Informatics in Medicine Unlocked 20 (2020) 100380

Contents lists available at ScienceDirect

Informatics in Medicine Unlocked

journal homepage: http://www.elsevier.com/locate/imu

Identification of differentially expressed genes present in the whole blood of Pulmonary Arterial Hypertension patients and control patients: An integrated bioinformatics approach

Chisom Ezenwa [a], Opeyemi S. Soremekun [b], Oyekanmi Nashiru [a,**], Segun Fatumo [a,c,d,*]

[a] Centre for Genomics Research and Innovation, National Biotechnology Agency, Abuja, Nigeria
[b] Molecular Bio-computation and Drug Design Laboratory, School of Health Science,University of KwaZulu-NatalWestville Campus, Durban, South Africa
[c] Uganda Medical Informatics Centre and MRC/UVRI LSHTM Uganda Research Unit, Entebbe, Uganda
[d] London School of Hygiene and Tropical Medicine, London, United Kingdom

ARTICLE INFO

Keywords:
DEGs
Enrichment analysis
Integrated bioinformatics
Pulmonary arterial hypertension
RNAseq

ABSTRACT

Introduction: While Pulmonary Arterial Hypertension (PAH) remains a commonly undiagnosed disease, it remains a life-threatening disease; it is characterized by pulmonary vascular remodelling subsequently leading to heart failure. Different researches have been at the forefront of exploring drugs for treatment and molecular biomarkers for early diagnosis of PAH.

Method: ology: In this study, we used an integrated bioinformatics approach to investigate Differentially Expressed Genes (DEG) in the whole blood of PAH patients relative to gender/age matching controls. Microarray dataset of the aforementioned experiment was retrieved from Gene Expression Omnibus (GEO). DEG analysis was carried out with the aid of the limma algorithm in-built in the GEO2R tool. Gene Ontology terms such as molecular function, biological process, cellular component, and pathway were investigated using the online tool Protein ANalysisTHrough Evolutionary Relationships (PANTHER). Protein-Protein interaction network was carried out using STRING.

Results: From the analysis, 191 genes were down-regulated while 5 were up-regulated. Some of these genes are implicated in pathways involved in adrenaline and noradrenaline biosynthesis, angiogenesis, EGF receptor signaling pathway, and VEGF signaling pathway. Furthermore, in the ontology of molecular function, these genes are involved in transport activity, catalytic activity, and molecular transducer activity. Interestingly, the angiogenesis, adrenaline, and noradrenaline biosynthetic pathways are heavily involved in the pathogenesis and progression of PAH. Furthermore, the gene products of these predicted genes were also explored. **Conclusion**: The gene products (proteins) of these DEGs can be further explored as potential drug targets in the treatment of PAH. This study has also been able to establish the interaction responsible for PAH which can be explored in gene therapy.

1. Introduction

Pulmonary hypertension (PH) is an aggregate of diseases possessing a resting mean pulmonary artery pressure (mPAP) $\geq$ 25 mmHg [1]. It is characterized by endothelial cell dysfunction and increased contractility of the small pulmonary arteries at the cellular level. Therefore, it can cause abnormal smooth muscle proliferation jointly with resistance to apoptosis [2,3]. PH is a hemodynamic diseased state which leads to unfavourable pulmonary and cardiovascular indicators. The anatomic site and disease aetiology have been factored into the classification of PH; this came about after the conclusion of the 1998 Second Symposium on Pulmonary Hypertension [4]. The updated classification that originated from the 3rd World Symposium on Pulmonary Hypertension reclassified PH classification. PAH classified under group 1 includes: drug- and toxin-induced PAH, PAH long-term responders to calcium channel blockers, and PAH with overt features of venous/capillaries involvement. The updated list in group 5 includes PH with unclear and/or multifactorial mechanisms [5]. Pulmonary Arterial

* Corresponding author. Uganda Medical Informatics Centre and MRC/UVRI LSHTM Uganda Research Unit, Entebbe, Uganda.
** Corresponding author.
E-mail addresses: oyekan.nash@gmail.com (O. Nashiru), segun.fatumo@lshtm.ac.uk (S. Fatumo).

https://doi.org/10.1016/j.imu.2020.100380
Received 13 April 2020; Received in revised form 16 June 2020; Accepted 19 June 2020
Available online 22 June 2020

Hypertension (PAH) is described by high arterial pressure due to the reshaping of pulmonary arteries [6]. PAH is classified to belong to group 1 of PH. Different classes of PAH include drug- and toxin-induced, Heritable (HPAH), and Idiopathic (IPAH). Comorbidity of PAH and other diseases such as schistosomiasis, portal hypertension, HIV infection, and connective tissue disease has been established [7]. The leading cause of death in PAH is vasoconstriction of pre-capillary arterioles leading to increased pulmonary vascular resistance (PVR), and subsequently right ventricular failure (RVF) [5]. The relationship between IPAH and gender, bone morphogenetic protein receptor type 2 mutation and family history has been observed [8]. Several associated germline gene mutations and an autosomal dominant pattern of inheritance have been established in hPAH [9]. These mutations are seen in BMPRII [10], ALK1 [11], SMAD9 and ENG1 [12]. Unlike mutations in ALK 1, SMAD9, and ENG, which occurs in only 5% of the hPAP population [13], about 70% of hPAH patients [9], and a (10–40% of patients with iPAH [14], possess mutations in the BMPRII gene. RNA sequencing and microarrays analysis facilitates the fast detection of genes, pathways involved in remodelling process, or groups of co-regulated genes [15]. These technologies have been used in the broad and unbiased investigation of the differentially expressed genes patterns in PAH [15]. Gene expression studies have been used in identifying genes and pathways previously not associated with PAH pathogenesis, detection of novel biomarkers, identifying high-risk PAH patients, and determining the effect of medication on disease progression [15]. A previous meta-analysis of seven different studies was carried out to determine the similarity of data between these studies and possible determination of transcriptomic signature [16]. Due to change in lifestyle, PAH is fast becoming a major concern; therefore in this present study, GSE131793 data were retrieved from the Gene Expression Omnibus (GEO) database to explore genes overexpressed and underexpressed in PAH relative to non-PAH individuals. The aim of this study is therefore to identify DEGs that can be used as diagnostic markers and also serve as molecular targets.

1.1. Methodology

1.1.1. Expression dataset

Expression data having accession number GSE131793 was retrieved from the Gene Expression Omnibus (GEO) database [17]. In this study, peripheral blood mononuclear cells (PBMCs) were isolated from whole blood using BD Vacutainer CPT cell preparation tubes. Total RNA was extracted from the PBMCs, the quality of the total RNA was accessed with the aid of RNA 6000 Nano LabChip (Agilent 2100 Bioanalyzer; Santa Clara, CA). Gene expression was carried out using Affymetrix GeneChip Human Gene 1.0 ST Array. There are 10 healthy control and 10 PAH samples (5 IPAH and 5APAH). There are 13 females and 7 males whose PBMC was taken through qtPCR. The minimum age was 30 while the maximum age was 77. Only two of these patients were vasoresponsive.

1.2. Dataset processing and differential expression analysis

In the identification of differentially expressed genes between PAH and non-PAH patients, we used GEO2R [17]. GEO2R is an interactive web tool that aids in the comparison between two or more groups of samples in a GEO series. GEO2R performs analysis by using GEOquery [17] and limma packages from Bioconductor. GEOquery parses the expression data into the R data structures and limma performs statistical tests to identify DEGs.

1.3. Pathway enrichment analysis, Gene Ontology, and PPI network analysis

Gene Ontology is a method used in the categorization of gene expression attributes [18]. Protein ANalysis THrough Evolutionary Relationships (PANTHER) [23] is a straightforward visualization tool that is employed by researchers in the analysis of the biological functions of gene sets [19]. The PANTHER Classification System and analysis tool was therefore employed to classify the DEGs according to biological process, protein class, pathway enrichment, and molecular function, to ascertain their overrepresentation [19]. Search Tool for the Retrieval of Interacting Genes (STRING) version 11.0, covering 24584628 proteins from 5090 species was employed to determine Protein-protein Interaction (PPIs). The interactions retrieved in STRING had a confidence score. During the prediction, we used experimentally validated interactions possessing a confidence score of 0.4 and maximum additional interactor set at 0 to develop the protein-protein interaction network with the aid of Cytoscape software version 3.8.0 [20]. String Enrichment plugin in Cytoscape was used to retrieve the functional enrichment. The statistically significant difference was set at $P < 0.05$.

2. Results

2.1. DEGs in PAH samples and normal samples

Analysis of the data obtained from GEO was carried out using the limma package [21]. We set our P-value at 0.05 and Log2 fold change (FC) > 1.0 was taken as the cut-off values. Our analysis was based on the top 250 genes being differentially expressed between the two conditions. False Detection Rate (FDR) of Benjamini and Hochberg was employed to adjust p-values for comparison between the PAH samples and normal control. The top 15 up-regulated and the top 15 down-regulated genes are represented in Table 1. Furthermore, subsequent analysis was carried out using these top 15 up-regulated and down-regulated genes (see Table 2).

2.2. DEGs and their functional classification

Functional classification categorizes genes based on the biological process, protein class, and molecular function they are involved in. We explored the classification of these dysregulated genes based on these ontologies. In the ontology of molecular function, the result indicated that the up-regulated genes are involved in binding activity, catalytic activity, and transporter activity, while the down-regulated genes are implicated in binding, catalytic activity, transcription regulator activity, and molecular function regulator Fig. 1A. Biological processes are described as processes paramount to the proper functioning of an organism as well as their ability to relate with their immediate

Table 1
Top 20 up-regulated and down-regulated genes differentially expressed between PAH and non-PAH samples.

Differentially Expressed Genes	
Up-Regulated Genes	Down-Regulated Genes
HP	FAT4
ITGB3	RPL36AP40
ITGA2B	DPP6
OSBP2	LOC730098///CCL27
MFSD2B	CNFN
VSTM1	BMP1
LANCL3	WDR59
RNASE2	TBC1D10A
CMTM5	CARD8
F13A1	LOC729080///GCSH
HIST1H3H	SDHD
CD151	MED17
CALD1	TSKU
MYL9	PXDN
PKHD1	PAX6
DMTN	SNORD116-28
LCN2	ZNF816
ABLIM3	FGF18
SMOX	ARNT2
MTURN	MUC17

Table 2
Gene-drug interaction of some of the DEGs.

Gene	Drug
HP	Streptozotocin, Estradiol, Pyridoxine
ITGB3	Tirofiban, Eptifibatide, Cilengitide
ITGA2B	Tirofiban, Aspirin, Enoxaparin
RNASE2	CHEMBL1161861, CHEMBL574817
F13A1	Propylene glycol
LCN2	2,3-Dihydroxybenzoic acid
SMOX	Azaserine, Spermine
SDHD	Hexachlorophene

environment. The regulation of the biological process is mediated by protein modification, gene expression, substrate molecule, and interaction with protein. In this ontology, the up-regulated genes are involved in biogenesis, cellular process, localization, biological regulation, response to stimulus, signaling, biological adhesion, locomotion, multicellular organismal process, and metabolic process. Similarly, the down-regulated genes were involved in 8 of these processes in Fig. 1B. In the classification of the genes in ontology of protein class, the genes are represented in a broad spectrum of classification that mediate various biological and molecular processes. The up-regulated genes are classified as carrier protein, immunity protein, cytoskeletal protein, metabolite interconversion enzyme, cell adhesion molecule, protein modifying enzyme, and membrane traffic protein. The down-regulated genes are

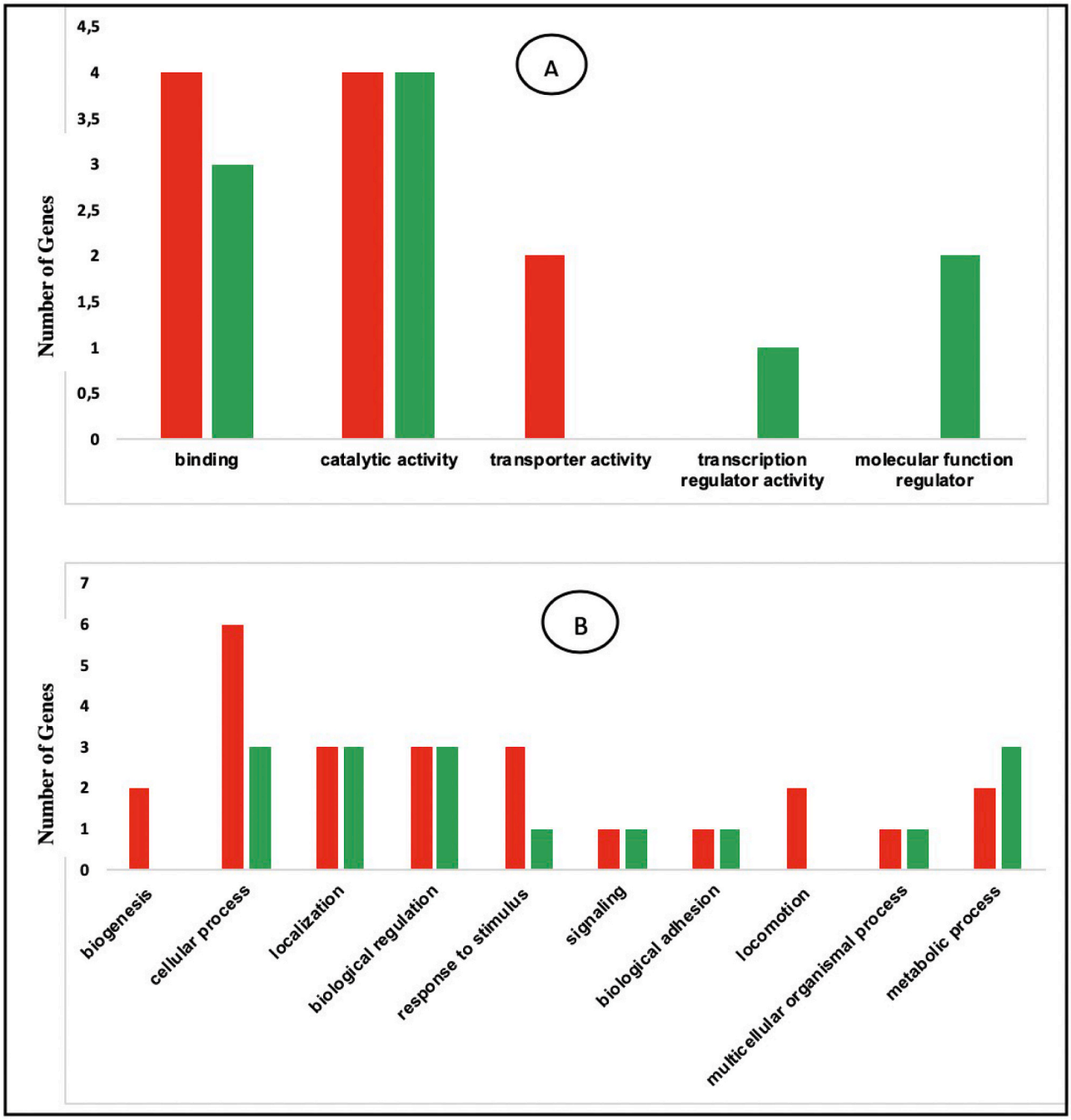

Fig. 1. Functional classification of DEGs in non-PAH and PAH samples based on PANTHER molecular function (**A**) and biological process (**B**). Up-regulated genes represented in (red) and down-regulated genes (green). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

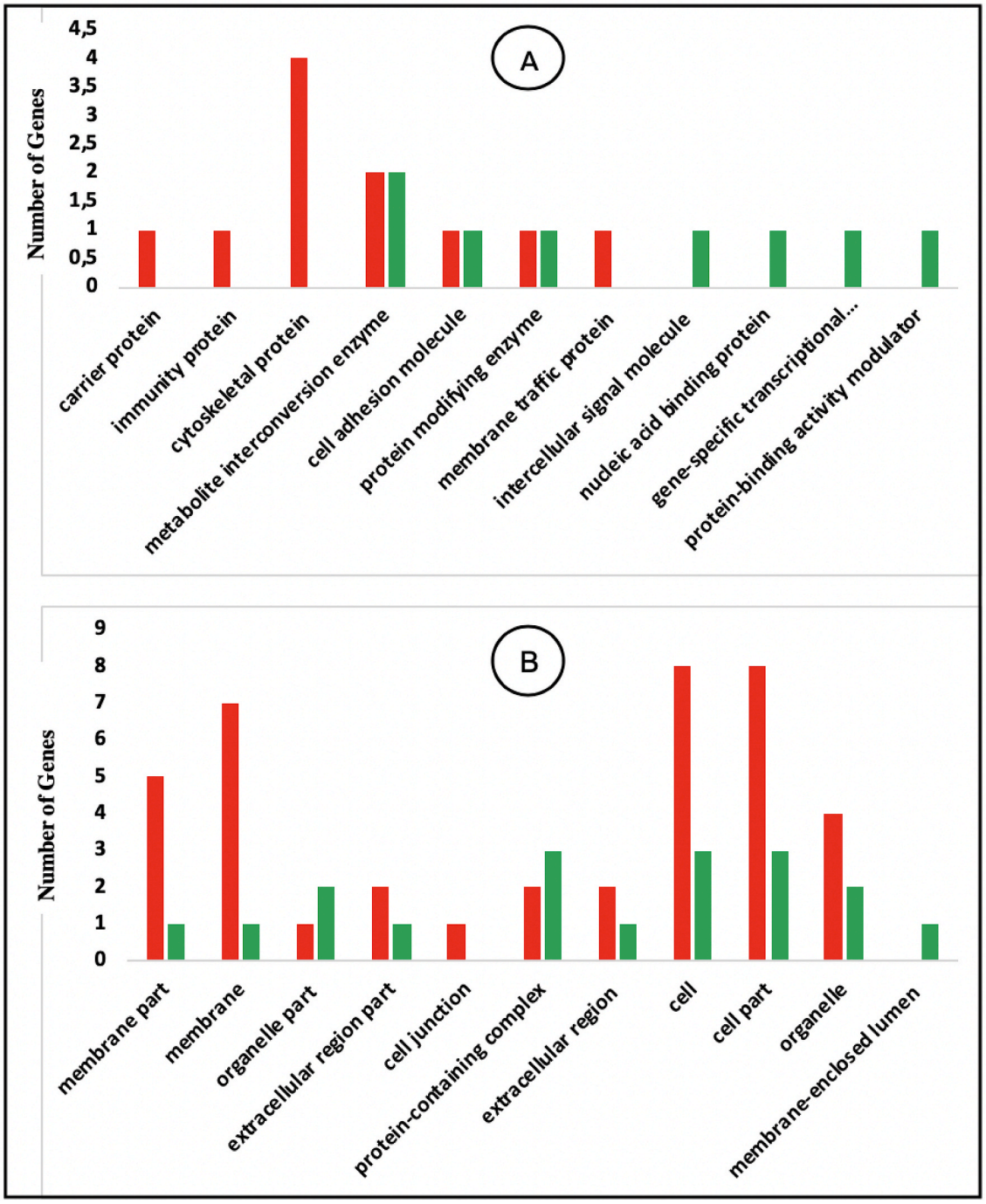

Fig. 2. Functional classification of DEGs in non-PAH and PAH samples based on PANTHER protein class (**A**) and cellular component (**B**). Up-regulated genes represented in (red) and down-regulated genes (green). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

classified as metabolite interconversion enzyme, cell adhesion molecule, protein modifying enzyme, intercellular signal molecule, nucleic acid-binding protein, gene-specific transcriptional regulator, and protein-binding activity modulator. Cellular component ontology outlines macromolecular complexes and subcellular structures; it is mostly used in the annotation of cellular locations of proteins. Cellular components represented among these genes are membrane part, extracellular region, protein-containing complex, membrane-enclosed lumen, etc (see Fig. 2).

2.3. Aberrant pathways identification and analysis of DEGs by PPI network

Pathway enrichment analysis enables researchers to have mechanistic insight into the gene list. To unravel aberrant pathways in PAH samples, pathway enrichment analysis was carried out with the aid of PANTHER. Thirty-six pathways having a P-value not greater than 0.05 were enriched in the down-regulated genes. The four-top pathway with the highest number of genes are Integrin signaling pathway (5 genes), Alzheimer's diseases-presenilin pathway (4 genes), Angiogenesis (4 genes), and Blood coagulation (4 genes) (Fig. 3). The up-regulated DEGs

were enriched in Inflammation mediated by chemokine and cytokine signaling pathway.

PPI network analysis is crucial in the understanding of the biological responses of cells to infection. Based on the interactions result from STRING, we built a protein-protein interaction network using Cytoscape [20]. A total of 20 nodes and 39 edges were identified in the up-regulated gene network, with a network diameter of 3, clustering coefficient of 0.382, and network density of 0.103. For the down-regulated genes, 25 nodes and 90 edges were identified with a network diameter of 1, clustering coefficient of 0.460, and network density of 0.150 (Fig. 4). The top 5 hub genes with the highest node degrees in the up-regulated genes are were FN1, ITGAV, SHC1, HIST1H4D and HIST1H3H. MED28, MED17, MED10, MED20, and NPRL2 were the top 5 hub genes identified in the down-regulated genes based on the number of nodes. Furthermore, these genes were investigated for betweenness centrality, MNC centrality, stress centrality, and degree centrality.

2.4. Gene products as drug targets

Selection of the right drug target is an important step in the development of novel drugs; the discovery of pharmacological targets is often hampered by limited information regarding the pathogenesis of the disease. Molecular modelling and computer-aided drug design have been employed in facilitating the drug discovery and design pipeline. We used the drug-gene interaction database (DGIdb) [22]. Some of the genes however do not have a direct drug target.

3. Discussion

The description of the genetic basis of PAH is through familial clustering, however; in the actual sense, it was the discovery of mutant BMPR2 protein that propagated the understanding of genetic susceptibility in PH [23]. Aside from this, different mutations have been implicated in the pathogenesis of PAH, for instance, the mutation in ALK1, Eng, and Smad9 [24]. Gene expression profiling has been employed in the identification of distinct mutations and potential genes present in PAH. Gene expression profiling has also been applied in the identification of genes or biomarker signature present in a cohort of patients [25]. We investigated the differentially expressed genes between normal control and PAH patients. Haptoglobin (HP) is a protein with high anti-oxidative property; due to its ability to bind to haemoglobin it prevents oxidation mediated by heme-iron. Dahan et al. hypothesized that Hp polymorphism may be responsible for the onset of PAH [26]. Furthermore, in a different study by Nakamura et al., serum HP was found to be expressed in PAH patients. This decreased HP level is suggested to be a reflection of microangiopathy and subclinical haemolysis [27]. The second upregulated gene ITGB3 (Integrin beta-3) is a protein found on thrombocytes [28].

4. Conclusion

Microarray analysis was employed to determine the DEGs and microRNAs; the analysis revealed that ITGB3 was up-regulated in PAH. Carrying studies of this nature is mostly characterized by some limitations, such as small sample size. Data retrieval from publicly available databases such as GEO could be associated with using data that are not properly normalized; some studies may contain data which were never intended to be compared to each other, and some do not have the replicate dataset for adequate statistical analysis. Despite various significant limitations, including the relatively small sample size used in this study, and lack of an independent validation cohort to support the dataset used, we have identified some genes that could be targeted in the prevention, management, and treatment of PAH.

Ethical statement

The Ethical Statement should include information where authors disclose any financial and personal relationships with other people or organizations that could inappropriately influence (bias) their work. Examples of potential competing interests include employment, consultancies, stock ownership, honoraria, paid expert testimony, patent applications/registrations, and grants or other funding

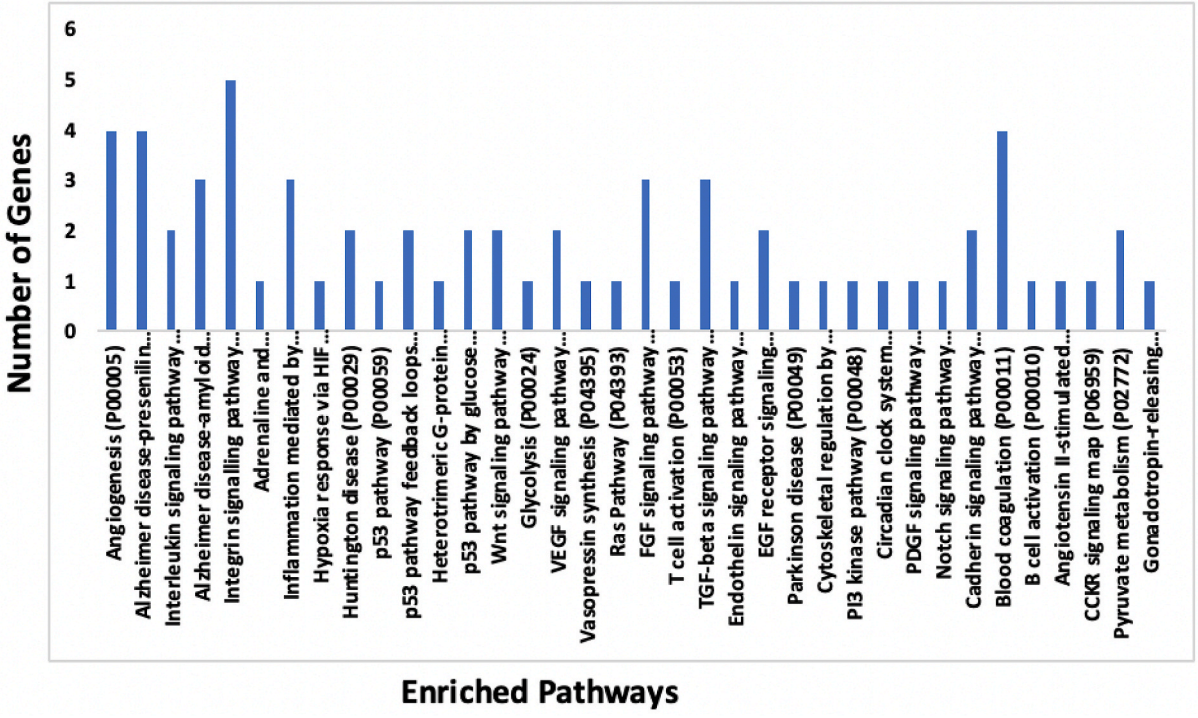

Fig. 3. Pathway enrichment analysis of DEGs in non-PAH and PAH samples.

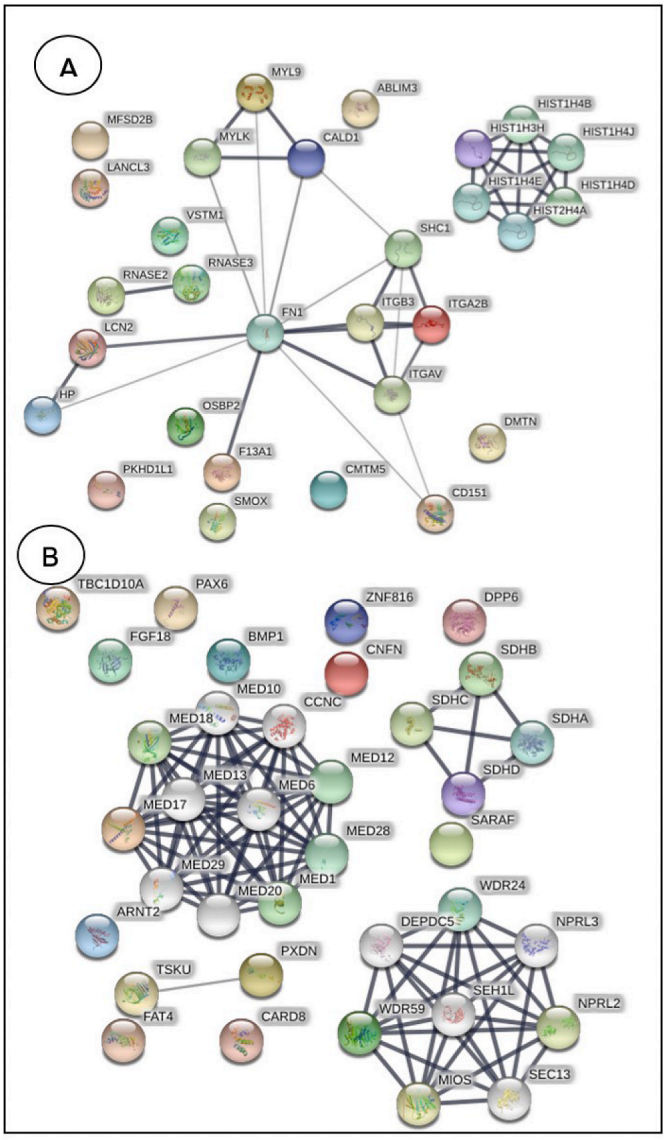

Fig. 4. Network analysis of the Differentially Expressed Genes based on betweenness centrality.

Funding

SF and ON were supported by H3Africa Bioinformatics Network (H3ABioNet). H3ABioNet was supported by the National Human Genome Research Institute (NHGRI) and the Office of The Director (OD), National Institutes of Health under award number U41HG006941. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Declaration of competing interest

Please state any conflicts of interest.

Acknowledgement

We have no one to acknowledge

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.imu.2020.100380.

References

[1] Hoeper MM, et al. Definitions and diagnosis of pulmonary hypertension. J Am Coll Cardiol 2013;62(25).
[2] Tuder RM, Stacher E, Robinson J, Kumar R, Graham BB. Pathology of pulmonary hypertension. Clin Chest Med Dec-2013;34(4):639–50. Clin Chest Med.
[3] Rabinovitch M. Molecular pathogenesis of pulmonary arterial hypertension. J Clin Invest Dec-2012;122(12):4306–13. J Clin Invest.
[4] Simonneau G, et al. Clinical classification of pulmonary hypertension. J Am Coll Cardiol 2004;43(12). S5–S12.
[5] Sajjadi S, Hejazi S, Azami G, Akbari M, Akbari F. Pulmonary arterial hypertension: a review article. Rev Clin Med Feb. 2018;5(1):1–7.
[6] Keshavarz A, Kadry H, Alobaida A, Ahsan F. Newer approaches and novel drugs for inhalational therapy for pulmonary arterial hypertension. Expet Opin Drug Deliv Feb. 2020:1–23.
[7] Pesto S, Begic Z, Prevljak S, Pecar E, Kukavica N, Begic E. "Pulmonary hypertension - new trends of diagnostic and therapy," *Medical archives (sarajevo, Bosnia and herzegovina)*70. The Academy of Medical Sciences of Bosnia and Herzegovina; Jul-2016. p. 303–7. 4.
[8] Badesch DB, et al. Pulmonary arterial hypertension: baseline characteristics from the REVEAL registry. Chest Feb. 2010;137(2):376–87.
[9] Machado RD, et al. Genetics and genomics of pulmonary arterial hypertension. J Am Coll Cardiol Jun-2009;54(1).
[10] Lane KB, et al. Heterozygous germline mutations in BMPR2, encoding a TGF-β receptor, cause familial primary pulmonary hypertension. Nat Genet Sep. 2000;26(1):81–4.
[11] Harrison RE, et al. Molecular and functional analysis identifies ALK-1 as the predominant cause of pulmonary hypertension related to hereditary haemorrhagic telangiectasia. J Med Genet Dec. 2003;40(12):865–71.
[12] Chaouat A, et al. Endoglin germline mutation in a patient with hereditary haemorrhagic telangiectasia and dexfenfluramine associated pulmonary arterial hypertension. Thorax May 2004;59(5):446–8.
[13] Lan N, Massam B, Kulkarni S, Lang C. Pulmonary arterial hypertension: pathophysiology and treatment. Diseases May 2018;6(2):38.
[14] Aldred MA, et al. BMPR2 gene rearrangements account for a significant proportion of mutations in familial and idiopathic pulmonary arterial hypertension. Hum Mutat 2006;27(2):212–3.
[15] Hoffmann J, Wilhelm J, Olschewski A, Kwapiszewska G. Microarray analysis in pulmonary hypertension. Eur Respir J 2016;48(1):229–41.
[16] Elinoff JM, et al. Meta-analysis of blood genome-wide expression profiling studies in pulmonary arterial hypertension. Am J Physiol Lung Cell Mol Physiol Jan. 2020;318(1):L98–111.
[17] Barrett T, et al. "NCBI GEO : archive for functional genomics data sets — update 2013;41:991–5. November 2012.
[18] Gene T, Consortium O. "The gene ontology Resource : 20 years and still GOing strong, vol. 47; 2019. p. 330–8. November 2018.
[19] Mi H, Muruganujan A, Thomas PD, Panther. Modeling the evolution of gene function , and other gene attributes , in the context of phylogenetic trees 2013;41:377–86. November 2012, 2013.
[20] Shannon Paul, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13(22):6.
[21] Smyth GK, Ritchie M, Thorne N. "Linear models for microarray data user ' s guide. Bioinformatics 2011;20:3705–6.
[22] Cotto KC, et al. DGIdb 3.0: a redesign and expansion of the drug-gene interaction database. Nucleic Acids Res 2018;46(D1):D1068–73.
[23] Newman JH, et al. Mutation in the gene for bone morphogenetic protein receptor II as a cause of primary pulmonary hypertension in a large kindred. N Engl J Med 2001;345(5):319–24.
[24] Cogan JD, et al. High frequency of BMPR2 exonic deletions/duplications in familial pulmonary arterial hypertension. Am J Respir Crit Care Med 2006;174(5):590–8.
[25] Fraidenburg DR, Machado RF. A review of transcriptome analysis in pulmonary vascular diseases. Methods Mol Biol 2018;1783:259–77. Mc 719.
[26] Dahan I, et al. Interaction between the haptoglobin 2 phenotype and diabetes mellitus on systolic pulmonary arterial pressure and nitric oxide bioavailability in hemodialysis patients. J. Diabetes Res. 2015:7–12. 2015.
[27] Nakamura H, et al. Decreased haptoglobin levels inversely correlated with pulmonary artery pressure in patients with pulmonary arterial hypertension. Med. (United States) 2017;96(43).
[28] Heemskerk JWM, Mattheij NJA, Cosemans JMEM. Platelet-based coagulation: different populations, different functions. J Thromb Haemostasis 2013;11(1):2–16.