Journal of Applied Pharmaceutical Science Vol. 12(07), pp 160-170, July, 2022
Available online at http://www.japsonline.com
DOI: 10.7324/JAPS.2022.120716
ISSN 2231-3354

Transcriptomics analysis of gene expression in pulmonary arterial hypertension and identification of hub proteins

Ruhi Hashma, Ruchi Yadav*
Amity Institute of Biotechnology, Amity University Uttar Pradesh, Lucknow, India.

ARTICLE INFO

Received on: 29/10/2021
Accepted on: 23/05/2022
Available Online: 05/07/2022

Key words:
Differential expression, microarray, network construction, Cytoscape, Affy.

ABSTRACT

Pulmonary arterial hypertension (PAH) is a disease of increased pressure in blood vessels of lungs which is caused by the blockage in blood vessels. It is a fatal chronic cardiopulmonary disease that affects both heart and lungs. PAH is a common global disease in which irreversible changes in blood vessels, resulting in long-term resistance of blood vessels and right ventricular failure, are caused. In recent decades, tremendous research has been done toward the understanding of basic pathobiology of PAH and its fundamental history, biomarker prognosis, and treatment options. However, studies providing PAH-related transcriptomic experiments and gene expression in PAH condition are rare. To identify the genes involved in PAH microarray, gene expression data was retrieved from NCBI Gene Expression Omnibus database with accession number: GSE113439 includes 15 PAH samples from patients and 11 from normal cell that is taken as controls data. Total of 100 differentially expressed genes (DEGs) were predicted using the Limma package of R and Bioconductor. Functional enrichment of DEGs was done using bioinformatics databases like Gene Ontology used for functional classification of genes and the Kyoto Encyclopedia of Genes and Genomes databases used for pathway study. Interaction Network was modeled using Cytoscape tool and further CytoHubba tool was used for the prediction of hub genes from the network of DEG. Total five genes, i.e., EIF5B, NCL, PNN, RIOK1, and RSL1D1 were identified as hub genes. Correlation analysis of these hub genes shows that they have a function in PAH disease and may involve in the cause and progression of PAH.

INTRODUCTION

Pulmonary hypertension is generally named essential (idiopathic) or secondary. It is presently clear, in any case, that there are circumstances inside the classification of optional pneumonic hypertension that look like essential pneumonic hypertension in their histopathological highlights and their reaction to treatment (Simonneau *et al*., 2019). Pulmonary blood vessel hypertension is characterized as a consistent raise of pulmonary arterial strain to more than 25 mm Hg very still or to in excess of 30 mm Hg with workout, with a mean pneumonic slender wedge weight and left ventricular end-diastolic weight of under 15 mm Hg (Frost *et al*., 2019).

Pneumonic blood vessel hypertension involves idiopathic aspiratory blood vessel hypertension (once in the past, essential aspiratory hypertension), pulmonary blood vessel hypertension in the setting of collagen vascular sickness [e.g., in restricted cutaneous foundational sclerosis, otherwise called the Calcinosis, Raynaud's, esophageal dysmotility, sclerodactyly, telangiectasia (CREST) condition, entryway hypertension, inborn left-to-right intracardiac shunts, and disease with the human immunodeficiency infection and determines pneumonic hypertension of the infant (Noordegraaf *et al*., 2019)]. Intimal fibrosis increased medial size, pulmonary arteriolar blockage, and plexiform lesions are all histological features of lung tissue in each of these patients. While the pathophysiology of many pulmonary arterial hypertension (PAH) procedures is unknown, numerous recent discoveries have been made, particularly in the area of genetic and biological cells in idiopathic PAH (Pagnesi *et al*., 2020).

*Corresponding Author
Ruchi Yadav, Amity Institute of Biotechnology, Amity University Uttar Pradesh, Lucknow, India.
E-mail: ryadav @ lko.amity.edu

Pulmonary hypertension (PAH) is a common global disease in which irreversible changes in blood vessels, resulting in long-term resistance of blood vessels and right ventricular failure, are caused. the last, death. Currently, researchers have been able to study the physiological behavior of this disease and its impact on human health (Delcroix *et al*., 2021). Recently progress has been made toward the understanding of basic pathological biology and symptoms of PAH disease, biomarker prognosis, and treatment options. However, studies are still scarce which provide PAH-related gene expression profiles. Consequently, identifying clinical molecular biomarkers and probing the underlying biology involved in PAH is an important mission that can help develop novel science-based diagnostics and adopt goal-treatment approaches in PAH patients (Vachiéry *et al*., 2018).

Microarray technology is the most widely used gene expression technology used to study simultaneous expression and regulation of thousands of genes at cellular, organ, or organism level. This novel strategy aids in a more systematic understanding of gene regulation and connections between genes. In microarray practice, however, many undesirable systematic changes are observed (Yadav and Srivastava, 2018). Even in the repeated experiment, some variations are observed. Normalization is the technique of removing specific variances that affect gene expression levels measurement. Despite the fact that several standard ways have been presented, determining which method works best is tough. In the early stages of microarray data analysis, synchronizing is crucial. The results of subsequent analyses are heavily reliant on normalization (Yadav and Srivastava, 2016).

A variety of novel experiments have been developed to study and perform comparison of two samples (e.g., tumor versus normal tissue) based on their gene expression. Microarray is such experiment that measures the intensity values that is proportional to gene expression and it is widely used for comparative study (Yadav and Srivastava, 2019). Certain aspects of microarray investigations enable the development of such novel methods: (i) the enormous number of genes that contribute to expression measurements, which much outweighs the number of samples (observations) available, and (ii) the fact that gene expression measurements are usually highly correlated when it comes to pathway / network interactions. These issues are exacerbated in regression situations, where the goal is to link the expression of several genes to an external outcome or phenotype at the same time. As a result, numerous approaches to addressing these concerns have lately been presented. This shows that gene collecting, without any further constraints, can take up to a year. Using a microarray-based method for evaluating cardiomyopathy in transgenic mice has been widely used for the identification of disease-causing gene and its functional behavior (Basavegowda and Dagnew, 2020).

Expression of genes, a glimpse of all transcriptional activity in a biological sample is provided by microarrays. Microarrays enable the discovery of new and unexpected gene roles, unlike most standard molecular biology methods, which typically allow the analysis of a single gene or a limited number of genes. These technologies have been used to find novel disease subgroups, build new diagnostic tools, and identify basic illness or treatment response mechanisms, among other things. However, because this technology inevitably generates a vast amount of data, we must interpret it using sophisticated statistical and computational tools (Leday *et al*., 2018).

Microarrays detect levels of expression in relation to measuring mRNA resistance in thousands of genes that are not embedded in a glass ("chip") gene, providing a unique way of finding the balance and integration produced by genetic sequencing efforts. From inverted northern blots on filters detected with radioactive probes to a high-tech field including minorized synthesis, multicolor fluorescence identification, and database-driven sample and data management, microarray technology has advanced (Mehmood *et al*., 2017). Whole genomes have recently been analyzed, as well as specific gene families. Gene selection, microarray synthesis, sample preparation, array hybridization, detection, and data processing are typical procedures in microarray tests, with the relevant controls required for each step. Microarray technology is also known as high-bandwidth technology used for the parallel analysis of gene expression for several genes of known and unknown function (Sayed *et al*., 2019).

MATERIALS AND METHODS

Microarray data of PAH disorders was retrieved from NCBI Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo) with the accession number GSE113439 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE113439). Microarray experiment data used for current research is of PAH diseases. This experiment has been conducted using GPL6244 Affymetrix Human Gene 1.0 ST Array platform and it includes 15 PAH samples from patients and 11 samples from normal controls persons. From 15, only 6 PAH patient's samples and these samples were labeled as Group 1 were taken for the research work. Out of 11 normal controls, 6 samples were taken as control samples and labeled as Group 2. Differential gene expression analysis was done between Group 1 (PAH patient's samples) and Group 2 samples (normal controls). The workflow used for the analysis and identification of hub genes has been shown in Figure 1. It shows the list of Bioconductor packages, tools, and databases used for the current study this pipeline can be used for the prediction of hub genes from microarray data and also for network construction.

R and Bioconductor packages (https://www.bioconductor.org/) were used for the statistical analysis of microarray data. Affy and AffyQCReport (Parman *et al*., 2020) is a package of R functions and analysis classes of oligonucleotide arrays produced by Affymetrix. The normalization of data was done using the Affy Bioconductor package. First the raw data, i.e., CEL files were read using the ReadAffy () command. The data is stored in an affybatch object which is further called in several analysis. Robust Multi-array Average (RMA) method of normalization was applied on the microarray data. RMA is a normalization procedure for microarrays that background corrects, normalizes, and summarizes the probe level information without the use of the information obtained in the MM probes. The expression file is generated after doing the normalization of data.

AffylmGUI (quality assessment) AffylmGUI Package (Harris *et al*., 2021 affylmGUI is a graphical user interface (GUI) in the integrated workflow of Affymetrix microarray data. This package was used for complete analysis of microarray data from quality control analysis to the application of linear fit model between different set of samples.

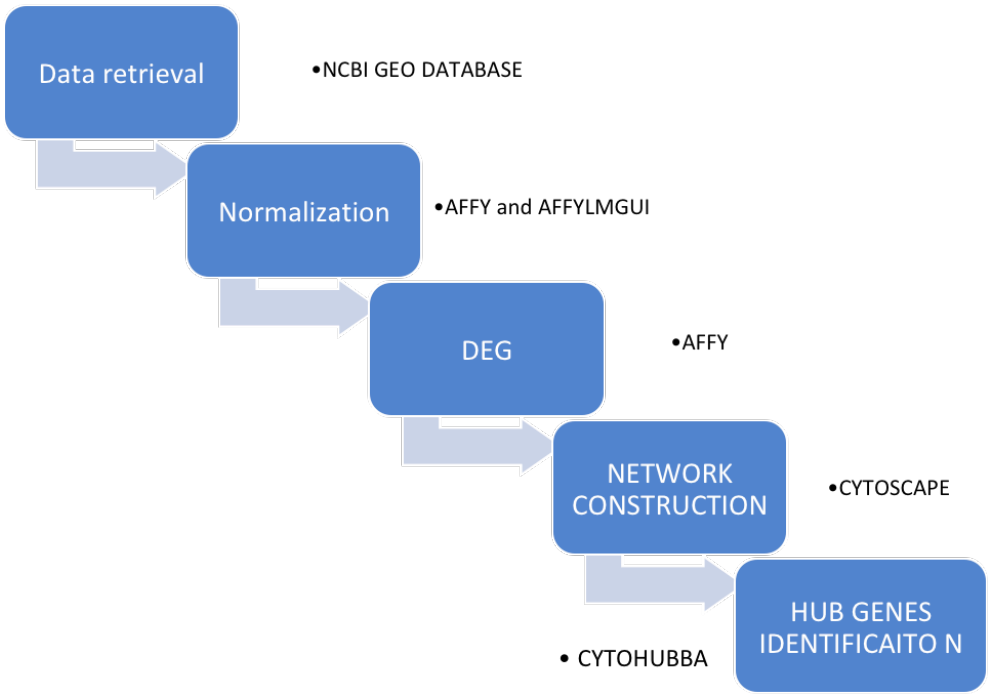

Figure 1. Workflow for construction of network from microarray gene expression data. Workflow represents the methodology used along with databases and tools used in each step.

Limma package Limma is a R / Bioconductor software package that can be used for the identification of differentially expressed genes (DEGs) using linear model (Ritchie *et al*., 2015). Limma package Limma was performed to identify the top DEG using the linear fit model. For the Limma analysis, different contrast between the samples is made so as to obtain the DEG between them. Top genes file was generated through Limma package and volcano plot was constructed through it. Statistical analysis of data was done using Limma package.

>AffyBatch object
>size of arrays = 1,050 × 1,050 features (23 kb)
>cdf = HuGene-1_0-st-v1 (32,321 affyids)
>number of samples = 12
>number of genes = 32,321
>annotation = hugene10stv1
> library (limma)
> colnames (design) <- c (“group1”, “group2”)
> fit = lmFit (eset, design)
> contrast.matrix <- makeContrasts (group2-group1, levels = design)
> fit21 = contrasts.fit (fit, contrast.matrix)
> fit21 = eBayes (fit21)
> table21 = topTable (fit21, coef=1, adjust=”BH”)
> table21
> top_genes <- topTable (fit2, number = 100, adjust = “BH”)
> top_genes

The network created through string database was visualized using Cytoscape (https://cytoscape.org/) (Otasek *et al*., 2019) and the protein—protein interaction (PPI) network studied thoroughly. CytoHubba tool (https://apps.cytoscape.org/apps/cytohubba) (Ma *et al*., 2021) was used for the identification of hub genes. All the network created for both the genes which includes significant genes of *T*-test analysis and the top 25 genes from Limma Bioconductor package were subjected to cytoHubba tool of Cytoscape.

RESULT AND DISCUSSION

Affy QC plots

The density plot and the Raw boxplot was obtained through affylmgui package (Figs. 2 and 3) these plots show the variation in intensity across chip of all probes. Density plots represent the log intensity of chip expression in in *x*-axis and *y*-axis shows the density of chip expression in all sample files. Box plot shows the variation in intensity measured across chip using five different values (minutes, first quartile, median, second quartile and maximum values). Box plot *x*-axis represent the sample files under consideration for current study compared against one another across five values represented in *y*-axis. Before statistical analysis normalization was done and normalized intensity Box Plot for each array has been shown in Figure 4. RNA degradation plot was also studied to visualize the intensity of probes from 5′ end to 3′ end and shown in Figure 5. RNA degradation plot shows the intensity difference of probes those located at the 5′ end and 3′ end of mRNA. This plot directly measures the quality of RNA probes over microarray chip and shows that intensity of probes degrades towards 3′ end. Hence, this parameter is important for differential gene expression analysis and to avoid biasness in interpretation of the results.

Limma package

Differentially expressed genes were identified by R and Bioconductor Limma package with default parameters. Top 25 genes were selected based on linear fit model analysis using

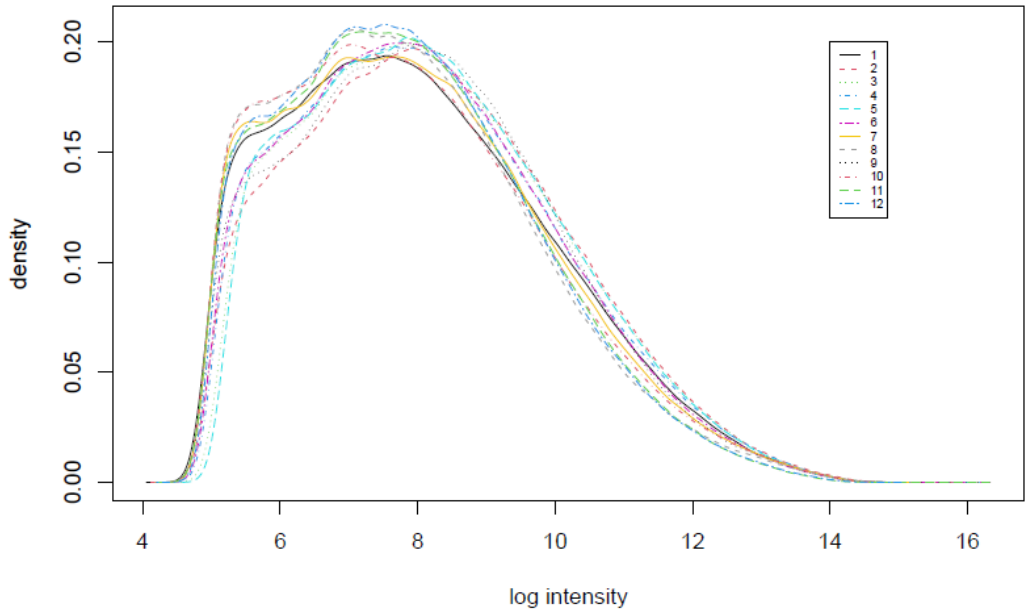

Figure 2. Density plot shows the variation between all sample's files.

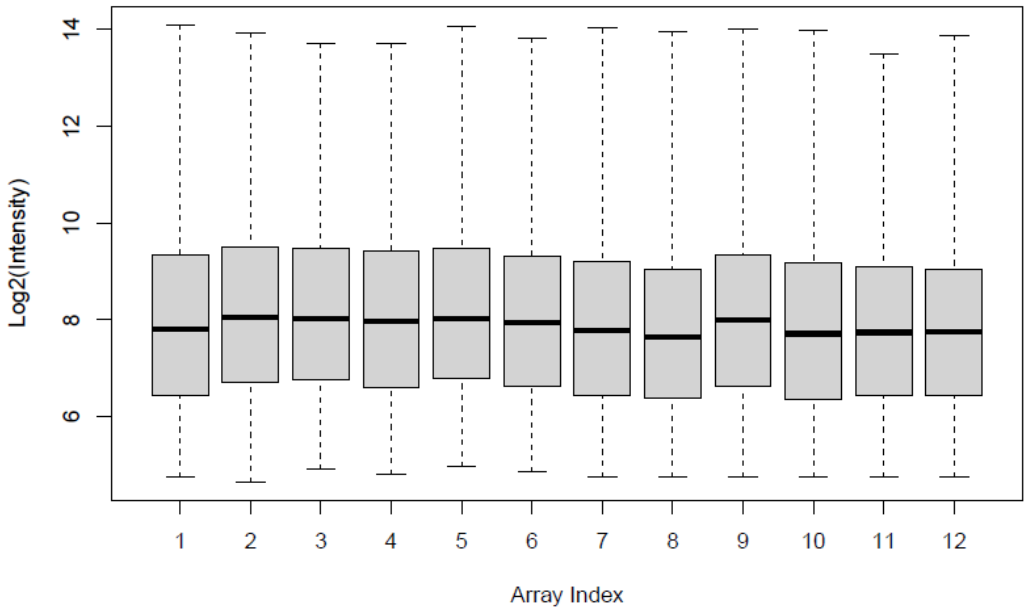

Figure 3. Raw boxplot shows the variation across all files using five different values (minutes, first quartile, median, second quartile and maximum values).

Limma package and ranking of genes according to *B* value. Further functional annotation, these genes were also used for the prediction of PPI and for the identification of hub genes. A volcano plot was visualized that shows the scatterplot between statistical significance which is *p*-value versus magnitude of change which fold change. Figure 6 shows the Volcano plot between two samples used for the identification of DEGs. Volcano plot analysis and according to *p*-value it is clear that there is no significant difference in gene expression between PAH samples and control samples. Hence, protein interaction and identification of hub genes becomes crucial in further investigation and analysis.

Detailed description of the top 25 DEGs used for further study has been shown in Table 1, it describes the gene name, statistical values like log FC values, *t* value, *p* value etc. These 25 genes have been selected on the basis of linear fit model statistics applied using LIMM package and ordered according to *B* value that is log -odds value of DEGs.

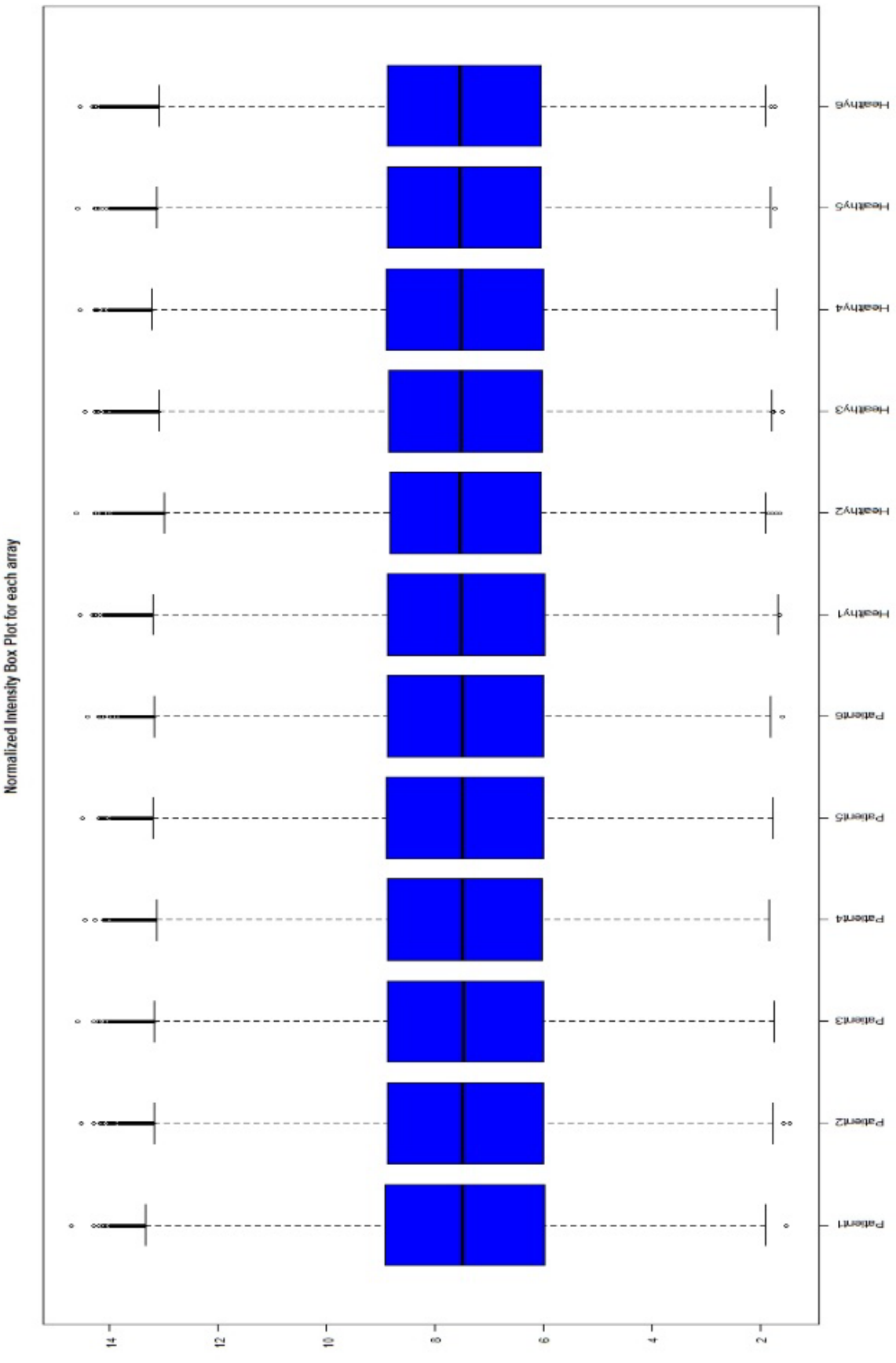

Figure 4. Normalized intensity Box Plot for each array through affImGUI shows all the samples files have been normalized across all five values.

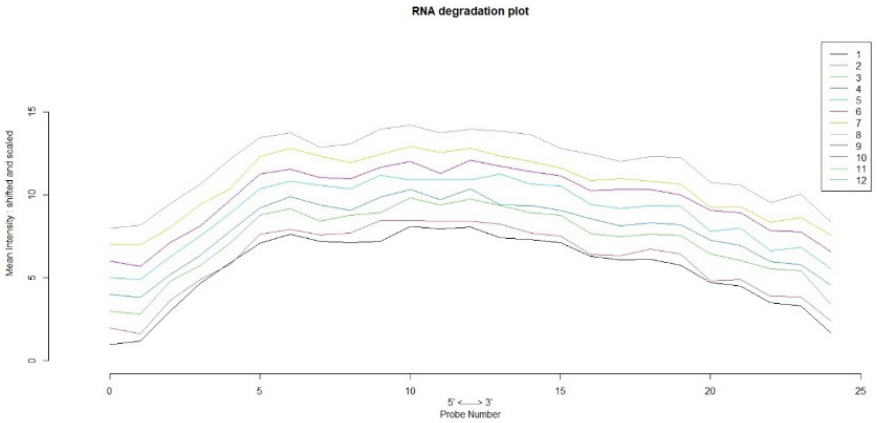

Figure 5. RNA degradation plot through affylmGUI. It shows the intensity from 5′ end to 3′ end of probes.

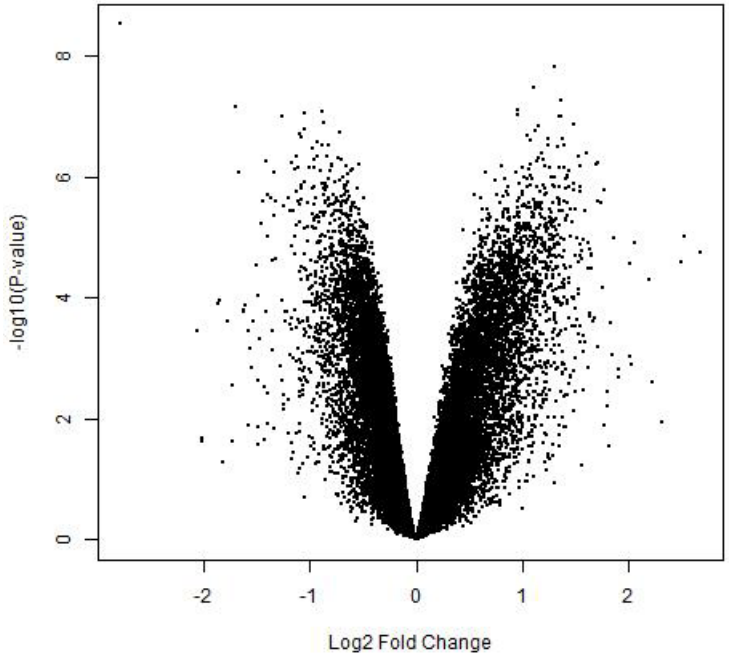

Figure 6. Volcano plot between two samples for the identification and visualization of DEGs.

For prediction of significance and mechanism of these 25 genes, further protein interaction and functional enrichment was done. PPI study of all 25 genes predicted from Limma package has been used to study protein interaction using STRING database (https://string-db.org/) (Szklarczyk *et al*., 2019). PPI result has been shown in Figure 7 with the following network properties nodes: 26, edges: 20, node degree (average): 1.54, local clustering coefficient: 0.307 and PPI enrichment *p*-value: 0.0582. Six proteins were identified outside the network and they do not have any edges these are the proteins that do not have any interacting partner.

Further PPI network was visualized in Cytoscape tool for further analysis and prediction of hub genes as shown in Figure 8. CytoHubba tool was used for the identification of hub genes as shown in Figure 9.

Five proteins were identified as hub genes out of the top 25 DEGs which was obtained from Limma package (Fig. 9). Functional annotation and enrichment of all five hub genes was done using Gene Ontology (GO) database (http://geneontology.org), Kyoto Encyclopedia of Genes and Genomes (KEGG) database (https://www.genome.jp/kegg/pathway.html), etc. Hub

Table 1. Top 25 DEGs identified from Limma Package. Table shows the affymetrix id, gene name along with the statistical values like log FC values, *t* value, etc.

S.No	Affymetrix exon ID	Gene name	logFC	*t*	*p* value	adj. *p* value
1	7920875	Small Cajal body-specific RNA 4 (SCARNA4)	−2.77886	−14.1057	2.94E-09	9.5E-05
2	7939805	Olute carrier family 39 member 13 (SLC39A13)	−0.93293	−9.20486	4.66E-07	0.000368
3	7941583	RAB1B, member RAS oncogene family (RAB1B)	−0.88595	−9.59034	2.92E-07	0.000331
4	7947245	Heat shock protein 90 pseudogene	1.61091	9.322391	4.04E-07	0.000368
5	7965200	Coiled-coil domain containing 59(CCDC59)	1.350658	10.51322	1E-07	0.000271
6	7952004	Sodium voltage-gated channel beta subunit 4 (SCN4B)	−1.08437	−9.83745	2.18E-07	0.000325
7	7974066	Pinin, desmosome associated protein (PNN)	1.08155	8.959517	6.34E-07	0.000368
8	7999520	Ribosomal L1 domain containing 1 (RSL1D1)	1.360134	11.08757	5.36E-08	0.000271
9	8006112	Nuclear speckle splicing regulatory protein 1 (NSRP1)	1.484717	10.26929	1.32E-07	0.000304
10	8004030	Ring finger protein 167 (RNF167)	−0.70254	−9.12322	5.16E-07	0.000368
11	8020903	Polypeptide N-acetylgalactosaminyltransferase 1 (GALNT1)	1.532747	9.213639	4.61E-07	0.000368
12	8023392	Small nucleolar RNA, H/ACA box 37 (SNORA37)	−1.09165	−9.95937	1.89E-07	0.000321
13	8038809	Natural killer cell granule protein 7 (NKG7)	−1.69087	−10.8504	6.92E-08	0.000271
14	8041015	Kanadaptin	1.358944	10.51216	1E-07	0.000271
15	8043861	Eukaryotic translation initiation factor 5B (EIF5B)	1.39114	9.583929	2.94E-07	0.000331
16	8058052	HSPD1 heat shock protein	1.700203	8.974133	6.22E-07	0.000368
17	8059689	Nucleolin (NCL)	1.331521	9.529218	3.14E-07	0.000338
18	8085954	5-azacytidine induced 2 (AZI2)	1.129046	10.05894	1.68E-07	0.00032
19	8105504	Pseudogene (LOC100421561)	1.244837	9.798323	2.28E-07	0.000325
20	8105714	SREK1 splicing proteins	1.175357	9.334732	3.98E-07	0.000368
21	8116760	RIO kinase 1 (RIOK1)	1.154624	9.169106	4.88E-07	0.000368
22	8131292	RB associated KRAB zinc finger (RBAK)	1.144675	10.21212	1.41E-07	0.000304
23	8146482	Trimethylguanosine synthase 1 (TGS1)	0.948904	10.59864	9.13E-08	0.000271
24	8159541	Chromosome 9 open reading frame 142 (C9orf142)	−0.79022	−8.95752	6.35E-07	0.000368
25	8166500	Zinc finger protein, X-linked (ZFX)	1.306693	12.32878	1.51E-08	0.000244

genes annotation with their Gene Name, GO ID, GO Molecular Function, and KEGG Pathways have been shown in Table 2.

Serine/threonine-protein kinase RIO1 (RIOK1)

Function

Involved in the final stages of the 40S ribosomal subunit's cytoplasmic formation and the conversion of 18S-E pre-rRNA to mature 18S rRNA. Reprocessing of NOB1 and PNO1 from the late 40S precursor is required. Prior to binding to the 60S ribosomal subunit, the connection with the late 40S subunit intermediate may comprise a translation-like checkpoint point cycle. Despite the fact that the protein kinase domain is thought to function primarily as an ATPase, (Lim *et al*., 2019).

Pinin (PNN)

Function

The E-box 1 core region of the E-cadherin promoter gene is bound by transcriptional activator. 5′CAGGTG-3′ is the core-binding sequence. CTBP1-mediated transcription repression was successfully removed. At the splice junction on mRNAs, an auxiliary constituent of the splicing-dependent multiprotein exon junction complex (EJC), is located. The EJC is a dynamic assembly of core proteins and various outlying nuclear and cytoplasmic associated components that only join the compound for a brief period of time, either during EJC construction or after mRNA metabolism. Involved in epithelia cell-cell attachment creation and maintenance. Renal cell carcinoma has a latent tumor suppressor (Meng *et al*., 2021).

Nucleolin (NCL)

Function

Growing eukaryotic cells have a significant nucleolar protein. It is shown to be linked to pre-ribosomal particles and intra-nucleolar chromatin. By binding histone H1, it promotes chromatin decondensation. It is assumed to be involved in the writing of pre-rRNA, ribosome assembly, and transcriptional writing development. It binds to RNA oligonucleotides with the configuration 5′-UUAGGG-3, which is twice as strong as

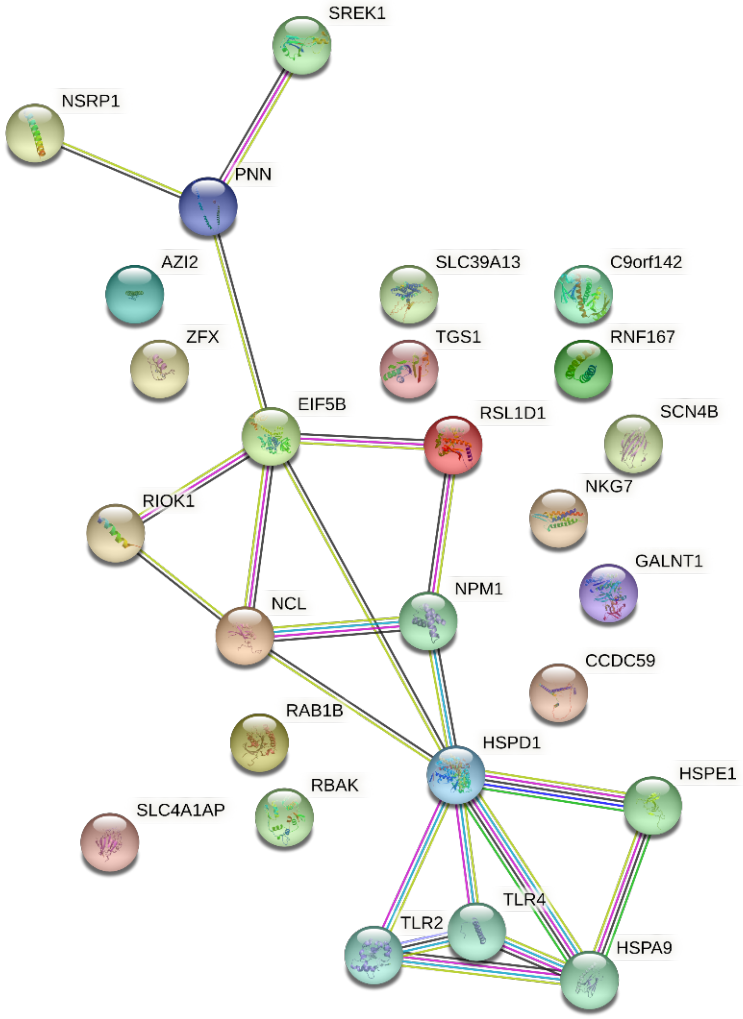

Figure 7. Protein-protein network construction of 25 DEGs using STRING database.

telomeric single-stranded DNA with the sequence 5′-TTAGGG-3′ (Wang *et al*., 2019a).

Eukaryotic translation initiation factor 5B (EIF5B)

Function

The translation process starts. Translational GTPase combines 40S and 60S subunits to form an 80S initiation complex with initiator methionine-tRNA in a P-site base bound to the original codon. This results in consistent changes to an enzyme due to GTP binding and -hydrolysis, resulting in the production of productive interactions with the ribosome. The release of the enzyme is required for the production of strong ribosomes after the initial structure has been formed (Wang *et al*., 2019b).

Ribosomal L1 domain-containing protein 1 (RSL1D1)

Function

The regulation of cellular ageing is done through inhibition of translation of PTEN. In response to DNA damage, it acts as a pro-apoptotic regulator. EIF5B is highly divergent and belongs to the common ribosomal protein uL1 family (Cheng *et al*., 2020).

DEGs identified through DEGs analysis of PAH and control samples were used for the construction of PPI map. PPI map and with was used for the prediction of functional and hub genes that are expressed in PAH disease condition. Five genes that are EIF5B, NCL, PNN, RIOK1, RSL1D1 were identified as hub genes and have functions related with PAH disease. Functional analysis of these five hub genes shows that they have function in cancer (RSL1D1), RNA transport (EIF5B, PNN) etc. Research has been done to study the mechanism and genes expressed in PAH condition, it has been shown that genes are expressed in early stage of embryo development and hence some fetal genes are also responsible (Lemay *et al*., 2021).

Genes associated with PAH condition has also been shown in connective tissue diseases and many genes has been identified like NOCH 1, TBX4 etc (Hernandez-Gonzalez *et al*., 2021). Microarray gene expression experiments has also been done to identify drug targets like CCR1, CCL5, SAA1 etc. (Xu *et al*., 2021). These findings along with the current study and

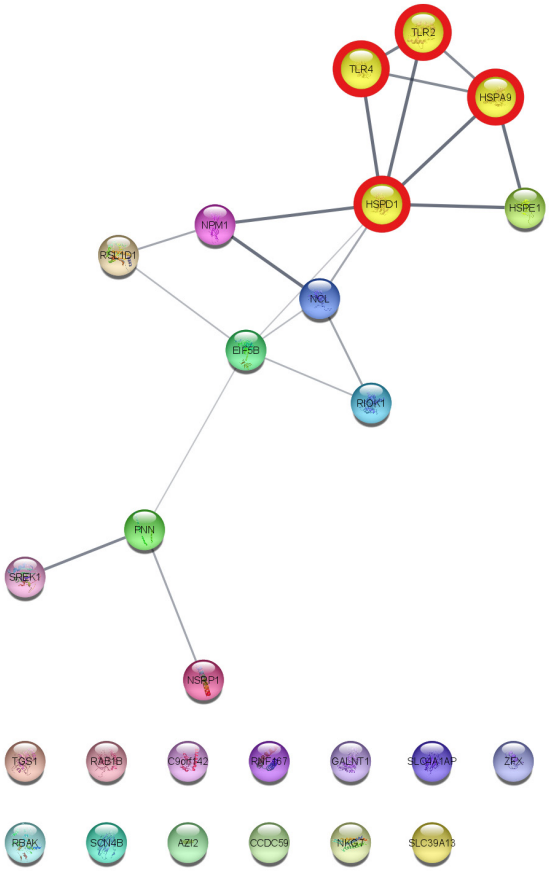

Figure 8. PPI visualized in Cytoscape tool and identification of hub genes highlighted in red circle with yellow nodes.

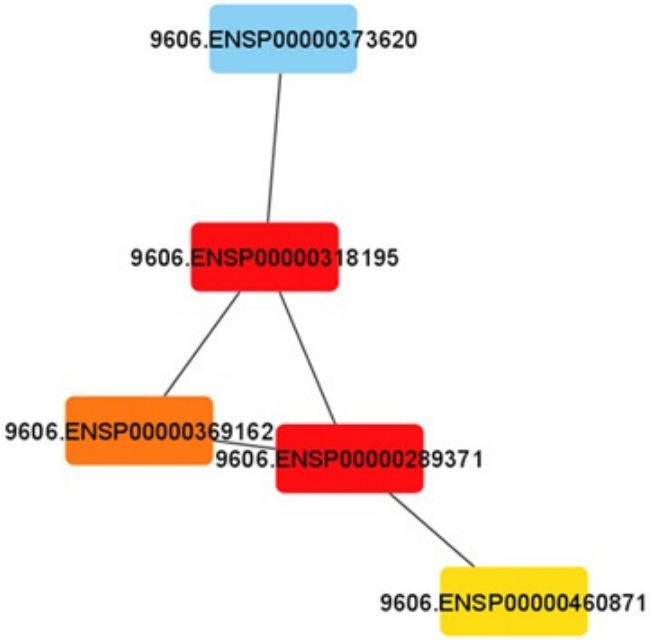

Figure 9. Top 5 hub genes identified using CytoHubba tool.

Table 2. GO, molecular function, and pathways of all five hub genes.

Gene name	GO ID	GO molecular function	KEGG pathways
EIF5B	GO:0005525	Binding to GTP	has03013 RNA transport
	GO:0003924	GDP binding	
	GO:0046872	RNA binding	
	GO:0005515	Glycoprotein binding	
	GO:0003723	GTP and GDP binding	
	GO:0003743	Ribosome-dependent translation of mRNA activity	
	GO:0044547	Protein kinase B bindi	
	GO:0042802	Signaling receptor binding	
NCL	GO:0048027	Telomeric DNA binding	hhas3013, hsa03015 RNA transport, mRNA surveillance pathway
	GO:0005515	Glycoprotein binding	
	GO:0008022	rRNA primary transcript binding	
	GO:0003723	ErbB-4 class receptor binding	
	GO:0042162	Sequence-specific DNA binding	
PNN	GO:0003677	Antigen bihasng	hsa03008 Ribosome biogenesis in eukaryotes
	GO:0005515	Glycoprotein binding	
	GO:0003723	Protein kinase B binding	
RIOK1	GO:0005198	Structural molecule binding activity	
	GO:0005524	Methionine adenosyltransferahasactivity	
	GO:0016787	Zinc ion binding	
	GO:0046872	Ion channel activity	
	GO:0005515	Glycoprotein binding	
RSL1D1	GO:0004674	MAP kinase activity	Senescence and autophagy in cancer
	GO:0045296	Dopamine receptor binding	
	GO:0003730	Translation repressor activity	
	GO:0048027	Ubiquitin protein ligase binding	
	GO:0003723	Fibroblast growth factor binding	

results provides more insight into the PAH diseases that can be used for molecular diagnostics and in pharma industry for the designing of drugs. Identified genes can be used as novel drug targets and towards the molecular understanding of PAH.

CONCLUSION

Microarray gene expression experiment is an important tool to study the differential gene expression in different state or functional category of cell. This experiment can be used in multidimensional analysis for the identification of DEGs, co-expressed genes, pathway analysis, protein interaction analysis etc. In this research we have used microarray data for the identification of hub genes that are expressed in PAH patients. The microarray data analysis was done on the data of PAH patients and their control retrieved from public database of microarray. DEGs were identified using R and Bioconductor packages like Limma, Affy package etc. Statistical test like *T*-test analysis was done using Limma package and top 25 DEGs were further used for the study. These 25 DEGs were used for the network construction through STRING database. Then the network was redirected to Cytoscape where the study of PPI was done and then the five hub genes were identified through cytoHubba. Hub genes identified out of 25 genes were: EIF5B, NCL, PNN, RIOK1, RSL1D1. Function annotation of these five hub genes shows that they have function in cancer (RSL1D1), RNA transport (EIF5B, PNN), etc. These genes can be useful for the identification of novel drug targets and function understanding of PAH.

ACKNOWLEDGMENT

The authors would like to acknowledge Amity Institute of biotechnology, Amity University Uttar Pradesh, Lucknow campus for providing us facilities to conducting this study. This research project is not funded by any specific grant from funding agencies in the public, commercial, or non-profit sectors.

CONFLICT OF INTEREST

Authors do not have any conflict of interests.

FUNDING

No funds were provided for this research.

AUTHOR CONTRIBUTIONS

All authors made substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data; took part in drafting the article or revising it critically for important intellectual content; agreed to submit to the current

journal; gave final approval of the version to be published; and agree to be accountable for all aspects of the work. All the authors are eligible to be an author as per the international committee of medical journal editors (ICMJE) requirements/guidelines.

ETHICAL APPROVALS

This study does not involve experiments on animals or human subjects.

DATA AVAILABILITY

All data generated and analyzed are included within this research article.

PUBLISHER'S NOTE

This journal remains neutral with regard to jurisdictional claims in published institutional affiliation.

REFERENCES

Basavegowda HS, Dagnew G. Deep learning approach for microarray cancer data classification. CAAI Trans Intell Technol, 2020; 5:22–33.

Cheng R, Qi L, Kong X, Wang Z, Fang Y, Wang J. Identification of the significant genes regulated by estrogen receptor in estrogen receptor-positive breast cancer and their expression pattern changes when tamoxifen or fulvestrant resistance occurs. Front Gene, 2020; 11.pp: 1117

Delcroix M, Torbicki A, Gopalan D, Sitbon O, Klok FA, Lang I, Jenkins D, Kim NH, Humbert M, Jais X, Noordegraaf AV. ERS statement on chronic thromboembolic pulmonary hypertension. Eur Respir J, 2021; 57: 1–37.

Frost A, Badesch D, Gibbs JS, Gopalan D, Khanna D, Manes A, Oudiz R, Satoh T, Torres F, Torbicki A. Diagnosis of pulmonary hypertension. Eur Respir J, 2019; 53: 1–12

Harris GM, Abbas S, Miles MF. GCSscore: an R package for differential gene expression analysis in Affymetrix/Thermo-Fisher whole transcriptome microarrays. BMC Genom, 2021; 22:1–2.

Hernandez-Gonzalez I, Tenorio-Castano J, Ochoa-Parra N, Gallego N, Pérez-Olivares C, Lago-Docampo M, Palomino Doza J, Valverde D, Lapunzina P, Escribano-Subias P. Novel genetic and molecular pathways in pulmonary arterial hypertension associated with connective tissue disease. Cells, 2021; 10:1488.

Leday GG, Vértes PE, Richardson S, Greene JR, Regan T, Khan S, Henderson R, Freeman TC, Pariante CM, Harrison NA, Bullmore E. Replicable and coupled changes in innate and adaptive immune gene expression in two case-control studies of blood microarrays in major depressive disorder. Biol Psych, 2018; 83:70–80.

Lemay SE, Awada C, Shimauchi T, Wu WH, Bonnet S, Provencher S, Boucherat O. Fetal gene reactivation in pulmonary arterial hypertension: good, bad, or both? Cells, 2021; 10:1473.

Lim H, He D, Qiu Y, Krawczuk P, Sun X, Xie L. Rational discovery of dual-indication multi-target PDE/Kinase inhibitor for precision anti-cancer therapy using structural systems pharmacology. PLoS Comput Biol, 2019; 15:e1006619.

Ma H, He Z, Chen J, Zhang X, Song P. Identifying of biomarkers associated with gastric cancer based on 11 topological analysis methods of CytoHubba. Sci Rep, 2021; 11:1.

Mehmood R, El-Ashram S, Bie R, Dawood H, Kos A. Clustering by fast search and merge of local density peaks for gene expression microarray data. Sci Rep, 2017; 7:1–7.

Meng XY, Zhang HZ, Ren YY, Wang KJ, Chen JF, Su R, Jiang JH, Wang P, Ma Q. Pinin promotes tumor progression via activating CREB through PI3K/AKT and ERK/MAPK pathway in prostate cancer. Am J Cancer Res, 2021; 11:1286.

Noordegraaf AV, Chin KM, Haddad F, Hassoun PM, Hemnes AR, Hopkins SR, Kawut SM, Langleben D, Lumens J, Naeije R. Pathophysiology of the right ventricle and of the pulmonary circulation in pulmonary hypertension: an update. Eur Respir J, 2019; 53: 1–13

Otasek D, Morris JH, Bouças J, Pico AR, Demchak B. Cytoscape automation: empowering workflow-based network analysis. Gen Biol, 2019; 20:1–5.

Pagnesi M, Baldetti L, Beneduce A, Calvo F, Gramegna M, Pazzanese V, Ingallina G, Napolano A, Finazzi R, Ruggeri A, Ajello S. Pulmonary hypertension and right ventricular involvement in hospitalised patients with COVID-19. Heart, 2020; 106:1324–31.

Parman C, Halling C, Gentleman R. *AffyQCReport: QC report generation for affyBatch objects*. R package version 1.68.0, 2020. It is bioconductor package and citation has been written as mentioned by developer https://129.217.206.11/packages/3.12/bioc/html/affyQCReport.html

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucl Acids Res*, 2015; 43:e47.

Sayed S, Nassef M, Badr A, Farag I. A nested genetic algorithm for feature selection in high-dimensional cancer microarray datasets. Exp Syst Appl, 2019; 1:233–43.

Simonneau G, Montani D, Celermajer DS, Denton CP, Gatzoulis MA, Krowka M, Williams PG, Souza R. Haemodynamic definitions and updated clinical classification of pulmonary hypertension. Eur Respir J, 2019;53: 1–13.

Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucl Acids Res, 2019; 47:D607–13.

Vachiéry JL, Delcroix M, Al-Hiti H, Efficace M, Hutyra M, Lack G, Papadakis K, Rubin LJ. Macitentan in pulmonary hypertension due to left ventricular dysfunction. Eur Resp J, 2018; 51.

Wang F, Zhou S, Qi D, Xiang SH, Wong ET, Wang X, Fonkem E, Hsieh TC, Yang J, Kirmani B, Shabb JB. Nucleolin is a functional binding protein for salinomycin in neuroblastoma stem cells. J Am Chem Soc, 2019a; 141:3613–22: 1–9.

Wang J, Johnson AG, Lapointe CP, Choi J, Prabhakar A, Chen DH, Petrov AN, Puglisi JD. eIF5B gates the transition from translation initiation to elongation. Nature, 2019b; 573:605–8.

Xu J, Yang Y, Yang Y, Xiong C. Identification of potential risk genes and the immune landscape of idiopathic pulmonary arterial hypertension via microarray gene expression dataset reanalysis. Genes, 2021; 12:125.

Yadav R, Srivastava P. Clustering, pathway enrichment, and protein-protein interaction analysis of gene expression in neurodevelopmental disorders. Adv Pharmacol Sci, 2018; 2018: p.3632159

Yadav R, Srivastava P. Significant analysis of microarray (SAM) to identify synergistic effect of RV and NGF in repairing damaged neuronal cells. Toxicol Int, 2019; 25:26–39.

Yadav R, Srivastava P. Visualization of high throughput genomic data using R and Bioconductor. J Data Mining Genom Proteom, 2016; 7:2153–0602.

How to cite this article:
Hashma R, Yadav R. Transcriptomics analysis of gene expression in pulmonary arterial hypertension and identification of hub proteins. J Appl Pharm Sci, 2022; 12(07):160–170.