Genetic Effects of Welding Fumes on the Development of Respiratory System Diseases

Humayan Kabir Rana[a,†], Mst. Rashida Akhtar[b,†], M. Babul Islam[c], Mohammad Boshir Ahmed[d], Pietro Lio'[e], Julian M.W. Quinn[f], Fazlul Huq[g], Mohammad Ali Moni[f,g,*]

[a]*Department of Computer Science and Engineering, Green University of Bangladesh*
[b]*Department of Computer Science and Engineering, Varendra University, Rajshahi, Bangladesh*
[c]*Department of Applied Physics and Electronic Engineering, University of Rajshahi, Bangladesh*
[d]*School of Civil and Environmental Engineering, University of Technology Sydney, NSW 2007, Australia*
[e]*Computer Laboratory, The University of Cambridge, 15 JJ Thomson Avenue, Cambridge, UK*
[f]*Bone Biology Division, Garvan Institute of Medical Research, Darlinghurst, NSW, Australia*
[g]*School of Biomedical Science, Faculty of Medicine and Health, The University of Sydney, Australia*

Abstract

Background: The welding process releases potentially hazardous gases and fumes, mainly composed of metallic oxides, fluorides and silicates. Long term welding fume (WF) inhalation is a recognized health issue that carries a risk of developing chronic health problems, particularly respiratory system diseases (RSDs). Aside from general airway irritation, WFs may drive direct cellular responses in the respiratory system which increase risk of RSD, but these are not well understood.

Methods: We developed a quantitative framework to identify gene expression effects of WFs that may affect RSD development. We analyzed gene expression microarray data from WF-exposed tissues and RSD-affected tissues, including chronic bronchitis (CB), asthma (AS), pulmonary edema (PE), lung cancer (LC) datasets. We built disease-gene (diseasome) association networks and identified dysregulated signaling and ontological pathways, and protein-protein interaction sub-network using neighborhood-based benchmarking and multilayer network topology.

Results: We observed many genes with altered expression in WF-exposed tissues were also among differentially expressed genes (DEGs) in RSD tissues; for CB, AS, PE and LC there were 34, 27, 50 and 26 genes respectively. DEGs analysis, with disease association networks, pathways, ontological analysis and protein-protein interaction sub-network suggest significant links between WF exposure and the development of CB, AS, PE and LC.

Conclusions: Our network-based analysis and investigation of the genetic links of WFs and RSDs confirm a number of genes and gene products are plausible participants in RSD development. Our results are a significant resource to identify causal influences on the development of RSDs, particularly in the context of WF exposure.

[†]These authors hold joint first authorship of this work.

Keywords: Welding fumes, Respiratory system diseases, Chronic Bronchitis, Asthma, Pulmonary Edema, Lung Cancer.

1. Introduction

Welding is the process of joining metal at their contacting surface by using high temperature melting. It can very hazardous because it exposes the welder and others nearby to a variety of gases, WFs and radiant energy. Among WFs some the most dangerous components [1] include are metallic oxides, fluorides and silicates including those of beryllium, aluminum, cadmium, chromium, copper, iron, lead, manganese, magnesium, nickel, vanadium, and zinc [2]. A known risk associated with large scale and long term WF exposure is the development of RSDs [1, 3]. The human respiratory system consists of many delicate tissues that transport air into the lungs for rapid uptake of oxygen in blood while eliminating carbon dioxide. The same exposed cell-rich surfaces that facilitate oxygen transport and uptake to the bloodstream also facilitate uptake of noxious gasses. In addition, such toxic chemicals can directly damage cells lining the respiratory system. This predisposes the respiratory system to develop RSDs when affected by number of factors and pathogens that lead to RSDs such as asthma (AS), chronic bronchitis (CB), pulmonary edema (PE) and even lung cancer (LC) [4]. How WF and other pollutants do this is unclear, as indeed are the influences of many pathogens on lung tissues. One way to investigate the WFs influences in causing or promoting such ill-health is to investigate gene expression effects of WFs on respiratory tissues and how this compares with tissues affected by CB, AS, PE and LC.

CB reduces the functionality of lung through damage to the cilia of the breathing passages and obstructs airflow inside the lung [5]. Iron, Aluminum and Magnesium oxide WFs may have particularly strong effects on this disease [6]. Asthma is a pulmonary system disorder that narrows and inflames the airways and is characterized by episodic wheezing, shortness of breath, chest tightness and coughing. It is often triggered by allergen exposure and air pollutants [7]. Nitrogen oxides, carbon monooxide and phosgene in WFs are components particularly linked to asthma [6]. PE is a chronic lung condition characterized by fluid buildup in the lung, leading to serious breathing problems, coughing and chest pain [8]. Many irritants can drive PE but nickel, beryllium and cobalt oxides of the WFs have been linked for to PE as well as and LCs [6]. The latter, including both small cell and non-small cell carcinomas, are some of the deadliest types of cancer and is a leading cause of death in the world [9]. Most cases are linked to tobacco smoke, though genetic factors and air pollutants play important roles. Such pollutants include many of the components of WFs, which can be taken up into, and affect, pulmonary tissues [6].

In this study, we developed a systematic and quantitative network-based approach to investigate how gene expression changes seen in WF exposure may interact with RSDs. Thus, we studied the gene expression in RSD affected tissues, namely CB, AS, PE and LC.

*Corresponding author:
Email address: mohammad.moni@sydney.edu.au (Mohammad Ali Moni)

To find the genetic effects of WFs on RSDs, we analyzed differentially expressed genes from these disease (found by comparing gene expression in WF and RSD affected tissues with controls), disease association network, signaling and ontology pathways, and protein-protein interaction networks. We also examined the validity of our study by employing the gold benchmark databases dbGAP and OMIM.

2. Materials and Methods

2.1. Datasets employed in this study

To understand the genetic effects of WFs on RSDs at the molecular level, we analyzed gene expression microarray data. In this study, we employed gene expression microarray data from the National Center for Biotechnology Information (NCBI) (*http://www.ncbi.nlm.nih.gov/geo/*). We analyzed 5 different datasets with accession numbers GSE62384, GSE22148, GSE69683, GSE68610 and GSE10072 [10, 11, 12, 13, 14]. The WFs dataset (GSE62384) is a microarray data of fresh welding fumes influence on upper airway epithelial cells (RPMI 2650). This data is collected from the people with spark-generated welding fumes at high (760 g/m3) and low (85 g/m3) concentrations. The donors inhaled welding fumes for 6 hours continuously, followed by zero hours or four hours post-exposure incubation. The CB dataset (GSE22148) is taken from the 480 ex-smokers and stage 2 - 4 chronic obstructive pulmonary disease patients from the United Kingdom. The asthma dataset (GSE69683) is from microarray data collected from severe and moderate asthmatic patients and healthy subjects. The PE dataset (GSE68610) is an Affymetrix gene expression array of 25 acute lung injury patients. The LC dataset (GSE10072) is a microarray data where RNA was collected from 28 current smokers, 26 former smokers and 20 never smokers. In this data, final gene expression values have been reported by comparing between 58 tumor and 49 non-tumor lung tissues.

2.2. Overview of analytical approach

We applied a systematic and quantitative framework to investigate the genetic effects of WFs to the development of the RSDs by using several sources of available microarray datasets. The graphical representation of this approach is shown in figure 1. This approach uses gene expression microarray data to identify dysregulated genes, and identified commonly dysregulated genes for each RSD with WFs. Further, these common dysregulated genes are employed to analyze signaling pathway, gene ontology (GO) and protein-protein interaction (PPI) sub-networks. Gold benchmark data is also used in this approach to verify the validation of our study.

2.3. Analysis methods

Microarray based gene expression analysis is a global and sensitive method to find and quantify the human disorders at the molecular level [15]. We used these technologies to analyze the gene expression profiles of CB, AS, PE and LC to find the genetic effects of WFs on the development of respiratory system diseases. To make uniform mRNA expression data of different platforms and to avoid the problems of experimental systems, we normalized the

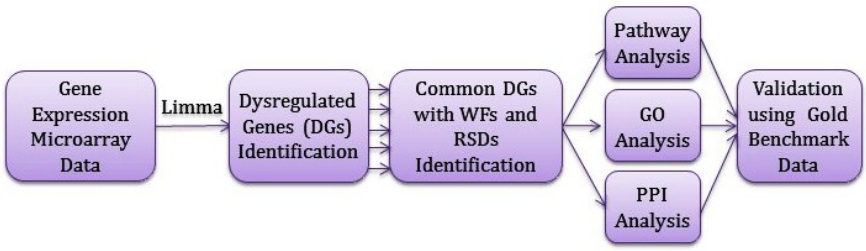


Figure 1: Flow-diagram of the analytical approach used in this study.

gene expression data by using the Z-score transformation (Z_{ij}) for each RSD gene expression profile using $Z_{ij} = \frac{g_{ij} - mean(g_i)}{SD(g_i)}$,

where SD denotes the standard deviation, g_{ij} denotes the value of the gene expression i in sample j. After this transformation gene expression values of different diseases at different platforms can be directly compared. We applied two conditions for t-test statistic. We applied unpaired T-test to find differentially expressed genes of each disease over control data and selected significant dysregulated genes. We have chosen a threshold of at least 1 log_2 fold change and a p-value for the t-tests of $<= 1 * 10^{-2}$.

We employed neighborhood-based benchmark and the multilayer topological methods to find gene-disease associations. We constructed gene-disease network (GDN) using the gene-disease associations, where the nods in the network represent either gene or disease. This network can also be recognized as a bipartite graph. The primary condition for a disease to be connected with other diseases in GDN is they should share at least one or more significant dysregulated genes. Let D is a specific set of diseases and G is a set of dysregulated genes, gene-disease associations attempt to find whether gene $g \in G$ is associated with disease $d \in D$. If G_i and G_j, the sets of significant dysregulated genes associated with diseases D_i and D_j respectively, then the number of shared dysregulated genes (n^g_{ij}) associated with both diseases D_i and D_j is as follows [15]:

$$n^g_{ij} = N(G_i \cap G_j) \tag{1}$$

The common neighbours are the based on the Jaccard Coefficient method, where the edge prediction score for the node pair is as [15, 16]:

$$E(i,j) = \frac{N(G_i \cap G_j)}{N(G_i \cup G_j)} \tag{2}$$

where G is the set of nodes and E is the set of all edges. We used R software packages "comoR" [17] and "POGO" [18] to cross check the genes-diseases associations.

To investigate the molecular determinants of WFs on several RSDs, we analyzed pathway and gene ontology using Enrichr ($https://amp.pharm.mssm.edu/Enrichr/$). We used KEGG, WikiPathways, Reactome and BioCarta databases for analyzing signaling pathway [19, 20, 21, 22]. We used GO Biological Process and Human Phenotype Ontology databases for ontological analysis [23, 24]. We also constructed a protein-protein interaction subnetwork for each RSD, using the STRING database ($https://string-db.org$), a biological

database and web resource of known and predicted protein-protein interactions. Furthermore, we examined the validity of our study by employing two gold benchmark databases OMIM ($https://www.omim.org$) and dbGAP ($https://www.ncbi.nlm.nih.gov/gap$).

3. Results

3.1. Gene Expression Analysis

We analyzed the gene expressing microarray data from the National Center for Biotechnology Information (NCBI) ($http://www.ncbi.nlm.nih.gov/geo/$) for investigating the genetic effect of WFs to the development of RSDs. We observed that WF exposure results in 903 DEGs for which the adjusted $p <= .01$ and $|logFC| >= 1$. The DEGs of WFs contain 392 genes showing up and 511 with down-regulated expression. Similarly, we identified the most significant dysregulated genes for each RSD after applying statistical analysis. In CB We identified 678 DEGs (463 up and 215 down), in asthma 602 DEGs (297 up and 305 down), in PE 759 DEGs (404 up and 355 down) and in LC 890 DEGs (673 up and 217 down). We also performed cross-comparative analysis to find the DEGs common between between the WF dataset and each RSD dataset. We observed that WFs shares a large number of DEGs with the other datasets, specifically 34, 27, 50 and 26 for CB, asthma, PE and LC respectively. To investigate the significant associations for DEGs from the RSDs with the WF datasets, we built two separate gene-disease association networks for up and down-regulated genes, centered on the WFs as shown in figure 2 and 3. The essential condition for two diseases to be associated with each is they should have at least one or more common genes in between them [25, 26]. Notably, 2 significant genes, PMAIP1 and SEC24A are commonly differentially expressed among WFs, PE, AS and LC; 6 genes, FGFR3, ID3, PROS1, AK4, TOX3 and MTHED2 are common to WFs, PE and LC. Similarly, PECR, ALDH3A2 and TPD52L1 are commonly differentially expressed in WFs, PE and CB; one gene HSD17BB is dysregulated among WFs, asthma and CB.

3.2. Pathway and Functional Association Analysis

Pathways are the important key to understand the reactions of an organism for its internal changes. The pathway-based analysis is a modern technique to understand how different complex diseases are related to each other by underlying molecular or biological mechanisms [27]. We analyzed pathways of the commonly dysregulated genes of WFs and each RSD using Enrichr, a comprehensive web-based tool for analyzing gene set enrichment [28]. Signaling pathways of the commonly differentially expressed genes in between WFs and each RSD were analyzed using four global databases includes KEGG, WikiPathways, Reactome and BioCarta. We collected and combined pathways from the mentioned four databases and selected the most significant pathways of each RSD after several steps of statistical analysis.

Four significant pathways associated with CB as shown in figure 4(a). Among these pathways, 'Integrated Pancreatic Cancer Pathway' contains some proteins that are known as important in human breast cancer development [29, 30]. The pathway 'Assembly of the motile cilium' is only known in cilia functions such as respiratory cilia [31]. 'Methionine

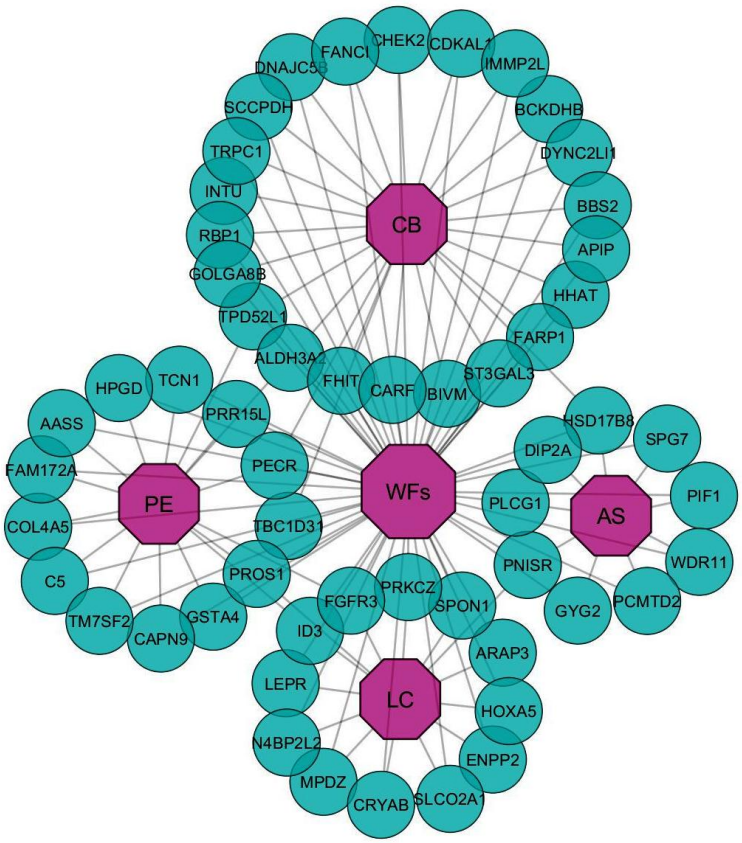


Figure 2: Gene-disease association network of Welding fumes (WFs) with Chronic Bronchitis (CB), Asthma (AS), Pulmonary Edema (PE) and Lung Cancer (LC). Octagon-shaped violet colored nodes represent different RSDs, and dark-cyan colored round-shaped nodes represent commonly up-regulated genes for WFs with the other respiratory system diseases. A link is placed between a disease and a gene if mutations in that gene lead to the specific disease.

salvage' is a pathway that performs six reactions in the human body and is responsible for the recycling of sulfur in the respiratory system [32]. The pathway 'Glyoxylate metabolism and glycine degradation' is responsible for overproduction of oxalate in human liver and lung. [33].

The top five significant pathways associated with AS are shown in figure 4(b). Among these pathways, 'Non-small cell lung cancer' is linked to development of adeno-squamous cell and large-cell carcinoma types that are responsible for approximately 75% of all lung cancer [34, 35]. 'IL-1 Signaling Pathway' is responsible to control pro-inflammatory responses, including inflammasome activity tissues [36]. 'Glycogen synthesis' is responsible for several reactions to produce glycogen in liver, muscle, lung and other tissues; glycogen serves as a major stored short-term fuel in muscle and other high metabolism cells [37]. 'Constitutive Signaling by EGFRvIII' is linked to the general development of tumors. 'Signaling by EGFRvIII in Cancer' is more responsible for cancer in several cell types found in the respiratory system.

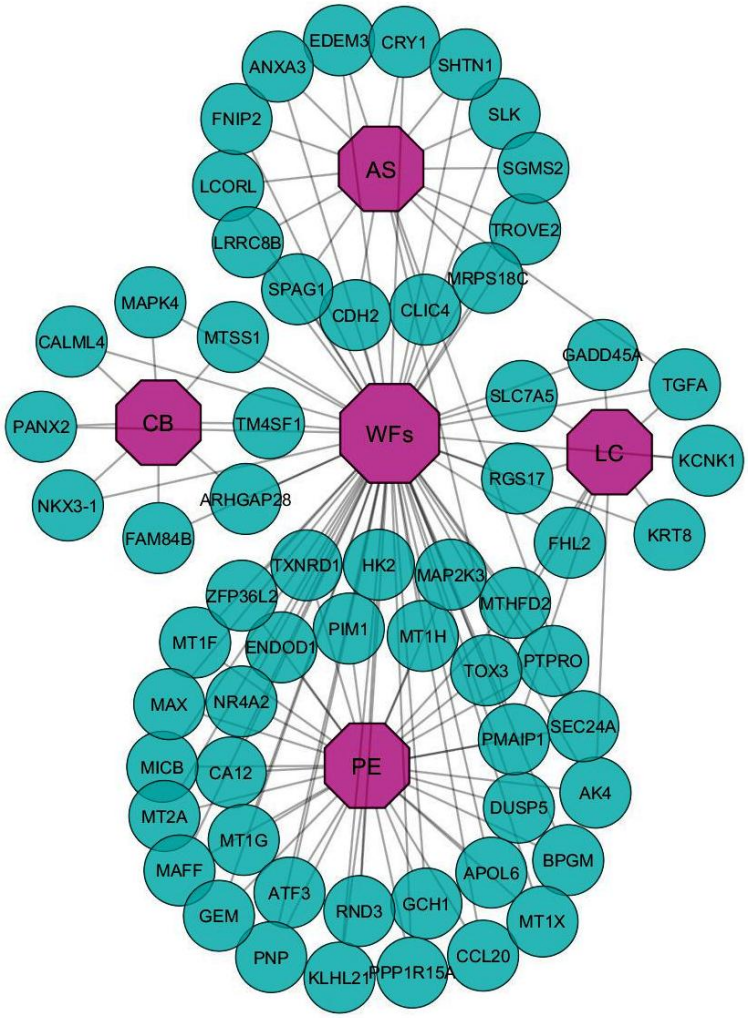
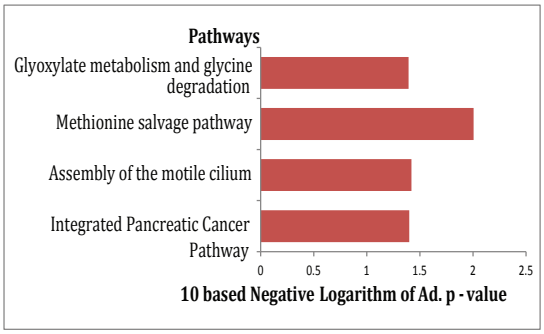


Figure 3: Gene-disease association network of Welding fumes (WFs) with Chronic Bronchitis (CB), Asthma (AS), Pulmonary Edema (PE) and Lung Cancer (LC). Octagon-shaped violet colored nodes represent different RSDs, and dark-cyan colored round-shaped nodes represent commonly down-regulated genes for WFs with the other respiratory system diseases. A link is placed between a disease and a gene if mutations in that gene lead to the specific disease.

Five significant pathways associated with PE are indicated in figure 4(c). These include 'Central carbon metabolism in cancer' which includes three transcription factors HIF-1, c-MYC and p53 that influence the regulate tumor cell growth [38]. 'Small cell lung cancer' pathway relates to a highly aggressive neoplasm responsible for approximately 25% of all lung cancer [39]. 'TNF signaling pathway' is involved in the control of inflammation, immunity and cell survival. 'One carbon pool by folate' is pathway linked to progression of several forms of cancer [40]. 'Response to metal ions' pathway may mediate cell toxicity effects of heavy metals such as zinc, copper, and iron [41].

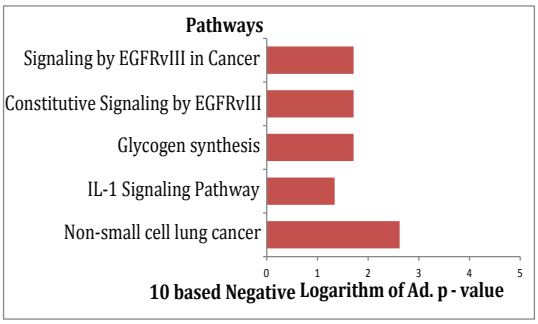
The top five significant pathways associated with LC are shown in figure 4(d). Among

Pathway	Genes in the pathway	Adjusted p-value
Integrated Pancreatic Cancer Pathway	CHEK2, MAPK4	3.98E-02
Assembly of the motile cilium	BBS2, DYNC2LI1	3.80E-02
Methionine salvage pathway	APIP	9.86E-03
Glyoxylate metabolism and glycine degradation	BCKDHB	4.05E-02



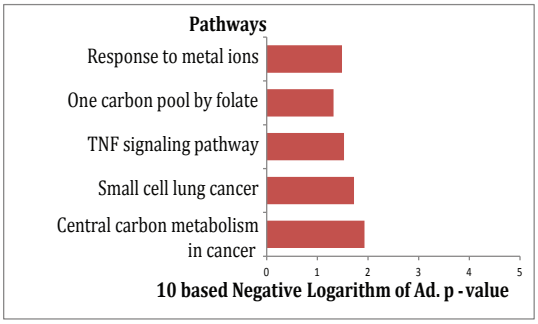
(a) Pathways associated with significantly common differentially expressed genes of the CB with WFs.

Pathway	Genes in the pathway	Adjusted p-value
Non-small cell lung cancer	TGFA, PLCG1	2.40E-03
IL-1 Signaling Pathway	PLCG1	4.58E-02
Glycogen synthesis	GYG2	1.93E-02
Constitutive Signaling by EGFRvIII	PLCG1	1.93E-02
Signaling by EGFRvIII in Cancer	PLCG1	1.93E-02



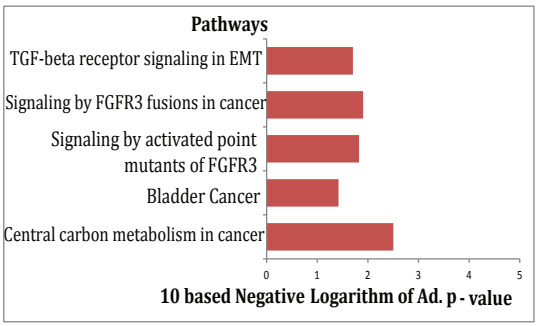
(b) Pathways associated with significantly common differentially expressed genes of the AS with WFs.

Pathway	Genes in the pathway	Adjusted p-value
Central carbon metabolism in cancer	FGFR3, HK2	1.17E-02
Small cell lung cancer	MAX, COL4A5	1.89E-02
TNF signaling pathway	MAP2K3, CCL20	2.98E-02
One carbon pool by folate	MTHFD2	4.79E-02
Response to metal ions	MT2A, MT1F, MT1G	3.27E-02



(c) Pathways associated with significantly common differentially expressed genes of the PE with WFs.

Pathway	Genes in the pathway	Adjusted p-value
Central carbon metabolism in cancer	SLC7A5, FGFR3	3.16E-03
Bladder Cancer	FGFR3	3.81E-02
Signaling by activated point mutants of FGFR3	FGFR3	1.49E-02
Signaling by FGFR3 fusions in cancer	FGFR3	1.24E-02
TGF-beta receptor signaling in EMT	PRKCZ	1.98E-02



(d) Pathways associated with significantly common differentially expressed genes of the LC with WFs.

Figure 4: Pathway analyses to identify significant pathways common to the WFs and the RSDs revealed by the commonly expressed genes.These include significant pathways common to WFs and (a) CB (b) AS (c) PE and (d) LC.

these pathways, 'Central carbon metabolism in cancer' is responsible for three transcription factors HIF-1, c-MYC and p53 that can coordinate regulation of tumor and cancer in cell [38]. 'Bladder Cancer' is responsible for developing several carcinomas in urinary tract, renal pelvis, ureter and bladder [42]. 'Signaling by activated point mutants of FGFR3' is responsible for several cancer including breast, prostate, bladder, cervical, neck and head [43]. 'Signaling by FGFR3 fusions in cancer' is specifically responsible for lung and bladder cancer [44]. 'TGF-beta receptor signaling in EMT' is responsible for the development of tumors in the early stage that can bring cancer in cells [45].

3.3. Gene Ontological Analysis

The Gene Ontology (GO) refers to a universal conceptual model for representing gene functions and their relationship in the domain of gene regulation. It is constantly expanded by accumulating the biological knowledge to cover regulation of gene functions and the relationship of these functions in terms of ontology classes and semantic relations between classes [46]. GO of the commonly dysregulated genes for each RSD and WFs were analyzed using two databases including GO Biological Process and Human Phenotype Ontology. We collected and concatenated gene ontologies from mentioned two databases, and performed several statistical analyses to identify the most significant ontologies of each respiratory system diseases. Notably, we found 15, 15, 24 and 17 gene ontology terms are associated with the CB, AS, PE and LC respectively as shown in table 1.

3.4. Protein-Protein Interaction Analysis

Protein-protein interaction network (PPIN) is the graphical representation of the physical connection of proteins in the cell. Protein-protein interactions (PPIs) are essential to every molecular and biological process in a cell, so PPIs is crucial to understand cell physiology in disease and healthy states [47]. We have used STRING database to analyze and construct protein-protein interaction subnetworks of the significantly commonly dysregulated genes of each RSD. We have clustered into four different groups of protein-protein interactions of four RSDs as shown in figure 5.

4. Discussion

We investigated the gene expression effects of WFs in the development of RSDs based on gene regulation analysis, gene-disease association networks, signaling pathways, gene expression ontologies and protein-protein interaction sub-networks. For the purpose of our study, we analyzed gene expression microarray data from WFs, CB, AS, PE, LC and control datasets. We identified a large number of genes that were commonly differentially expressed in WFs and RSD detatsets by gene expression analysis. As there were many significantly DEGs in common between WFs and RSDs it indicates that WFs may induce gene expression mediated effects in lung cells that increase the incidence and/or severity of RSDs. We constructed two separate gene-disease association networks for up and down-regulated genes showed a strong evidence that WFs are linked to the development of RSDs as shown in Figure 2 and 3. The pathway-based analysis is useful to understand how different complex

Table 1: Gene ontology analyses to identify significant ontological pathways common to the WFs and the RSDs revealed by the commonly expressed genes.These include significant ontological pathways common to WFs and (a) CB (b) AS (c) PE and (d) LC.

GO Term	Pathway	Genes in the Pathway	Adjusted p-value
GO:2001233	Regulation of apoptotic signaling pathway	TPD52L1, NKX3-1	8.37E-04
GO:0060562	Epithelial tube morphogenesis	MTSS1, NKX3-1	1.52E-03
GO:0050678	Regulation of epithelial cell proliferation	MTSS1, NKX3-1	6.45E-03
GO:0046467	Membrane lipid biosynthetic process	ALDH3A2, SCCPDH	4.55E-03
GO:0051225	Spindle assembly	GOLGA8B, CHEK2	7.89E-03
GO:2001235	Positive regulation of apoptotic signaling pathway	TPD52L1, NKX3-1	7.15E-03
HP:0100631	Neoplasm of the adrenal gland	CHEK2	2.61E-02
HP:0002858	Meningioma	CHEK2	2.77E-02
HP:0007707	Congenital primary aphakia	BBS2	2.93E-02
HP:0100641	Neoplasm of the adrenal cortex	CHEK2	1.64E-02
HP:0002476	Primitive reflexes	ST3GAL3	1.48E-02
HP:0012126	Stomach cancer	CHEK2	1.48E-02

(a) Gene ontologies associated with the significantly common dysregulated genes of the CB with WFs.

GO Term	Pathway	Genes in the Pathway	Adjusted p-value
GO:2000323	Negative regulation of glucocorticoid receptor signaling pathway	CRY1	9.07E-03
GO:0031334	Positive regulation of protein complex assembly	PMAIP1, FNIP2	5.94E-03
GO:1904018	Positive regulation of vasculature development	ANXA3, PLCG1	8.17E-03
GO:0010634	Positive regulation of epithelial cell migration	ANXA3, PLCG1	4.25E-03
GO:0032147	Activation of protein kinase activity	SLK, TGFA, PLCG1	3.37E-03
GO:0043067	Regulation of programmed cell death	SLK, PMAIP1, DIP2A	4.98E-03
HP:0002492	Abnormality of the corticospinal tract	SPG7	1.29E-02
HP:0004469	Chronic bronchitis	SPAG1	1.55E-02
HP:0012261	Abnormal respiratory motile cilium physiology	SPAG1	2.44E-02
HP:0200073	Respiratory insufficiency due to defective ciliary clearance	SPAG1	1.16E-02
HP:0012384	Rhinitis	SPAG1	2.44E-02
HP:0000458	Anosmia	WDR11	2.82E-02
HP:0012387	Bronchitis	SPAG1	3.58E-02

(b) Gene ontologies associated with the significantly common dysregulated genes of the AS with WFs.

GO Term	Pathway	Genes in the Pathway	Adjusted p-value
GO:0071294	Cellular response to zinc ion	MT2A, MT1F, MT1G	1.08E-09
GO:0071276	Cellular response to cadmium ion	MT1G, MT1H, MT1X	9.74E-09
GO:0006882	Cellular zinc ion homeostasis	MT1F, MT1H, MT1X	1.37E-08
GO:0071280	Cellular response to copper ion	MT1X, MT1G, MT2A	1.83E-09
GO:0045926	Negative regulation of growth	MT2A, MT1F, MT1G	4.89E-07
GO:0046916	Cellular transition metal ion homeostasis	MT2A, MT1F, MT1G, MT1H, MT1X	3.34E-06
GO:1905898	Positive regulation of response to endoplasmic reticulum stress	PPP1R15A, PMAIP1	7.81E-04
GO:0071248	Cellular response to metal ion	MT2A, MT1F, MT1G	4.75E-06
HP:0002970	Genu varum	HPGD, FGFR3	2.61E-03
HP:0006297	Hypoplasia of dental enamel	ALDH3A2, FGFR3	3.32E-03

(c) Gene ontologies associated with the significantly common dysregulated genes of the PE with WFs.

GO Term	Pathway	Genes in the Pathway	Adjusted p-value
GO:1903708	Positive regulation of hemopoiesis	N4BP2L2, HOXA5	4.18E-05
GO:0009132	Nucleoside diphosphate metabolic process	AK4	9.96E-03
GO:0045647	Negative regulation of erythrocyte differentiation	HOXA5	8.72E-03
GO:2000665	Regulation of interleukin-13 secretion	PRKCZ	8.72E-03
GO:0044320	Cellular response to leptin stimulus	LEPR	9.96E-03
GO:0045837	Negative regulation of membrane potential	PMAIP1	8.72E-03
GO:2000394	Positive regulation of lamellipodium morphogenesis	ENPP2	9.96E-03
GO:0050730	Regulation of peptidyl-tyrosine phosphorylation	ENPP2, PRKCZ	5.14E-03
GO:0032754	Positive regulation of interleukin-5 production	PRKCZ	9.96E-03
GO:0043065	Positive regulation of apoptotic process	GADD45A, PMAIP1	6.47E-03
HP:0000975	Hyperhidrosis	SLCO2A1, FGFR3	5.38E-03
HP:0000522	Alacrima	FGFR3	1.24E-02
HP:0010662	Abnormality of the diencephalon	LEPR	1.86E-02
HP:0000495	Recurrent corneal erosions	FGFR3	1.49E-02
HP:0001413	Micronodular cirrhosis	KRT8	1.86E-02

(d) Gene ontologies associated with the significantly common dysregulated genes of the LC with WFs.

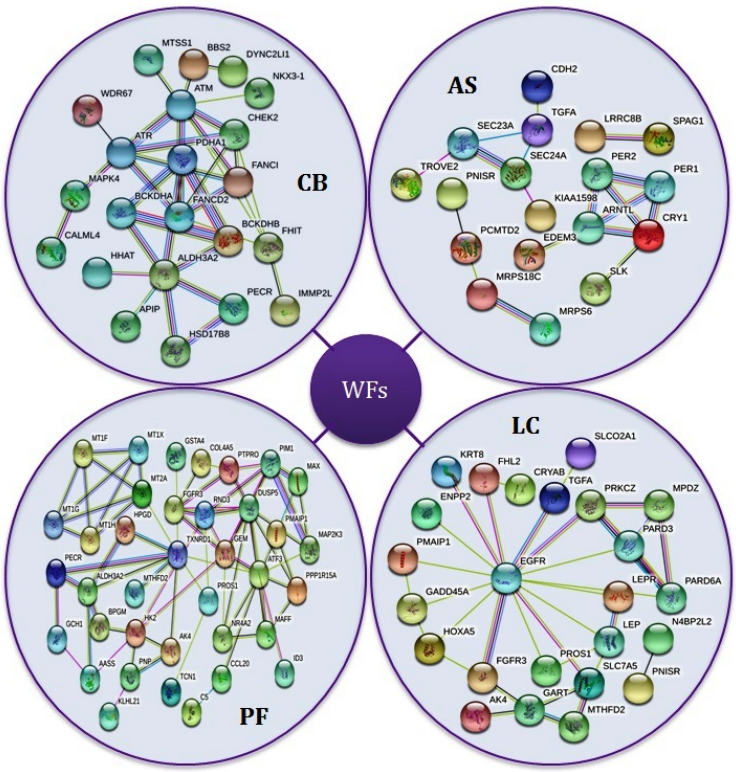


Figure 5: Protein-protein interaction network of RSDs using STRING.

diseases are related to each other by underlying molecular or biological mechanisms. We identified significant pathways common to the dysregulated geneset of each RSD and the WF data.Similarly, gene expression ontologies and protein-protein interaction sub-networks of the commonly differentially expressed genes suggest that WFs may be a major stimulus for progression of several RSDs.

We have used the gold benchmark databases (dbGAP and OMIM) to verify the outcome of our study and observed that there are a number of shared genes between the WFs and RSDs as shown in figure 6. For cross checking the validity of our study, we collected disease and genes names from OMIM Disease, OMIM Expanded and dbGap databases using the observed dysregulated genes seen in WF exposure. We concatenated the list of diseases as well as genes from the mentioned three databases and selected only respiratory system diseases (RSDs) after several steps of statistical analysis [48]. Interestingly, we found our selected four RSDs among the list of collected RSDs from the mentioned databases as shown in figure 6. These genes and their products are strong candidate mediators for the deletarious effects of WFs and lung that predispose welders to develop CB, AS, PE and LC.

5. Conclusions

In this study, we have considered gene expression microarray data from welding fumes (WFs), CB, AS, PE, and LC datasets to analyze and investigate the genetic effects of WFs

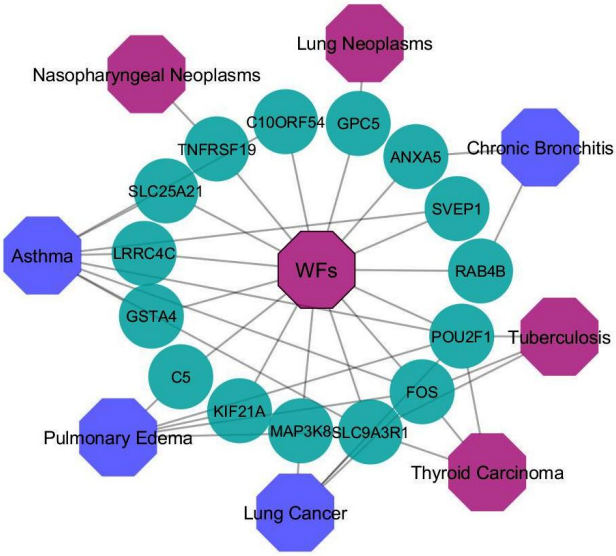


Figure 6: Gene-disease association network of Welding fumes (WFs) with several RSDs. Violet colored octagon-shaped nodes represent different RSDs, blue colored octagon-shaped nodes represent our selected four RSDs and round-shaped dark-cyan colored nodes represent differentially expressed genes for WFs. A link is placed between a disease and a gene if mutations in that gene lead to the specific disease.

on RSDs. We analyzed gene regulation, built gene-disease association networks, identified signaling pathways, identified gene expression ontologies and protein-protein interaction sub-networks of WFs and each RSDs. Our findings showed the nature of the association between WFs and RSDs at the molecular and cellular level. This types of study are likely to be useful for making gene-based evidence based recommendations about accurate disease prediction, identification and therapeutic treatments.

References

[1] J. M. Antonini, Health effects of welding, Critical reviews in toxicology 33 (1) (2003) 61–103.

[2] H. K. Rana, M. R. Akhtar, M. B. Ahmed, P. Lio, J. Quinn, F. Huq, M. A. Moni, Genetic effects of welding fumes on the progression of neurodegenerative diseases, bioRxiv (2018) 480806.

[3] Welding fumes and brain damage — kemper america, `https://kemperamerica.com/welding-fumes-brain-damage/`, (Accessed on 11/27/2018).

[4] Respiratory system: Facts, function and diseases, `https://www.livescience.com/22616-respiratory-system.html`, (Accessed on 11/27/2018).

[5] L. V. Wain, N. Shrine, M. S. Artigas, A. M. Erzurumluoglu, B. Noyvert, L. Bossini-Castillo, M. Obeidat, A. P. Henry, M. A. Portelli, R. J. Hall, et al., Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential druggable targets, Nature genetics 49 (3) (2017) 416.

[6] Hazardous substances in welding fumes: How they affect the human body - safe welding, `https://safe-welding.com/hazardous-substances-in-welding-fumes-how-they-affect-the-human-body/`, (Accessed on 11/27/2018).

[7] I. V. Yang, C. A. Lozupone, D. A. Schwartz, The environment, epigenome, and asthma, Journal of Allergy and Clinical Immunology 140 (1) (2017) 14–23.

[8] F. Cortellaro, E. Ceriani, M. Spinelli, C. Campanella, I. Bossi, D. Coen, G. Casazza, C. Cogliati, Lung ultrasound for monitoring cardiogenic pulmonary edema, Internal and emergency medicine 12 (7) (2017) 1011–1017.

[9] A. F. Gazdar, P. A. Bunn, J. D. Minna, Small-cell lung cancer: what we know, what we need to know and the path forward, Nature Reviews Cancer 17 (12) (2017) 725.

[10] Geo accession viewer, `https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE62384.`, (Accessed on 11/27/2018).

[11] D. Singh, S. M. Fox, R. Tal-Singer, J. Plumb, S. Bates, P. Broad, J. H. Riley, B. Celli, Induced sputum genes associated with spirometric and radiological disease severity in copd ex-smokers, Thorax 66 (6) (2011) 489–495.

[12] J. Bigler, M. Boedigheimer, J. P. Schofield, P. J. Skipp, J. Corfield, A. Rowe, A. R. Sousa, M. Timour, L. Twehues, X. Hu, et al., A severe asthma disease signature from gene expression profiling of peripheral blood from u-biopred cohorts, American journal of respiratory and critical care medicine 195 (10) (2017) 1311–1320.

[13] X. Fang, J. Abbott, L. Cheng, J. K. Colby, J. W. Lee, B. D. Levy, M. A. Matthay, Human mesenchymal stem (stromal) cells promote the resolution of acute lung injury in part through lipoxin a4, The Journal of Immunology (2015) 1500244.

[14] M. T. Landi, T. Dracheva, M. Rotunno, J. D. Figueroa, H. Liu, A. Dasgupta, F. E. Mann, J. Fukuoka, M. Hames, A. W. Bergen, et al., Gene expression signature of cigarette smoking and its role in lung adenocarcinoma development and survival, PloS one 3 (2) (2008) e1651.

[15] M. A. Moni, P. Lio, Genetic profiling and comorbidities of zika infection, The Journal of infectious diseases 216 (6) (2017) 703–712.

[16] M. A. Moni, H. Xu, P. Lio, Cytocom: a cytoscape app to visualize, query and analyse disease comorbidity networks, Bioinformatics 31 (6) (2014) 969–971.

[17] M. A. Moni, P. Liò, comor: a software for disease comorbidity risk assessment, Journal of clinical bioinformatics 4 (1) (2014) 8.

[18] M. A. Moni, P. Liò, How to build personalized multi-omics comorbidity profiles, Frontiers in cell and developmental biology 3 (2015) 28.

[19] M. Kanehisa, M. Furumichi, M. Tanabe, Y. Sato, K. Morishima, Kegg: new perspectives on genomes, pathways, diseases and drugs, Nucleic acids research 45 (D1) (2016) D353–D361.

[20] M. Kutmon, A. Riutta, N. Nunes, K. Hanspers, E. L. Willighagen, A. Bohler, J. Mélius, A. Waagmeester, S. R. Sinha, R. Miller, et al., Wikipathways: capturing the full diversity of pathway knowledge, Nucleic acids research 44 (D1) (2015) D488–D494.

[21] A. Fabregat, K. Sidiropoulos, G. Viteri, O. Forner, P. Marin-Garcia, V. Arnau, P. DEustachio, L. Stein, H. Hermjakob, Reactome pathway analysis: a high-performance in-memory approach, BMC bioinformatics 18 (1) (2017) 142.

[22] D. Nishimura, Biocarta, Biotech Software & Internet Report: The Computer Software Journal for Scient 2 (3) (2001) 117–120.

[23] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, et al., Gene ontology: tool for the unification of biology, Nature genetics 25 (1) (2000) 25.

[24] S. Köhler, N. A. Vasilevsky, M. Engelstad, E. Foster, J. McMurry, S. Aymé, G. Baynam, S. M. Bello, C. F. Boerkoel, K. M. Boycott, et al., The human phenotype ontology in 2017, Nucleic acids research 45 (D1) (2016) D865–D876.

[25] M. A. Moni, P. Liò, Network-based analysis of comorbidities risk during an infection: Sars and hiv case studies, BMC bioinformatics 15 (1) (2014) 333.

[26] M. H. Rahman, S. Peng, C. Chen, M. A. Moni, et al., Genetic effect of type 2 diabetes to the progression of neurological diseases.

[27] L. Jin, X.-Y. Zuo, W.-Y. Su, X.-L. Zhao, M.-Q. Yuan, L.-Z. Han, X. Zhao, Y.-D. Chen, S.-Q. Rao, Pathway-based analysis tools for complex diseases: a review, Genomics, proteomics & bioinformatics 12 (5) (2014) 210–220.

[28] M. V. Kuleshov, M. R. Jones, A. D. Rouillard, N. F. Fernandez, Q. Duan, Z. Wang, S. Koplev, S. L. Jenkins, K. M. Jagodnik, A. Lachmann, et al., Enrichr: a comprehensive gene set enrichment analysis web server 2016 update, Nucleic acids research 44 (W1) (2016) W90–W97.

[29] Integrated breast cancer pathway (homo sapiens) - wikipathways, https://www.wikipathways.org/index.php/Pathway:WP1984, (Accessed on 11/27/2018).

[30] H. Xu, M. A. Moni, P. Liò, Network regularised cox regression and multiplex network models to predict disease comorbidities and survival of cancer, Computational biology and chemistry 59 (2015) 15–31.

[31] Reactome — search results for assembly of the motile cilium_homo sapiens, https://reactome.org/content/query?q=Assembly+of+the+motile+cilium_Homo+sapiens, (Accessed on 11/27/2018).

[32] Reactome — methionine salvage pathway, https://reactome.org/content/detail/R-HSA-1237112., (Accessed on 11/27/2018).

[33] Reactome — glyoxylate metabolism and glycine degradation, https://reactome.org/content/detail/R-HSA-389661., (Accessed on 11/27/2018).

[34] Kegg pathway: map05223, https://www.genome.jp/dbget-bin/www_bget?pathway:map05223, (Accessed on 11/27/2018).

[35] M. A. Hossain, S. M. S. Islam, J. Quinn, F. Huq, M. A. Moni, Identification of ovarian cancer gene expression patterns associated with disease progression and mortality, bioRxiv (2018) 473165.

[36] K. Kandasamy, S. S. Mohan, R. Raju, S. Keerthikumar, G. S. S. Kumar, A. K. Venugopal, D. Telikicherla, D. J. Navarro, S. Mathivanan, C. Pecquet, et al., Netpath: a public resource of curated signal transduction pathways, Genome biology 11 (1) (2010) 1.

[37] P. J. Roach, A. A. Depaoli-Roach, T. D. Hurley, V. S. Tagliabracci, Glycogen and its metabolism: some new developments and old themes, Biochemical Journal 441 (3) (2012) 763–787.

[38] Kegg pathway: Central carbon metabolism in cancer - homo sapiens (human), https://www.genome.jp/kegg-bin/show_pathway?hsa05230, (Accessed on 11/27/2018).

[39] Kegg pathway: Small cell lung cancer - homo sapiens (human), https://www.genome.jp/kegg-bin/show_pathway?hsa05222, (Accessed on 11/27/2018).

[40] One carbon metabolism (homo sapiens) - wikipathways, https://www.wikipathways.org/index.php/Pathway:WP241, (Accessed on 11/27/2018).

[41] G. Dong, H. Chen, M. Qi, Y. Dou, Q. Wang, Balance between metallothionein and metal response element binding transcription factor 1 is mediated by zinc ions, Molecular medicine reports 11 (3) (2015) 1582–1586.

[42] Kegg pathway: Bladder cancer - homo sapiens (human), https://www.genome.jp/kegg-bin/show_pathway?hsa05219, (Accessed on 11/27/2018).

[43] M. V. Dieci, M. Arnedos, F. Andre, J. C. Soria, Fibroblast growth factor receptor inhibitors as a cancer treatment: from a biologic rationale to medical perspectives, Cancer discovery 3 (3) (2013) 264–279.

[44] S. G. Burgess, I. Peset, N. Joseph, T. Cavazza, I. Vernos, M. Pfuhl, F. Gergely, R. Bayliss, Aurora-a-dependent control of tacc3 influences the rate of mitotic spindle assembly, PLoS genetics 11 (7) (2015) e1005345.

[45] B. Ozdamar, R. Bose, M. Barrios-Rodiles, H.-R. Wang, Y. Zhang, J. L. Wrana, Regulation of the polarity protein par6 by tgfß receptors controls epithelial cell plasticity, Science 307 (5715) (2005) 1603–1609.

[46] Gene regulation ontology - summary — ncbo bioportal, https://bioportal.bioontology.org/ontologies/GRO, (Accessed on 11/27/2018).

[47] Network analysis of protein interaction data: an introduction — embl-ebi train online, https://www.ebi.ac.uk/training/online/course/network-analysis-protein-interaction-data-introduction, (Accessed on 11/27/2018).

[48] N. Sakib, U. N. Chowdhury, M. B. Islam, J. M. Quinn, M. A. Moni, A system biology approach to identify the genetic markers to the progression of parkinson's disease for aging, lifestyle and type 2 diabetes, bioRxivdoi:10.1101/482760.