Identification of a Novel Genetic Mechanism Involved in Repeated Lipopolysaccharide (LPS) Enabled Carcinogenesis of Lung Cancer

M. Manasa¹; D. Macrin²*

¹Research Scholar, Department of Bioinformatics, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India.
²Project Guide, Department of Bioinformatics, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India.
²danniem.sse@saveetha.com

Abstract

Aim: To identify the gene expression of Lipopolysaccharide (LPS) in lung cancer by analysing and comparing the data retrieved from the GEO database. Materials and methods: The Microarray dataset was retrieved from NCBI gene expression omnibus (GEO) with ID GSE132661 to find significant genes and upregulated and downregulated genes. STRING analysis is used to find the relations between upregulated and downregulated genes to find possible interactions in lung cancer. Enrichr analyses were used to compare input gene sets with annotated gene sets to find pathways and ontologies. ‘CytoHubba’ plugin of ‘Cytoscape’ used to analyze the data from STRING to identify HUB genes. Results: We observed 11753 genes were significantly overexpressed and upregulated genes (UG) (log FC ≥ 1) 73 genes and downregulated genes (DG) (log FC ≤ −1) 56 genes co-expressed in lung cancer. STRING analysis is used to find 33 possible interactions in lung cancer. Enrichr analysis were used to find cellular biological pathways involved in inflammatory responses. ‘CytoHubba’ is used to identify RELN, NTRK2, MYCN, DMD, WT1, PIK3C2A, PLCZ1, PIP5K11 as HUB genes. ‘ClueGO’ identified pathways like PIP kinase activity and IP metabolism involved in inflammatory responses. Conclusion: We identified the gene expression of eight hub genes that are involved in different pathways. Furthermore, our results suggest that these novel gene expressions could be targeted for the further drug discovery process.

Key-words: Lipopolysaccharide (LPS), Phosphate-buffered Salt (PBS), Upregulated and downregulated Genes, Novel Genetic Mechanism, Expression Patterns, Molecular Biology, Genetic Analysis, Gene Expression.
1. Introduction

My research focuses on identifying the gene expressions and pathways of lipopolysaccharide, a carcinogen that causes lung cancer. The usage of tobacco and other lifestyle carcinogens has been increasing day by day. The chemical present in tobacco smokers is Lipopolysaccharide (LPS) (Bagaitkar, Demuth, and Scott 2008). Lipopolysaccharide (LPS) is a carcinogen that is known to cause inflammatory responses in lung cancer (S. Sun, Schiller, and Gazdar 2007). In Mus musculus, lipopolysaccharide (LPS) mediates chronic inflammation-induced T-cell fatigue, increases programmed cell death-1, regulates cell proliferation, promotes alveolar cell development and increases the pathogenicity and dysfunctions of lung cancer. Phosphate-buffered salt (PBS) is non-toxic which prevents the rupturing of a hepatic cell. The applications of my study were applied to identify the differentially expressed genes in LPS treated cells and normal cells. These genes were enriched to deduce their functions and the pathways that pertain to them.

Previous research has shown that nicotine significantly promotes tumour progression and metastasis in a mouse (Lambert et al. 2005). In a recent study, researchers discovered an immune gene signature that predicts treatment responses and survival in patients with tobacco carcinogen-induced lung cancer who have been receiving immune checkpoint blockade therapy (C.-H. Liu et al. 2021). Lung carcinogenesis and lung metastasis from other primary tumours have both been related to the lung microbiome (“The Influence of Lung Microbiota on Lung Carcinogenesis, Immunity, and Immunotherapy” 2020; Cheng et al. 2016). Smoking reduction of 50% among people who smoke 15 or more cigarettes per day significantly reduces the risk of lung cancer (Godtfredsen, Prescott, and Osler 2005).

Previously our team has a rich experience in working on various research projects across multiple disciplines (Sathish and Karthick 2020; Varghese, Ramesh, and Veeraiyan 2019; S.R. Samuel, Acharya, and Rao 2020; Venu, Raju, and Subramani 2019; M. S. Samuel et al. 2019; Venu, Subramani, and Raju 2019; Mehta et al. 2019; Sharma et al. 2019; Malli Sureshbabu et al. 2019; Krishnaswamy et al. 2020; Muthukrishnan et al. 2020; Gheena and Ezhilarasan 2019; Vignesh et al. 2019; Ke et al. 2019; Vijayakumar Jain et al. 2019; Jose, Ajitha, and Subbaiyan 2020). Now the growing trend in this area motivated us to pursue this project.

Computational biological techniques enable us to compute the results of a comparative analysis of differentially expressed genes and their pathways involved in inflammatory responses but the mechanism of adverse effects of Lipopolysaccharide (LPS) involved in causing lung cancer is not studied. This study aims to identify gene expressions of Lipopolysaccharide (LPS) in lung cancer by
analyzing and comparing the data retrieved from Gene Expression Omnibus (GEO) that leads to drug discovery.

2. Materials and Methods

2.1 Dataset

The expressional gene dataset of tobacco carcinogens like Lipopolysaccharide (LPS) induces lung cancer (Udhaya Kumar et al. 2020). The Microarray dataset retrieved from the NCBI Gene Expression Omnibus (GEO) with gene id GSE132661 consists of 4 expressional profiles (grouped into 2) of 12 samples in *Mus musculus* treated with Phosphate buffered salt (PBS) (Therriault et al. 2003). Each group consists of three samples of tobacco carcinogens, which induces lung cancer. Here, by inducing two group Lipopolysaccharide (LPS) carcinogen with Phospate buffered salt (PBS), controlled into 'Mus musculus', we analyze the differential expressed genes involved in lung cancer. (Table. 1).

<table>
<thead>
<tr>
<th>Accession</th>
<th>Title</th>
<th>Source name</th>
<th>Strains</th>
<th>Treatment</th>
<th>Tissue</th>
<th>Treatment regimen</th>
<th>No of samples in each group</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSM3885092</td>
<td>PBS_Fvbn_1</td>
<td>Lung tissue</td>
<td>FVB/NJ</td>
<td>PBS</td>
<td>Lung</td>
<td>weekly treatment for 16 weeks</td>
<td>3</td>
</tr>
<tr>
<td>GSM3885093</td>
<td>PBS_Fvbn_2</td>
<td>Lung tissue</td>
<td>FVB/NJ</td>
<td>PBS</td>
<td>Lung</td>
<td>weekly treatment for 16 weeks</td>
<td>3</td>
</tr>
<tr>
<td>GSM3885094</td>
<td>PBS_Fvbn_3</td>
<td>Lung tissue</td>
<td>FVB/NJ</td>
<td>PBS</td>
<td>Lung</td>
<td>weekly treatment for 16 weeks</td>
<td>3</td>
</tr>
<tr>
<td>GSM3885095</td>
<td>LPS_Fvbn_1</td>
<td>Lung tissue</td>
<td>FVB/NJ</td>
<td>LPS</td>
<td>Lung</td>
<td>weekly treatment for 16 weeks</td>
<td>3</td>
</tr>
<tr>
<td>GSM3885096</td>
<td>LPS_Fvbn_5</td>
<td>Lung tissue</td>
<td>FVB/NJ</td>
<td>LPS</td>
<td>Lung</td>
<td>weekly treatment for 16 weeks</td>
<td>3</td>
</tr>
<tr>
<td>GSM3885097</td>
<td>LPS_Fvbn_6</td>
<td>Lung tissue</td>
<td>FVB/NJ</td>
<td>LPS</td>
<td>Lung</td>
<td>weekly treatment for 16 weeks</td>
<td>3</td>
</tr>
</tbody>
</table>
2.2. Identification of Significant Genes

The Microarray dataset retrieved from the NCBI gene expression omnibus GEO database with ID GSE132661 ranked the significant genes based on the criteria $p$-value > 0.05 (Clough and Barrett 2016).

2.3. Classification of Upregulated and Down-regulated Genes

In GEO2R analysis the upregulated genes (UG) (log FC ≥ 1) and downregulated genes (DG) (log FC ≤ −1) were co-expressed in lung cancer.

2.4. Gene-gene Interaction

STRING analysis is to relate the upregulated and downregulated genes to find possible interactions (Szklarczyk et al. 2019) involved in the inflammatory response of Lipopolysaccharide (LPS) carcinogen with Phosphate buffered salt (PBS), controller.

2.5. Gene Enrichment Analysis

Analyze the data from string analysis to identify the gene ontologies of biological process, molecular function, cellular component, and pathways like KEGG and Reactome in differential expressed genes.

2.6. Refined Gene Enrichment Analysis

Enrichr analysis is used to compare the input gene sets to find pathways and ontologies involved in inflammatory responses of Lipopolysaccharide (LPS) carcinogen in causing lung cancer (Chen et al. 2013). In enrichment analysis contains cellular components (CC), molecular function (MF), biological process (BP), biological pathways, transcription, cell types, disease/drugs. Input the common genes for enrichment analysis.
2.7.Hub Gene Identification

Import the file from the STRING database to Cytoscape and download cytohubba from the app manager in Cytoscape. 'CytoHubba' plugin of 'Cytoscape' is to visualize the molecular interaction networks of eight HUB genes with gene expression profiles involved in lung cancer (Otasek et al. 2019).

2.8.Integrated Gene Enrichment

'ClueGO' plugin of 'Cytoscape' is to visualize the pathways of eight HUB genes involved in inflammatory responses of Lipopolysaccharide (LPS) carcinogen lung cancer (In-Yee Lee, Jan-Ming Ho, and Ming-Syan Chen 2005). It includes gene ontologies like biological process, Cellular component, Immune system process, Molecular function, Protein domains, and pathways like KEGG, Reactome, and wiki pathways.

3. Results

3.1.Identification of Significant Genes

The Microarray dataset was retrieved from NCBI gene expression omnibus (GEO) with ID GSE132661. The upregulated genes -73 and downregulated genes -56 were identified and we observed 11753 genes were significant, showing inflammatory responses in lung cancer. (Table 2).

<table>
<thead>
<tr>
<th>Table 2 - Data Retrieved from GEO Database</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of significant genes</td>
</tr>
<tr>
<td>Total number of upregulated genes</td>
</tr>
<tr>
<td>Total number of downregulated genes</td>
</tr>
</tbody>
</table>

3.2.Gene-Gene Interaction

STRING analysis is used for relating the upregulated (UR) and downregulated (DR) genes, we found 33 possible interactions in lung cancer. (Fig. 1) (Costantini, Ferrara, and Cortesi 2011).
3.3. Refined Gene Enrichment Analysis

Enrichr analysis is used to compare the input gene sets with annotated gene sets (Kuleshov et al. 2016) to find pathways like the Calcium signalling pathway, (Fig. 2a) T-cell receptor signalling pathway, (Fig. 2b) (Ping, Liu, and Zhang 2018). Interleukin-17 signalling pathway, (Fig. 2c), and ontologies like T-cell receptor complex, (Fig. 2d). cellular response of cytokine stimulus, (Fig. 2e) neutrophil-mediated immunity, (Fig. 2f) (Reimand et al. 2019). These pathways and ontologies were mostly involved in inflammatory responses of Lipopolysaccharide (LPS) in causing lung cancer.
T helper cell surface molecules
Interleukin-17 signaling pathway
Inositol phosphate metabolism
Arrhythmogenic right ventricular cardiomyopathy (ARVC)
Striated muscle contraction
Dilated cardiomyopathy
Activation of TRKA receptors
Cytochrome P450 metabolism of vitamins
Muscle contraction
T cell receptor signaling in naive CD8+ T cells

Arrhythmogenic right ventricular cardiomyopathy (ARVC)
Inositol phosphate metabolism
Hypertrophic cardiomyopathy (HCM)
Dilated cardiomyopathy (DCM)
Adrenergic signaling in cardiomyocytes
Gastric acid secretion
Cardiac muscle contraction
Nitrogen metabolism
Taste transduction
Calcium signaling pathway

Spermatogenesis (GO:0007283)
Male gamete generation (GO:0048232)
Regulation of male gonad development (GO:2000018)
Neuron migration (GO:0001764)
Neurogenesis (GO:0042063)
Negative regulation of response to stimulus (GO:0048585)
Generation of neurons (GO:0048659)
Anterograde trans-synaptic signaling (GO:0098916)
Positive regulation of axonogenesis (GO:0030772)
Actin-myosin filament sliding (GO:0033275)
3.4. HUB Genes Identification

‗CytoHubba‘ plugin ‗Cytoscape‘ is used to visualize the molecular interaction networks of eight HUB genes with gene expression profiles from the whole network (Shannon et al. 2003), we identified CD81, CD247, CD1D2, NANOS2, CYP26B1, SST, CCKBR, GPR132 as HUB genes overexpressed in inflammatory responses leads to the lung cancer in Mus musculus (Fig. 3).
3.5. Integrated Gene Enrichment Analysis

‗ClueGO‘ plugin of ‗Cytoscape‘ is to visualize the pathways of eight HUB genes involved in inflammatory responses (Shen et al. 2017), we analyzed the pathway like TCR complex interacts with peptide antigen-presenting MHC Class 1 (Fig. 4).

Fig. 4 - Pathway Enrichment of Eight HUB Genes Showed Unilateral Dysregulation of Genes Involved in the Inflammatory Response
4. Discussion

The microarray dataset is used to find upregulated and downregulated genes and we identified 11753 genes as significant genes which cause inflammatory responses. STRING analysis is used to relate the upregulated and downregulated genes to find possible interactions involved in the inflammatory response of Lipopolysaccharide (LPS) in causing lung cancer. Enrichr analysis is used to compare the input gene sets to find pathways and ontologies involved in inflammatory responses like the Calcium signalling pathway, (Fig. 2a) Calcium, as a second messenger, is an important signal transduction element in cell growth, including cell cycle, differentiation, proliferation, and apoptosis. Calcium signalling is activated in a pathological cell, triggering the intracellular environment to switch the cell to react abnormally (Yang et al. 2010) T-cell receptor signalling pathway, (Fig. 2b) T cells are a new type of cell that can be used in adoptive cell therapy. Which is used to treat a variety of advanced cancers (Ping, Liu, and Zhang 2018). Interleukin-17 signalling pathway, (Fig. 2c), IL-17 and the Wnt signalling pathway were suspected of being involved in the treatment of ischemic stroke and the promotion of lung cancer. (L. Sun et al. 2020) and ontologies like T-cell receptor complex, (Fig. 2d) TCRs on each T cell are distinct and differ between individuals and populations, allowing immune responses to a wide variety of foreign antigens. (Y.-Y. Liu et al. 2019). cellular response of cytokine stimulus, (Fig. 2e) neutrophil-mediated immunity, (Fig.2f) (Reimand et al. 2019). These pathways and ontologies were mostly involved in inflammatory responses of Lipopolysaccharide (LPS) in causing lung cancer.

‘CytoHubba’ plugin ‘Cytoscape’ is used to identified CD81, CD247, CD1D2, NANOS2, CYP26B1, SST, CCKBR, GPR132 as HUB genes overexpressed in inflammatory responses. CD81 - Leukocytosis and impaired transendothelial neutrophil emigration are observed in Mus musculus with targeted mutations of CD18, the common 2 subunits of CD11/CD18 integrins (Wu et al. 2003). CYP26B1- CYP26B1 overexpression may disrupt T cell trafficking and differentiation in the gut and lymphoid organs(Shannon et al. 2003). GPR132 - deletion of GPR132 slows inflammation and cancer growth, but it also negates the anti-tumour effects of PPAR and rosiglitazone (Cheng et al. 2016). Somatostatin receptors, sst1 and sst2, were mainly expressed in the majority of neuroblastomas at the initial diagnosis, and upregulation of functional sst1 or sst2 in neuroblastoma cell lines suppresses tumorigenicity in a xenograft model. (Albers et al. 2000).

‘CluGO’ plugin of ‘Cytoscape’ is used to analyze the pathways like TCR complex interacts with peptide antigen-presenting MHC Class 1 (Fig. 4). When T lymphocytes with alpha-beta T cell receptors (TCRs) interact with a molecular complex composed of a peptide bond to an MHC-encoded
class I or class II molecule on the surface of an antigen-presenting cell (Corr et al. 1994). Antigen-presenting cells can activate CD4+ T cells, resulting in the coordination and regulation of effector cells. In all cases, a clonotypic T cell receptor interacts with a specific pMHC complex, potentially resulting in the sustained cell-cell contact formation, and T cell activation leads to cell proliferation and the mounting of a specific cellular immune response (Wieczorek et al. 2017).

CD81 (TAPA-1) is a widely expressed cell-surface protein involved in a wide range of biologic responses (Levy, Todd, and Maecker 1998). The CD247 gene is involved in T-cell signalling and has been linked to blood pressure in human genetic studies (Rudemiller et al. 2014). CD1d2 molecules present different sets of self-antigen(s) in the mouse thymus, influencing the development of invariant NKT cells (Sundararaj et al. 2018). The presence of CYP26 in the sebaceous gland epithelium lends support to the theory that altered RA metabolism may play a role in the pathogenesis of acne (Heise et al. 2006)

Our institution is passionate about high quality evidence based research and has excelled in various fields ((Vijayashree Priyadharsini 2019; Ezhilarasan, Apoorva, and Ashok Vardhan 2019; Ramesh et al. 2018; Mathew et al. 2020; Sridharan et al. 2019; Pc, Marimuthu, and Devadoss 2018; Ramadurai et al. 2019). We hope this study adds to this rich legacy.

The major limitation of our study is that these genes and pathways need to be confirmed by wet lab techniques such as western blot and RT-PCR before they can be clinically applied. We identified the gene expression of ten hub genes with gene expression profiles involved in inflammatory responses of LPS carcinogen in causing lung cancer. Furthermore, our results suggest that these differentially expressed genes could be targeted for the further drug discovery process.

5. Conclusion

People who smoke cigarettes were already 15 to 30 times more likely to develop lung cancer. Lipopolysaccharide is one of the carcinogens present in tobacco and acts as a cofactor for many types of carcinogenesis. We identified overexpressed genes that were upregulated and downregulated in lung cancer. We observed possible interactions of overexpressed genes. Hence, it is desirable to collect and integrate all types of gene-gene interactions under one framework. We identified CD81, CD247, CD1D2, NANOS2, CYP26B1, SST, CCKBR, GPR132 as overexpressed genes, BC048507, CDCA8, NTRK2, RELN, PIK3C2A, PIP5KLI as downregulated genes, and we analyzed the pathway like TCR complex interacts with peptide antigen-presenting MHC Class-I. Furthermore, our results
provide for data analysis pipelines in diverse areas ranging from disease module identification to novel biomarker discovery.

Declarations

Conflict of Interest

The authors of this paper declare no conflict of interest.

Author Contribution

Author MM was involved in data collection, data analysis, manuscript writing. Author MD was involved in conceptualization, guidance and critical review of manuscript.

Acknowledgments

The authors would like to express their gratitude towards Saveetha School of engineering, Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for providing the necessary infrastructure to carry out this work successfully.

Funding

We thank the following organizations for providing financial support that enabled us to complete the study.

1. Finura Bioteks
2. Saveetha University
3. Saveetha Institute of Medical and Technical Sciences
4. Saveetha School of engineering
References


In-Yee Lee, Jan-Ming Ho, and Ming-Syan Chen. 2005. “CLUGO: A Clustering Algorithm for Automated Functional Annotations Based on Gene Ontology.” In Fifth IEEE International Conference on Data Mining (ICDM’05), 4 pp. –. https://doi.org/10.1109/ICDM.2005.42.


