on the topological parameter of degree. The highly connected genes are colored red to yellow. The red colored nodes are the nodes with the highest scores and as the score lowers the color changes to yellow.

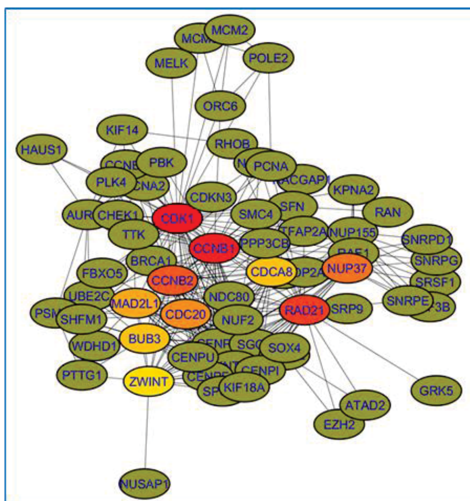

Figure 4. A network of hub genes from non-treatment network created using Cytoscape app, CytoHubba.

Tables 4 and 5 show top hub genes along with their regulation and description in non-treatment and treatment studies based on three topological parameters - degree, closeness and betweenness. Some hub genes induce oncogenes or tumor suppression through other known oncogenes and tumor suppressor genes. There are no hub gene in common between non-treatment and treatment studies.

Table 4. Top hub genes from non-treatment studies along with their regulation and description.

Gene	Regulation	Description
CDK1	DOWN	Cyclin-dependent kinase 1
CCNB1	UP	Cyclin B1
RAD21	UP	RAD21 homolog
CCNB2	DOWN	Cyclin B2
CDC20	DOWN	Cell division cycle 20
NUP37	DOWN	Nucleoporin 37kDa
MAD2L1	UP	MAD2 mitotic arrest deficient-like 1
BUB3	UP	BUB3 mitotic checkpoint protein
CDCA8	DOWN	Cell division cycle associated 8
BRCA1	UP	Breast cancer 1, early onset

Table 5. Top hub genes from treatment studies along with their regulation and description.

Gene	Regulation	Description
MAPK14	Down	Mitogen-activated protein kinase 14
JUN	Up	Jun proto-oncogene, AP-1 transcription factor subunit
MAPK8	Up	Mitogen-activated protein kinase 8
GNB1	Down	G protein subunit beta 1
ETS1	Down	ETS proto-oncogene 1, transcription factor
TP63	Down	Tumor protein p63

3.3 Enriched GO Terms

The top ten enriched GO terms for non-treatment and treatment studies at a significant level, $P < 0.05$, are shown in Figures 5a and 5b, respectively.

Enriched Go Terms in Non-Treatment Studies: The biological processes that are significantly enriched with DEGs are - nuclear division, mitosis, cell division, mitotic cell cycle, cell cycle phase and cell cycle process. The cellular components that are significantly enriched are - chromosome, centromeric region, non-membrane bounded organelle, intracellular non-membrane bounded organelle, chromosomal part, chromosome, spindle protein complex, condensed chromosome, macromolecular complex and condensed chromosome kinetochore. The molecular functions that are significantly enriched are - protein binding, enzyme binding, motor activity, structure specific DNA binding, troponin C binding, double stranded DNA binding, transmembrane receptor protein serine/ threonine kinase binding, microtubule motor activity, receptor serine/threonine kinase binding and actin filament binding.

Enriched Go Terms in Treatment Studies: The enriched biological processes are - negative regulation of cellular

process, negative regulation of biological process negative regulation of calcium ion transport, regulation of apoptosis, regulation of programmed cell death, regulation of cell death, positive regulation of apoptosis, regulation of synaptic plasticity , positive regulation of apoptosis, negative regulation of cellular metabolic process, positive regulation of programmed cell death, and cellular biogenic amine metabolic process. The enriched cellular components are - receptor complex, nucleoplasm, protein complex, nuclear lumen, plasma membrane part, nucleoplasm part, cell surface, nicotine acetylcholine-gated receptor channel complex, growth cone and site of polarized growth. The enriched molecular functions are - nuclear localization sequence binding and prostaglandin-endoperoxide synthase activity.

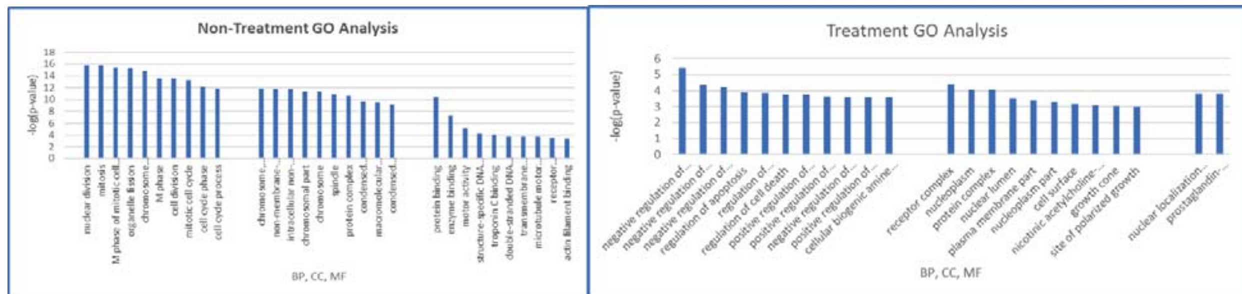

Figure 5. The Top Ten Enriched GO Terms in a) Non-Treatment Studies and b) Treatment Studies.

3.4 Granularity of DEGs

Table 6 shows the granularity of discovered DEGs at different level of analysis. The raw data from NCBI GEO had a total of 16,876 DEGs.

Table 6. Level of Granularity of Number of DEGs

Granularity	No of DEGs	
	Non-Treatment	Treatment
GEO	16876	
GEO2R	2467	1834
Reactome (Network)	407	259
Reactome (Pathways)	168	59
Cytohubba	30	30
Hub Genes (Common)	10	6

Based on their study type it was categorized into non-treatment and treatment. The non-treatment group had 2,467 DEGs and treatment group had 1,834 DEGs. Different Cytoscape apps were used for data visualization and analysis. Reactome, a Cytoscape app, was used for visualizing the network of DEGs. There were 407 non-treatment DEGs and 259 treatment DEGs connected in a network based on Reactome pathways database. Reactome app was also used for pathway analysis. The enriched pathways in non-treatment and treatment groups had 168 and 59 DEGs respectively. Cytohubba was used for hub gene analysis based on three parameters of degree, closeness and betweenness. There were 10 hub genes for each category. Hub genes common in at least two categories were selected for each group. Hence, a total of 10 and 6 hub genes for non-treatment and treatment were obtained.

4 Discussion

Lung cancer is the second highest cause of mortality in men and women [10]. The traditional methods of lung cancer treatment are surgery, chemotherapy and radiation or a combination of two or more of these methods. However, the low survival rate of cancer indicates a need for improvement towards cancer diagnosis and treatment approaches. The advancement in biotechnological tools and bioinformatics has changed the course of lung cancer research towards identifying underlying causes of cancer such as molecular pathways and biomarkers. This study used gene expression profiles from GEO datasets to identify DEGs based on two study groups - Non-treatment (healthy samples as control and disease samples as case) and Treatment (disease sample without medicine as control and disease samples with medicine as case). Cytoscape app, CytoHubba was used to identify hub genes from the PPI network of non-treatment and treatment studies.

Table 7 shows the top hub genes from Non-Treatment and Treatment Studies along with their regulation. It also shows the hub genes identified as oncogenes and tumor suppressor genes in literature. The DEGs CDK1, CCNB1, CCNB2, CDC20, CDCA8, JUN and ETS1 are oncogenes and BRCA1, TP63 are tumor suppressor genes. It is clear from this table that most of the oncogenes are down regulated with the exception of JUN. One of two tumor suppressor genes are upregulated and the other one is downregulated. Two potential biomarker genes- CCNB1-which is an oncogene from non-treatment studies, is upregulated. This gene supports our hypothesis, which states that both oncogenes that are upregulated and tumor suppressor genes that are downregulated could be target for cancer treatment. CCNB1 is also the only mismatched DEG in non-treatment pathways. It is present in 5 out of 10 enriched pathways. It is also one of the hub genes for non-treatment studies. A study by Soria et al

showed the overexpression of CCNB1 in both NSCLC and SCC [11]. CCNB1 is a known regulatory protein in mitosis and is necessary for control of G2/M transition phase in cell cycle. This is also confirmed by the pathway analysis which shows CCNB1 is present in the pathways - Mitotic prometaphase, cell cycle, cell cycle checkpoints, Mitotic G1-G1/S phases and APC/C-mediated degradation of cell cycle proteins.

Table 7. Probable Biomarkers for Lung Cancer.

Hub Genes	Gene Regulation		Oncogene/Tumor Suppressor Gene	Treatment
	Present Study	TCGA		
CDK1	DOWN	UP	Oncogene	No
CCNB1	UP	UP	Oncogene	No
RAD21	UP	-	-	No
CCNB2	DOWN	UP	Oncogene	No
CDC20	DOWN	-	Oncogene	No
NUP37	DOWN	-	-	No
MAD2L1	UP	-	-	No
BUB3	UP	-	-	No
CDCA8	DOWN	-	Oncogene	No
BRCA1	UP	-	Tumor Sup	No
MAPK14	DOWN	-	-	Yes
JUN	UP	-	Oncogene	Yes
MAPK8	UP	-	-	Yes
GNB1	DOWN	-	-	Yes
ETS1	DOWN	-	Oncogene	Yes
TP63	DOWN	UP	Tumor Sup	Yes

JUN is an oncogene from treatment studies but is upregulated so this suggests the treatment is not working. If treatment works, this oncogene should have been downregulated in treatment studies. JUN, a signal transducing transcription factor of the AP-1 family is a proto oncogene that is associated with apoptosis [12]. Levresse et al.[13] and other studies have reported the protective response of c-Jun and JNK pathway in SCLC cells. This is also confirmed in table 3, which shows JUN is present in the pathways - Th1 and Th2 cell differentiation, Osteopontin mediated events, CD40/CD40L signaling, BCR signaling pathway, TGF beta signaling pathway and HTLV-I infection.

TP63 is a member of tumor suppressor gene p53 that plays vital role in cellular differentiation and responsiveness to cellular stress [14][15]. It is present in treatment studies but is downregulated. This also suggests that treatment is not working. Tumor suppressor genes should be upregulated with treatment.

All the hub genes from non-treatment studies except BRCA1 are present in the top two enriched pathways Mitotic prometaphase and cell cycle of non-treatment studies as shown in table 4.

CDK1, a cyclin dependent kinase (CDKs), that plays an important role in cell progression [16]. It is present in the enriched pathways - Mitotic prometaphase, cell cycle and cell cycle checkpoints and Mitotic G1-G1/S phases. All of these pathways are related to cell cycle.

GNB1, gene encoding guanine nucleotide-binding protein (G-protein), is composed of an alpha, beta and gamma subunit that are essential for signaling function of G-protein coupled receptors (GPCRs). GNB1 integrates the signals between receptors and effectors [17].

MAPKs (Mitogen-activated protein kinases) are protein Ser/Thr kinases that convert external stimuli into a wide range of cellular responses [18]. MAPK14 is present in Rapid glucocorticoid signaling, Th1 and Th2 cell differentiation and CD40/CD40 signaling. MAPK8, also known as c-Jun N-terminal kinase (JNK), plays a role in various biological processes such as proliferation differentiation and transcriptional regulation [12]. In this study, MAPK8 is present in Rapid glucocorticoid signaling, Th1 and Th2 cell differentiation, Ospeopontinmediated events, CD40/CD40L signaling, BCR signaling, TGF-beta signaling and HTLV-I infection pathways.

ETS, proto oncogene member of ETS family of transcription factors, is considered as an activator of transcription that is expressed in various cancers [19]. It is present in AP-1 transcription factor network, BCR signaling and HTLV-I infection pathways.

5 Conclusion

In this study, the hub genes for non-treatment and treatment were isolated. For non-treatment group, upregulated DEGs are CDK1, RAD21, CCNB2, MAD2L1, BUB3, and BRCA1 and downregulated DEGs are CDK1, CCNB2, CDC20, NUP37 and CDCA8. For treatment group, upregulated DEGs are JUN and MAPK8 and downregulated DEGs are GNB1, ETS1 and TP63.

CCNB1 is an oncogene in non-treatment that is upregulated and this is the only gene that supports our hypothesis. This could be the potential biomarker for lung cancer.

The hub genes common in both non-treatment and treatment groups could not be found. However, the hub genes found in this study could be potential biomarker for lung cancer. Further research and biological experiments are required for validation of the results.

Acknowledgement: This work was supported by NSF CAREER Award #1651917 (Transferred to #1901628) to AMM.

References

[1] J. E. Larsen *et al.*, “Expression profiling defines a recurrence signature in lung squamous cell carcinoma.,” *Carcinogenesis*, vol. 28, no. 3, pp. 760–766, Mar. 2007.

[2] H. Guo, J. Chen, and F. Meng, “Identification of novel diagnosis biomarkers for lung adenocarcinoma from the cancer genome atlas,” *Int. J. Clin. Exp. Med.*, vol. 9, no. 5, pp. 7908–7918, 2016.

[3] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2018," *CA. Cancer J. Clin.*, vol. 68, no. 1, pp. 7–30, Jan. 2018.

[4] B. Kim *et al.*, "Clinical validity of the lung cancer biomarkers identified by bioinformatics analysis of public expression data," *Cancer Res.*, vol. 67, no. 15, pp. 7431–7438, 2007.

[5] G. Wu, E. Dawson, A. Duong, R. Haw, and L. Stein, "ReactomeFIViz: a Cytoscape app for pathway and network-based data analysis," *F1000Research*, 2014.

[6] S. Maere, K. Heymans, and M. Kuiper, "BiNGO: a Cytoscape plugin to assess overrepresentation of Gene Ontology categories in Biological Networks," *Bioinformatics*, vol. 21, no. 16, pp. 3448–3449, Aug. 2005.

[7] C.-H. Chin, S.-H. Chen, H.-H. Wu, C.-W. Ho, M.-T. Ko, and C.-Y. Lin, "cytoHubba: identifying hub objects and sub-networks from complex interactome," *BMC Syst. Biol.*, vol. 8, no. Suppl 4, pp. S11–S11, Dec. 2014.

[8] D. Koschützki and F. Schreiber, "Centrality Analysis Methods for Biological Networks and Their Application to Gene Regulatory Networks," *Gene Regul. Syst. Bio.*, vol. 2, pp. 193–201, May 2008.

[9] G. A. Pavlopoulos *et al.*, "Using graph theory to analyze biological networks," *BioData Min.*, vol. 4, p. 10, Apr. 2011.

[10] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA. Cancer J. Clin.*, vol. 68, no. 6, pp. 394–424, Nov. 2018.

[11] J. C. Soria *et al.*, "Overexpression of cyclin B1 in early-stage non-small cell lung cancer and its clinical implication.," *Cancer Res.*, vol. 60, no. 15, pp. 4000–4004, Aug. 2000.

[12] S. Zhang *et al.*, "PAR1-mediated c-Jun activation promotes heat stress-induced early stage apoptosis of human umbilical vein endothelial cells," *Mol. Med. Rep.*, vol. 15, no. 5, pp. 2595–2603, May 2017.

[13] V. Levresse, L. Marek, D. Blumberg, and L. E. Heasley, "Regulation of platinum-compound cytotoxicity by the c-Jun N-terminal kinase and c-Jun signaling pathway in small-cell lung cancer cells.," *Mol. Pharmacol.*, vol. 62, no. 3, pp. 689–697, Sep. 2002.

[14] E. R. Flores, "The Roles of p63 in Cancer," *Cell Cycle*, vol. 6, no. 3, pp. 300–304, Feb. 2007.

[15] L. Zhang *et al.*, "Quantitative Assessment of the Influence of TP63 Gene Polymorphisms and Lung Cancer Risk: Evidence Based on 93,751 Subjects," *PLoS One*, vol. 9, no. 1, p. e87004, Jan. 2014.

[16] K. J. Barnum and M. J. O'Connell, "Cell Cycle Regulation by Checkpoints," *Methods Mol. Biol.*, vol. 1170, pp. 29–40, 2014.

[17] S. Petrovski *et al.*, "Germline De Novo Mutations in GNB1 Cause Severe Neurodevelopmental Disability, Hypotonia, and Seizures," *Am. J. Hum. Genet.*, vol. 98, no. 5, pp. 1001–1010, May 2016.

[18] M. Cargnello and P. P. Roux, "Activation and Function of the MAPKs and Their Substrates, the MAPK-Activated Protein Kinases," *Microbiol. Mol. Biol. Rev.*, vol. 75, no. 1, pp. 50–83, Mar. 2011.

[19] J. P. Plotnik, J. A. Budka, M. W. Ferris, and P. C. Hollenhorst, "ETS1 is a genome-wide effector of RAS/ERK signaling in epithelial cells," *Nucleic Acids Res.*, vol. 42, no. 19, pp. 11928–11940, Oct. 2014.