Genetic Effects of Welding Fumes on the Progression of Neurodegenerative Diseases

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

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

Abstract

Background: Welding involves exposure to fumes, gases and radiant energy that can be hazardous to human health. Welding fumes (WFs) comprise a complex mixture of metallic oxides, silicates and fluorides that may result in different health effects. Inhalation of WFs in large quantities over a long periods may pose a risk of developing neurodegenerative diseases (NDGDs), but the nature of this risk is poorly understood. To address this we performed transcriptomic analysis to identify links between WF exposure and NDGDs.

Methods: We developed quantitative frameworks to identify the gene expression relationships of WF exposure and NDGDs. We analyzed gene expression microarray data from fume-exposed tissues and NDGDs including Parkinson's disease (PD), Alzheimer's disease (AD), Lou Gehrig's disease (LGD), Epilepsy disease (ED) and multiple sclerosis disease (MSD) datasets. We constructed disease-gene relationship networks and identified dysregulated pathways, ontological pathways and protein-protein interaction sub-network using multilayer network topology and neighborhood-based benchmarking.

Results: We observed that WF associated genes share 18, 16, 13, 19 and 19 differentially expressed genes with PD, AD, LGD, ED and MSD respectively. Gene expression dysregulation along with relationship networks, pathways and ontologic analysis indicate that WFs may be linked to the progression of these NDGDs.

Conclusions: Our developed network-based approach to analysis and investigate the genetic effects of welding fumes on PD, AD, LGD, ED and MSD neurodegenerative diseases could be helpful to understand the causal influences of WF exposure for the progression of the NDGDs.

Keywords: Welding fumes, Alzheimer's disease, Parkinson's disease, Epilepsy disease, Neurodegenerative diseases.

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

1. Introduction

Welding processes can expose an operator fumes, gases and radiant energy, often in a confined space. Thus, welding fumes (WFs) are recognised as a particular health hazard [1], comprising complex mixtures of metallic oxides, silicates and fluorides as well as heavy metal contaminants such as cadmium, aluminium, chromium, copper and lead [2, 3]. A welder may inhale these fumes in significant quantities over an extended period of time, and an in addition to the risk of developing pulmonary ailments other very significant disease risks are emerging, notably neurodegenerative diseases (NDGDs) [1, 4].

NDGDs are a collective term for a heterogeneous group of disorders that are incurable and characterized by the progressive degeneration of the function and structure of the central nervous system [5]. NDGDs primarily affect the neurons of the central nervous system and progressively damage their function. Neurons are very vulnerable to injury and normally do not divide or replace themselves directly, making damage repair slow [6, 7]. For this reason NDGDs can be devastating and permanent with few options for treatment. To understand how this such damage can occur we used a bioinformatic approach to investigate how welding fume (WF) actions on tissues may influence development of PD, AD, LGD, ED and MSD .

PD is the second-most common neurologic disease that affects neural cells in the brain which produce dopamine in the substantia nigra [8, 9]. There are several symptoms of PD include tremors, muscle rigidity, and changes in gait and speech. Welding fumes contain Manganese that can develop Parkinsons disease [10, 11]. The AD is the most common type of incurable dementia that causes problems with progressive memory loss and other cognitive abilities. Existing medical treatments for AD produce only a modest improvement of symptoms but there is currently no cure [12]. Aluminum exposure to welding is a risk factor to produce AD. LGD also cognizant as Amyotrophic lateral sclerosis (ALS), is a neurodegenerative disease that progressively damages motor neurons and muscle atrophy controlling voluntary muscle movement. The initial symptoms of LGD are muscle weakness or stiffness, can bring death by progressive muscular paralysis and respiratory system failure within 2 to 5 years. US Food and Drug Administration (FDA) approved Riluzole and Edaravone drugs that may prolong LGD survival. Nevertheless, there is no effective cure or prevention for this devastating disease [13, 14]. ED is a heterogeneous group of neurodegenerative disorder that affects neural cells in the brain which are recognized by recurrent seizures or unusual behavior, awareness and sensations suffering over 60 million people in the world. AEDs are Current anti-epileptic drugs that can minimize symptoms but there is no permanent cure or prevention of ED [15]. MSD is a severe neurodegenerative disorder that attacks the neurons of the central nervous system in the spinal cord and brain, on young adults most commonly [16]. The symptoms of MSD include muscle weakness, trouble with sensation and blindness

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

. Medical treatments of MSD can prolong patient survival but there is no permanent cure or prevention strategy for MSD. Manganese exposure to welding is thought to be an important risk factor on the progression of LGD, ED and MSD [17].

Our study employed a systematic and quantitative approach to identify welding fume-responding genes (WFGs) which may indicate a link to development of NDGDs. For these purposes, we studied several NDGDs including PD, AD, LGD, ED and MSD. To understand the effects of WFs on NDGDs, we examined gene expression dysregulation, disease association network, dysregulated pathway, gene expression ontology and protein-protein interaction. We also investigated the validation of our study by using the gold benchmark databases (dbGAP and OMIM).

2. Materials and Methods

2.1. Datasets employed in this study

To investigate the effects of WFs on NDGDs at the molecular level, we used gene expression microarray data. In this study, we used Gene Expression Omnibus from the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/geo/). We analyzed 6 different datasets for our study with accession numbers GSE62384, GSE19587, GSE28146, GSE833, GSE22779 and GSE38010 [18, 19, 20, 21, 22, 23]. The WF dataset (GSE62384) is a result of gene expression analysis 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 PD dataset (GSE19587) is taken from 6 postmortem brains of PD patients and from 5 control brains. The AD dataset (GSE28146) is a microarray data on RNA from fresh frozen hippocampal tissue blocks that contain both white and gray matter, potentially obscuring region-specific changes. The LGD dataset (GSE833) is an Affymetrix Human Full Length HuGeneFL [Hu6800] Array. In this data, postmortem spinal cord grey matter from sporadic and familial ALGD patients are compared with controls. The ED dataset (GSE22779) is a gene expression profiles of 4 non-leukemic individuals (1 healthy and 3 with epilepsy) is generated from the mononuclear cells isolated from the peripheral blood samples before, and after 2, 6, and 24 hours of in-vivo glucocorticoid treatment. The MSD dataset (GSE38010) is a microarray data of multiple sclerosis (MS) patients brain lesions compared with control brain samples.

2.2. Overview of analytical approach

We used systematic and quantitative approach to identify the effect of WFs on the progression of the NDGDs using different sources of available microarray datasets. The graphical representation of this approach is shown in figure 1. This approach included gene expression, signaling pathway, Gene Ontology (GO) and protein-protein interaction analyses. This approach also used Gold benchmark data to verify the validity of our study.

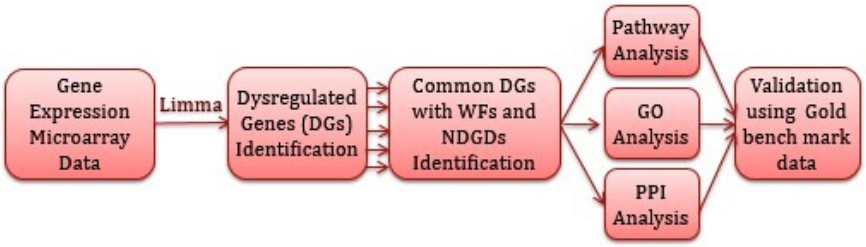

Figure 1: Flow-diagram of the analytical approach used in this study.

2.3. Analysis methods

Gene expression analysis using microarrays is a global and popular method to develop and refine the molecular determinants of human disorders that have proven to be a sensitive method. We used these technologies to analyze the gene expression profiles of Parkinson's disease (PD), Alzheimer's disease (AD), Lou Gehrig's disease (LGD), Epilepsy disease (ED) and Multiple Sclerosis disease (MSD) to find the effects of welding fumes on them [24, 25]. To uniform the mRNA expression data of different platforms and to avoid the problems of experimental systems, we normalized the gene expression data (disease state or control data) by using the Z-score transformation (Z_{ij}) for each NDGD gene expression profile using $Z_{ij} = \frac{g_{ij} - mean(g_i)}{SD(g_i)}$,

where SD implies the standard deviation, g_{ij} represents the value of the gene expression i in sample j. After this transformation we can directly compare of gene expression values of various diseases under different platforms. We applied two conditions for t-test statistic. We performed unpaired T-test to identify differentially expressed genes in patients over control data and selected significant genes. We have chosen a threshold of at least 1 log_2 fold change and a p-value of $<= 1 * 10^{-2}$.

We applied the topological and neighborhood based benchmark methods to find gene-disease associations. Gene-disease network (GDN) was constructed by using the gene-disease associations, where the nods in the network represent either gene or disease [26, 27]. This network can also be characterized as a bipartite graph. The diseases are connected in GDN when they share at least one or more significant differentially expressed genes. These topological and neighborhood based benchmark methods were adopted from our previous studies [28].

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_{ij}^g) associated with both diseases D_i and D_j is as follows [29]:

$$n_{ij}^g = N(G_i \cap G_j) \tag{1}$$

The common neighbours are the based on the Jaccard Coefficient method, where the edge prediction score for the node pair is as [29]:

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

where G is the set of nodes and E is the set of all edges. We used R software packages "comoR" [30] and "POGO" [28] to cross check the genes-diseases associations.

To find molecular pathways of several NDGDs, we have analyzed pathway and gene ontology using Enrichr ($https://amp.pharm.mssm.edu/Enrichr/$), a comprehensive gene set enrichment analysis web-based tool [31]. We used STRING ($https://string-db.org.$) for analyzing protein-protein interactions [32].

3. Results

3.1. Gene Expression Analysis

To investigate the potential effects of WFs on NDGDs, we analyzed the gene expression microarray data from the National Center for Biotechnology Information (NCBI) ($http://www.ncbi.nlm.nih.gov/geo/$). We found that 903 genes were differentially expressed for WF exposure with adjusted $P <= .01$ and $|logFC| >= 1$. Among them, 392 and 511 were up and down regulated respectively. Similarly, our analysis identified the most significant differentially expressed genes for each NDGD after various steps of statistical analysis. We identified differentially expressed genes, 774 (263 up and 511 down) in PD, 565 (291 up and 274 down) in AD, 501 (296 up and 205 down) in LGD, 725 (350 up and down) in ED and 834 (455 up and 388 down) in MSD. The cross-comparative analysis was also performed to find the common differentially expressed genes between WFs and each NDGD. We observed that WFs shares 18, 16, 13, 19 and 19 differentially expressed genes with PD, AD, LGD, ED and MSD respectively. To find the significant associations among these NDGDs with WF exposure, we built two separate disease relationships networks for up and down-regulated genes, centered on the WF-affected genes as shown in figure 2 and 3. Two diseases are associated with each if there exist one or more common genes in between these diseases [27]. Noticeably, 2 significant genes, N4BD2L2 and NAAA are commonly differentially expressed among WF exposure, LGD and WD; one gene DAAM1 is commonly dysregulated among WF exposure, ED and MSD.

3.2. Pathway and Functional Association Analysis

Pathways are the key to know how an organism reacts to perturbations in 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 [33]. We analyzed pathways of the common differentially expressed genes using Enrichr, a comprehensive gene set enrichment analysis web-based tool. Pathways of the commonly dysregulated genes in between WFGs and each NDGD were analyzed using four databases includes KEGG, WikiPathways, Reactome and BioCarta. We combined pathways from four mentioned databases and identified the most significant pathways of each disease after various steps of statistical analysis.

We observed that PD has five significant pathways as shown in table 1. Among these pathways, 'Glutamate Neurotransmitter Release Cycle' is responsible to release the glutamate from the presynaptic neuron and its binding to glutamate receptors on the postsynaptic cell to generate a series of events that lead to the propagation of the synaptic transmission

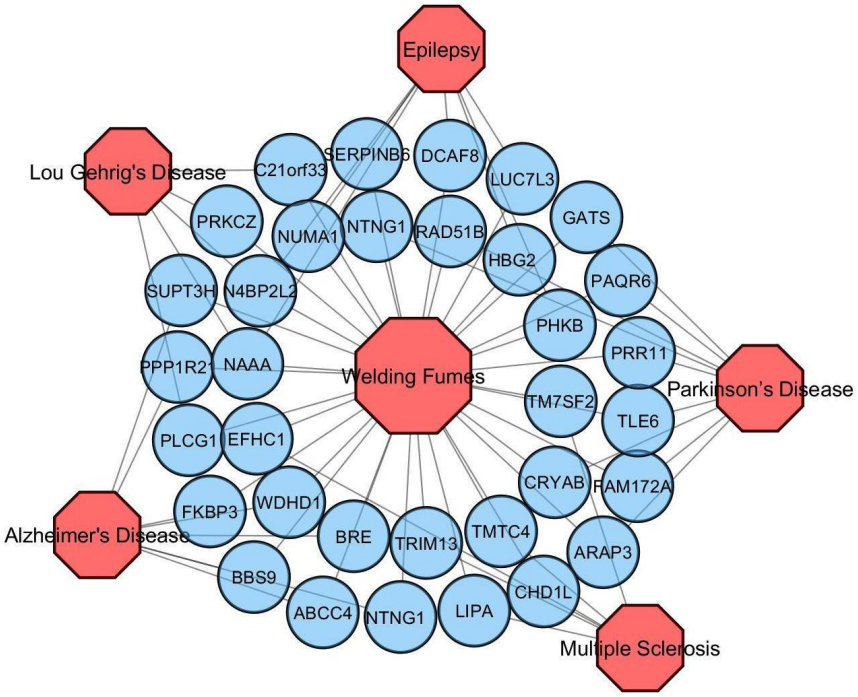

Figure 2: Disease network of welding fume-responding genes (WFGs) with Parkinson's disease (PD), Alzheimer's disease (AD), Lou Gehrig's disease (LGD), Epilepsy disease (ED) and multiple sclerosis disease (MSD). Red colored octagon-shaped nodes represent different categories of disease, and round-shaped sky blue colored nodes represent commonly up-regulated genes among WFGs with the other NDGDs. A link is placed between a disorder and a disease gene if mutations in that gene may lead to (or otherwise has an association with) the specific disorder.

[34]. The pathway 'Sphingolipid de novo biosynthesis' is responsible to provide signals in molecules that regulate various biological functions [35]. The pathway 'Intrinsic Pathway for Apoptosis' is responsible to manage a variety of intracellular stress signal including DNA damage, growth factor withdrawal, unfolding stresses in the endoplasmic reticulum and death receptor stimulation [36]. Kinesins are a super-group of motor proteins based on microtubule that has various functions in the transport of vesicles, organelles, chromosomes, and regulate microtubule dynamics [37]. The pathway 'Neurotransmitter Release Cycle' is responsible to control electrical signals passing through the axons in the form of action potential.

We observed that AD has four significant pathways as shown in table 2. Among these pathways, 'Circadian rhythm pathway' is responsible to feed and influence clocks in other tissues by hormone secretion and nervous stimulation from the brain [38]. Sphingomyelin synthesis appears to be regulated primarily at the level of this transport process through the reversible phosphorylation of CERT (Saito et al. 2008). 'Amyotrophic lateral sclerosis (ALS)' is responsible for most common motor neuron disease [39]. 'MAPKinase Signaling Pathway' is responsible for manage signals of reactions that regulate cell proliferation and apoptosis [40].

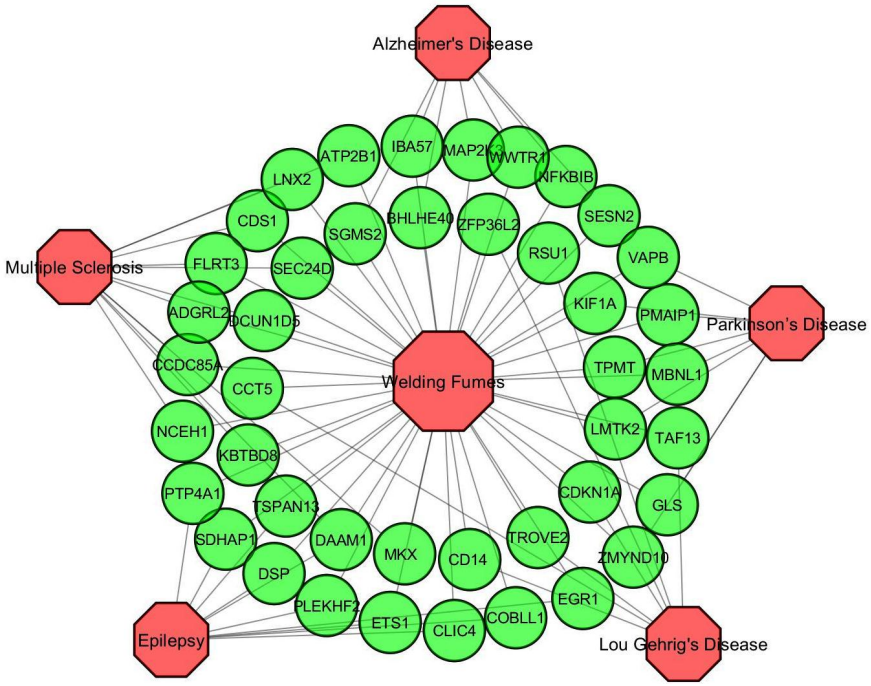

Figure 3: Disease network of welding fume-responding genes (WFGs) with Parkinson's disease (PD), Alzheimer's disease (AD), Lou Gehrig's disease (LGD), Epilepsy disease (ED) and Multiple Sclerosis disease (MSD). Red colored octagon-shaped nodes represent different categories of disease, and round-shaped green colored nodes represent commonly down-regulated genes among WFGs and ND. A link is placed between a disorder and a disease gene if mutations in that gene is linked to the specific disorder.

Table 1: Pathways Associated with Significantly Common Differentially Expressed Genes of the PD with WFs.

Pathway	Genes in the pathway	Adjusted p-value
Glutamate Neurotransmitter Release Cycle	GLS	2.02E-02
Sphingolipid De Novo Biosynthesis	VAPB	2.77E-02
Intrinsic Pathway for Apoptosis	PMAIP1	3.51E-02
Kinesins Pathway	KIF1A	3.68E-02
Neurotransmitter Release Cycle	GLS	4.25E-02

Table 2: Pathways Associated with Significantly Common Differentially Expressed Genes of the AD with WFs.

Pathway	Genes in the pathway	Adjusted p-value
Circadian rhythm pathway	BHLHE40	2.23E-02
Sphingolipid metabolism pathway	SGMS2	3.47E-02
Amyotrophic lateral sclerosis (ALS)	MAP2K3	3.76E-02
MAPKinase Signaling Pathway	MAP2K3	4.12E-02

We observed that LGD has six significant pathways as shown in table 3. Among these

pathways, 'Rap1 signaling pathway' is responsible for controlling a variety of processes, such as cell adhesion, cell polarity and cell-cell junction formation [41]. 'P53 signaling pathway' manages various stress signals, including activated oncogenes, oxidative stress and DNA damage.

Table 3: Pathways Associated with Significantly Common Differentially Expressed Genes of the LGD with WFs.

Pathway	Genes in the pathway	Adjusted p-value
Signaling Pathways in Glioblastoma	CDKN1A, PLCG1, PRKCZ	1.48E-05
MAPK Signaling Pathway	CD14, PRKCZ	4.38E-03
TRIF-mediated programmed cell death	CD14	5.99E-03
EPO Signaling Pathway	PLCG1	6.58E-03
Rap1 signaling pathway	PLCG1, PRKCZ	6.82E-03
P53 signaling pathway	CDKN1A	4.06E-02

We observed that ED has five significant pathways as shown in table 4. Among these pathways, 'Neurotransmitter Release Cycle' is responsible to control electrical signals passing through the axons in the form of action potential. 'Glycogen Metabolismserves' serves as a major stored fuel for several tissues. The keratinocytes function is to form a barrier against environmental damage by fungi pathogenic bacteria, parasites, viruses, and UV radiation.

Table 4: Pathways Associated with Significantly Common Differentially Expressed Genes of the ED with WFs.

Pathway	Genes in the pathway	Adjusted p-value
Ectoderm Differentiation	NUMA1, SERPINB6	6.82E-03
NR3C Signaling	EGR1	1.70E-02
Glycogen Metabolism	PHKB	3.19E-02
Neurotransmitter Release Cycle	NAAA	4.49E-02
Keratinocyte Differentiation	ETS1	4.67E-02

We observed that MSD has five significant pathways as shown in table 5. Among these pathways, 'Endocrine and other factor-regulated calcium reabsorption' is essential for numerous physiological functions including muscle contraction, intracellular signalling processes, neuronal excitability and bone formation [42]. 'Mineral absorption' provides mineral in the neural cell to sustain life. 'Cholesterol biosynthesis' controls cholesterol to the nucleus and activating genes.

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 [43]. GO of the significantly dysregulated genes were analyzed using Enrichr, a

Table 5: Pathways Associated with Significantly Common Differentially Expressed Genes of the MSD with WFs.

Pathway	Genes in the pathway	Adjusted p-value
Steroid biosynthesis	LIPA, TM7SF2	1.44E-04
Metabolism of lipids and lipoproteins	CDS1, NCEH1, SEC24D, TM7SF2	2.47E-03
Cholesterol biosynthesis	TM7SF2	2.05E-02
Endocrine and other factor-regulated calcium reabsorption	ATP2B1	4.15E-02
Mineral absorption	ATP2B1	4.49E-02

comprehensive gene set enrichment analysis web-based tool [31]. GO of the commonly differentially expressed genes (i.e., dysregulated genes common to WFGs and each NDGD) for each NDGD and WFGs were analyzed using two databases of Enrichr including GO Biological Process and Human Phenotype Ontology. We combined ontologies from two mentioned databases and identified the most significant GO term of each disease after various steps of statistical analysis. We observed that 15, 15, 24, 19 and 17 gene ontology classes are associated with the significantly commonly dysregulated (i.e., Dysregulated genes linkning WFGs and each NDGD) genes for WFs with the PD, AD, LGD, Ed and MSD respectively as shown in table 6-10.

Table 6: Gene Ontologies Associated with the Significantly Common Dysregulated Genes of the PD with WFs.

GO Term	Pathway	Genes in the pathway	Adjusted p-value
GO:0001844	Protein insertion into mitochondrial membrane involved in apoptotic signaling pathway	PMAIP1	5.94E-03
GO:1902043	Positive regulation of extrinsic apoptotic signaling pathway via death domain receptors	PMAIP1	5.94E-03
GO:0006987	Activation of signaling protein activity involved in unfolded protein response	VAPB	6.78E-03
GO:0001881	Receptor recycling	LMTK2	8.47E-03
GO:0045837	Negative regulation of membrane potential	PMAIP1	5.94E-03
GO:0043029	T cell homeostasis	PMAIP1	9.31E-03
GO:0032463	Negative regulation of protein homooligomerization	CRYAB	7.63E-03
GO:0032075	Positive regulation of nuclease activity	VAPB	9.31E-03
GO:0016192	Vesicle-mediated transport	LMTK2, ARAP3, KIF1A	4.73E-03
GO:1905906	Regulation of amyloid fibril formation	CRYAB	7.63E-03
HP:0003677	Slow progression	KIF1A, CRYAB	3.08E-03
HP:0007210	Lower limb amyotrophy	KIF1A	8.47E-03
HP:0200073	Respiratory insufficiency due to defective ciliary clearance	ZMYND10	7.63E-03
HP:0003323	Progressive muscle weakness	VAPB	9.31E-03
HP:0003555	Muscle fiber splitting	CRYAB	6.78E-03

3.4. Protein-Protein Interaction Analysis

Protein-protein interaction networks (PPINs) are the mathematical representation of the physical contacts of proteins in the cell. Protein-protein interactions (PPIs) are essential to every molecular and biological process in a cell, so PPIs are crucial to properly understand cell physiology in disease and healthy states [44]. PPIs of the differentially expressed genes were analyzed using STRING, a biological database and web resource of known and predicted protein-protein interactions. We constructed protein-protein interaction network of significantly commonly dysregulated genes (i.e., dysregulated genes common to WFGs and each NDGD) of all NDGDs using STRING. We clustered into five different groups of interactions of five NDGDs as shown in figure 4.

Table 7: Gene Ontologies Associated with the Significantly Common Dysregulated Genes of the AD with WFs.

GO Term	Pathway	Genes in the pathway	Adjusted p-value
GO:0043619	Regulation of transcription from RNA polymerase II promoter in response to oxidative stress	SESN2	6.73E-03
GO:0032055	Negative regulation of translation in response to stress	SESN2	5.24E-03
GO:0006684	Sphingomyelin metabolic process	SGMS2	7.48E-03
GO:0035414	Negative regulation of catenin import into nucleus	WWTR1	7.48E-03
GO:1990253	Cellular response to leucine starvation	SESN2	8.22E-03
GO:0032309	Icosanoid secretion	ABCC4	8.97E-03
GO:0045859	Regulation of protein kinase activity	MAP2K3, WWTR1	5.36E-03
HP:0001156	Brachydactyly syndrome	NTNG1, BBS9	8.87E-03
HP:0002141	Gait imbalance	BBS9	1.34E-02
HP:0007707	Congenital primary aphakia	BBS9	1.34E-02
HP:0010747	Medial flaring of the eyebrow	BBS9	1.56E-02
HP:0009806	Nephrogenic diabetes insipidus	BBS9	1.49E-02
HP:0002370	Poor coordination	BBS9	1.71E-02
HP:0006829	Severe muscular hypotonia	IBA57	1.79E-02
HP:0001827	Genital tract atresia	BBS9	1.93E-02

Table 8: Gene Ontologies Associated with the Significantly Common Dysregulated Genes of the LGD with WFs.

GO Term	Pathway	Genes in the pathway	Adjusted p-value
GO:2000737	Negative regulation of stem cell differentiation	N4BP2L2, ZFP36L2	1.81E-05
GO:0034128	Negative regulation of MyD88-independent toll-like receptor signaling pathway	CD14	4.79E-03
GO:0071364	Cellular response to epidermal growth factor stimulus	PLCG1, ZFP36L2	1.06E-04
GO:1901988	Negative regulation of cell cycle phase transition	ZFP36L2	4.19E-03
GO:1903708	Positive regulation of hemopoiesis	N4BP2L2	4.79E-03
GO:1901991	Negative regulation of mitotic cell cycle phase transition	CDKN1A, ZFP36L2	1.02E-03
GO:0071363	Cellular response to growth factor stimulus	PLCG1, ZFP36L2	3.07E-03
GO:0050821	Protein stabilization	CDKN1A, CCT5	3.65E-03
HP:0001738	Exocrine pancreatic insufficiency	CDKN1A	1.13E-02
HP:0010832	Abnormality of pain sensation	CCT5	1.49E-02
HP:0002717	Adrenal overactivity	CDKN1A	1.79E-02
HP:0003431	Decreased motor nerve conduction velocity	CCT5	1.79E-02
HP:0001258	Spastic paraplegia	CCT5	1.84E-02
HP:0002936	Distal sensory impairment	CCT5	3.66E-02

Table 9: Gene Ontologies Associated with the Significantly Common Dysregulated Genes of the ED with WFs.

GO Term	Pathway	Genes in the pathway	Adjusted p-value
GO:0086069	Bundle of His cell to Purkinje myocyte communication	DSP	8.07E-03
GO:0086073	Bundle of His cell-Purkinje myocyte adhesion involved in cell communication	DSP	6.28E-03
GO:0030575	Nuclear body organization	ETS1	7.18E-03
GO:0003223	Ventricular compact myocardium morphogenesis	DSP	6.28E-03
GO:0002934	Desmosome organization	DSP	8.07E-03
GO:0051639	Actin filament network formation	COBLL1	8.07E-03
GO:1903708	Positive regulation of hemopoiesis	N4BP2L2	7.18E-03
HP:0011902	Abnormal hemoglobin	HBG2	1.25E-02
HP:0001663	Ventricular fibrillation	DSP	8.97E-03
HP:0003445	EMG: neuropathic changes	DCAF8	1.79E-02
HP:0011663	Right ventricular cardiomyopathy	DSP	8.07E-03
HP:0001730	Progressive hearing impairment	SERPINB6	1.70E-02

4. Discussion

We investigated the gene expression relationship of WF exposure and neurodegenerative diseases (NDGDs) based on the associations of genetics, signaling pathways, gene expression ontologies and protein-protein interactions network. For the purpose of our study, we

Table 10: Gene Ontologies Associated with the Significantly Common Dysregulated Genes of the MSD with WFs.

GO Term	Pathway	Genes in the pathway	Adjusted p-value
GO:0021795	Cerebral cortex cell migration	EFHC1	6.28E-03
GO:0048678	Response to axon injury	FLRT3	1.07E-02
GO:1902187	Negative regulation of viral release from host cell	TRIM13	1.25E-02
GO:0014033	Neural crest cell differentiation	KBTBD8	1.43E-02
GO:0006293	Nucleotide-excision repair, preincision complex stabilization	CHD1L	1.96E-02
HP:0004311	Abnormality of macrophages	LIPA	1.70E-02
HP:0001433	Hepatosplenomegaly	LIPA	2.23E-02
HP:0100639	Erectile abnormalities	FLRT3	2.84E-02
HP:0002612	Congenital hepatic fibrosis	LIPA	3.37E-02
HP:0001522	Death in infancy	LIPA	3.98E-02

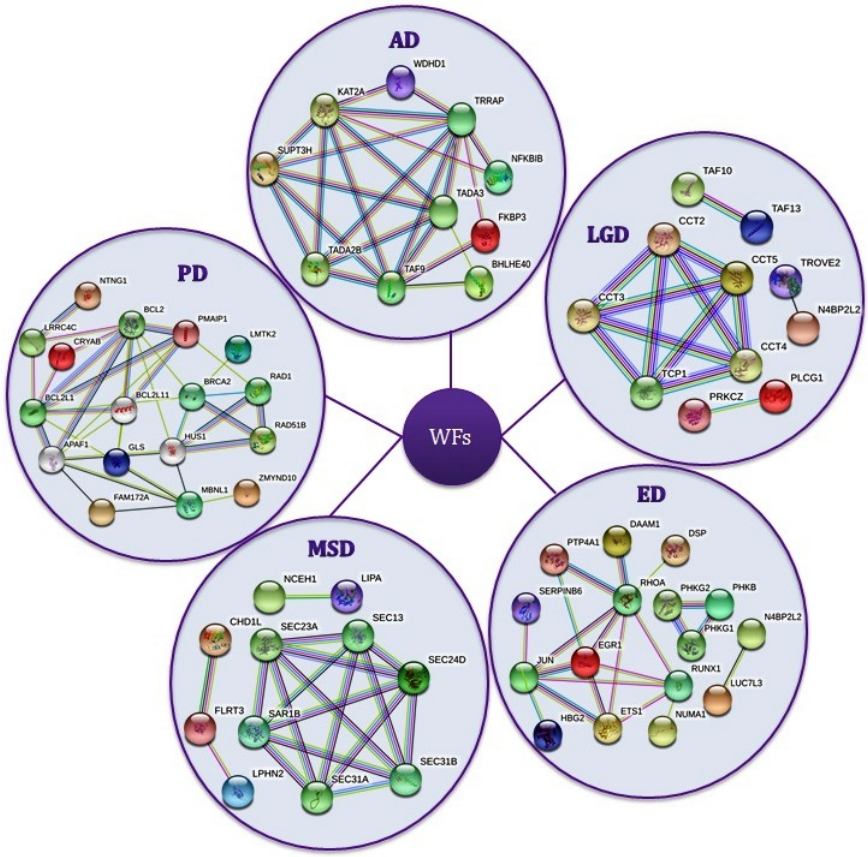

Figure 4: Protein-Protein Interaction Network of the significant genes dysregulated by the NDGDs and WF exposure.

analyzed Gene Expression Omnibus (GEO) microarray data from WFs, Parkinson's disease (PD), Alzheimer's disease (AD), Lou Gehrig's disease (LGD), Epilepsy disease (ED), Multiple Sclerosis disease (MSD) and control datasets [45]. We found a good number of significantly commonly dysregulated genes found among both WFGs and NDGDs by gene expression analysis [46]. As there have a good number of significantly commonly dysregu-

lated genes among WFGs and NDGDs, it indicates that WFGs are likely to have influences on NDGD risk. Our two separate disease relationships networks for up and down-regulated genes strongly indicated that WFGs are linked to NDGDs as shown in Figure 2 and 3. The pathway-based analysis is a new approach to understand how different complex conditions can be related to each other through underlying molecular or biological mechanisms [47, 48]. We identified pathways among dysregulated genes common to WFGs and each NDGD. These identified pathways accorded that WF exposure could have a strong association with NDGDs. Similarly, gene expression ontologies and protein-protein interactions of common differentially expressed genes determine that WF exposure may be a risk factor for several NDGDs that may affect a welder's long term health.

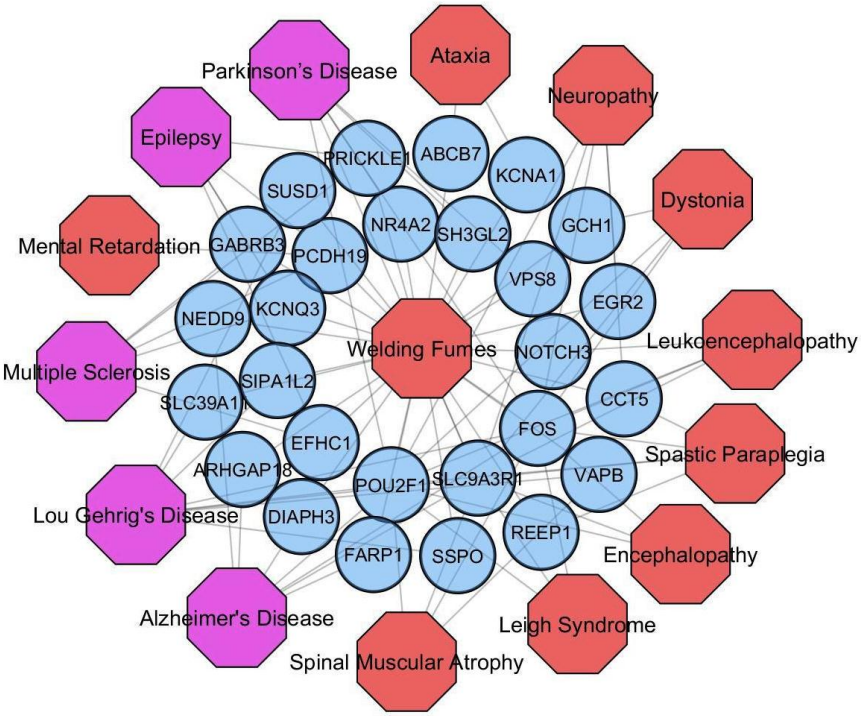

Figure 5: Disease network of WFGs with several NDGDs. Red colored octagon-shaped nodes represent different categories of NDGDs, Violet colored octagon-shaped nodes represent our selected five NDGDs and round-shaped sky blue colored nodes represent differentially expressed genes among WFGs. A link is placed between a disorder and a disease gene if mutations in that gene is known to lead to the specific disorder.

We verified our results using the gold benchmark databases (dbGAP and OMIM) and found that there are some shared genes between the WFGs and NDGDs as shown in figure 5. We collected genes and disease names from OMIM Disease, OMIM Expanded and dbGap databases using differentially expressed genes among WFGs for cross checking the validity of our study. We combined the diseases from three mentioned databases and selected only neurodegenerative diseases (NDGDs) after various steps of statistical analysis. Interestingly, we found our selected five NDGDs among the list of collected NDGDs from the gold-benchmark databases as shown in figure 5. Moreover, we found our identified genes in figure 5 had been

shown in other studies to be associated with disease progression in NDGDs. Specifically, Liu H et al. has shown a link between NR4A2 and PD [49]; Taguchi K et al. found POU2F1 to be linked to AD incidence [50]; A Hggmark et al. found SLC9A3R1 and SLC39A11 are associated with AD [51]; J Wang et al. found SCN1A and FOS is associated with ED [52]; J Wang et al. showed SCN1A and FOS to be linked to ED [52]; and Mahurkar et al. showed GPC5 to be linked to MSD [53]. In summary, we have found that the welding fume associated genes have strong associations with progression of PD, AD, LGD, ED and MSD and these findings are supported by previous work on these neurodegenerative diseases by other researchers.

5. Conclusions

In this study, we have considered GEO microarray data from WFs, PD, AD, LGD, ED, MSD and control datasets to analyze and investigate the gene expression effects of WFs on neurodegenerative diseases (NDGDs). We analyzed dysregulated genes, disease relationship networks, dysregulated pathways, gene expression ontologies and protein-protein interactions of WFs and NDGDs. Our findings showed that WFs have a strong association with genes dysregulated in NDGDs. This kind of study will be useful for making genomic evidence based recommendations about the accurate disease prediction, identification and therapeutic treatments. This study also will be useful indicate the nature of dangerous effects that welding may pose to human health.

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. Islam, P. Lio, F. Huq, M. A. Moni, Genetic effects of welding fumes on the development of respiratory system diseases, bioRxivdoi:10.1101/480855.

[3] M. I. Khan, Welding science and technology, New Age International, 2007.

[4] K. America, Are there links between hazardous welding fumes and brain damage?, `https://kemperamerica.com/welding-fumes-brain-damage/`, accessed: 2018-08-07 (2015).

[5] R. Sharma, S.-Y. Kim, A. Sharma, Z. Zhang, S. P. Kambhampati, S. Kannan, R. M. Kannan, Activated microglia targeting dendrimer–minocycline conjugate as therapeutics for neuroinflammation, Bioconjugate chemistry 28 (11) (2017) 2874–2886.

[6] M. Golpich, E. Amini, Z. Mohamed, R. Azman Ali, N. Mohamed Ibrahim, A. Ahmadiani, Mitochondrial dysfunction and biogenesis in neurodegenerative diseases: pathogenesis and treatment, CNS neuroscience & therapeutics 23 (1) (2017) 5–22.

[7] M. S. Satu, K. C. Howlader, T. M. N. U. Akhund, F. Haq, J. M. Quinn, M. A. Moni, Bioinformatics approach to identify diseasome and comorbidities effect of mitochondrial dysfunctions on the progression of neurological disorders, bioRxivdoi:10.1101/483065.

[8] W. Poewe, K. Seppi, C. M. Tanner, G. M. Halliday, P. Brundin, J. Volkmann, A.-E. Schrag, A. E. Lang, Parkinson disease, Nature reviews Disease primers 3 (2017) 17013.

[9] M. A. Moni, H. K. Rana, M. B. Islam, M. B. Ahmed, P. Lio', M. A. M. Hasan, F. Huq, J. Quinn, Early detection of neurological dysfunction using blood cell transcript profiles, bioRxivdoi:10.1101/483016.

[10] B. A. Racette, S. R. Criswell, J. I. Lundin, A. Hobson, N. Seixas, P. T. Kotzbauer, B. A. Evanoff, J. S. Perlmutter, J. Zhang, L. Sheppard, et al., Increased risk of parkinsonism associated with welding exposure, Neurotoxicology 33 (5) (2012) 1356–1361.

[11] 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.
[12] J. Wang, B. J. Gu, C. L. Masters, Y.-J. Wang, A systemic view of alzheimer diseaseinsights from amyloid-β metabolism beyond the brain, Nature Reviews Neurology 13 (10) (2017) 612.
[13] E. Tokuda, Y. Furukawa, Abnormal protein oligomers for neurodegeneration, Oncotarget 8 (25) (2017) 39943.
[14] J. Sun, Commentary: target intestinal microbiota to alleviate disease progression in amyotrophic lateral sclerosis, Journal of neurology & neuromedicine 2 (6) (2017) 13.
[15] M. Alves, E. Beamer, T. Engel, The metabotropic purinergic p2y receptor family as novel drug target in epilepsy, Frontiers in pharmacology 9 (2018) 193.
[16] K. Sapko, A. Szczepańska-Szerej, A. Jamroz-Wiśniewska, M. Kulczyński, M. Marciniec, K. Rejdak, Progressive forms of multiple sclerosis: disease-modifying therapy review, World Scientific News 105 (2018) 157–167.
[17] O. Proudfoot, et al., Manganese in manganism, parkinson's disease, huntington's disease, amyotrophic lateral sclerosis, and batten disease: A narrative review, Neurology India 65 (6) (2017) 1241.
[18] A. Stanam, Gene expression omnibus, `https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE62384`, accessed: 2018-08-08 (2018).
[19] N. M. Lewandowski, S. Ju, M. Verbitsky, B. Ross, M. L. Geddie, E. Rockenstein, A. Adame, A. Muhammad, J. P. Vonsattel, D. Ringe, et al., Polyamine pathway contributes to the pathogenesis of parkinson disease, Proceedings of the National Academy of Sciences 107 (39) (2010) 16970–16975.
[20] E. M. Blalock, H. M. Buechel, J. Popovic, J. W. Geddes, P. W. Landfield, Microarray analyses of laser-captured hippocampus reveal distinct gray and white matter signatures associated with incipient alzheimer's disease, Journal of chemical neuroanatomy 42 (2) (2011) 118–126.
[21] F. Dangond, D. Hwang, S. Camelo, P. Pasinelli, M. P. Frosch, G. Stephanopoulos, G. Stephanopoulos, R. H. Brown Jr, S. R. Gullans, Molecular signature of late-stage human als revealed by expression profiling of postmortem spinal cord gray matter, Physiological genomics 16 (2) (2004) 229–239.
[22] M. Carlet, K. Janjetovic, J. Rainer, S. Schmidt, R. Panzer-Grümayer, G. Mann, M. Prelog, B. Meister, C. Ploner, R. Kofler, Expression, regulation and function of phosphofructo-kinase/fructose-biphosphatases (pfkfbs) in glucocorticoid-induced apoptosis of acute lymphoblastic leukemia cells, BMC cancer 10 (1) (2010) 638.
[23] M. H. Han, D. H. Lundgren, S. Jaiswal, M. Chao, K. L. Graham, C. S. Garris, R. C. Axtell, P. P. Ho, C. B. Lock, J. I. Woodard, et al., Janus-like opposing roles of cd47 in autoimmune brain inflammation in humans and mice, Journal of Experimental Medicine (2012) jem–20101974.
[24] M. R. Rahman, T. Islam, T. Zaman, M. Shahjaman, M. R. Karim, D. Holsinger, M. A. Moni, Blood-based molecular biomarker signatures in alzheimer's disease: Insights from systems biomedicine perspective, bioRxivdoi:10.1101/481879.
[25] M. H. Rahman, S. Peng, C. Chen, M. A. Moni, et al., Genetic effect of type 2 diabetes to the progression of neurological diseases.
[26] 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.
[27] 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.
[28] M. A. Moni, P. Liò, How to build personalized multi-omics comorbidity profiles, Frontiers in cell and developmental biology 3 (2015) 28.
[29] M. A. Moni, P. Lio, Genetic profiling and comorbidities of zika infection, The Journal of infectious diseases 216 (6) (2017) 703–712.
[30] M. A. Moni, P. Liò, comor: a software for disease comorbidity risk assessment, Journal of clinical bioinformatics 4 (1) (2014) 8.
[31] 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.
[32] D. Szklarczyk, A. Franceschini, S. Wyder, K. Forslund, D. Heller, J. Huerta-Cepas, M. Simonovic, A. Roth, A. Santos, K. P. Tsafou, et al., String v10: protein–protein interaction networks, integrated over the tree of life, Nucleic acids research 43 (D1) (2014) D447–D452.
[33] 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.
[34] Reactome, Neuronal system (homo sapiens), https://reactome.org/content/detail/R-HSA-210500, accessed: 2018-08-14 (2008).
[35] L. Sasset, Y. Zhang, T. M. Dunn, A. Di Lorenzo, Sphingolipid de novo biosynthesis: a rheostat of cardiovascular homeostasis, Trends in Endocrinology & Metabolism 27 (11) (2016) 807–819.
[36] Wikipathways, Intrinsic pathway for apoptosis (homo sapiens), https://www.wikipathways.org/index.php/Pathway:WP1841, accessed: 2018-08-14 (2018).
[37] C. J. Lawrence, R. K. Dawe, K. R. Christie, D. W. Cleveland, S. C. Dawson, S. A. Endow, L. S. Goldstein, H. V. Goodson, N. Hirokawa, J. Howard, et al., A standardized kinesin nomenclature, The Journal of cell biology 167 (1) (2004) 19–22.
[38] Reactome, Circadian rhythm pathway, https://reactome.org/content/query?q=Circadian+rhythm+pathway&species=Homo+sapiens&species=Entries+without+species&cluster=true, accessed: 2018-08-14 (2008).
[39] A. J. Pratt, E. D. Getzoff, J. J. P. Perry, Amyotrophic lateral sclerosis: update and new developments, Degenerative neurological and neuromuscular disease 2012 (2) (2012) 1.
[40] Reactome, Signaling by hippo, https://reactome.org/content/detail/R-HSA-2028269, accessed: 2018-08-14 (2012).
[41] Wikipathways, p53 signaling pathway - homo sapiens (human), https://www.genome.jp/kegg-bin/show_pathway?map=hsa04115&show_description=show, accessed: 2018-08-15 (2018).
[42] S. Boros, R. J. Bindels, J. G. Hoenderop, Active ca2+ reabsorption in the connecting tubule, Pflügers Archiv-European Journal of Physiology 458 (1) (2009) 99–109.
[43] Bioportal, Gene regulation ontology, https://reactome.org/content/detail/R-HSA-2028269, accessed: 2018-08-15 (2016).
[44] Network analysis of protein interaction data: An introduction/ protein-protein interaction networks, https://www.ebi.ac.uk/training/onlkine/course/network-analysis-protein-protein-interaction-data-introduction/, accessed: 2018-08-19 (2017).
[45] U. N. Chowdhury, M. B. Islam, S. Ahmad, M. A. Moni, Network-based identification of genetic factors in ageing, lifestyle and type 2 diabetes that influence the progression of alzheimer's disease, bioRxivdoi:10.1101/482844.
[46] M. R. Rahman, T. Islam, M. Shahjaman, J. M. Quinn, D. Holsinger, M. A. Moni, Common molecular biomarker signatures in blood and brain of alzheimers disease, bioRxivdoi:10.1101/482828.
[47] 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.
[48] M. A. Hossain, T. A. Asa, J. M. W. Quinn, M. M. Rahman, F. Huq, M. A. Moni, Network-based genetic profiling, and therapeutic target identification of thyroid cancer, bioRxivdoi:10.1101/480632.
[49] H. Liu, H. Liu, T. Li, J. Cui, Y. Fu, J. Ren, X. Sun, P. Jiang, S. Yu, C. Li, Nr4a2 genetic variation and parkinson's disease: Evidence from a systematic review and meta-analysis, Neuroscience letters 650 (2017) 25–32.
[50] K. Taguchi, H. D. Yamagata, W. Zhong, K. Kamino, H. Akatsu, R. Hata, T. Yamamoto, K. Kosaka, M. Takeda, I. Kondo, et al., Identification of hippocampus-related candidate genes for alzheimer's disease, Annals of neurology 57 (4) (2005) 585–588.
[51] A. Häggmark, M. Mikus, A. Mohsenchian, M.-G. Hong, B. Forsström, B. Gajewska, A. Barańczyk-Kuźma, M. Uhlén, J. M. Schwenk, M. Kuźma-Kozakiewicz, et al., Plasma profiling reveals three proteins associated to amyotrophic lateral sclerosis, Annals of clinical and translational neurology 1 (8) (2014) 544–553.

[52] J. Wang, Z.-J. Lin, L. Liu, H.-Q. Xu, Y.-W. Shi, Y.-H. Yi, N. He, W.-P. Liao, Epilepsy-associated genes, Seizure 44 (2017) 11–20.

[53] S. Mahurkar, M. Moldovan, V. Suppiah, M. Sorosina, F. Clarelli, G. Liberatore, S. Malhotra, X. Montalban, A. Antigüedad, M. Krupa, et al., Response to interferon-beta treatment in multiple sclerosis patients: a genome-wide association study, The pharmacogenomics journal 17 (4) (2017) 312.