Pak. J. Biochem. Mol. Biol., 2022, 55(4):205-225

Research paper

Identification of dysregulated genes in Parkinson's Disease based on Bioinformatics Analysis

Hira Wasi, Ruba Iqbal, Syeda Hira Seemab*

Department of Biosciences, Mohammad Ali Jinnah University, Karachi, Pakistan.

*Corresponding Author: Syeda Hira Seemab (hira.seemab@jinnah.edu)

ABSTRACT

Parkinson's disease is increasing among elderly people around the world. This disorder has symptoms like akinesia, rigidity, and/or tremor. Parkinson's disease is characterized by dopaminergic neuron degeneration. The purpose of this study is to identify dysregulated pathways and genes included and affected in the human body by extracting the microarray expression dataset from dataset NCBI Gene Expression Omnibus (GEO) will be retrieved then data analysis, functional annotation, and protein interaction network will be done to develop blood-based biomarker by bioinformatics analysis. An entire of 180 Differentially Expressed Genes (DEGs) were detected in the gene expression profile of Parkinson's disease containing 5 control and 5 test samples were studied using t-test (the cut-off is taken $P<0.05$) of accession number **GSE54536** which included whole transcriptome analysis, in which 23 genes consider as up-regulated and 157 genes considered as down-regulated. The outcome of this study provided deeper insights into mechanisms of Parkinson's disease on the molecular level, also cellular and biological levels by pathway enrichment analysis. Recently findings of this study have shown the involvement of DEGs in the Cancer pathway with a p-value of 4.50E-02, which is selected as it is the most significantly enriched pathway having the highest gene count as compared to other principles pathways. The STRING tool was applied to create a protein-protein interaction network, which was then visualized in Cytoscape. A total of 13 hub proteins were identified with minimum of 6 and a maximum of 9 degrees. Finally, a Connectivity Map (CMap) analysis was carried out to identify 56 small compounds capable of altering the expression of genes in Parkinson's disease, which could be used as a prognostic biomarker to assist individuals be diagnosed earlier and undergoing more effective treatment.

KEYWORDS Parkinson's disease, DEGs, Cancer pathway, expression profile, bioinformatics analysis, whole transcriptome analysis.

INTRODUCTION

Parkinson's disorder (PD) is a progressive neurological disorder considered by way of means of a huge range of motor and non- motor functions which can affect one characteristic to a variable degree. This assessment describes the medical traits of PD whit emphasis on the functions that differentiate the disorder from different parkinsonian disorders [1]. The precise categories of Parkinsonism are primary, secondary, and undetermined, which are described together with a five-stage severity grading system. Two-thirds of all patients with primary Parkinsonism develop the disease between the ages of fifty and sixty- nine. [2]. The signs and symptoms of Parkinson's disease are thought to be caused by a decrease in dopamine levels inside the central nervous system. This discount in dopamine has been related to numerous anatomical and functional changes in the brain. Parkinsonism is a complicated network disorder that happens

abnormal activity with inside the basal ganglia impacts the excitability, sensory, and synchronization of the brain's activities' [3].

Pakinsonism's most common form(s), as well as secondary Parkinsonism. Parkinsonism is characterized by akinesia, rigidity, and/or tremor during rest, as well as an uneven onset. Akinesia is characterized by bradykinesia, hypokinesia, and a lack of spontaneous instinctive and voluntary movements (akinesia). It can also appear as a barely covered face (hypomimia), repeating low volume, occasionally scratchy speech, decreased movement of one arm while walking, bending to drag one leg, or a small-stepped gait. Axial akinesia is characterized by the aberrant movement of the spinal and posterior muscles. Patients report having difficulty turning or handing over the bed whether standing or walking, rising from either a low chair or sitting down afterward. [4].

There is currently no medication to stop the neurodegenerative system since the processes behind the development of Parkinson's disease require further inquiry [5]. The purpose of this study is to use gene expression bioinformatics methodologies to determine molecular biomarkers for Parkinson's disease initiation. The Gene Expression Omnibus was used to construct a PD gene expression profile, which found differentially co-expressed genes (DCGs) and dysfunctional pathways in PD patients versus controls. The goal of this study was to collect data that may be used to investigate possible diagnostic biomarkers in Parkinson's disease.

MATERIALS AND METHODS

Data Collection of Microarray Expression Data

The keyword "PD" was used to search for the expression profile of "Parkinson's Disease" in the NCBI Gene Expression Omnibus (GEO) database, which has a vast set of microarray expression data. For this research, we use the gene expression profile GSE54536. All 10 PD samples available were divided into samples (5 control, 5 test) Healthy samples and PD disease samples. PD samples were then analyzed based on GPL10558 [Illumina Human HT-12 V4.0 expression bead chip] [8].

Identification of Differentially Expressed Genes (DEGs)

The log2 transformation was used to preprocess the raw expression datasets in R language. [9]. The "limma" linear model package was used to examine the microarray expression datasets. Differentially expressed genes were isolated from Parkinson's disease patients and healthy persons. For multiple testing modifications, the Benjamini and Hochberg approach (false Discover rate FDR) was used. The DEGs criterion was set at FDR less than 0.05 [10].

Gene Ontology (GO) Analysis of DEGs

DAVID was used to investigate DEG functions and classify genes based on cellular components, biological processes, and molecular functions [11].

Pathway Enrichment Analysis

The dysregulated biochemical pathways in Parkinson's disease were investigated to determine the functional alteration. DAVID inputs for Pathway enrichment analysis were metabolic and non-metabolic pathways from the Kyoto Encyclopedia of Genes and Genomes database [12].

Protein-Protein Network Interaction

The interaction between the up - regulatory and down-regulated DEGs can be studied using STRING TOOL, a database of known protein interactions, which was used to determine the PPI interaction among DEGs, and the PPI interaction network was visualized using Cytoscape tool [13,14].

Identification of Small Molecules

The Connectivity Map (CMap) is a database to assemble the entire transcriptome profile from cultured human cells and applies simple pattern- matching algorithms; both facilitate the detection of functional connections between genes, diseases, and drugs along with the transient feature of changes in common gene expression [15]. Using CMap, 204 DEGs were characterized into upregulated and down- regulated genes. Furthermore, the topmost ten up and down-regulated genes were chosen for enrichment analysis by GSEA (gene set enrichment analysis). The enrichment value lies between -1 to 1 (If the value near -1, the small molecule will mimic the normal status of cell and could be a good drug, whereas a value near to 1, small molecule will mimic the disease status) [16].

RESULTS

Differentiating DEGs into up-regulated gene and down-regulated genes

The microarray expression datasets were analyzed using the "limma" linear model package of "R-language". Genes that were differentially expressed between Parkinson's patients and healthy people were discovered. There were 180 DEGs in total, which were divided into Up-regulated and Down- regulated genes. 23 genes were extracted as up-regulated genes and 157 genes were extracted as down-regulated genes. In Tables 1 & 2, the false discovery rate value (as reported by Hosch and Benjamini) below 0.05 i.e. Adj. P-value was set as the DEGs threshold. In the microarray expression datasets that were evaluated to distinguish between genes that are up and down- regulated, the log-FC value was considered. Positive values were selected to represent up-regulated genes, while negative values represent down-regulated genes.

Diagrammatic Representation of Differentially Expressed Genes:

In contrast to visualizing the expression of Up and Down-regulated genes using the R Language's UMAP library, three main graphical plots including the Volcano plot, the Mean-difference plot and Box-plot were generated.

Volcano Plot

In figure1, the level of significance of each represented by its –log10 (p-values), and the variation in each gene's level of expression is shown by its log2 (fold change). In a volcano plot, the genes that were most up- regulated displayed on the right (shown by red) and and the most down-regulated genes were on the left (shown by blue). The genes that are most statistically significant were at the top and the remaining genes are non- significant because they are neither up- regulated nor down-regulated (shown with grey).

Mean-Difference Plot

In figure 2, the log2 fold-change depicts the mean expression level for each gene. Each of the dark grey dots, representing a different gene shows that there were no significant differences between the control and test groups. Points (dots) that were statistically significant with a false

discovery rate value less than 0.05 depict by blue and red dots, which indicate down and up-regulated genes, respectively.

In figure 3, the samples display comparative results between two groups that were "control vs test". The assigned samples were distributed with different colors as green ones represent the control group whereas the blue ones represent test samples according to groups to know whether the chosen samples were appropriate for differential expression study. The median-centered values were representative of normalized and cross-comparable data for the **"GSE54536"** gene expression value distribution. In the following box plot, one patient sample was depicted by each box.

Gene ontology (GO) analysis OF DEGs in Parkinson's disease

The DEGs were characterized according to biological processes, cellular component and molecular function. The gene expression profile of Parkinson's disease containing 5 normal control and 5 diseased samples were studied using classical t-test in limma package (the cut-off is taken $P<0.05$). A total of 180 differentially expressed genes were identified in Parkinson's disease. There were 176 David IDs of input DEGs present in the record of the specie "Homo sapiens" out of 180.

DEGs Clustering Based on Biological Process

Table 03 of gene ontology analysis explains DEGs that were characterized according to biological processes. There is 176 David IDs in the record of the specie "Homo sapiens" out of 180 (84 genes from the list are not in the output). The threshold count is 02. The gene count was higher in 10 biological processes as compared to the other processes. The biological processes in which more gene count were observed include: Positive Regulation of Transcription from RNA Polymerase II Promoter (23), Regulation of Catalytic Activity (10), Positive Regulation of Transcription, DNA- Templated (12), Vesicle-Mediated Transport (8), Phosphorylation (5), Autophagy (5), Peptidyl-Serine Phosphorylation (5), Response to Hypoxia (5), Positive Regulation of I-Kappa b Kinase/Nf-Kappab Signaling (5), Wnt Signaling Pathway (5).

DEGs Clustering Based on Cellular Component

Table 4 of gene ontology analysis explicates DEGs that are characterized according to cellular component.in the record for the Homo sapiens species, there were 176 DAVID IDs. The threshold was set to the count of 02. Ten of the cellular components had a significantly greater gene count than the other cellular components. The cellular component with the highest gene count include: Cytosol (75), Nucleus (73), Cytoplasm (62), Nucleoplasm (50), Membrane (38), Mitochondrion(18), Endoplasmic Reticulum (15), Golgi Membrane(12), Microtubule(10), Trans- Golgi Network(8).

DEGs Clustering Based on Molecular Function

Table 5 of gene ontology analysis explains DEGs that were characterized according to Molecular functions. There was 176 David IDs in the record of the specie Homo sapiens. The threshold was set on the count of 02. The gene count was much higher in

10 molecular functions as compared to remaining molecular functions. The molecular functions having highest gene count includes: Protein Binding (141), Identical Protein Binding (25), ATP-Binding (24), RNA-Binding (22), Zinc-Ion Binding (13), Protein Kinase Binding(10), Transcription Factor Activity, Sequence-Specific DNA Binding(10), GTPase Activator Activity(9), RNA Polymerase II Sequence-Specific DNA Binding Transcription Factor Binding(6), Transcription Regulatory Region Sequence- Specific DNA Binding(6).

Pathway enrichment analysis of DEGs in Parkinson's disease

A total of 13 pathways were found on the basis of selected criteria. In Table 5, based on the threshold P-value (< 0.05) and gene count was set on > 5, the principle pathways among the included enriched pathways were chosen as demonstrated in Table 3. These principles pathways include: Salmonella infection with a p-value= 4.00E-03), Endocytosis with a P- value= 1.50E-02), Influenza A with a P-value=3.10E-02), and Pathways in cancer with a P-value=4.50E-02).

DEGs involved in Cancer pathway enrichment analysis (KEGG pathway analysis)

In table 6, the involvement of DEGs in **Cancer Pathways** with a P-value = 4.50E-02 (Demonstrated in **Table 5** selected as it is the most significant enriched pathway having highest gene count i.e., 11) as compared to other principle pathways. There were 11 Cancer genes involved in pathway enrichment analysis namely C-X-C Motif Chemokine Receptor 4(CXCR4), Fos Proto-Oncogene, AP-1 Transcription Factor Subunit(FOS), Ras association , Domin family member 1(RASSF1),Rho Guanine Nucleotide Exchange Factor 1(ARHGEF1),frizzled Class Receptor 2(FZD2), Interleukin 15(IL15), Junction Plakoglobin (JUP), Phosphatidylinositol-4,5-Bisphosphate 3- Kinase Catalytic Subunit Delta(PIK3CD), retinoid X Receptor Alpha(RXRA), Signal Transducer and Activator of Transcription 2(STAT2) & Transcription Factor 7 Like 2(TCF7L2).

Protein-Protein Network Analysis of 180 DEGs:

To assess the relationship among DEGs from the obtained datasets, the PPI (Protein-Protein Interaction) network of the Differentially Expressed Genes in Parkinson's disease was constructed using the online interactive search engine STRING (Search Tool for the Retrieval of Interacting Genes Database). As every biological activity involves protein-protein interaction (PPI), which is crucial for understanding how cell behave in an organism. The PPI network built through combining genomic and proteomic data to enable identification of genes and proteins linked to various diseases (Tables 7 and 8).

In figures 4 and 5, the obtained protein-protein interaction (PPI) network of DEGs built via STRING (Search Tool for the Retrieval of Interacting Genes Database) which was then further visualized by Cytoscape. By adjusting definite parameters, the network was examined by Molecular Complex Detection (MCODE), a Cytoscape plug-in to detect the intersecting clusters from the resulting PPI network. The expression levels and prognostic values of hub genes were subsequently analyzed using different plug-ins including CytoNCA (Network Centrality Analysis) and Metscape (Visualization and Interpretation of Metabolomics Data).

Identification of Small Molecules

To identify the related small active molecules against Parkinson's disease, we carried out bioinformatics analysis of DEGs using the CMap database called Touchstone which would enable users to query these data for annotated compounds. The data were analyzed according to Pertubagen Class (PCL) which refers to CMap designated grouping of compounds and identified as first-group compounds sharing the same mechanism of action as per studies from different literature.

The query was 180 DEGs, which was entered as a gene set pair of Up-regulated and down- regulated mutually. The results showed that the Pertubagen compounds from the L-1000 dataset (consisting of about 1 million gene-expression profiles on the responses of about 50 human cell lines to one of approximately 20,000 compounds with various targets as well as the mechanism of action. The CMap connectivity score (Tau) which was a standard enrichment score, ranged between a positive score of +100 and a negative score of -100. The positive score indicates similarity between the query gene set and the Pertubagen class compound while the negative score indicates the opposing effect between the query gene set and the Pertubagen class compounds.

Up and Down-regulated genes among 180 DEGs involved in Connectivity Map Analysis:

The present analysis identified various small molecules among which a total of 56 DEGs were found as Pharmacological class compounds (PCLs) out of 180 DEGs. The identified compounds were sorted from high to low ranked and enrichment scores having a different mechanism of action and various targets. There were 9 high ranked compounds having enrichment scores of 90 or greater including, Rosiglitazone (99.89), Doconexent (99.26), Ambroxol (99.01), SB-203580 (98.93), Staurosporine (98.77), GW-9508 (97.13), XL- 147 (96.03), Alitretinoin (91.42) and LY- 294002 (90.99) show the strongest connectivity to query genes as compared to other compounds. It was concluded that the Pertubagen class compounds having high enrichment scores will be a potential therapeutic target for Parkinson’s disease

Table 1: Differentiating DEGs into Up- regulated genes.

ID	Adj.P-Value	Gene symbol	Gene title	Gene ID
ILMN_1719905	0.0285	TLR10	Toll-like receptor 10	81793
ILMN_1669497	0.0285	OSBPL10	oxysterol binding protein like 10	114884
ILMN_2401714	0.0349	MS4A1	membrane spanning 4-domains A1	931
ILMN_1691071	0.0349	FCRLA	Fc receptor like A	84824
ILMN_2409395	0.0349	CCNC	cyclin C	892
ILMN_1658743	0.0349	CCNDBP1	cyclin D1 binding protein 1	23582
ILMN_1786532	0.037	CNIH1	cornichon family AMPA receptor auxiliary protein 1	10175
ILMN_2411076	0.0398	MATR3	matrin 3	9782
ILMN_1661646	0.0398	BANK1	B-cell scaffold protein with ankyrin repeats 1	55024
ILMN_2357770	0.0398	TCEA1	transcription elongation factor A1	6917
ILMN_2060413	0.0406	CD24	CD24 molecule	1E+08
ILMN_2075189	0.0406	SLC35F2	solute carrier family 35 member F2	54733
ILMN_2168217	0.0418	GPR183	G protein-coupled receptor 183	1880
ILMN_1801584	0.0461	CXCR4	C-X-C motif chemokine receptor 4	7852
ILMN_2325574	0.0461	CASC4	cancer susceptibility candidate 4	113201
ILMN_2052598	0.0461	ARMC10	armadillo repeat containing 10	83787
ILMN_2271894	0.0461	ZNF654	zinc finger protein 654	55279
ILMN_1671905	0.0461	SFR1	SWI5 dependent homologous recombination repair protein 1	119392
ILMN_1745075	0.0461	RPLP0	ribosomal protein lateral stalk subunit P0	6175
ILMN_1684445	0.0461	FCRL5	Fc receptor like 5	83416
ILMN_1776939	0.0468	MS4A1	membrane spanning 4-domains A1	931
ILMN_1675117	0.0468	HSD17B11	hydroxysteroid 17-beta dehydrogenase 11	51170
ILMN_2395974	0.0468	PRDX3	peroxiredoxin 3	10935

Table 2: Differentiating DEGs into Down- regulated genes

ID	Adj.P-Value	Gene symbol	Gene title	Gene ID
ILMN_1669523	0.0285	FOS	Fos proto-oncogene, AP-1 transcription factor subunit	2353
ILMN_1745034	0.0285	SLC11A2	solute carrier family 11 member 2	4891
ILMN_3245066	0.0285	DENND4B	DENN domain containing 4B	9909
ILMN_1672356	0.0285	ANKRD13D	ankyrin repeat domain 13D	338692
ILMN_2065606	0.0285	TOMM40L	translocase of outer mitochondrial membrane 40 like	84134
ILMN_1721575	0.0285	VPS18	VPS18, CORVET/HOPS core subunit	57617
ILMN_1762899	0.0285	EGR1	early growth response 1	1958
ILMN_1751607	0.0285	FOSB	FosB proto-oncogene, AP-1 transcription factor subunit	2354
ILMN_1682792	0.0285	BYSL	bystin like	705
ILMN_1655987	0.0285	STAB1	stabilin 1	23166
ILMN_1654612	0.0285	ZNF589	zinc finger protein 589	51385
ILMN_1718558	0.0317	PARP12	poly(ADP-ribose) polymerase family member 12	64761
ILMN_2396020	0.0317	DUSP6	dual specificity phosphatase 6	1848
ILMN_1690566	0.0317	RASSF4	Ras association domain family member 4	83937
ILMN_1651776	0.0317	FHOD1	formin homology 2 domain containing 1	29109
ILMN_1657771	0.0317	CRTC2	CREB regulated transcription coactivator 2	200186
ILMN_1668639	0.0317	TBC1D10B	TBC1 domain family member 10B	26000
ILMN_1779857	0.0317	KLF4	Kruppel like factor 4	9314
ILMN_3239225	0.0317	RNY3	RNA, Ro-associated Y3	6085
ILMN_1757106	0.0339	6-Mar	membrane associated ring-CH-type finger 6	10299
ILMN_1693630	0.0349	VPS9D1	VPS9 domain containing 1	9605
ILMN_1799467	0.0349	SAMD9L	sterile alpha motif domain containing 9 like	219285
ILMN_3235096	0.0349	SNORA28	small nucleolar RNA, H/ACA box 28	677811
ILMN_2216838	0.0349	DKFZP586I1420	uncharacterized protein DKFZp586I1420	222161
ILMN_2090802	0.0349	TMEM79	transmembrane protein 79	84283
ILMN_1733869	0.0349	OGDH	oxoglutarate dehydrogenase	4967
ILMN_1775074	0.0349	TUBGCP2	tubulin gamma complex associated protein 2	10844
ILMN_1795949	0.0349	CORO7	coronin 7	79585
ILMN_1802380	0.0349	RERE	arginine-glutamic acid dipeptide repeats	473
ILMN_1701621	0.0349	SCO2	SCO2 cytochrome c oxidase assembly protein	9997
ILMN_1669645	0.0349	PKD1	polycystin 1, transient receptor potential channel interacting	5310
ILMN_1733811	0.0349	JUP	junction plakoglobin	3728
ILMN_1805979	0.0349	GAB3	GRB2 associated binding protein 3	139716
ILMN_2410975	0.0349	ATG9A	autophagy related 9A	79065
ILMN_1683462	0.0349	GSS	glutathione synthetase	2937
ILMN_1778845	0.0349	FAM111A	family with sequence similarity 111 member A	63901
ILMN_1723895	0.0349	GTF3C5	general transcription factor IIIC subunit 5	9328
ILMN_1704253	0.0349	C6orf106	chromosome 6 open reading frame 106	64771
ILMN_1753712	0.0349	STX10	syntaxin 10	8677
ILMN_1664012	0.0349	CANT1	calcium activated nucleotidase 1	124583
ILMN_1675709	0.0349	ARFGAP1	ADP ribosylation factor GTPase activating protein 1	55738
ILMN_2052163	0.0349	YIPF1	Yip1 domain family member 1	54432
ILMN_1793386	0.0349	MED12	mediator complex subunit 12	9968
ILMN_1767253	0.0349	RRP12	ribosomal RNA processing 12 homolog	23223

ILMN_1763828	0.0349	MTF1	metal regulatory transcription factor 1	4520
ILMN_1780302	0.0366	DYNC1H1	dynein cytoplasmic 1 heavy chain 1	1778
ILMN_1725700	0.037	MOV10	Mov10 RISC complex RNA helicase	4343
ILMN_1689968	0.0392	PLEKHO2	pleckstrin homology domain containing O2	80301
ILMN_1669607	0.0398	PHKG2	phosphorylase kinase catalytic subunit gamma 2	5261
ILMN_1690921	0.0398	STAT2	signal transducer and activator of transcription 2	6773
ILMN_1727709	0.0398	GPBAR1	G protein-coupled bile acid receptor 1	151306
ILMN_1880280	0.0461	FOXP1-IT1	FOXP1 intronic transcript 1	1.01E+08
ILMN_1702487	0.0461	SGK1	serum/glucocorticoid regulated kinase 1	6446
ILMN_1715968	0.0461	KMT2B	lysine methyltransferase 2B	9757
ILMN_1803560	0.0462	LAT2	linker for activation of T-cells family member 2	7462
ILMN_1726636	0.0463	HECTD2	HECT domain E3 ubiquitin protein ligase 2	143279
ILMN_1763036	0.0468	CLCN6	chloride voltage-gated channel 6	1185
ILMN_1657495	0.0468	MLEC	malectin	9761
ILMN_1753498	0.0468	COASY	Coenzyme A synthase	80347
ILMN_2095840	0.0468	KAT6A	lysine acetyltransferase 6A	7994
ILMN_2320964	0.0468	ADAR	adenosine deaminase, RNA specific	103
ILMN_1775579	0.0468	ACAD9	acyl-CoA dehydrogenase family member 9	28976
ILMN_1672486	0.0468	TCF7L2	transcription factor 7 like 2	6934
ILMN_2350421	0.0468	CACTIN	cactin, spliceosome C complex subunit	58509
ILMN_1697554	0.0468	SASH3	SAM and SH3 domain containing 3	54440
ILMN_1656868	0.0468	NPIPB3	nuclear pore complex interacting protein family member B3	23117
ILMN_1703866	0.0468	SUPT5H	SPT5 homolog, DSIF elongation factor subunit	6829
ILMN_1724181	0.0468	IL15	interleukin 15	3600
ILMN_1752197	0.0468	PDDC1	Parkinson disease 7 domain containing 1	347862
ILMN_1799381	0.0468	SNORD14A	small nucleolar RNA, C/D box 14A	26822
ILMN_1710078	0.047	TMEM181	transmembrane protein 181	57583
ILMN_1687315	0.047	RXRA	retinoid X receptor alpha	6256
ILMN_1812557	0.047	CDK5RAP3	CDK5 regulatory subunit associated protein 3	80279
ILMN_1665212	0.0477	EDC4	enhancer of mRNA decapping 4	23644
ILMN_1811955	0.0477	PRMT5	protein arginine methyltransferase 5	10419
ILMN_1675612	0.0479	BLCAP	bladder cancer associated protein	10904
ILMN_1765649	0.0482	IRF3	interferon regulatory factor 3	3661
ILMN_1815205	0.0488	LYZ	lysozyme	4069
ILMN_1795918	0.0488	AGAP3	ArfGAP with GTPase domain, ankyrin repeat and PH domain 3	116988
ILMN_1772370	0.0488	ARHGEF1	Rho guanine nucleotide exchange factor 1	9138
ILMN_1651788	0.0488	MAP3K11	mitogen-activated protein kinase kinase kinase 11	4296
ILMN_1673069	0.0488	DPP9	dipeptidyl peptidase 9	91039
ILMN_1708946	0.0488	VPS4A	vacuolar protein sorting 4 homolog A	27183
ILMN_1758963	0.049	NADK	NAD kinase	65220
ILMN_2393573	0.0491	CSRNP1	cysteine and serine-rich nuclear protein 1	64651
ILMN_1703123				
ILMN_1705602	0.0491	KLHL17	kelch like family member 17	339451
ILMN_2317751	0.0492	REC8	REC8 meiotic recombination protein	9985
ILMN_1672662	0.0495	SLC20A1	solute carrier family 20-member 1	6574

ILMN_1692539	0.0443	SH3BP1	SH3 domain binding protein 1	23616
ILMN_1758831	0.0449	RNF31	ring finger protein 31	55072
ILMN_2126344	0.0449	SEC16A	SEC16 homolog A, endoplasmic reticulum export factor	9919
ILMN_3238196	0.0456	CYTH4	cytohesin 4	27128
ILMN_1811104	0.0461	POGLUT1	protein O-glucosyltransferase 1	56983
ILMN_1746012	0.0461	MBD6	methyl-CpG binding domain protein 6	114785
ILMN_1666122	0.0461	HEG1	heart development protein with EGF like domains 1	57493
ILMN_1656676	0.0461	ZYG11B	zyg-11 family member B, cell cycle regulator	79699
ILMN_1795400	0.0461	TBCD	tubulin folding cofactor D	6904
ILMN_2196479	0.0461	XRN2	5'-3' exoribonuclease 2	22803
ILMN_2140700	0.0461	CRIPAK	cysteine rich PAK1 inhibitor	285464
ILMN_2278729	0.0461	EIF3B	eukaryotic translation initiation factor 3 subunit B	8662
ILMN_1743205	0.0461	ABCA7	ATP binding cassette subfamily A member 7	10347
ILMN_1696021	0.0461	KPNA6	karyopherin subunit alpha 6	23633
ILMN_1699603	0.0461	MRPL12	mitochondrial ribosomal protein L12	6182
ILMN_1880280	0.0461	FOXP1-IT1	FOXP1 intronic transcript 1	1.01E+08
ILMN_1702487	0.0461	SGK1	serum/glucocorticoid regulated kinase 1	6446
ILMN_1715968	0.0461	KMT2B	lysine methyltransferase 2B	9757
ILMN_1803560	0.0462	LAT2	linker for activation of T-cells family member 2	7462
ILMN_1726636	0.0463	HECTD2	HECT domain E3 ubiquitin protein ligase 2	143279
ILMN_1763036	0.0468	CLCN6	chloride voltage-gated channel 6	1185
ILMN_1657495	0.0468	MLEC	malectin	9761
ILMN_1753498	0.0468	COASY	Coenzyme A synthase	80347
ILMN_2095840	0.0468	KAT6A	lysine acetyltransferase 6A	7994
ILMN_2320964	0.0468	ADAR	adenosine deaminase, RNA specific	103
ILMN_1775579	0.0468	ACAD9	acyl-CoA dehydrogenase family member 9	28976
ILMN_1672486	0.0468	TCF7L2	transcription factor 7 like 2	6934
ILMN_2350421	0.0468	CACTIN	cactin, spliceosome C complex subunit	58509
ILMN_1697554	0.0468	SASH3	SAM and SH3 domain containing 3	54440
ILMN_1656868	0.0468	NPIPB3	nuclear pore complex interacting protein family member B3	23117
ILMN_1703866	0.0468	SUPT5H	SPT5 homolog, DSIF elongation factor subunit	6829
ILMN_1724181	0.0468	IL15	interleukin 15	3600
ILMN_1752197	0.0468	PDDC1	Parkinson disease 7 domain containing 1	347862
ILMN_1799381	0.0468	SNORD14A	small nucleolar RNA, C/D box 14A	26822
ILMN_1710078	0.047	TMEM181	transmembrane protein 181	57583
ILMN_1687315	0.047	RXRA	retinoid X receptor alpha	6256
ILMN_1812557	0.047	CDK5RAP3	CDK5 regulatory subunit associated protein 3	80279
ILMN_1665212	0.0477	EDC4	enhancer of mRNA decapping 4	23644
ILMN_1811955	0.0477	PRMT5	protein arginine methyltransferase 5	10419
ILMN_1675612	0.0479	BLCAP	bladder cancer associated protein	10904
ILMN_1765649	0.0482	IRF3	interferon regulatory factor 3	3661
ILMN_1815205	0.0488	LYZ	lysozyme	4069
ILMN_1795918	0.0488	AGAP3	ArfGAP with GTPase domain, ankyrin repeat and PH domain 3	116988
ILMN_1772370	0.0488	ARHGEF1	Rho guanine nucleotide exchange factor 1	9138
ILMN_1651788	0.0488	MAP3K11	mitogen-activated protein kinase kinase kinase 11	4296
ILMN_1673069	0.0488	DPP9	dipeptidyl peptidase 9	91039
ILMN_1708946	0.0488	VPS4A	vacuolar protein sorting 4 homolog A	27183
ILMN_1758963	0.049	NADK	NAD kinase	65220

ILMN_1703123	0.0491	CSRNP1	cysteine and serine-rich nuclear protein 1	64651
ILMN_1705602	0.0491	KLHL17	kelch like family member 17	339451
ILMN_2317751	0.0492	REC8	REC8 meiotic recombination protein	9985
ILMN_1672662	0.0495	SLC20A1	solute carrier family 20-member 1	6574

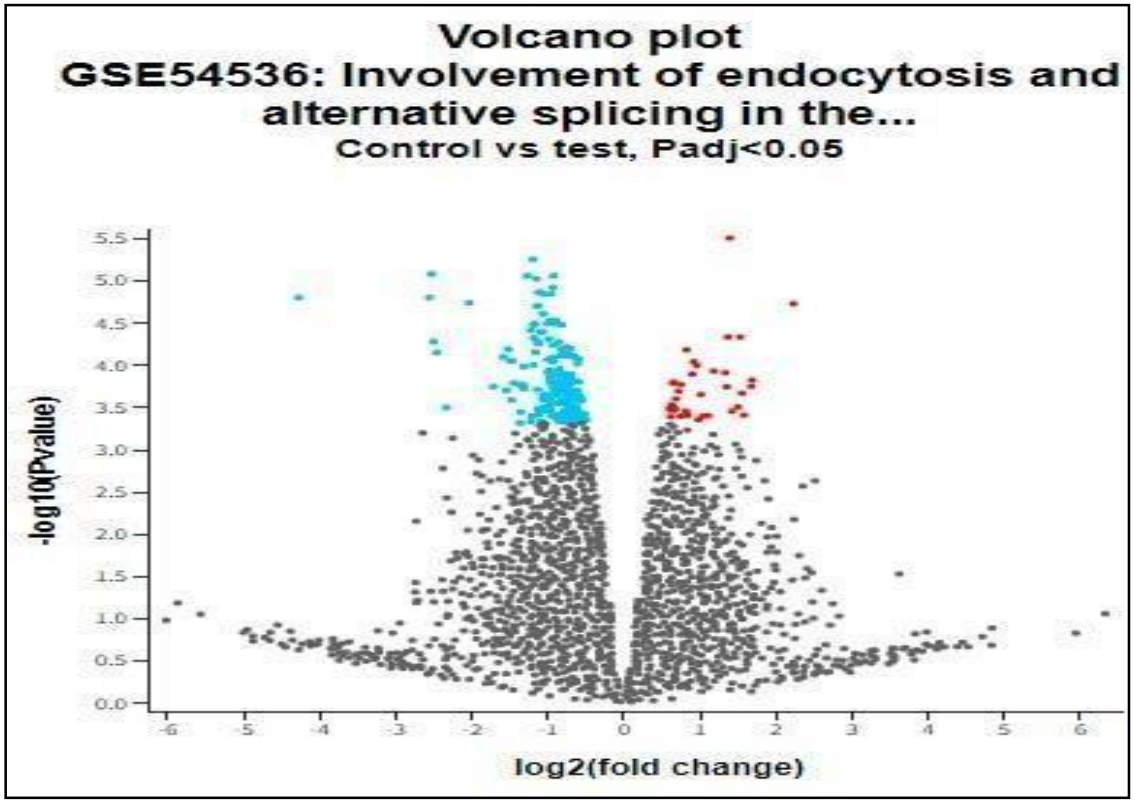

Figure 1: Volcano plot of DEGs

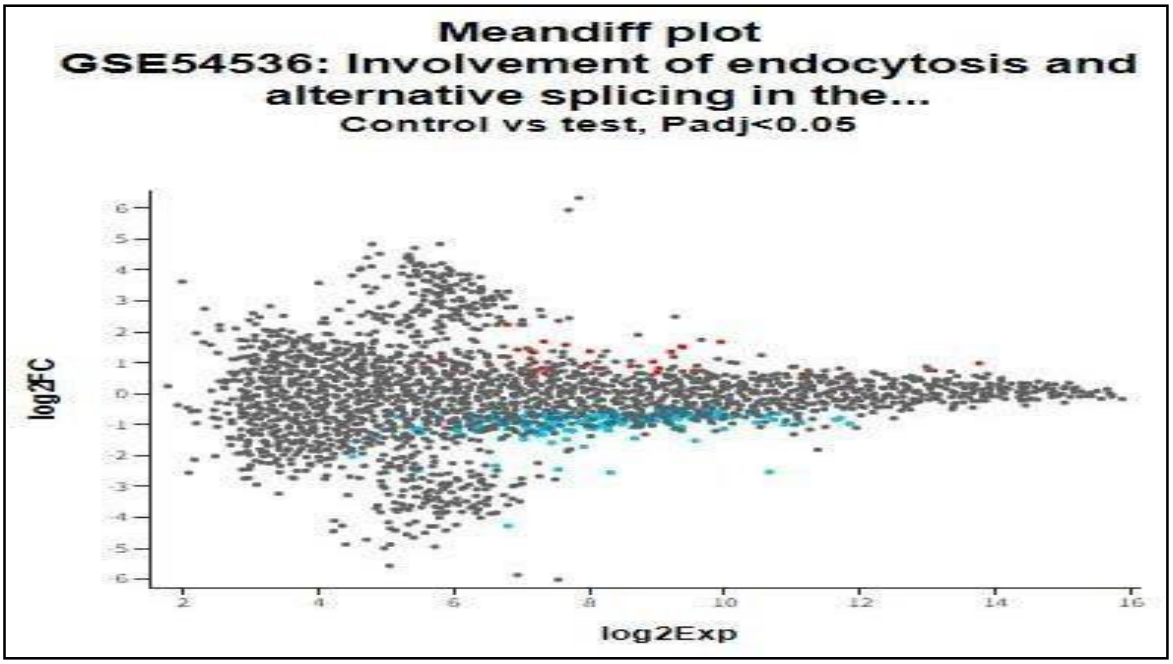

Figure 2: Mean-Difference Plot of DEGs

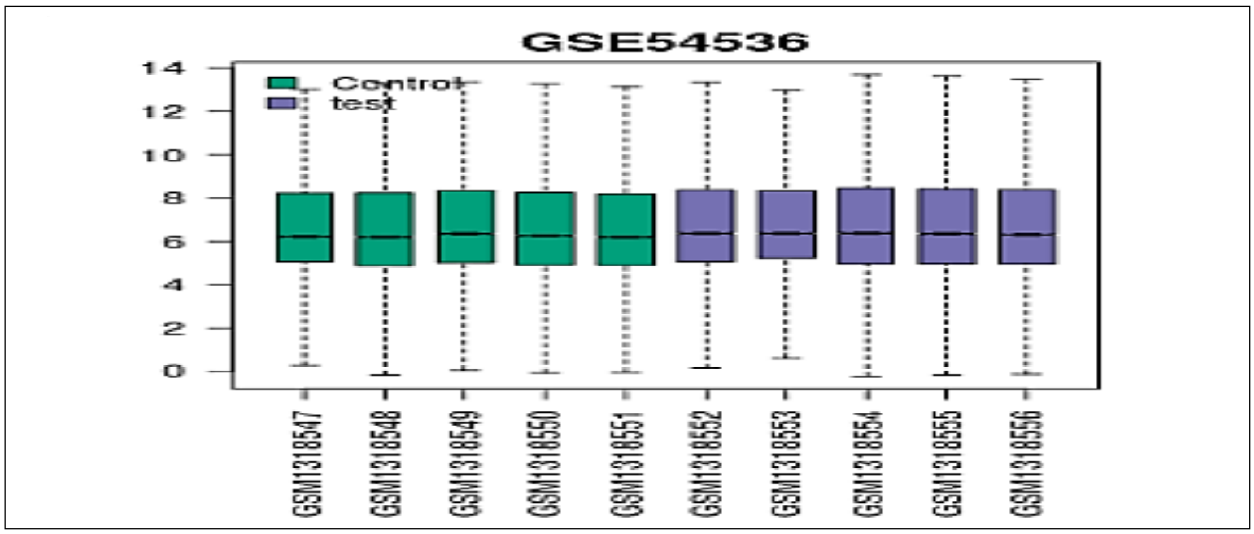

Figure 3: BOX-PLOT of DEGs

Table 3: Clustering of DEGs on the basis of Biological Process

GO TERM	GENE COUNT	P-VALUE	BENJAMANI SCORE
Positive Regulation of Transcription from Rna Polymerase Ii Promoter	23	5.90 E-04	3.50 E-01
Vesicle-Mediated Transport	8	2.60 E-03	9.90 E-01
Regulation Of Catalytic Activity	10	8.00 E-03	1.00 E+00
Phosphorylation	5	2.50 E-02	2.50 E-02
Positive Regulation of Transcription,DnaTemplated	12	2.90 E-02	1.00 E+00
Autophagy	5	3.90 E-02	1.00 E+00
Peptidyl-Serine Phosphorylation	5	6.40 E-02	1.00 E+00
Response To Hypoxia	5	7.20 E-02	1.00 E+00
Positive Regulation Of I-KappabKinase/NfKappab Signaling	5	8.50 E-02	1.00 E+00
Wnt Signaling Pathway	5	8.50 E-02	1.00 E+00

Table 4: Clustering of DEGs on the basis of Cellular Component.

GO TERM	GENE COUNT	P-VALUE	BENJAMANI SCORE
Cytosol	75	2.00E-06	5.50E-04
Nucleus	73	1.60E-04	2.20E-02
Membrane	38	5.40E-04	4.80E-02
Trans-Golgi Network	8	1.10E-03	7.70E-02
Nucleoplasm	50	1.60E-03	8.30E-02
Microtubule	10	1.80E-03	8.30E-02
Cytoplasm	62	5.80E-03	2.20E-01
Late Endosome	6	7.70E-03	2.60E-01
Early Endosome	8	1.10E-02	3.30E-01
Golgi Membrane	12	1.80E-02	4.80E-01
Spindle Pole	5	2.30E-02	5.30E-01
Late Endosome Membrane	5	2.70E-02	5.60E-01
Lysosome	7	4.40E-02	7.80E-01
Endosome	7	5.00E-02	8.40E-01
Cytoplasmic Vesicle	7	5.60E-02	8.70E-01
Nuclear Speck	8	6.70E-02	9.50E-01
Lysosomal Membrane	7	8.90E-02	9.80E-01
Centrosome	9	9.00E-02	9.80E-01
Mitochondrion	18	9.00E-02	9.80E-01
Endoplasmic Reticulum	15	9.10E-02	9.80E-01

Table 5: Clustering of DEGs on the basis of Molecular Function.

GO TERM	GENE COUNT	P-VALUE	BENJAMANI SCORE
Protein Binding	141	1.90 E-08	7.30 E-06
GTPase Activator Activity	9	3.40 E-03	6.70 E-01
ATP Binding	24	7.40 E-03	8.90 E-01
Ubiquitin Binding	5	9.50 E-03	8.90 E-01
Identical Protein Binding	25	1.10 E-02	8.90 E-01
RNA Binding	22	1.50 E-02	9.40 E-01
RNA Polymerase II Sequence-Specific DNA Binding Transcription Factor Binding	6	2.40 E-02	1.00 E+00
Protein Kinase Binding	10	3.00 E-02	1.00 E+00
Transcription Factor Activity, Sequence Specific DNA Binding	10	5.20 E-02	1.00 E+00
Transcription Regulatory Region Sequence Specific DNA Binding	6	5.50 E-02	1.00 E+00
Zinc Ion Binding	13	7.40 E-02	1.00 E+00

Table 6: DEGs involved in cancer Pathway, the most enriched pathway

ENTERZ_GENE_ID	Gene Name	Species
7852	C-X-C Motif Chemokine Receptor 4(CXCR4)	Homo Sapiens
2353	Fos Proto-Oncogene, AP-1 Transcription Factor Subunit(FOS)	Homo Sapiens
11186	Ras Association Domain Family Member 1(RASSF1)	Homo Sapiens
9138	Rho Guanine Nucleotide Exchange Factor 1(ARHGEF1)	Homo Sapiens
2535	Frizzled Class Receptor 2(FZD2)	Homo Sapiens
3600	Interleukin 15(IL15)	Homo Sapiens
3728	Junction Plakoglobin(JUP)	Homo Sapiens
5293	Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Delta(Pik3cd)	Homo Sapiens
6256	Retinoid X Receptor Alpha(RXRA)	Homo Sapiens
6773	Signal Transducer and Activator of Transcription 2(STAT2)	Homo Sapiens
6934	Transcription Factor 7 Like 2(TCF7L2)	Homo Sapiens

Table 7: Network statistics of 180 DEGs using STRING tool. There were 157 nodes and 143 edges with PPI enrichment vale i.e. P-value was (0.0000389), typically ≤ 0.05

NETWORK STATISTIC	
Number of nodes	157
Number of edges	143
Average node degree	1.82
Expected number of edges	101
PPI enrichment value	3.89 e -05

Table 8: Proteins having High node degrees. The threshold set to < 5. Out of 171 genes 13 genes were identified as hub proteins.

PROTEIN NAME	NODE DEGREES
EGR1	9.0
STAT2	9.0
FOS	8.0
FARSA	7.0
IRF3	7.0
PRMT5	7.0
RPLP0	7.0
UBE2L6	7.0
XRN2	7.0
ADAR	6.0
CXCR4	6.0
MOV10	6.0
SUPT5H	6.0

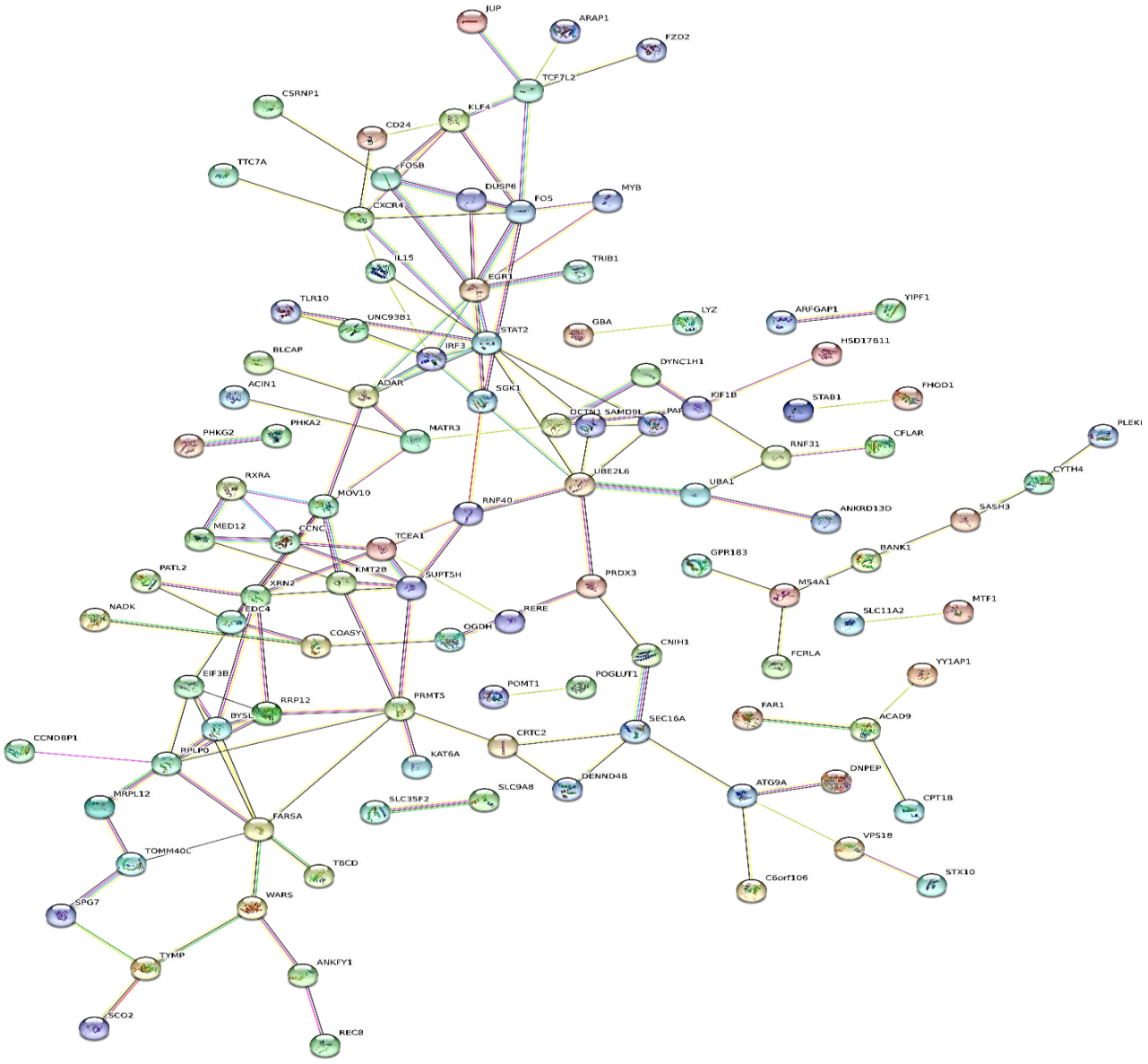

Figure 4: PPI association network of DEGs from GSE54536 datasets. The nodes and edges were retrieved using STRING Tool which represents the interactions between them. There were 157 out of 180 DEGs identified by STRING tool.

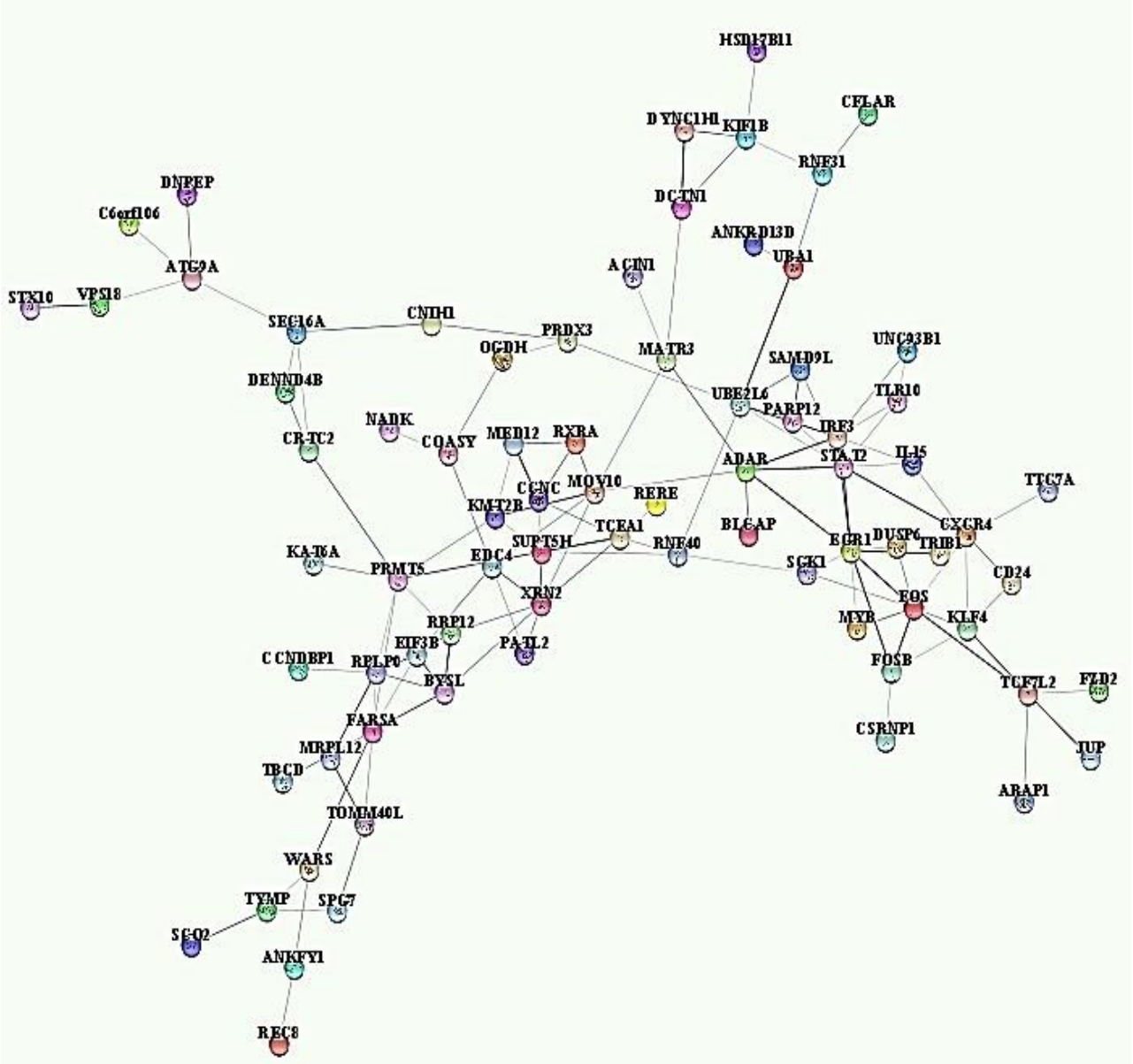

Figure 5: illustrates the PPI interaction network for 180 DEGs imported from theSTRING database visualized by Cytoscape.

DISCUSSION

Parkinson's disease (PD) is a degenerative neurological condition in which dopaminergic neurons of the substantia nigradegenerate, which is the main cause of Parkinson's disease. None of the current cures for Parkinson's disease prevent or delay loss of dopaminergic neurons. Limitedresearch and understanding of the important molecular processes that causes neurodegeneration is the greatest barrier to the development of neuro-protective medicines [18]. In current study, bioinformatics approach was applied to investigate molecular mechanism of Parkinson's disease through identification of total 180 DEGs. These DEGs are significant to develop blood-based biomarkers for Parkinson's disease.

The results of GEO were used to verify normal distribution of gene expression values and contain each of the 10 accessible PD samples was examined as a user-defined sample (5 control, 5 test), healthy sample, and sample with PD disease were studied using classical t-test in limma package (the cut-off is taken P<0.05) of accession number **GSE54536** includes whole transcriptome analysis. The GPL10558 [Illumine Human HT-12 V4.0 expression bead chip] is then used to examine PD samples. Analysis of these datasets on PD gene expression yielded 180 differentially expressed genes in which 23 genes are extracted as up-regulated and 157 genes are extracted as down-regulated, these genes pass cut-off criteria for P-value (> 0.05 or 5%) that are found across study. In contrast to previous studies, 147 DEGs are present in one of the previous studies taking three gene expression profile datasets (GSE6613, GSE54563 and GSE72267) of Parkinson's disease were used and 94 test and 45 normal controls were analyzed [17]. While in other study, 204 DEGs were present taking one gene expression profile of dataset (GSE48280) were used and 19 Idiopathic inflammatory myopathy samples were analyzed [16], demonstrating how useful this study can be and highlighted the significance by suing the gene expression profile. RbcL and Mat-k markers have the potential for the identification and authentication of flowering plants. In our study we extracted DNA of wild grasses collected across different regions of Quetta (Figure 2), the amplified products of conserved regions of both primers were required to qualify as barcodes for the species discrimination. Results obtained by DAVID 180 genes analyzed for gene ontology analysis. Bunches of gene were obtained when these significant enrichments of DEGs were characterized according to biological processes, molecular function, and cellular component. There is 176 David IDs in the record of the specie "Homo sapiens" out of 180 (84 genes from the list are not in the output). The gene count was higher in 10 biological processes, 25 cellular components, and 18 molecular functions. In contrast to, one of the previous studies, the gene count was higher in 8 biological processes, and 5 cellular components and the change in genes expression affects mainly 3 molecular functions, the cut-off criteria of P-value (< 0.001) were also considered. [16]

The KEGG analysis translates gene-level information into pathway-level information, which may then be applied to differentiate between healthy and diseased samples. [19]. The DEGs in Parkinson's disease were expressively enriched in 4 KEGG pathways and the cancer pathway was observed to be the most significantly enriched pathway in the occurrence of Parkinson's disease.

Eleven genes were involved in the cancer pathway, namely CX- C motif chemokine receptor 4(CXCR4), Fos proto-oncogene, AP-1 transcription factor subunit(FOS), Ras association domain family member 1(RASSF1), Rho guanine nucleotide exchange factor 1(ARHGEF1), frizzled class receptor 2(FZD2), interleukin 15(IL15), junction Plakoglobin (JUP),

phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit delta(PIK3CD), retinoid X receptor alpha(RXRA), signal transducer and activator of transcription 2(STAT2) and transcription factor 7 like 2(TCF7L2).

STRING employed 125 overlapping DEGs between GSE54536 and GSE100054 to create a protein-protein (PPI) network with 125 nodes and 93 edges. They then used Cytoscape to evaluate the STRING results, and 20 genes in the PPI network were determined as hub genes in Parkinson's disease. [20]. We extract 13 hub proteins that were identified in PD patients namely ADAR(double-stranded RNA-specific adenosine deaminase isoform a [Homo sapiens], CXCR4(C-X-C chemokine receptor type 4 isoform a [Homo sapiens], EGR1(early growth response protein 1 [Homo sapiens], FARSA(phenylalanine--tRNA ligase alpha subunit [Homo sapiens], FOS(protein c-Fos [Homo sapiens], IRF3 interferon regulatory factor 3 isoform 2 [Homo sapiens], MOV10(helicase MOV-10 isoform 1 [Homo sapiens], PRMT5(protein arginine N-methyltransferase 5 isoform X1 [Macaca fascicularis], STAT2(signal transducer and activator of transcription 2 isoform 1 [Homo sapiens]), RPLP0(acidic ribosomal protein P0 [Homo sapiens], SUPT5H(transcription elongation factor SPT5 isoform X1 [Gorilla gorilla], UBE2L6 (ubiquitinISG15-conjugating enzyme E2 L6 isoform 1 [Homo sapiens], XRN2(exoribonuclease 2 [Marmota monax].The hub proteins have highest node degree i.e. 9.0 as compared to others including EGR1 (early growth response 1) and STAT2 (signal transducer and activator of transcription). The others were having node degrees ranging from 8.0 to 6.0 including: FOS, FARSA, IRF3, PRMT5, RPLP0, UBE2L6, XRN2, ADAR, CXCR4, MOV10 and SUPT5H, respectively. Our study revealed that the protein-protein relationships among dysregulated DEGs.

REFERENCES

1. Jankovic, Joseph. "Parkinson's disease: clinical features and diagnosis." Journal of neurology, neurosurgery & psychiatry 79.4 (2008): 368-376.
2. Hoehn, Margaret M., and Melvin D. Yahr. "Parkinsonism: onset, progression and mortality." Neurology (2001).
3. Galvan, Adriana, and Thomas Wichmann. "Pathophysiology of parkinsonism." Clinical neurophysiology 119.7 (2008): 1459-1474.
4. Oertel, Wolfgang H., and Stanley Fahn. "Parkinsonism." Neurological Disorders. Academic Press, 2003. 1021- 1079.
5. Krokidis, Marios G. Identification of biomarkers associated with Parkinson's disease bygene expression profiling studies and bioinformatics analysis. AIMS neuroscience 6.4 (2019): 333.
6. Diao, Hongyu, et al. "Gene expression profiling combined with bioinformatics analysis identify biomarkers for Parkinson disease." PloS one 7.12 (2012): e52319.
7. R. Hao, H. Du, L. Guo, F. Tian, N. An, T. Yang, C. Wang, B. Wang, Z. Zhou,Identification of dysregulated genes in rheumatoid arthritis based on bioinformatics analysis,PeerJ 5 (2017) e3078.
8. R. Edgar, M. Domrachev, A.E. Lash, Gene Expression Omnibus: NCBI gene expressionand hybridization array data repository, Nucleic Acids Res 30(1) (2002) 207-10.
9. M. Consortium, L. Shi, L.H. Reid, W.D. Jones, R. Shippy, J.A. Warrington, S.C. Baker, P.J. Collins, F. de Longueville, E.S. Kawasaki, K.Y. Lee, Y. Luo, Y.A. Sun, J.C. Willey, R.A. Setterquist, G.M. Fischer, W. Tong, Y.P.

Dragan, D.J. Dix, F.W. Frueh, Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements, Nat Biotechnol 24(9) (2006) 1151-61.

10. A. Reiner-Benaim, FDR control by the BH procedure for two- sided correlated tests with implications to gene expression data analysis, Biom J 49(1) (2007) 107-26.
11. Huang, D. W., Sherman,T., Tan, Q., Collins, J. R., Alvord,W. G., Roayaei, J., ... & Lempicki, R. A. (2007). The DAVID GeneFunctional 8 (9), 1-16.
12. M. Kanehisa, S. Goto, S. Kawashima, A. Nakaya, The KEGG databases at GenomeNet,Nucleic Acids Res 30(1) (2002) 42-6.
13. Hooper, S. D., & Bork, P. (2005). Medusa: a simple tool for interaction graph analysis. Bioinformatics, 21(24), 4432-4433.
14. Udhaya Kumar, S., Thirumal Kumar, D., Bithia, R., Sankar, S., Magesh, R., Sidenna, M., ... & Zayed, H. (2020). Analysis of differentially expressed genes and molecular pathways in familial hypercholesterolemia involved in atherosclerosis: a systematic and bioinformatics approach. Frontiers in Genetics, 11, 734.
15. Xiao, J., & Yiqing, Q. (2013). Bioinformatics analysis of the gene expression profile in Bladder carcinoma. Genetics and molecular biology, 36, 287-291.
16. Seemab, S. H., Naveed, K., Saleem, M., Majid, A., & Ali, S. M. (2018). Identification of dysregulated genes in idiopathic inflammatory myopathy mRNA data using computational methods. International Journal of Biology and Biotechnology, 15(4), 605-612.
17. Gao, A. (2020). Identification of blood-based biomarkers for early stage Parkinson's disease. MedRxiv.
18. Dauer, W., & Przedborski, S. (2003). Parkinson's disease: mechanisms and models.Neuron, 39(6), 889-909.
19. Shen, J., Chen, X. C., Li, W. J., Han, Q., Chen, C., Lu, J. M., ... & Xue,
20. S. R. (2020). Identification of Parkinson's disease-related pathways and potential risk factors. *Journal of International Medical Research*, *48*(10), 0300060520957197.
21. Liu, X., Chen, J., Guan, T., Yao, H., Zhang, W., Guan, Z., & Wang,Y. (2019). miRNAs and target genes in the blood as biomarkers for the early diagnosis of Parkinson's disease. *BMC systems biology*, *13*(1), 1-8.