

Bioinformatics enriched analysis for identification of novel candidate genes and pathways for cystic fibrosis

Aurobinda Das^{1*}, Suresh Kumar Panda¹, Debakanta Sahoo²

¹Department of Zoology, Centurion University of Technology and Management, Odisha, India

¹Department of Ecology and Environmental Sciences, Pondicherry University, Puducherry, India

²Department of Zoology, GIET University, Rayagada, Odisha, India

*Corresponding author: aurobindadas1@gmail.com

To Cite this Article

Aurobinda Das, Suresh Kumar Panda, Debakanta Sahoo, “*Bioinformatics enriched analysis for identification of novel candidate genes and pathways for cystic fibrosis*”, *Journal of Science and Technology*, Vol. 06, Issue 05, Sep-Oct 2021, pp16-32.

Article Info

Received: 11.09.21

Revised: 18.09.21

Accepted: 21.09.2021

Published: 25.09.2021

Abstract

Cystic Fibrosis, an inherited disorder, is primarily found in white American society. It arises due to alteration in the CFTR gene. Several pieces of evidence insight oxidative stress as one of the primary reasons behind this disorder. Thus, with the aim of finding the physiopathology of this disorder and its association with oxidative stress, a comparative transcriptomic study of CF with control was performed. A transcriptomic dataset (GSE ID: 39843) was accessed through Gene Expression Omnibus. Differentially Expressed Genes were enriched, and novel DEGs were sorted out. A protein-protein interaction network was constructed, followed by Gene Ontology and KEGG pathway enrichment. Our analysis revealed a total of 704 DEGs with 556 up-regulated and 148 downregulated genes. Comparison with the reported gene list (obtained through OMIM and Gene Cards) lead to the identification of 156 novel DEGs (125 up-regulated and 31 downregulated). PPI network constructed for the respective DEGs consists of 233 nodes and 1304 edges. POLR2D, CUL1, RAN, MANF, LNX1, GMNN, and STOM were identified as hub nodes through networking. GO and KEGG enrichment study identifies some major significant processes and pathways like Pyrimidine metabolism, Glutathione metabolism, Pentose phosphate pathway, Ubiquitin mediated proteolysis. ABC transporters up-regulated DEGs while cytokine receptor, Cell adhesion molecules, tight junction, leukocyte transendothelial for down-regulated DEGs. POLR2D, CUL1, RAN, MANF, LNX1, GMNN, and STOM genes may represent novel candidate biomarkers related to CF disease, as also revealed through our present findings. The current work reveals a new outlook to detect potential targets behind the cause of CF occurrence.

Abstract

Background: Cystic Fibrosis, an inherited disorder, is mainly found in white American society. It arises due to alteration in the CFTR gene. Several pieces of evidence insight oxidative stress as one of the primary reasons behind this disorder. Thus, with the aim of finding the physiopathology of this disorder and its association with oxidative stress, a comparative transcriptomic study of CF with control was performed.

Methods: A transcriptomic dataset (GSE ID: 39843) was accessed through Gene Expression Omnibus. Differentially Expressed Genes were enriched, and novel DEGs were sorted out. A protein-protein interaction network was constructed, followed by Gene Ontology and KEGG pathway enrichment.

Results: Our analysis revealed a total of 704 DEGs with 556 up-regulated and 148 downregulated genes. Comparison with the reported gene list (obtained through OMIM and Gene Cards) lead to the identification of 156 novel DEGs (125 up-regulated and 31 downregulated). PPI network constructed for the respective DEGs consists of 233 nodes and 1304 edges. POLR2D, CUL1, RAN, MANF, LNX1, GMNN, and STOM were identified

as hub nodes through networking. GO and KEGG enrichment study identifies some major significant processes and pathways like Pyrimidine metabolism, Glutathione metabolism, Pentose phosphate pathway, Ubiquitin mediated proteolysis, and ABC-transporters for up-regulated DEGs while cytokine receptor, Cell adhesion molecules, tight junction, leukocyte transendothelial for down-regulated DEGs.

Conclusions:

POLR2D,CUL1,RAN,MANF,LNX1,GMNN, and STOM genes may represent novel candidate biomarkers related to CF disease, as also revealed through our present findings. The current work reveals a new outlook to detect potential targets behind the cause of CF occurrence.

INTRODUCTION

Cystic Fibrosis (CF), an autosomal recessive disorder, is a disease associated mainly with the respiratory system. It is commonly found in Caucasians [1]. It occurs due to mutation in the cystic fibrosis transmembrane conductance regulator (CFTR) gene. CFTR is situated on the long arm of chromosome 7 and encodes a transmembrane protein (CFTR), an epithelial ion channel that plays an essential role in regulating epithelial ion and water transport and fluid homeostasis [2]. Moreover, being a lethal genetic disease, it is related to shortening (40-50 years) of expected life span[3]. Symptoms like respiratory and pulmonary infection are often at the forefront of the clinical scenario and cause morbidity [4]. Also, earlier steatorrhea persists during this disorder and is dominated upon the beginning of digestive problems [5].

Many facts reveal that oxidative stress is a characteristic of CF [6]and may involve in Nitric Oxide bioavailability reduction and followed by endothelial abnormalities [7]. The variation between the production of free radicals and radicals neutralization via antioxidants occurs due to constant, elevated immune stimulation [8] and, together with diet deficiency and exogenous antioxidants malabsorption [9]. Though antioxidants administration orally has been performed to decrease oxidative stress and vascular function improvement temporarily was found [10];however, the oxidative stress role in CF patients with vascular abnormalities is still undetermined [6].

Thus, to find the physiopathology of this disorder and its association with oxidative stress, a comparative study of CF and non CF was performed. Since fewer genes have been reported with CF diseases' pathogenicity, our aim was to find more novel genes and reveal the overall molecular mechanism via in silico method using various bioinformatics tools. The main aim behind the current study is screening potent candidate genes of CF disorder. In this study, the gene expression profile (ID: GSE39843) analysis was performed to find differential gene expressions (DEGs). The Protein-Protein Interactions (PPI) network based on the combined scores was constructed. Moreover, a Functional and gene enrichment-based study was also carried out.

2 Materials and methods

2.1 Samples retrieval

Raw gene expression profile (GSE ID GSE39843) was acquired from the GEO NCBI database. The raw dataset which was retrieved for the expression profiling was based on CF air duct cell lines provided with oxidative stress condition. The array profile consists of twelve samples and one platform ([GPL570](#)[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array) [11]. Several in silico tools were used for the present work to find the DEGs of CF samples and downstream processing of DEGs.

2.2 Pre-processing of samples

Using Benjamini and Hochberg algorithm, pre-processing of raw expression profile was performed to compute the adjusted p-value. Also, eluding force normalization and Limma precision weights log transformation was applied to the expression profile[12]. The significant cutoff level was adjusted to 0.05. Probe level conversion to gene-level symbols was also accomplished using GEO2R.

2.3 DEGs retrieval

The retrieval of DEGs from the present dataset was carried out depending on the parameters with adjustable p values <0.05 and log fold change values >0.1 for overexpressed and <0.1 for under-expressed genes GEO2R. Visualization of the selected DEGs, highlighting overexpressed and underexpressed genes via volcano plot, log₂ fold variation versus log₂ average gene expression values through mean difference (MD) graph, and selected genes that coincide among various available contrasts through Venn Diagram was implemented. Construction of all these graphs through GEO2R with the available LIMMA package[13].

2.4 PPI network and subnetwork formation

Through the online tool STRING v 10.5 (<https://www.string-db.org/>), functional relation among proteins were found [14]. This software computes the combined scores within gene pairs for protein-protein interactions. So for the current study, DEGs selected lists were provided to the software, and a combined score > 0.4 was applied as a parameter for this analysis. Then Cytoscape v 3.2.1 (<http://www.cytoscape.org>) and in silico software package were used to construct numerous network and sub-networks [15]. Degree and edge betweenness criteria were initiated for the construction of networks.

2.5 GO and KEGG pathway interpretation

DAVID (Database for Annotation, Visualisation, And Integrated Discovery) (<https://david.abcc.ncifcrf.gov>) server involves an extensive list of genes dataset to merge reasonable sets of functional annotation [16]. Gene Ontology (GO) analysis reveals cellular component (CC), molecular function (MF), and biological process (BP) through DAVID v 6.8. Depending on the hypergeometric distribution, DAVID constructs complete gene datasets with related or nearby functions.

3. RESULTS

3.1 DEGs enrichment for CF.

Using GEO2R, normalization of gene expression dataset was performed. Box plot shows that all data have median centred value (Fig. 1A). Uniform Manifold Approximation and Projection shows the distribution of the samples and the number of nearest neighbors used in the calculation i.e. nbrs=5 (Fig. 1B). DEGs screening identified a total of 704 DEGs with 556 up-regulated and 148 downregulated genes were found (Supplementary file 1). For determining DEGs, p-value < 0.05 and $|\log_2 FC| > 0.1$ were used as parameters. On the basis of mean values of average gene expression, DEGs were screened. Furthermore, comparisons of DEGs with gene list obtained from OMIM and Gene Cards leads to the identification of a total of 156 reported DEGs (125 up-regulated and 31 down-regulated) while 548 novel DEGs (117 up-regulated and 431 down-regulated) (Fig. 2A, 2B). Construction of Volcano plot and MD graph of determined DEGs with over and under-expressed genes were visualized (Fig. 2C, 2D).

3.2 Construction of Principal Component and Heatmap plot.

Analysis of Principal Component for CF through scatter graph represents a total 58.9% variance on x-axis equivalent to principal component 1, and 11.2% variance on the y axis equivalent to principal component 2 (Fig. 3A). Representation of heat map through numerical differences for DEGs through data matrix with the colorful prototype was also produced (Fig. 3B).

3.3. Creation of PPI network

All DEGs were used to create a protein-protein interaction network (Fig. 4A). A separate PPI network was constructed for DEGs with the combined score greater than 0.9 generated by STRING (Fig. 4B). PPI network constructed for the respective DEGs consists of 233 nodes and 1304 edges. Blue and Red colors constitute the down and up-regulated genes of the generated DEGs. Degree and edge betweenness were implemented for hub node creation. Genes involved in forming hub nodes were UBC (Ubiquitin C), POLR2D (RNA Polymerase II Subunit D), CUL1 (Cullin 1), HSP90 (Heat Shock Protein 90), RAN (Ras-related nuclear protein), PLAUR (Plasminogen Activator Urokinase Receptor), MANF (Mesencephalic Astrocyte Derived Neurotrophic Factor) as up-regulated genes while NOTCH1 (NOTCH Receptor 1), STAT3 (Signal Transducer And Activator Of Transcription 3), LNX1 (Ligand Of Numb-Protein X 1), GMNN (Geminin DNA Replication Inhibitor), ADAM10 (ADAM Metallopeptidase Domain 10) and, STOM (Stomatin) as downregulated genes association of these hub nodes with genes of recognized disease and being novel were considered as potential candidate genes for CF.

Some DEGs established five separate networks from the major network, which were considered sub-networks (Fig. 4C). The first sub-network consists of 6 nodes and 12 edges (Fig. 4C1). The second subnetwork consists of 4 nodes and six edges (Fig. 4C2). The third subnetwork consists of 4 nodes and eight edges (Fig. 4C3). The fourth subnetwork consists of 9 nodes and 72 edges (Fig. 4C4). Furthermore, the fifth subnetwork consists of 5 nodes and 20 edges (Fig. 4C5). Depending on the degree and combined scores, a total of ten DEGs were screened out as novel biomarkers.

3.4. Construction of GO Enrichment and KEGG pathways.

In the network, all the DEGs were intensified for their processes, component, and functions. The present study reveals that most overexpressed DEGs with significant values ($p < 0.05$) regulates the processes of induction of apoptosis, cell cycle, response to oxidative stress, negative regulation of cellular protein, positive regulation of protein ubiquitination, positive regulation of ligase activity, cell redox homeostasis, and proteasomal protein catabolic process (Fig. 5A). Also, under-expressed DEGs with significant values ($p < 0.05$) regulates positive regulation of epidermis development, positive regulation of apoptosis, somatic stem cell division, regulation of cell growth, cell junction, negative regulation of cell adhesion, response to hormone stimulus, regulation of secretion, regulation of cell growth, calcium-independent cell cell adhesion processes (Fig. 5B).

Through KEGG enrichment study, it was found that Pyrimidine metabolism, Glutathione metabolism, Pentose phosphate pathway, Ubiquitin mediated proteolysis, ABC transporters for up-regulated genes and cytokine receptor, Cell adhesion molecules, tight junction, leukocyte transendothelial for down-regulated genes were the main pathways for this DEGs enrichment (Fig 5C).

Discussion

Cystic Fibrosis is a progressive genetic disorder that leads to frequent infections in the lungs. Despite the increasing CF occurrence, there is still a lack of suitable therapeutic targets [5]. In the current study, we have applied numerous bioinformatics tools to find CF's mechanism at the molecular level. In the present in silico approach, UBC, POLR2D, CUL1, HSP90, RAN, PLAUR, MANF, NOTCH1, STAT3, LNX1, GMNN, ADAM10, and STOM were found as novel DEGs in the samples of CF disorder with oxidative stress.

The UBC gene depicts the Ubiquitin gene. It encodes proteins that are involved in DNA repair and cell cycle regulation. These protein impairments lead to CF. U.Griesenbach et al. revealed through their studies the crucial role of the UBC gene in chronic lung disorders, cancer, and CF and related genetic therapy [17]. Since few studies disclosed UBC gene association with CF, thus provides a potent target for CF.

POLR2D genes are responsible for RNA pol II formation that induces Eukaryotic mRNA synthesis. R Marcotte et al. suggested through their work mutational POLR2D role in breast cancer progression [18]. However, to date, the POLR2D gene role was not found in CF, but its role in the transcription process, as shown in our present work, may prove to be a novel candidate for CF treatment.

CUL1 gene mediates its role with ubiquitination, cell cycle progression, and an important role in the WNT signaling pathway. Ye-Fei Huang et al. through gene expression profiling CUL1 overexpression in breast carcinoma [19]. Lan Chen et al. investigated through their studies overexpression of CUL1 in the proliferation of melanoma cells [20]. Though CUL1 gene association with CF was not found until now, its association with the cell cycle process and ubiquitination revealed through our work may prove to be a novel target for CF.

Kaisheng Liu et al. revealed the involvement of the HSP90 gene with CFTR membrane and regulating phosphorylation of the AKT pathway and its crucial role in colorectal carcinoma [21]. RT Youker et al. disclosed the onset of CF due to deletion of Phe 508 in the Hsp90 chaperone [22]. Thus, Hsp90 relation with CF unveils an essential marker for therapeutics.

Ran gene-encoded protein associated with DNA synthesis, cell cycle progression, and induces microtubules nucleation. Zied Boudhraa et al. reported that Ran GTPase tumorigenic role in cancer [23]. However, the Ran gene's role in CF is still unknown, but its involvement in a mitotic cell cycle as represented in our work may serve as a novel biomarker for CF disorder.

PLAUR gene encodes a urokinase protein that promotes the formation of plasmin. S Shetty et al., through DEGs analysis, disclosed the PLAUR gene as one of the overexpressed genes of urokinase receptors in fibroblast [24]. Sheila et al. investigated through their studies PLAUR polymorphism with lung dysfunction and asthma [25]. Varrie Ogilvie et al. PLAUR gene plays a crucial role in the regulation of TGF beta in tissue repairment and fibrosis [26]. C E Stewart et al. suggested PLAUR gene polymorphism and its involvement in CF [27], which indicates a promising candidate for curing CF.

MANF gene up-regulation leads to ER stress. Y Kim et al. disclosed through their work that overexpression of the MANF gene leads to genetic skeletal disorders [28]. Apostolou A et al. revealed MANF gene dysfunction inhibits cell proliferation and ER stress leading to cellular death [29]. Although MANF gene association with CF disease was not found to date, its involvement with cell proliferation which is also proved through our work, may provide a potent key for treatment.

NOTCH1 gene encodes a transmembrane protein that regulates adjacent cell interaction and is involved in the survival and proliferation of cells. According to Biao Hu et al., NOTCH1 is involved in signaling pathways of

airways fibrosis[30]. Y C Wang et al. concluded through their work the overexpression of the NOTCH1 gene in the pulmonary duct leading to fibrosis[31]. However, some works suggested the role of the NOTCH1 gene in promoting CF, which may prove to be one of the biomarkers of CF disease.

STAT3 gene in the phosphorylated form is found involved in cellular growth and apoptotic-related pathways. M. Pedroza et al. found through their studies phosphorylated STAT3 elevated levels in lung fibroids [32]. However, D. Chakraborty et al. suggested STAT3 gene association with fibroblast formation and any change leads to tissue fibrosis [33]. AC Tang et al. indicated that STAT3 increased level leads to ER stress, causing inflammation in the airways, leading to CF [34] since the STAT3 gene association with CF was represented through some works, which may provide a clue for CF treatment.

LNX1 gene encodes a proteasomal protein involved in degradation and ubiquitination. R Park et al. revealed LNX1 gene downregulates stable p53, promoting tumorigenesis[35]. S. Baisiwala et al. indicated the role of LNX1 gene modulating NOTCH1, which leads to glioma stem cell expansion [36]. Although the LNX1 gene role in CF is still unknown, its involvement in proteasomal degradation, as also suggested through our work, can provide a novel candidate for CF disorder.

GMNN gene is found in regulation of cell cycle, especially in metaphase to the anaphase transition phase. Christelle de Renty et al. found GMNN gene role in the replication process and DNA ablation [37]. Also, P.P Kushwaha et al. suggested through their work GMNN gene association with developmental process and cancer pathobiology [38]. M R Salbat et al. investigated GMNN overexpression in pancreatic carcinoma [39]. However, CF's GMNN gene role is unknown to date, but its involvement in cell cycle regulation may prove to be a novel candidate for CF treatment.

ADAM 10 gene encodes a protein involved in proteasomal degradation and is found involved in the cell adhesion process. P R Manzine et al. revealed ADAM 10 as a crucial player in proteolytic processing and cell adhesion in Alzheimer's disease [40]. M Muller et al. found that ADAM10 protein cleaves Ephrin B2, leading to fibrosis in pulmonary tracts[41]. Thus, ADAM10's relation with CF through some studies has proven to be potential CF disease targets.

The STOM gene encodes proteins that are found integrated into the RBC membrane and regulates ion channels. M T Landi et al. proposed through their work the overexpression of STOM gene in lung carcinoma [42]. M Nikpay et al. found the STOM gene's crucial role in coronary artery diseases [43]. Although STOM gene relation with CF is unknown until now, their function in ion channel regulation, which was also indicated through our current study, may be assigned as a novel lead for CF diseases.

In conclusion, POLR2D, CUL1, RAN, MANF, LNX1, GMNN, and STOM genes may represent novel candidate biomarkers related to CF disease as these genes have been found to play an essential role in apoptosis, cell cycle, response to oxidative stress. The current work reveals a new outlook to detect potential targets behind the cause of CF occurrence and the role of oxidative stress in the progression of the disease. However, in vivo and in vitro works are needed to be carried out for validation of results.

REFERENCES

- [1] M.-L. Kleme, E. Levy, Cystic Fibrosis-Related Oxidative Stress and Intestinal Lipid Disorders, *Antioxidants & Redox Signaling*. 22 (2015) 614–631. <https://doi.org/10.1089/ars.2014.6012>.
- [2] S.A. Molina, W.R. Hunt, Cystic Fibrosis, in *Lung Epithelial Biology in the Pathogenesis of Pulmonary Disease*, Elsevier, 2017: pp. 219–249. <https://doi.org/10.1016/B978-0-12-803809-3.00012-9>.
- [3] I. Durieu, R. Nove Josserand, La mucoviscidose en 2008, *La Revue de Médecine Interne*. 29 (2008) 901–907. <https://doi.org/10.1016/j.revmed.2007.12.020>.
- [4] J.M. Pilewski, R.A. Frizzell, Role of CFTR in Airway Disease, *Physiological Reviews*. 79 (1999) S215–S255. <https://doi.org/10.1152/physrev.1999.79.1.S215>.
- [5] M.-L. Kleme, E. Levy, Cystic Fibrosis-Related Oxidative Stress and Intestinal Lipid Disorders, *Antioxidants & Redox Signaling*. 22 (2015) 614–631. <https://doi.org/10.1089/ars.2014.6012>.
- [6] M.A. Tucker, B.M. Fox, N. Seigler, P. Rodriguez-Miguel, J. Looney, J. Thomas, K.T. McKie, C. Forseen, G.W. Davison, R.A. Harris, Endothelial Dysfunction in Cystic Fibrosis: Role of Oxidative Stress, *Oxidative Medicine and Cellular Longevity*. 2019 (2019) 1–8. <https://doi.org/10.1155/2019/1629638>.
- [7] A. García-Sánchez, A.G. Miranda-Díaz, E.G., Cardona-Muñoz, The Role of Oxidative Stress in Physiopathology and Pharmacological Treatment with Pro- and Antioxidant Properties in Chronic

- Diseases, Oxidative Medicine and Cellular Longevity. 2020 (2020) 1–16. <https://doi.org/10.1155/2020/2082145>.
- [8] LG Wood, D.A. Fitzgerald, A.K. Lee, M.L. Garg, Improved antioxidant and fatty acid status of patients with cystic fibrosis after antioxidant supplementation is linked to improved lung function1–3, *The American Journal of Clinical Nutrition*. 77 (2003) 150–159. <https://doi.org/10.1093/ajcn/77.1.150>.
- [9] R.K. Brown, H. Wyatt, J.F. Price, F.J. Kelly, Pulmonary dysfunction in cystic fibrosis is associated with oxidative stress, *European Respiratory Journal*. 9 (1996) 334–339. <https://doi.org/10.1183/09031936.96.09020334>.
- [10] R.A. Harris, S.K. Nishiyama, D.W. Wray, V. Tedjasaputra, D.M. Bailey, R.S. Richardson, The effect of oral antioxidants on brachial artery flow-mediated dilation following 5 and 10 min of ischemia, *Eur J Appl Physiol*. 107 (2009) 445–453. <https://doi.org/10.1007/s00421-009-1147-x>.
- [11] G. Voisin, G.F. Bouvet, P. Legendre, A. Dagenais, C. Massé, Y. Berthiaume, Oxidative stress modulates the expression of genes involved in cell survival in $\Delta F508$ cystic fibrosis airway epithelial cells, *Physiological Genomics*. 46 (2014) 634–646. <https://doi.org/10.1152/physiolgenomics.00003.2014>.
- [12] Y. Benjamini, D. Drai, G. Elmer, N. Kafkafi, I. Golani, Controlling the false discovery rate in behavior genetics research, *Behavioural Brain Research*. 125 (2001) 279–284. [https://doi.org/10.1016/S0166-4328\(01\)00297-2](https://doi.org/10.1016/S0166-4328(01)00297-2).
- [13] M.E. Ritchie, B. Phipson, D. Wu, Y. Hu, C.W. Law, W. Shi, G.K. Smyth, limma powers differential expression analyses RNA-sequencing and microarray studies, *Nucleic Acids Research*. 43 (2015) e47–e47. <https://doi.org/10.1093/nar/gkv007>.
- [14] A. Franceschini, D. Szklarczyk, S. Frankild, M. Kuhn, M. Simonovic, A. Roth, J. Lin, P. Minguez, P. Bork, C. von Mering, L.J. Jensen, STRING v9.1: protein-protein interaction networks, with increased coverage and integration, *Nucleic Acids Research*. 41 (2012) D808–D815. <https://doi.org/10.1093/nar/gks1094>.
- [15] M. Kohl, S. Wiese, B. Warscheid, Cytoscape: Software for Visualization and Analysis of Biological Networks, in M. Hamacher, M. Eisenacher, C. Stephan (Eds.), *Data Mining in Proteomics*, Humana Press, Totowa, NJ, 2011: pp. 291–303. https://doi.org/10.1007/978-1-60761-987-1_18.
- [16] D.W. Huang, B.T. Sherman, R.A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources, *Nat Protoc*. 4 (2009) 44–57. <https://doi.org/10.1038/nprot.2008.211>.
- [17] U. Griesenbach, D.M. Geddes, EFWF Alton, Gene therapy for cystic fibrosis: an example for lung gene therapy, *Gene Ther*. 11 (2004) S43–S50. <https://doi.org/10.1038/sj.gt.3302368>.
- [18] R. Marcotte, K.R. Brown, F. Suarez, A. Sayad, K. Karamboulas, P.M. Krzyzanowski, F. Sircoulomb, M. Medrano, Y. Fedyshyn, J.L.Y. Koh, D. van Dyk, B. Fedyshyn, M. Luhova, G.C. Brito, F.J. Vizeacoumar, F.S. Vizeacoumar, A. Datti, D. Kasimer, A. Buzina, P. Mero, C. Misquitta, J. Normand, M. Haider, T. Ketela, J.L. Wrana, R. Rottapel, B.G. Neel, J. Moffat, Essential Gene Profiles in Breast, Pancreatic, and Ovarian Cancer Cells, *Cancer Discovery*. 2 (2012) 172–189. <https://doi.org/10.1158/2159-8290.CD-11-0224>.
- [19] Y.-F. Huang, Z. Zhang, M. Zhang, Y.-S. Chen, J. Song, P.-F. Hou, H.-M. Yong, J.-N. Zheng, J. Bai, CUL1 promotes breast cancer metastasis through regulating EZH2-induced, the autocrine expression of the cytokines CXCL8 and IL11, *Cell Death Dis*. 10 (2019) 2. <https://doi.org/10.1038/s41419-018-1258-6>.
- [20] L. Chen, T. Liu, Y. Tu, D. Rong, Y. Cao, Cul1 promotes melanoma cell proliferation by promoting DEPTOR degradation and enhancing cap-dependent translation, *Oncology Reports*. 35 (2016) 1049–1056. <https://doi.org/10.3892/or.2015.4442>.
- [21] K. Liu, H. Jin, Y. Guo, Y. Liu, Y. Wan, P. Zhao, Z. Zhou, J. Wang, M. Wang, C. Zou, W. Wu, Z. Cheng, Y. Dai, CFTR interacts with Hsp90 and regulates the phosphorylation of AKT and ERK1/2 in colorectal cancer cells, *FEBS Open Bio*. 9 (2019) 1119–1127. <https://doi.org/10.1002/2211-5463.12641>.
- [22] R.T. Youker, P. Walsh, T. Beilharz, T. Lithgow, J.L. Brodsky, Distinct Roles for the Hsp40 and Hsp90 Molecular Chaperones during Cystic Fibrosis Transmembrane Conductance Regulator Degradation in Yeast, *MBoC*. 15 (2004) 4787–4797. <https://doi.org/10.1091/mbc.e04-07-0584>.
- [23] Z. Boudhraa, E. Carmona, D. Provencher, A.-M. Mes-Masson, Ran GTPase: A Key Player in Tumor Progression and Metastasis, *Front. Cell Dev. Biol*. 8 (2020) 345. <https://doi.org/10.3389/fcell.2020.00345>.
- [24] S. Shetty, A. Kumar, A.R. Johnson, S. Pueblitz, D. Holiday, G. Raghun, S. Idell, Differential expression of the urokinase receptor in fibroblasts from normal and fibrotic human lungs., *Am J Respir Cell Mol Biol*. 15 (1996) 78–87. <https://doi.org/10.1165/ajrcmb.15.1.8679225>.
- [25] S.J. Barton, G.H. Koppelman, J.M. Vonk, C.A. Browning, I.M. Nolte, C.E. Stewart, S. Bainbridge, S. Mutch, M.J. Rose-Zerilli, D.S. Postma, N. Maniatis, A.P. Henry, I.P. Hall, S.T. Holgate, P. Tighe, J.W. Holloway, I. Sayers, PLAUR polymorphisms are associated with asthma, PLAUR levels, and lung

- function decline, *Journal of Allergy and Clinical Immunology*. 123 (2009) 1391-1400.e17. <https://doi.org/10.1016/j.jaci.2009.03.014>.
- [26] V. Ogilvie, M. Passmore, L. Hyndman, L. Jones, B. Stevenson, A. Wilson, H. Davidson, R.R. Kitchen, R.D. Gray, P. Shah, E.W. Alton, J.C. Davies, D.J. Porteous, A.C. Boyd, Differential global gene expression in cystic fibrosis nasal and bronchial epithelium, *Genomics*. 98 (2011) 327–336. <https://doi.org/10.1016/j.ygeno.2011.06.008>.
- [27] C.E. Stewart, I. Sayers, Characterisation of urokinase plasminogen activator receptor variants in human airway and peripheral cells, *BMC Mol Biol*. 10 (2009) 75. <https://doi.org/10.1186/1471-2199-10-75>.
- [28] Y. Kim, S.-J. Park, Y.M. Chen, Mesencephalic astrocyte-derived neurotrophic factor (MANF), a new player in endoplasmic reticulum diseases: structure, biology, and therapeutic roles, *Translational Research*. 188 (2017) 1–9. <https://doi.org/10.1016/j.trsl.2017.06.010>.
- [29] A. Apostolou, Y. Shen, Y. Liang, J. Luo, S. Fang, Armet, a UPR-upregulated protein, inhibits cell proliferation and ER stress-induced cell death, *Experimental Cell Research*. 314 (2008) 2454–2467. <https://doi.org/10.1016/j.yexcr.2008.05.001>.
- [30] B. Hu, S.H. Phan, Notch in fibrosis and as a target of anti-fibrotic therapy, *Pharmacological Research*. 108 (2016) 57–64. <https://doi.org/10.1016/j.phrs.2016.04.010>.
- [31] Y.-C. Wang, Q. Chen, J.-M. Luo, J. Nie, Q.-H. Meng, W. Shuai, H. Xie, J.-M. Xia, H. Wang, Notch1 promotes the pericyte-myofibroblast transition in idiopathic pulmonary fibrosis through the PDGFR/ROCK1 signal pathway, *Exp Mol Med*. 51 (2019) 1–11. <https://doi.org/10.1038/s12276-019-0228-0>.
- [32] M. Pedroza, D.J. Schneider, H. Karmouty-Quintana, J. Coote, S. Shaw, R. Corrigan, J.G. Molina, J.L. Alcorn, D. Galas, R. Gelinis, M.R. Blackburn, Interleukin-6 Contributes to Inflammation and Remodeling in a Model of Adenosine Mediated Lung Injury, *PLoS ONE*. 6 (2011) e22667. <https://doi.org/10.1371/journal.pone.0022667>.
- [33] D. Chakraborty, B. Šumová, T. Mallano, C.-W. Chen, A. Distler, C. Bergmann, I. Ludolph, R.E. Horch, K. Gelse, A. Ramming, O. Distler, G. Schett, L. Šenolt, J.H.W. Distler, Activation of STAT3 integrates common profibrotic pathways to promote fibroblast activation and tissue fibrosis, *Nat Commun*. 8 (2017) 1130. <https://doi.org/10.1038/s41467-017-01236-6>.
- [34] A.C. Tang, A. Saferali, G. He, A.J. Sandford, L.J. Strug, S.E. Turvey, Endoplasmic reticulum stress regulates chemokine production in cystic fibrosis airway cells through STAT3 modulation, *J Infect Dis*. (2016) jiw516. <https://doi.org/10.1093/infdis/jiw516>.
- [35] R. Park, H. Kim, M. Jang, D. Jo, Y.-I. Park, S. Namkoong, J.I. Lee, I.-S. Jang, J. Park, LNX1 contributes to tumor growth by down-regulating p53 stability, *The FASEB Journal*. 33 (2019) 13216–13227. <https://doi.org/10.1096/fj.201900366R>.
- [36] S. Baisiwala, R.R. Hall, M.R. Saathoff, J. M. Shireman, C. Park, S. Budhiraja, C. Goel, L. Warnke, C. Hardiman, J.Y. Wang, K. McCortney, C.M. Horbinski, A.U. Ahmed, LNX1 Modulates Notch1 Signaling to Promote Expansion of the Glioma Stem Cell Population during Temozolomide Therapy in Glioblastoma, *Cancers*. 12 (2020) 3505. <https://doi.org/10.3390/cancers12123505>.
- [37] C. de Renty, K.J. Kaneko, M.L. DePamphilis, The dual roles of geminin during trophoblast proliferation and differentiation, *Developmental Biology*. 387 (2014) 49–63. <https://doi.org/10.1016/j.ydbio.2013.12.034>.
- [38] P.P. Kushwaha, K.C. Rapalli, S. Kumar, Geminin a multi-task protein involved in cancer pathophysiology and developmental process: A review, *Biochimie*. 131 (2016) 115–127. <https://doi.org/10.1016/j.biochi.2016.09.022>.
- [39] M.R. Salabat, L.G. Melstrom, M.J. Strouch, X.-Z. Ding, B.M. Milam, M.B. Ujiki, C. Chen, J.C. Pelling, S. Rao, P.J. Grippo, T.J. McGarry, D.J. Bentrem, Geminin is overexpressed in human pancreatic cancer and downregulated by the bioflavonoid apigenin in pancreatic cancer cell lines, *Mol. Carcinog*. 47 (2008) 835–844. <https://doi.org/10.1002/mc.20441>.
- [40] P.R. Manzine, M. Ettcheto, A. Cano, O. Busquets, E. Marcello, S. Pelucchi, M. Di Luca, K. Endres, J. Olloquequi, A. Camins, M.R. Cominetti, ADAM10 in Alzheimer's disease: Pharmacological modulation by natural compounds and its role as a peripheral marker, *Biomedicine & Pharmacotherapy*. 113 (2019) 108661. <https://doi.org/10.1016/j.biopha.2019.108661>.
- [41] M. Müller, S. Wetzel, J. Köhn-Gaone, K. Chalupsky, R. Lüllmann-Rauch, R. Barikbin, J. Bergmann, B. Wöhner, O. Zbodakova, I. Leuschner, G. Martin, G. Tiegs, S. Rose-John, R. Sedlacek, J.E.E. Tirnitz-Parker, P. Saftig, D. Schmidt-Arras, A disintegrin and metalloprotease 10 (ADAM10) is a central regulator of murine liver tissue homeostasis, *Oncotarget*. 7 (2016) 17431–17441. <https://doi.org/10.18632/oncotarget.7836>.

- [42] M.T. Landi, T. Dracheva, M. Rotunno, J.D. Figueroa, H. Liu, A. Dasgupta, F.E. Mann, J. Fukuoka, M. Hames, A.W. Bergen, S.E. Murphy, P. Yang, A.C. Pesatori, D. Consonni, P.A. Bertazzi, S. Wacholder, J.H. Shih, N.E. Caporaso, J. Jen, Gene Expression Signature of Cigarette Smoking and Its Role in Lung Adenocarcinoma Development and Survival, PLoS ONE. 3 (2008) e1651. <https://doi.org/10.1371/journal.pone.0001651>.
- [43] M. Nikpay, R. McPherson, Convergence of biomarkers and risk factor trait loci of coronary artery disease at 3p21.31 and HLA region, Npj Genom. Med. 6 (2021) 12. <https://doi.org/10.1038/s41525-021-00174-z>.

Figure Legends:

Fig.1: Microarray data normalization for GSE39843. A. Box plot. It shows that all datasets have median centered values distribution of values data for the selected samples. The lines in the box are coincident, indicating that these chips have been highly normalized. **B. UMAP plot.** It shows the distribution of data.

Fig. 2: Reported and novel DEGs. A. Venn diagram: All DEGs were compared with reported gene list obtained from the Gene Cards and OMIM to reveal the novel DEGs. **B. Out of total DEGs identified, 548 genes were found to be novel while 156 were found to be reported.** **C. MD plot:** Total identified DEGs were visualised through **C. MD plot and D. Volcano plot.**

Fig. 3: Principal component analysis and Heat map of DEGs. A. PCA plot shows a scatter plot with principal component 1 (x-axis) and principal component 2 (y-axis) showing total variance of ---% and ---% in ULM. ClustVis tool was used for this. **B. Heat map** showing the average gene expression of differentially expressed genes (DEGs) between cystic fibrosis cells under stress (CF_DMNQ) and normal cells with stress (nonCF_DMNQ). The blue to orange gradation represents the gene expression values change from small to large. ClustVis tool was used to draw heat map.

Fig. 4 : Protein-Protein interaction (PPI) of differentially expressed genes. A. A PPI network showing the interaction for all DEGs were constructed. **B. DEGs with combined score > 0.9** were used to construct a separate network showing hub nodes based on degree and edge betweenness. **C. Some of the DEGs** were forming separate network from the main-network and extracted to create five sub-networks (C1, C2, C3, C4, C5). **Red Circle and Red Diamond** up-regulated genes, **Blue Circle and Blue Diamond** down-regulated genes. **Lines** the correlation between genes **Thickness of lines (edges)** is proportional to the combined score. Cytoscape v 3.2.1 was used to construct the network.

Fig. 5: Gene Ontology and KEGG Pathway analysis for DEGs in PPI network. Bar graph showing significant processes, function and cellular component enriched in diabetic mothers for **A. up-regulated genes and B. Down-regulated genes.** **C. Pathway enrichment** for DEGs lead to identification of 4 significant pathways. DAVID v 6.7 was used for annotation.

Table 1: GO enrichment for UP-regulated DEGs

Category	Term	P Value
GOTERM_BP_FAT	GO:0051351~positive regulation of ligase activity	4.48E-07
GOTERM_BP_FAT	GO:0010498~proteasomal protein catabolic process	4.95E-07
GOTERM_BP_FAT	GO:0031398~positive regulation of protein ubiquitination	1.33E-05
GOTERM_BP_FAT	GO:0045454~cell redox homeostasis	4.03E-05
GOTERM_BP_FAT	GO:0051603~proteolysis involved in cellular protein catabolic process	6.40E-05
GOTERM_BP_FAT	GO:0000278~mitotic cell cycle	3.80E-04
GOTERM_BP_FAT	GO:0032269~negative regulation of cellular protein metabolic process	0.001063
GOTERM_BP_FAT	GO:0022402~cell cycle process	0.004747
GOTERM_BP_FAT	GO:0006979~response to oxidative stress	0.012397
GOTERM_BP_FAT	GO:0007049~cell cycle	0.025419

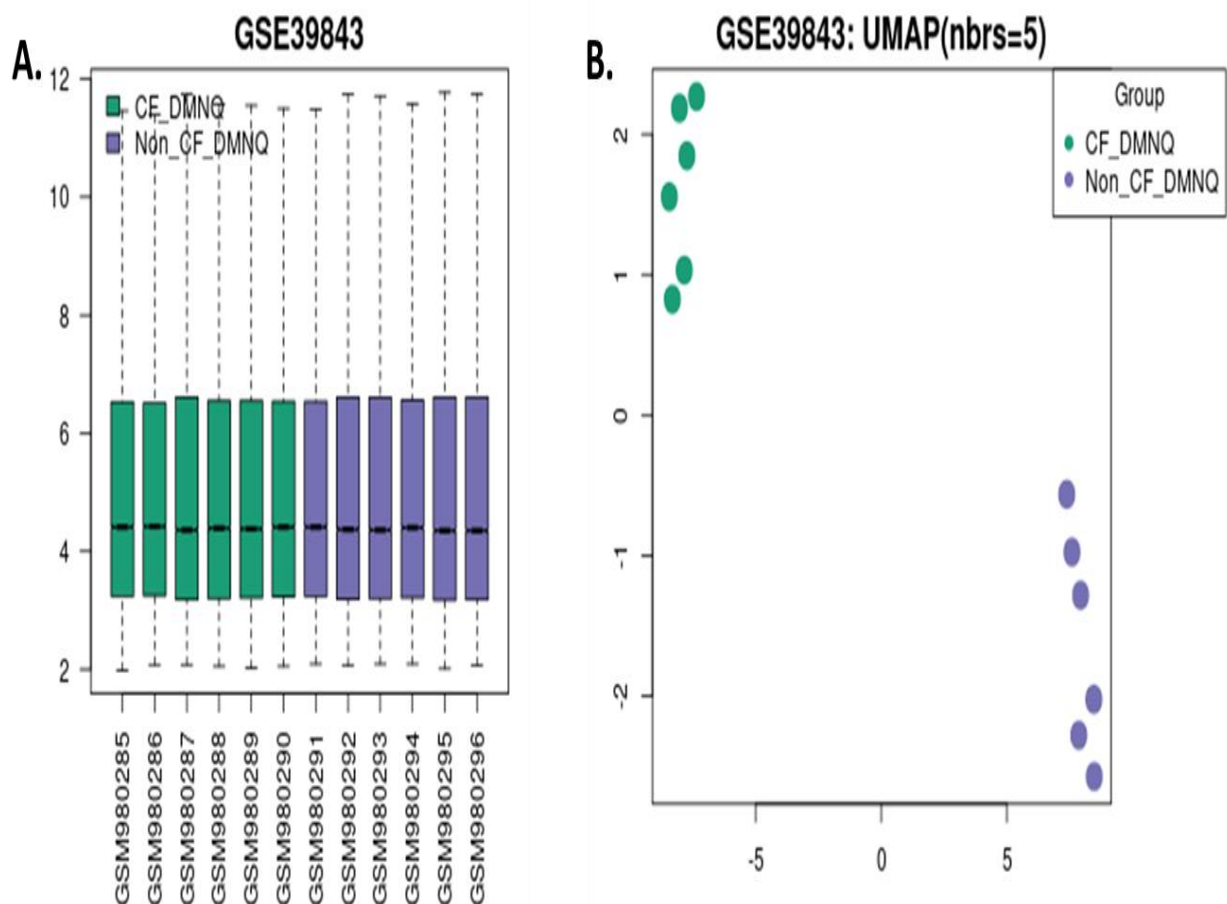
GOTERM_BP_FAT	GO:0006917~induction of apoptosis	0.047836
GOTERM_MF_FAT	GO:0008536~Ran GTPase binding	8.14E-06
GOTERM_MF_FAT	GO:0016667~oxidoreductase activity, acting on sulfur group of donors	5.93E-05
GOTERM_MF_FAT	GO:0016209~antioxidant activity	2.03E-04
GOTERM_MF_FAT	GO:0031072~heat shock protein binding	6.82E-04
GOTERM_MF_FAT	GO:0009055~electron carrier activity	0.00207
GOTERM_MF_FAT	GO:0015171~amino acid transmembrane transporter activity	0.004326
GOTERM_MF_FAT	GO:0001882~nucleoside binding	0.004604
GOTERM_MF_FAT	GO:0016875~ligase activity, forming carbon-oxygen bonds	0.00745
GOTERM_MF_FAT	GO:0019899~enzyme binding	0.010651
GOTERM_MF_FAT	GO:0005524~ATP binding	0.023499
GOTERM_CC_FAT	GO:0005829~cytosol	4.14E-08
GOTERM_CC_FAT	GO:0000502~proteasome complex	1.41E-05
GOTERM_CC_FAT	GO:0048770~pigment granule	5.39E-05
GOTERM_CC_FAT	GO:0005635~nuclear envelope	0.001414
GOTERM_CC_FAT	GO:0012505~endomembrane system	0.00232
GOTERM_CC_FAT	GO:0005730~nucleolus	0.010319

Table 2: GO enrichment for down-regulated DEGs

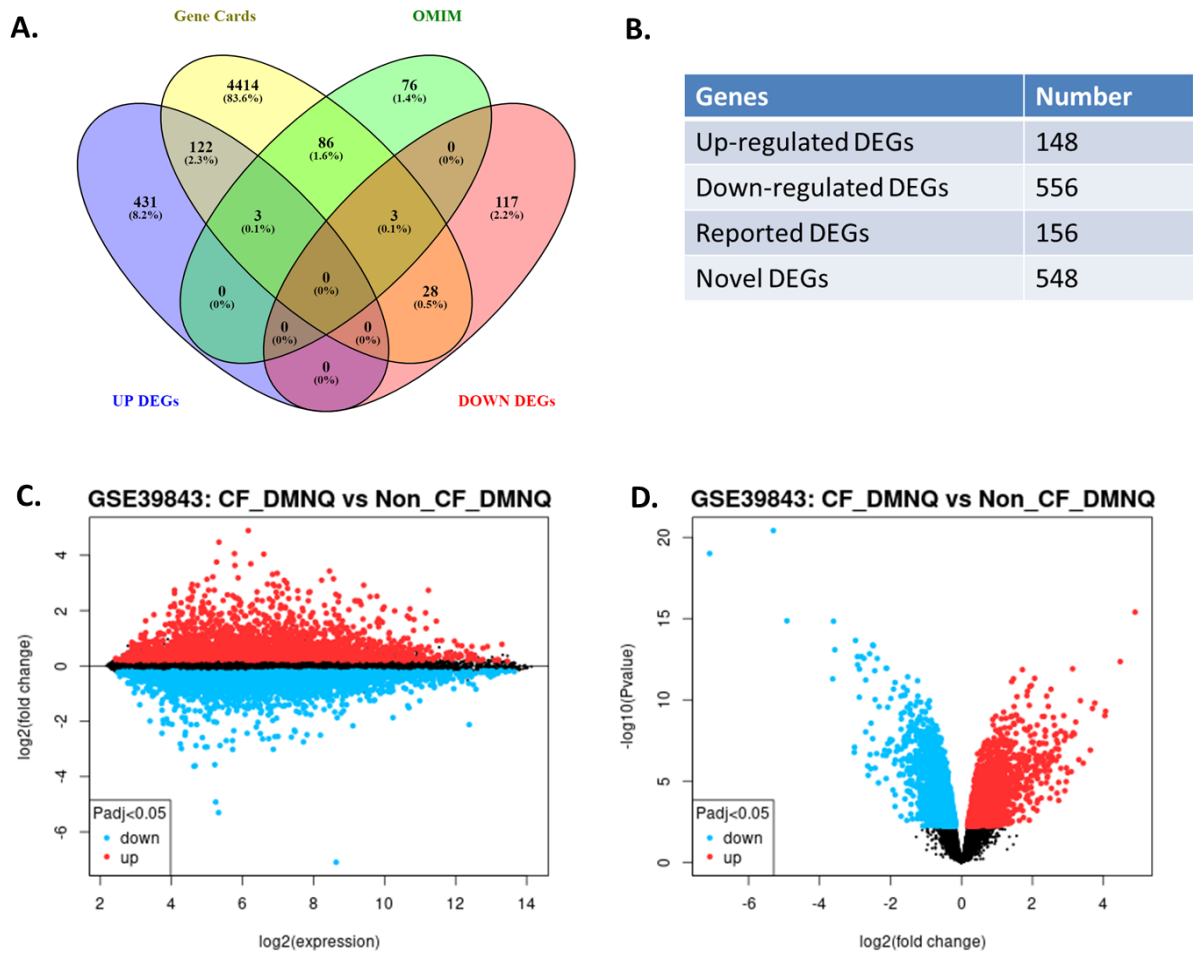
Category	Term	PValue
GOTERM_BP_FAT	GO:0016338~calcium-independent cell-cell adhesion	2.45E-09
GOTERM_BP_FAT	GO:0045216~cell-cell junction organization	0.001137
GOTERM_BP_FAT	GO:0018958~phenol metabolic process	0.001137
GOTERM_BP_FAT	GO:0016337~cell-cell adhesion	0.001217
GOTERM_BP_FAT	GO:0030155~regulation of cell adhesion	0.001295
GOTERM_BP_FAT	GO:0007162~negative regulation of cell adhesion	0.002073
GOTERM_BP_FAT	GO:0042127~regulation of cell proliferation	0.002159
GOTERM_BP_FAT	GO:0002684~positive regulation of immune system process	0.002828
GOTERM_BP_FAT	GO:0008284~positive regulation of cell proliferation	0.003043
GOTERM_BP_FAT	GO:0050671~positive regulation of lymphocyte proliferation	0.004188
GOTERM_BP_FAT	GO:0032946~positive regulation of mononuclear cell proliferation	0.004407
GOTERM_BP_FAT	GO:0070665~positive regulation of leukocyte proliferation	0.004407
GOTERM_BP_FAT	GO:0034330~cell junction organization	0.004632
GOTERM_BP_FAT	GO:0032535~regulation of cellular component size	0.005328
GOTERM_BP_FAT	GO:0009725~response to hormone stimulus	0.005943
GOTERM_BP_FAT	GO:0042417~dopamine metabolic process	0.006124
GOTERM_BP_FAT	GO:0009719~response to endogenous stimulus	0.009967
GOTERM_BP_FAT	GO:0007169~transmembrane receptor protein tyrosine kinase signaling pathway	0.010475
GOTERM_BP_FAT	GO:0030888~regulation of B cell proliferation	0.011817
GOTERM_BP_FAT	GO:0050670~regulation of lymphocyte proliferation	0.013036
GOTERM_BP_FAT	GO:0032944~regulation of mononuclear cell proliferation	0.013462
GOTERM_BP_FAT	GO:0070663~regulation of leukocyte proliferation	0.013462
GOTERM_BP_FAT	GO:0040008~regulation of growth	0.015473
GOTERM_BP_FAT	GO:0007167~enzyme linked receptor protein signaling pathway	0.015677

GOTERM_BP_FAT	GO:0010627~regulation of protein kinase cascade	0.015944
GOTERM_BP_FAT	GO:0045665~negative regulation of neuron differentiation	0.016196
GOTERM_BP_FAT	GO:0034311~diol metabolic process	0.017142
GOTERM_BP_FAT	GO:0009712~catechol metabolic process	0.017142
GOTERM_BP_FAT	GO:0006584~catecholamine metabolic process	0.017142
GOTERM_BP_FAT	GO:0010740~positive regulation of protein kinase cascade	0.017145
GOTERM_BP_FAT	GO:0030308~negative regulation of cell growth	0.017158
GOTERM_BP_FAT	GO:0033630~positive regulation of cell adhesion mediated by integrin	0.017638
GOTERM_BP_FAT	GO:0006576~biogenic amine metabolic process	0.019727
GOTERM_BP_FAT	GO:0051251~positive regulation of lymphocyte activation	0.019727
GOTERM_BP_FAT	GO:0045792~negative regulation of cell size	0.020812
GOTERM_BP_FAT	GO:0007155~cell adhesion	0.022058
GOTERM_BP_FAT	GO:0022610~biological adhesion	0.02224
GOTERM_BP_FAT	GO:0043627~response to estrogen stimulus	0.024258
GOTERM_BP_FAT	GO:0002696~positive regulation of leukocyte activation	0.02486
GOTERM_BP_FAT	GO:0048545~response to steroid hormone stimulus	0.026971
GOTERM_BP_FAT	GO:0045926~negative regulation of growth	0.027352
GOTERM_BP_FAT	GO:0001558~regulation of cell growth	0.027876
GOTERM_BP_FAT	GO:0050867~positive regulation of cell activation	0.027994
GOTERM_BP_FAT	GO:0009967~positive regulation of signal transduction	0.030424
GOTERM_BP_FAT	GO:0051046~regulation of secretion	0.031679
GOTERM_BP_FAT	GO:0008361~regulation of cell size	0.03369
GOTERM_BP_FAT	GO:0050864~regulation of B cell activation	0.036533
GOTERM_BP_FAT	GO:0034622~cellular macromolecular complex assembly	0.04003
GOTERM_BP_FAT	GO:0048103~somatic stem cell division	0.040677
GOTERM_BP_FAT	GO:0033628~regulation of cell adhesion mediated by integrin	0.040677
GOTERM_BP_FAT	GO:0043065~positive regulation of apoptosis	0.0417
GOTERM_BP_FAT	GO:0065003~macromolecular complex assembly	0.042401
GOTERM_BP_FAT	GO:0043068~positive regulation of programmed cell death	0.042894
GOTERM_BP_FAT	GO:0010942~positive regulation of cell death	0.043701
GOTERM_BP_FAT	GO:0010647~positive regulation of cell communication	0.045211
GOTERM_BP_FAT	GO:0002902~regulation of B cell apoptosis	0.046353
GOTERM_BP_FAT	GO:0045785~positive regulation of cell adhesion	0.049027
GOTERM_BP_FAT	GO:0030097~hemopoiesis	0.051116
GOTERM_BP_FAT	GO:0017145~stem cell division	0.051996
GOTERM_BP_FAT	GO:0045684~positive regulation of epidermis development	0.051996
GOTERM_BP_FAT	GO:0045321~leukocyte activation	0.055097
GOTERM_BP_FAT	GO:0042063~gliogenesis	0.056544
GOTERM_BP_FAT	GO:0051249~regulation of lymphocyte activation	0.057315
GOTERM_BP_FAT	GO:0043933~macromolecular complex subunit organization	0.058145
GOTERM_MF_FAT	GO:0042802~identical protein binding	7.24E-04
GOTERM_MF_FAT	GO:0019901~protein kinase binding	0.014538
GOTERM_MF_FAT	GO:0019900~kinase binding	0.027655
GOTERM_MF_FAT	GO:0031490~chromatin DNA binding	0.043912

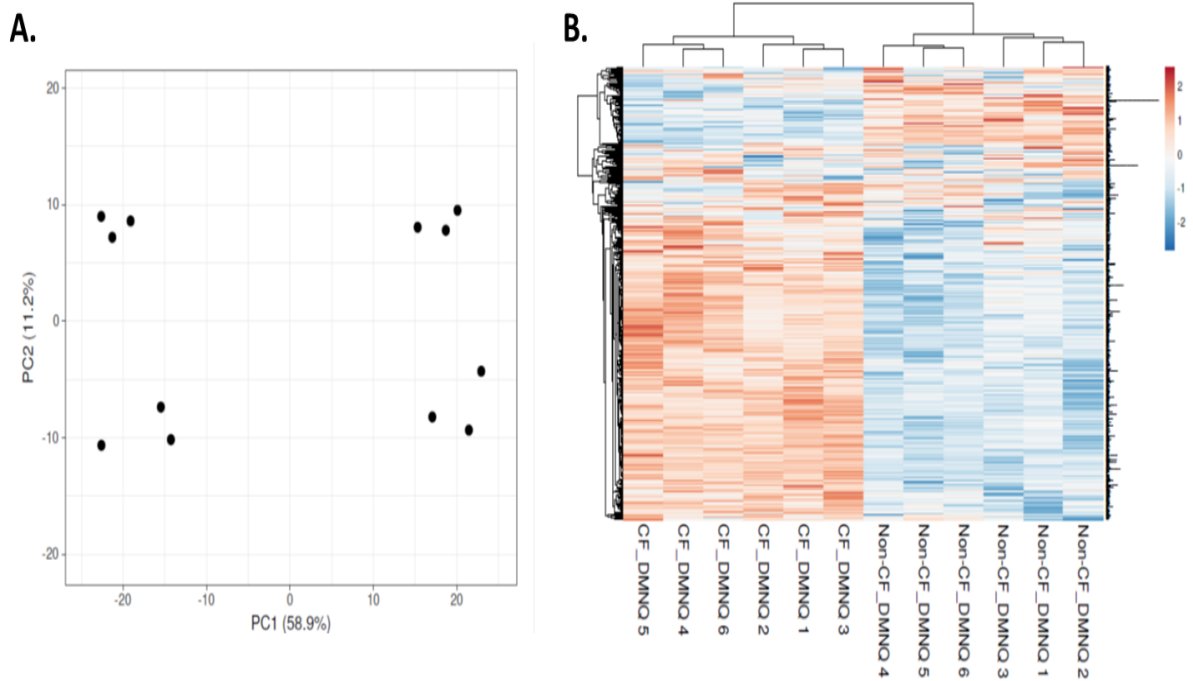
GOTERM_MF_FAT	GO:0004896~cytokine receptor activity	0.04821
GOTERM_MF_FAT	GO:0005198~structural molecule activity	0.04926
GOTERM_MF_FAT	GO:0030296~protein tyrosine kinase activator activity	0.050027
GOTERM_MF_FAT	GO:0046983~protein dimerization activity	0.057969
GOTERM_CC_FAT	GO:0070160~occluding junction	1.90E-09
GOTERM_CC_FAT	GO:0005923~tight junction	1.90E-09
GOTERM_CC_FAT	GO:0043296~apical junction complex	2.92E-08
GOTERM_CC_FAT	GO:0016327~apicolateral plasma membrane	3.79E-08
GOTERM_CC_FAT	GO:0005911~cell-cell junction	8.48E-07
GOTERM_CC_FAT	GO:0030054~cell junction	9.99E-04
GOTERM_CC_FAT	GO:0044459~plasma membrane part	0.004143
GOTERM_CC_FAT	GO:0005886~plasma membrane	0.035738
GOTERM_CC_FAT	GO:0031224~intrinsic to membrane	0.046696



(Fig. 1)

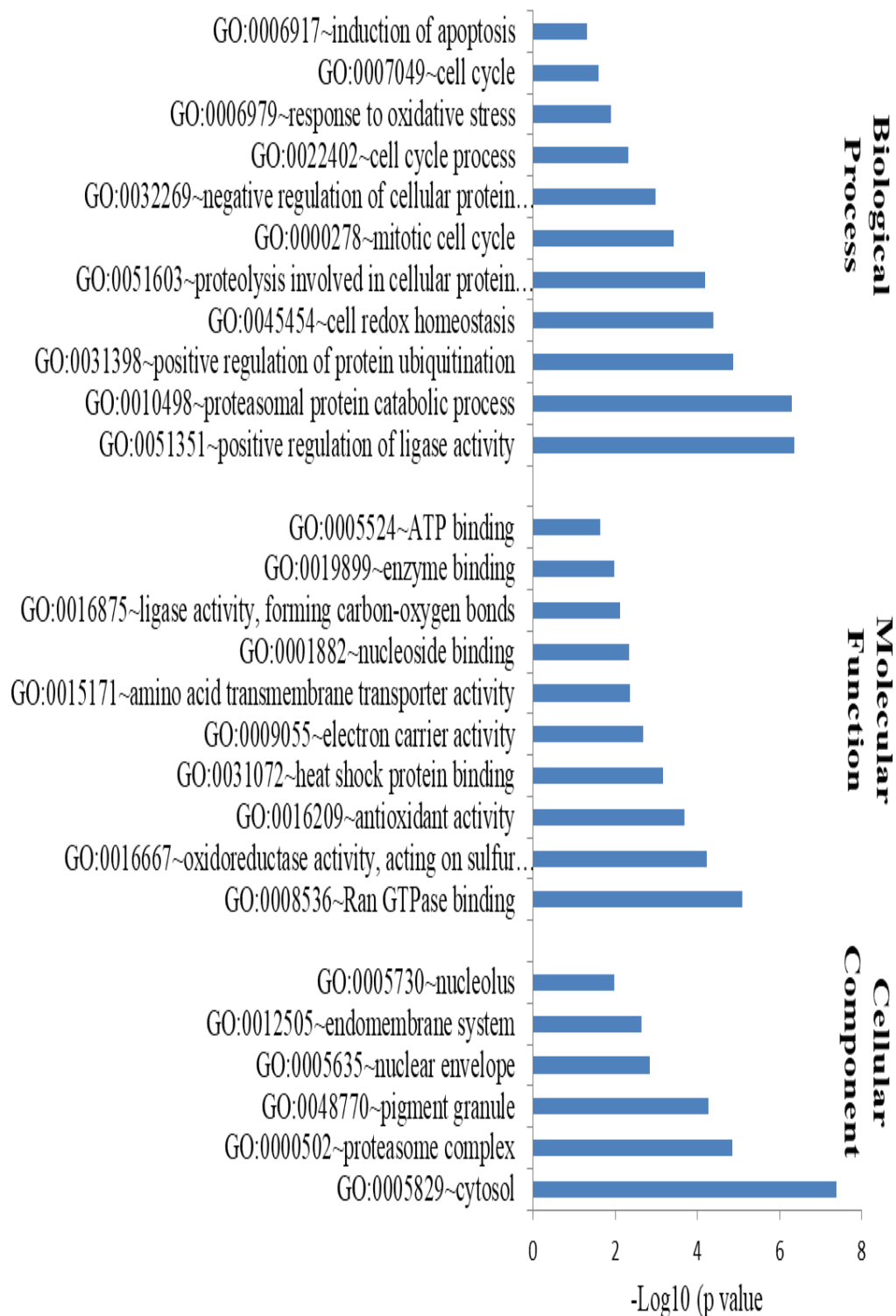


(Fig. 2)

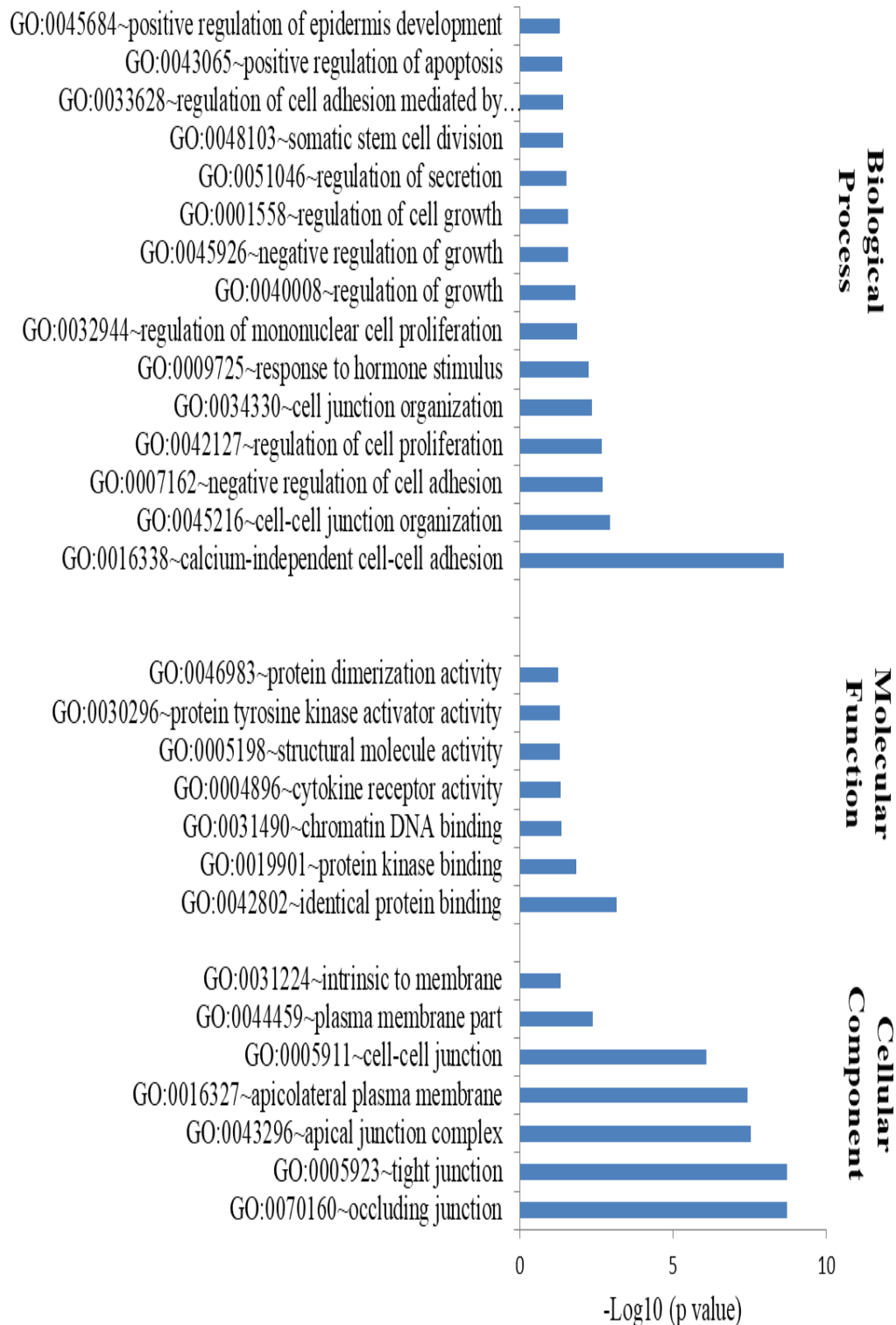


(Fig. 3)

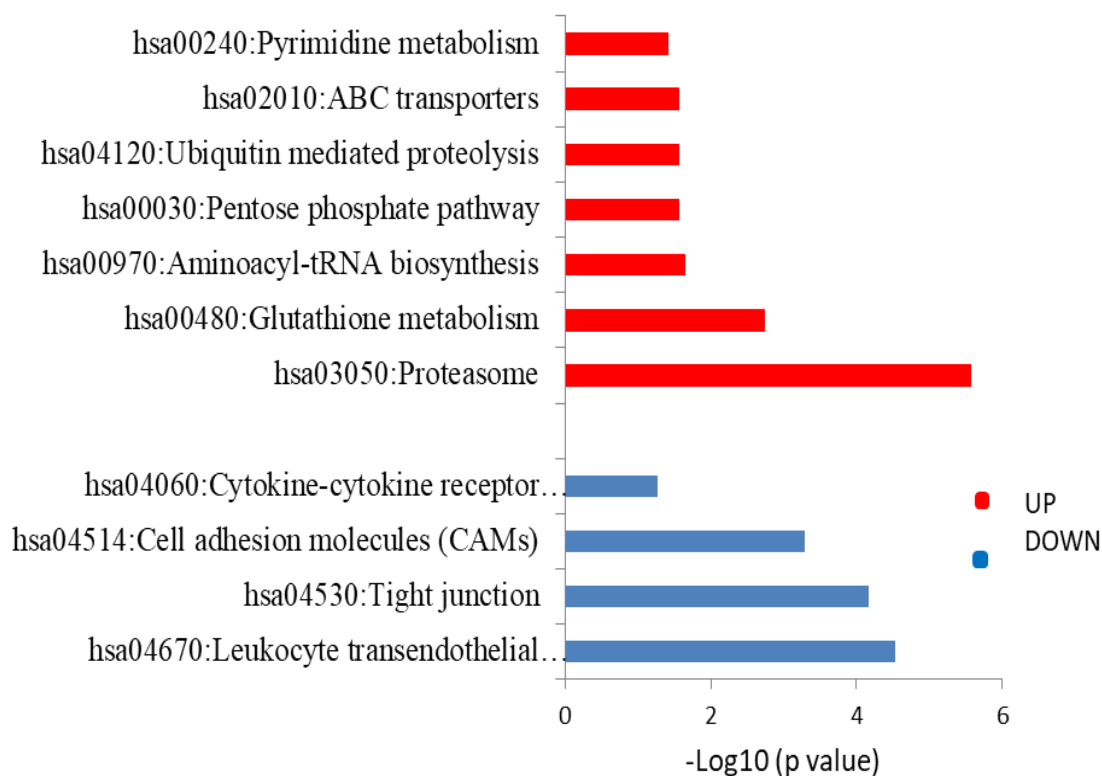
5A GENE ONTOLOGY (GO) enrichment analysis for up-DEGs



5B. GENE ONTOLOGY (GO) enrichment analysis for down DEGs



5C. KEGG pathway enrichment for DEGs



(Fig. 5C)