Article

In silico search for regulatory genes associated with lung and liver disease in Cystic Fibrosis

Rafat Ali, **Safia Tazyeen**, **Mohd Murshad Ahmed**, **Aftab Alam**, **Nikhat Imam**, **Shahnawaz Ali**, **Md Zubbair Malik**, **Romana Ishrat**

Computational Biology Lab, Centre for Interdisciplinary Research in Basic Sciences, Jamia Millia Islamia (A Central University), New Delhi 110025, India

E-mail: rishrat@jmi.ac.in

Received 6 March 2021; Accepted 15 April 2021; Published 1 September 2021

Abstract

The pathogenesis of Cystic Fibrosis (CF) airway disease is not well understood. CF is an autosomal recessive monogenic genetic disease. It affects the exocrine glands, which normally produce thin secretions such as mucus, sweat and tears. In CF, the mucus is thick and sticky which interferes with certain normal organs. A broad knowledge of the genes which are involved in the regulation or co-regulation of affected organs in the CF is required to get a better understanding of its pathophysiological mechanisms. DNA microarray approaches have made it possible to get an insight on gene expression across the genome. In the current study, microarray data related to CF and CF-associated affected organs were retrieved from the NCBS Gene Expression Omnibus database and were subjected to gene regulatory network analysis. We constructed two separated networks of up and down regulated genes from six microarray datasets. The power-law obeying topological properties showed scale-free hierarchical nature of the both networks. Density and compactness of both networks at each level was calculated by modularity and Hamiltonian energy. From all the leading hubs we found four key genes namely GSTT1, ANKRD7, PBX1, and TGFB2 deeply rooted in up and down regulated networks respectively. Conclusively these genes may have prognostic significance.

Keywords Cystic Fibrosis; microarray; modularity; Hamiltonian Energy; GSTT1; ANKRD7; PBX1; TGFB2.

Network Biology
ISSN 2220-8879
URL: http://www.iaees.org/publications/journals/nb/online-version.asp
RSS: http://www.iaees.org/publications/journals/nb/rss.xml
E-mail: networkbiology@iaees.org
Editor-in-Chief: WenJun Zhang
Publisher: International Academy of Ecology and Environmental Sciences

1 Introduction

Cystic fibrosis (CF) is a progressive, inherited disorder (Kulkarni et al., 2019). CF causes severe lung infections and bounds the breath capability over time (De Boeck, 2020). It also infects the digestive system and other organs in the body including pancreas, lungs, liver, kidneys and intestine (Gibson-Corley et al., 2016). An estimated 1 in 29 Caucasian Americans have the CF gene, but it is rare in people of Asian and Middle Eastern origin (In India-1 in 40,000-100,000) (Sanders and Fink, 2016). Diversity of CF around the

world is shown in Fig 1. CF affects the cells that produce different fluids such as mucus, perspiration and digestive fluids which are usually thin and greasy. But in CF patients, the fluids become adhesive and dense. Complications associated with Cystic Fibrosis are categorized according to the different CFTR mutations, modifier genes and environmental factors (Patel et al., 2020). Primary complications associated with CF are-respiratory problems, pancreatic insufficiency, gastro-intestinal abnormalities, hepatobiliary involvement, reproductive system problems, and reduction in nutritional status (Aris et al., 2005; Augarten et al., 2008; Cheng et al., 2000; Cohn et al., 1993; Colombo et al., 2004; Declercq et al., 2016; Fuchs et al., 1994; Gibson et al., 2003; Gorter et al., 2010; Haeusler et al., 1994; Herrmann et al., 2010; Lyon et al., 2002; Rowe et al., 2014). Secondary complications include CF-associated diabetes, low bone mineral density, gastro-intestinal problems, and psychological problems (Haeusler et al., 1994; Bruzzese et al., 2004; Elkins et al., 2006; Kelly et al., 2013; Ooi et al., 2012; Rose et al., 2013).

CF is a multisystem genetic disease (Ideozu et al., 2019a). CFis caused by mutations in the cystic fibrosis conductance regulator (CFTR) gene (Ideozu et al., 2019; Madácsy et al., 2018; Riordan et al., 1989). A number of mutations in cystic fibrosis transmembrane conductance regulator (CFTR) gene are responsible for this condition (Al Balushi et al., 2021; Cui et al., 2020; Lopes-Pacheco, 2020). The most common mutation found in CFTR protein in CF patients is F508del (deletion of phenylalanine coding bp at position 508)(Trouvé et al., 2017). CFTR protein belongs to the family of ATP-binding-cassette (ABC) transporter proteins. CFTR is present at apical membrane of epithelial cells and act as an ion channel (Gadsby et al., 2006; Ramjeesingh et al., 2003). It contains two cytoplasmic nucleotide-binding domains (NBDs) along with a regulatory domain (RD) (Hwang et al., 2013). CFTR protein basically regulates concentration gradient and ATP dependent flow of ions (He et al., 2008). It is well documented that multitude of CF disease phenotypes arise from over 2000types of mutational variations in CFTR gene (Al Balushi et al., 2021), but phenotypic variability presented among patients with the same CFTR genotype remains a major therapeutic challenge (Ideozu et al., 2019). This has driven an intense search for novel molecular drivers associated with CF pathophysiology that may hold promise as biomarkers or therapeutic targets (Ideozu et al., 2019b).

CF is posing pharmacological issues and great challenge for the co-management and treatment of CF-associated affected organs. Computational biology involves uniquely suited approach based on theoretical paradigm and methodological tools to research, describe, explore, and understand structural and relational aspects of human health and diseases (Luke et al., 2007; Sultan et al., 2021). Network-based studies are emerging as an important tool to determine the disease susceptibility genes and their relationship with different diseases. These studies have also improved our understanding of drug targets and effect of drugs and suggested new drug targets and approaches for therapeutics and therapeutic management in severe diseases (Berger and Iyengar, 2009; Freshour et al., 2021). Analysis of networks is significantly contributing to the genesis of systems pharmacology.

The present study was aimed to identify key regulatory genes associated with pathophysiology of Cystic Fibrosis. We implemented computational biology approaches involving related microarray data retrieval from Gene Expression Omnibus (NCBI), differentiating the genes according to their expressions, and construction and analysis of up and down regulatory networks to find out the key regulatory genes. We also carried out the gene ontology enrichment and pathway enrichment analysis to explore the physiological relevance of associated key genes (Trouvé et al., 2017).

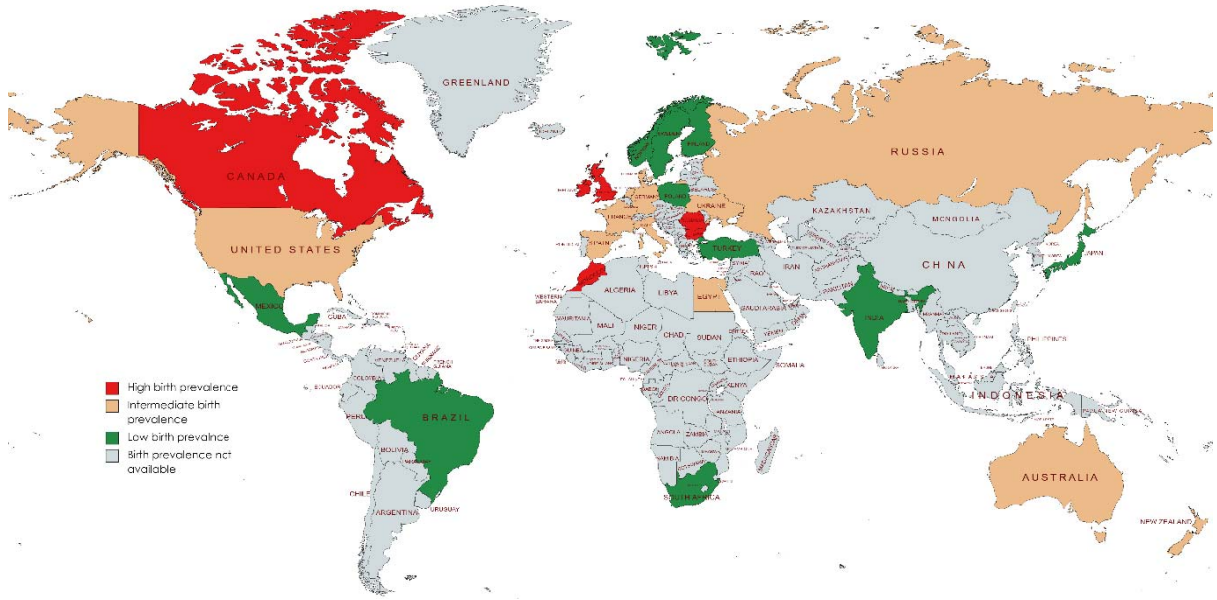

Fig. 1 Geographical distribution of Cystic Fibrosis.

2 Methods

Microarray data related to CF and CF-associated affected organs were retrieved from the NCBS Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo). We found six microarray datasets: GSE15568[27], GSE38956[28], GSE39843[29], GSE48452[30], GSE70442[31]and GSE78914[32] associated with Affymetrix platform. All the six datasets were preprocessed using GEO2R (https://www.ncbi.nlm.nih.gov/geo/geo2r) within built limma (Linear Models for Microarray Analysis) package. Default parameters of Limma were selected during the processing of microarray data.

2.1 Identification of differentially expressed genes

To identify differentially expressed genes, we analyzed all datasets through GEO2R.It is an inbuilt package of R web-based application that helps to analyze GEO data. In analysis p-value<0.05 was retained as threshold during data transformation through logarithmic (log_2FC) scale (Zhu et al., 2018). The fold change method is commonly used for identifying differentially expressed genes to evaluate the between two scenarios. We consider all genes that differ by more than a threshold value to be differentially expressed. As an example, if the threshold value taken is a two-fold difference, genes are differentially expressed if the expression under one condition is over two-fold greater of less than that under other condition (Cui et al., 2003).

2.2 Identification of enriched GO terms

Gene enrichment analysis is commonly used approach to describe gene terms. It provides controlled vocabulary to define gene products. Identification of overpresented GO terms in an input list of genes that could help users to better understand the functionality of these terms (Zheng et al., 2008). There are several freely available web-based tools which are used to carry out GO term enrichment analysis. Some of them are also implemented to carry out pathway analysis or other functions. Here, we conducted analysis using DAVID database (http://david.abcc.ncifcrf.gov/), the database for annotation, visualization and integrated discovery) (Dennis et al., 2003). The significance threshold value of all differentially expressed genes in ontology enrichment analysis was retained as P< 0.05 (Farooqui et al., 2018).

2.3 Network construction and their properties

The first step for building a new biological network is pluging an app with Cytoscape. We used String that is in built in Cytoscape (version 3.3.0). This plugin operates on target set gene given by a user (Farooqui et al., 2018). The input data in the form of gene name can be directly provided or imported in the concerned format. In this biological era, vast amount of molecular interaction data is being produced by several experimental techniques and computational approaches. In order to gain deep insight in to the complex network organization and their structural behavior formed by interacting molecules, we have used the versatile Cytoscape plugin Network Analyzer. It operates and shows a comprehensive set of topological properties of complex network. The parameters include how many numbers of nodes, edges, radius, density, the network diameter, connected components, cluster coefficient, the characteristic path length, centralization, heterogeneity and the distributions of node degrees, neighborhood connectivity's, shortest path length and centrality parameters (Zhang, 2016, 2018). The Network Analyzer is an interactive and highly customizable application, that does not require to expertise in network theory by the user (Assenov et al., 2008; Su et al., 2014).

2.4 Community detection/finding: Leading eigenvector method

Several community finding algorithms have been created to uncover the properties of complex network. However, algorithm evaluation remains still open in terms of computing time and accuracy (Newman et al., 2006). In this study, to detect the communities in the network, we used Leading eigenvector method from R package 'igraph'. The heart of this method is the spectral optimization of modularity by using the eigenvalues. In the beginning it calculates the eigenvalue and then network is split into further parts in a way that modularity improvement is maximized based on LEV. Furthermore, the modularity contribution is calculated at each level of network. It terminates once the value of modularity contribution is negative or it stops at motif level (Xie et al., 2011; Yang et al., 2016).

2.5 Modularity and energy (HE) calculation of network

Tightly connected groups of nodes in a complex network represent individuals belonging to communities, while modules in a biological network are somehow associated with functional modules. Modules sometime called community structures in biological science are tightly linked subcommunities of a complex network, i.e., subsets of nodes within the network connections are dense, and between which connections are sparser (Zhang, 2018). Nodes indeed belonging to such tight-knit modules, constitute units that separately contribute to the collective functioning of the network (Gary et al., 1979). To understand the compactness and organization of network at some state/level of network, we calculated the Hamiltonian energy (HE) of that gene network at that level. HE provides an understanding of the stability at global as well as community level (Traag et al., 2011).

2.6 Tracing of genes

One primary goal of gene network analysis is to identify fundamental gene regulators/gene of modules or sub-modules with respect to various biological contexts. For this, the Cystic Fibrosis genes of our interest were traced at each level of the extracted modules or sub-modules by using communities finding method. At each level genes were filtered and the genes that reached their motif level at the sixth level were considered the fundamental genes of our main network.

3 Results and Discussion

3.1 Retrieval of deferential expressed genes

First a small-scale analysis was performed with human epithelial cells and liver cells (F508 Del homozygous patients vs controls) (Clarke et el., 2013). Normalization of datasets (for noise data or redundancy in datasets) was carried out by using Limma package in R. Our analysis was to establish the set of fundamental genes by comparing analysis across data sets that is relevant in understanding gene functions. In our study we have

included total six GSE microarray series (GSE15568, GSE38956, GSE39843, GSE48452, GSE70442, and GSE78914) associated with CF. Gene name conversion, statistical analysis etc. was done by GEO2R (https://www.ncbi.nlm.nih.gov/geo/geo2r), DAVID (http://david.abcc.ncifcrf.gov/), and Panther (http://www.pantherdb.org/) gene annotation tools. Each series has different number of genes. The up and down regulated genes were filtered through Log fold change. Finally, 787 common DEGs (467 up-regulated and 320 down-regulated) were identified from the six microarray datasets. All the DEGs are Listed in Venn diagram (Fig. 2) and Table 1.

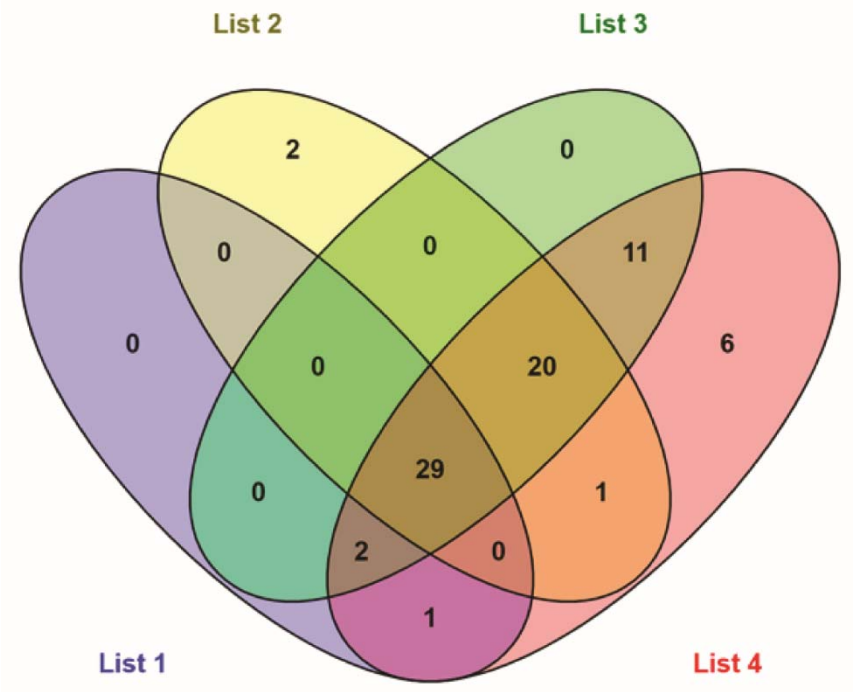

Fig. 2 Venn diagram of DEGs.

Table 1 Data sets and retrieved DEGs.

GSE Series	Samples	Number of probe/genes	UpR genes	DownR genes
GSE15568	29	22283	1006	1008
GSE38956	15	33297	704	535
GSE39843	12	54675	1073	1069
GSE48452	73	33297	988	706
GSE70442	8	54675	764	802
GSE78914	24	54675	807	1004

3.2 Gene ontology and pathway association of the DEGs

To better understand the function of these regulatory genes, we analyzed these DEGs for their gene ontology and annotated pathways, established by DAVID, Panther for GO and KEGG (https://www.genome.jp/kegg/) for pathway annotations. We search gene ontology and pathway enrichment for all common DEGs (up and down regulated genes) and found most of the genes involved in cellular process and metabolic processes as depicted in Fig. 3. The KEGG pathway enrichment analysis was also conducted to further evaluate the biological pathways in that common up and down regulated genes. The results of KEGG pathways analysis

indicated several most significant enriched pathways including metabolism of xenobiotics by cytochrome P450, chemical carcinogenesis, drug metabolism, drug metabolism –cytochrome P450, and retinol metabolism.

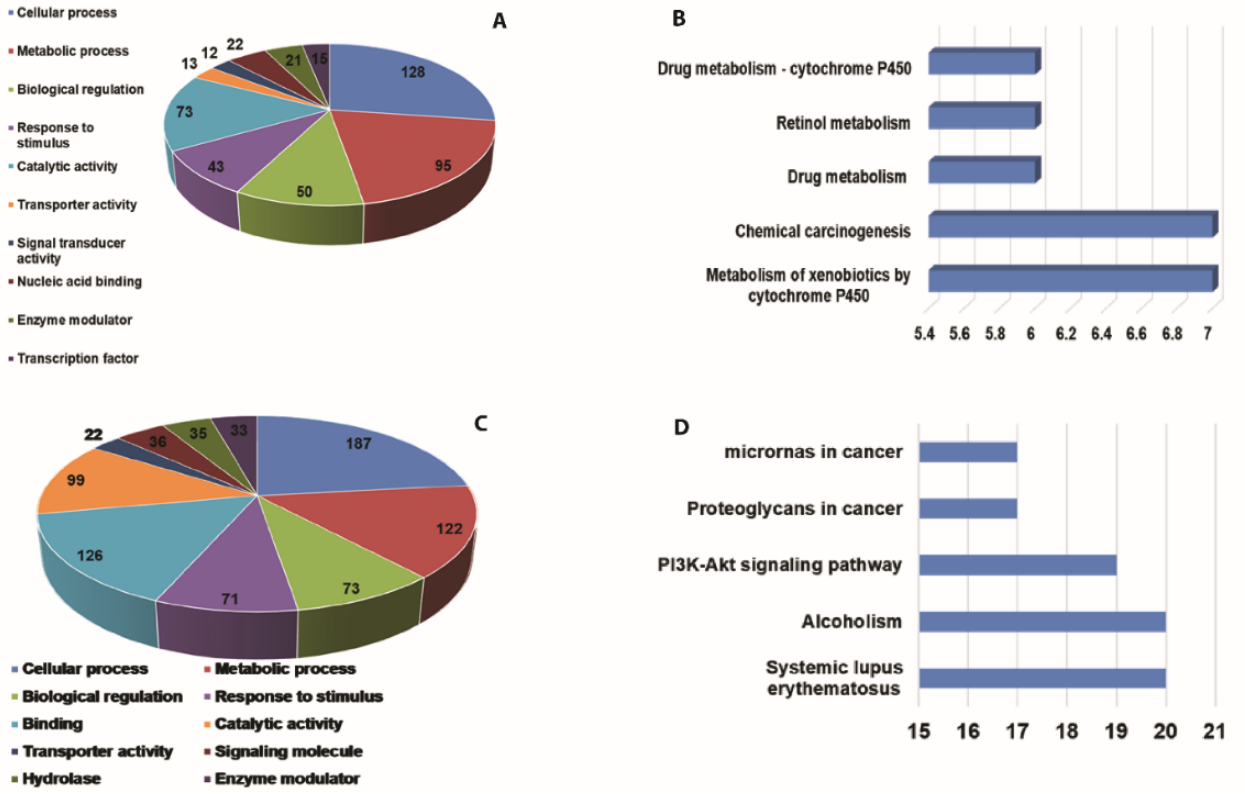

Fig. 3 Pie chart and bar graph representing the Gene Ontology (GO) terms and KEGG pathways. **A:** Pie chart showing GO of up-regulated genes; **B:** Bar graph showing pathway annotation of up-regulated genes; **C:** Pie chart showing GO of down-regulated genes; **D:** Bar graph is showing pathway annotation of down-regulated genes.

3.3 PPI networks association of the new potential key genes

To elucidate proteins which are encoded by our candidate potential key genes, we applied network theory approach for their potential interactions using the STRING (Szklarczyk et al., 2015). The default parameters/units were used in string for functional association between proteins. We constructed two separate regulatory networks of up and down-regulated common genes. The network of up-regulated genes is comprised of 691 nodes, 23312 edges, and 128 leading hubs (Fig. 4). Similarly, down-regulated gene's network showed 993 nodes, 30141 edges and 182 leading hubs (Fig. 5). All the hubs or nodes may not behave as the fundamental regulators for clinical and drug targets genes. To prove this, we have traced the complete network communities and found few genes that are deeply rooted from top to bottom in both networks and the vice-versa and provide the backbone of the network organization. So, these genes are important and can be termed as fundamental regulators. These leading genes in the cystic fibrosis networks take part at each level of organization of the network from starting to fundamental, regulating each unit i.e. motif. Out of these genes two genes ANKRD7 (Ankyrin repeat domain-containing protein 7) and GSTT1 (Glutathione S-transferase theta-1) expressed as up-regulated and other two genes PBX1 (Pre-B-cell leukemia transcription factor 1) and TGFB2 (Transforming growth factor-beta 2) expressed as down-regulated genes.

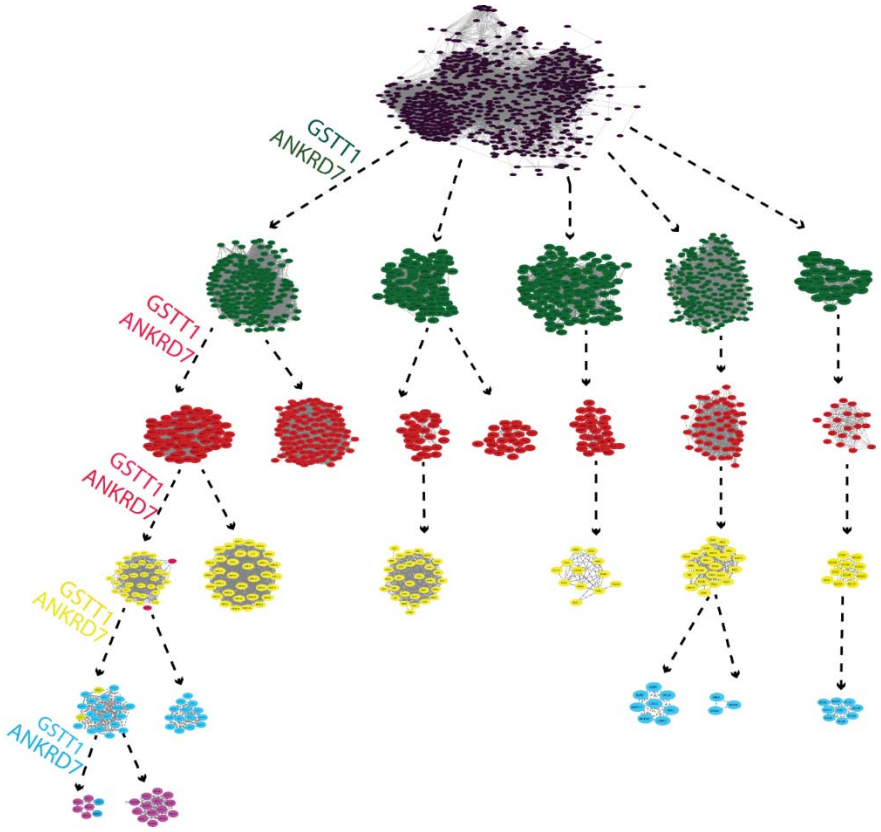

Fig. 4 Network, modules or sub modules at different levels of up-regulated genes.

The gene TGFB2 encodes a secreted ligand of the TGF-beta (transforming growth factor-beta) superfamily of proteins (Starling, 2019). Ligands of this family bind various TGF-beta receptors leading to recruitment and activation of SMAD family transcription factors that regulate gene expression (Kubiczkova et al., 2012; Starling, 2019). The encoded preproprotein is proteolytically processed to generate a latency-associated peptide (LAP) and a mature peptide, and is found in either a latent form composed of a mature peptide homodimer, a LAP homodimer, and a latent TGF-beta binding protein, or in an active form consisting solely of the mature peptide homodimer (Lebrun, 2012). The mature peptide may also form heterodimers with other TGF-beta family members. TGF beta pathway has been studied extensively and disruption in this pathway is linked to cause a variety of human cancers (Lebrun, 2012). Dietz syndrome is associated with mutated tgfb2 gene (Alansari et al., 2002). PBX1 is a well-known transcription factor associated with TALE class and is responsible for number of embryonic processes, including organogenesis, morphogenic patterning and haematopoiesis. It has been observed that growth and survival of the ovarian, melanoma and breast cancer cells depend on PBX1 (McCarthy et al., 2002). The GSTT1gene is involved in the regulation of conjugation of carcinogenic substances to excretable metabolites, because of this it is susceptible to cause cancer (Dahabreh, et al., 2010). A cytosolic enzyme is encoded by this gene which carries the catalytic cleavage of the epoxides, aliphatic aromatic and heterocyclic radicals, and arenoxides to glutathione. Its involvement is also associated with Alzheimer disease (Dahabreh et al., 2010). ANKRD7 is a ankyrin repeat domain, testis specific ankyrin motif containing protein (Fagerberg et al., 2014). This gene has no specific pathway in KEGG (Kyoto Encyclopedia of Gene and Genome).The term fundamental genes can be defined as the genes that are found to show the differential expression irrespective to the condition, state of diseases and pattern of their expression

(Lee and Loscalzo, 2019). They are found embedded deeper inside the network as backbone genes and if traced at different level of organization it can be found to from basic structural unit of the network called motif (Barabási et al., 2011; Lee and Loscalzo, 2019; Sonawane et al., 2019). Finally, by tracking down these genes, we found that the eleven more genes which are interacting with signature genes at motif level NCOR1, HOXB1, HIST2H2AA3, HIST2H2AC, VEGFB, FGB, RPL10L, GNB2L1, RPLS, WDR31 and RPL10. Apart from fundamental genes these interacting genes can be used in further research in future.

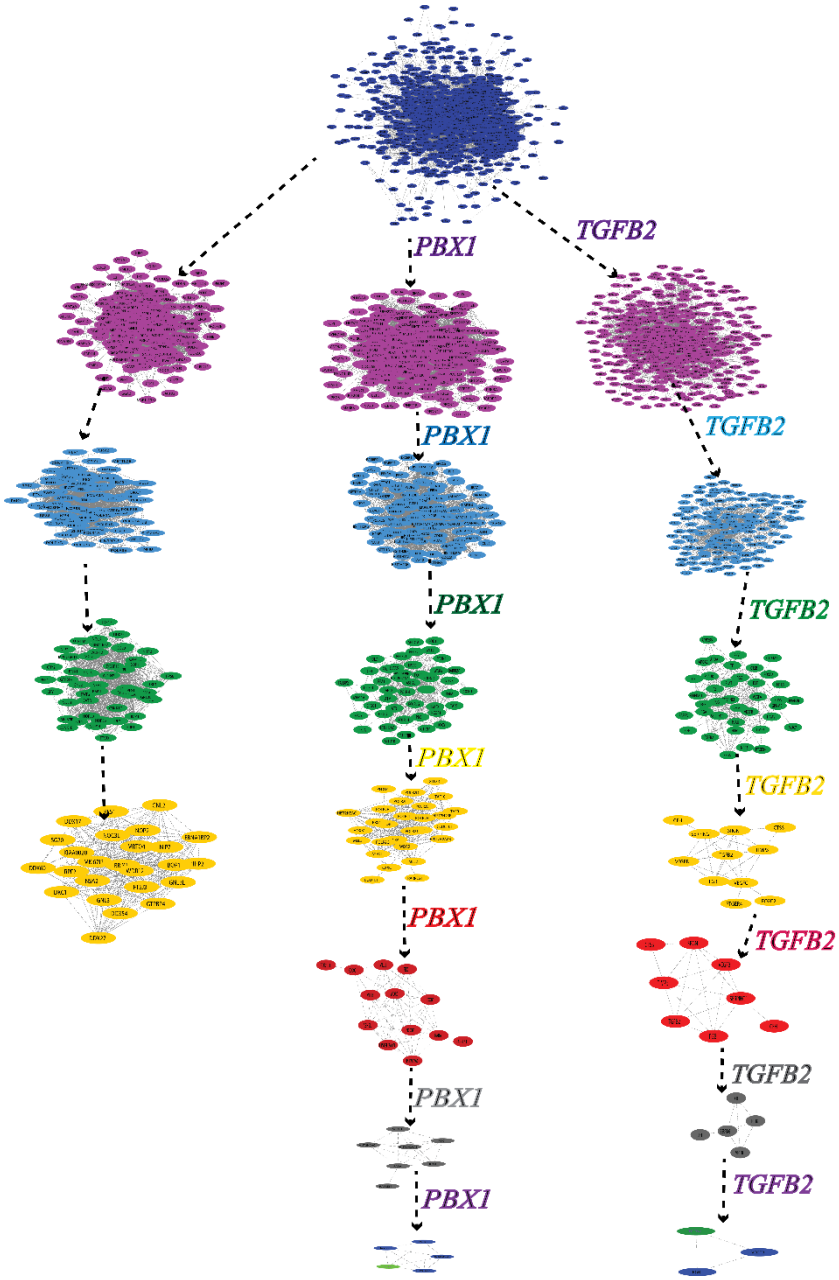

Fig. 5 Network, modules or sub modules at different levels of down-regulated genes.

3.4 Topological properties and organization of networks

The topology properties of the networks are probability of degree distribution P(k), clustering coefficient C(k) and neighbourhood connectivity CN(k) which exhibits power law nature with respect to degree k (Singh et al.,

2016; Zhang, 2018). The power law on the data distribution are confirmed and verified by following a standard statistical fitting procedure proposed by Clauset et al (2009). We summarized the networks as follows; the negative value of γ of degree distribution shows availability of each node in the network and the Positive value of β indicates assortivity nature of the network which means the importance of few hubs forming a cluster (rich-club formation). The negative value of α of clustering parameter shows disassociation in the communication between the nodes in the network (Fig. 6).

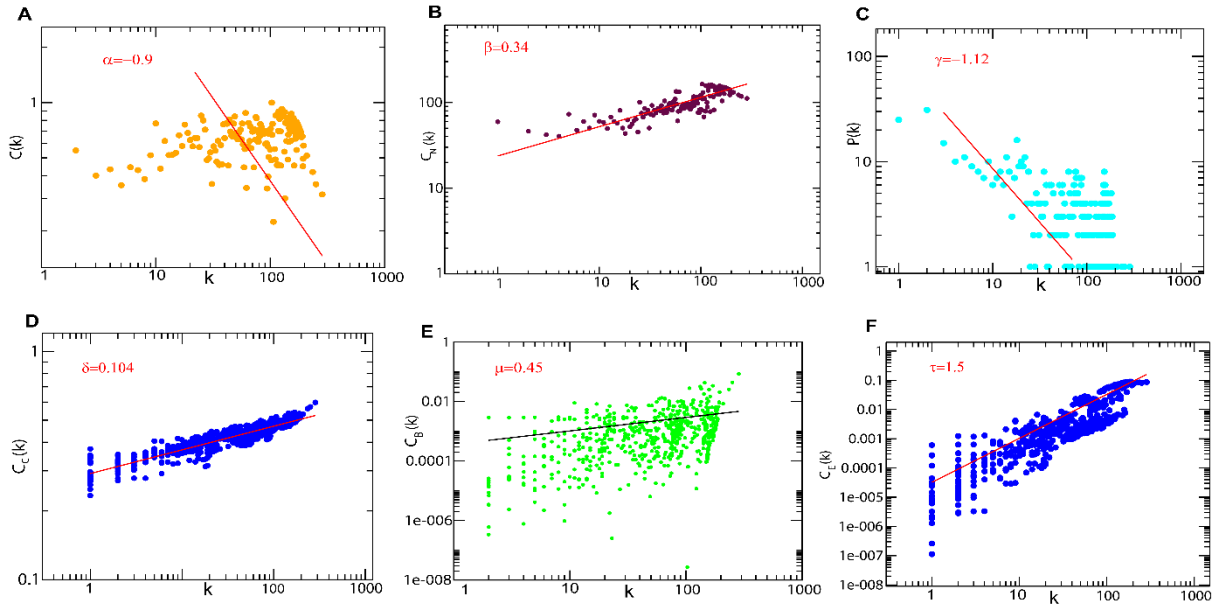

Fig. 6 Topological properties of UP-regulated network. A: Degree Distribution; B: Clustering coefficient; C: Neighborhood connectivity D: Betweenness centrality; E: Eigenvector centrality F: Closeness Centrality. The centrality measurement are betweenness centrality CB(k), closeness centrality CC(k) and eigen value centrality CE(k) which displays the importance of the hubs, their regulating mechanisms and obeys power law behaviours as follows:

$$\textbf{Equation:} \begin{bmatrix} P \\ C \\ C_N \end{bmatrix} \sim \begin{bmatrix} k^{\gamma} \\ K^{\alpha} \\ K^{\beta} \end{bmatrix} \rightarrow \begin{bmatrix} -1.12 \\ -0.9 \\ 0.34 \end{bmatrix}$$

$$\textbf{Equation:} \begin{bmatrix} C_B \\ C_C \\ C_E \end{bmatrix} \sim \begin{bmatrix} k^{\mu} \\ K^{\delta} \\ K^{\tau} \end{bmatrix} \rightarrow \begin{bmatrix} 0.45 \\ 0.104 \\ 1.5 \end{bmatrix}$$

Topological properties of up regulated gene's network

The positive value of exponents of these centrality parameters shows the strong regulatory role of the leading hubs in the network (Fig. 7). Therefore, we can say that the network follows hierarchical scale free properties.

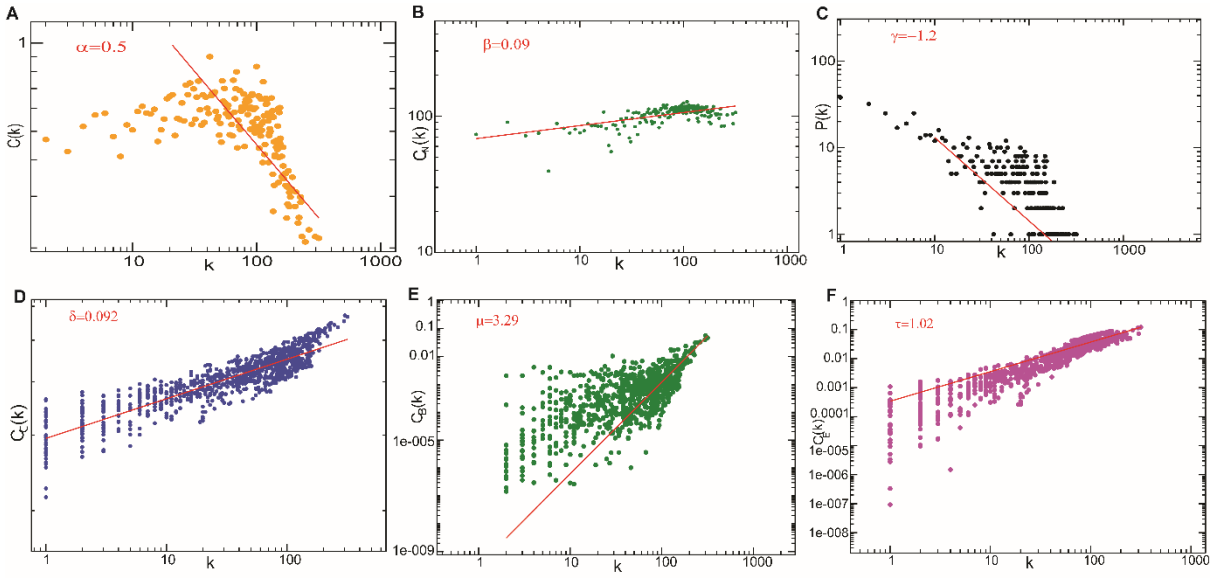

Fig. 7 Topological properties of down-regulated network. a: degree distribution; b: clustering co-efficient; c: neighborhood connectivity d: betweenness centrality; e: eigenvector centrality f: closeness centrality.

$$\textbf{\textit{Equation}}: \begin{bmatrix} \boldsymbol{P} \\ \boldsymbol{C} \\ \boldsymbol{C_N} \end{bmatrix} \sim \begin{bmatrix} \boldsymbol{k^{\gamma}} \\ \boldsymbol{K^{\alpha}} \\ \boldsymbol{K^{\beta}} \end{bmatrix} \rightarrow \begin{bmatrix} -1.2 \\ 0.5 \\ 0.09 \end{bmatrix}$$

$$\textbf{\textit{Equation}}: \begin{bmatrix} \boldsymbol{C_B} \\ \boldsymbol{C_C} \\ \boldsymbol{C_E} \end{bmatrix} \sim \begin{bmatrix} \boldsymbol{k^{\mu}} \\ \boldsymbol{K^{\delta}} \\ \boldsymbol{K^{\tau}} \end{bmatrix} \rightarrow \begin{bmatrix} 3.29 \\ 0.092 \\ 1.5 \end{bmatrix}$$

Topological properties of down regulated gene's network

Next, we calculated the distribution of energy for complete network, modules and sub modules. It describes the energy distribution for each node (Singh et al., 2016). Hamiltonian energy provides energy distribution not only at the global level of a network, but also at modular level (Haider et al., 2019). It is the ratio of Hamiltonian energy (Fig. 8, 9) of a module and sub modules at individual level. The distribution of energy for these genes shows how effectively they participate in the organization of network that has direct correlation to communication.

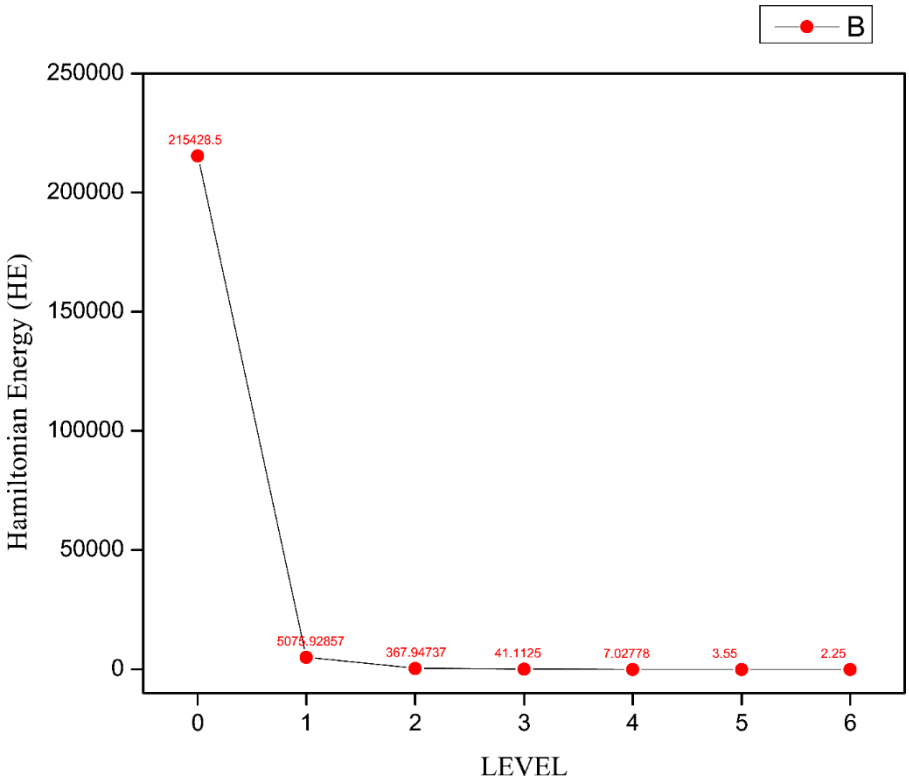

Fig. 8 Plots of distribution of Hamiltonian energy of Up-regulated genes network.

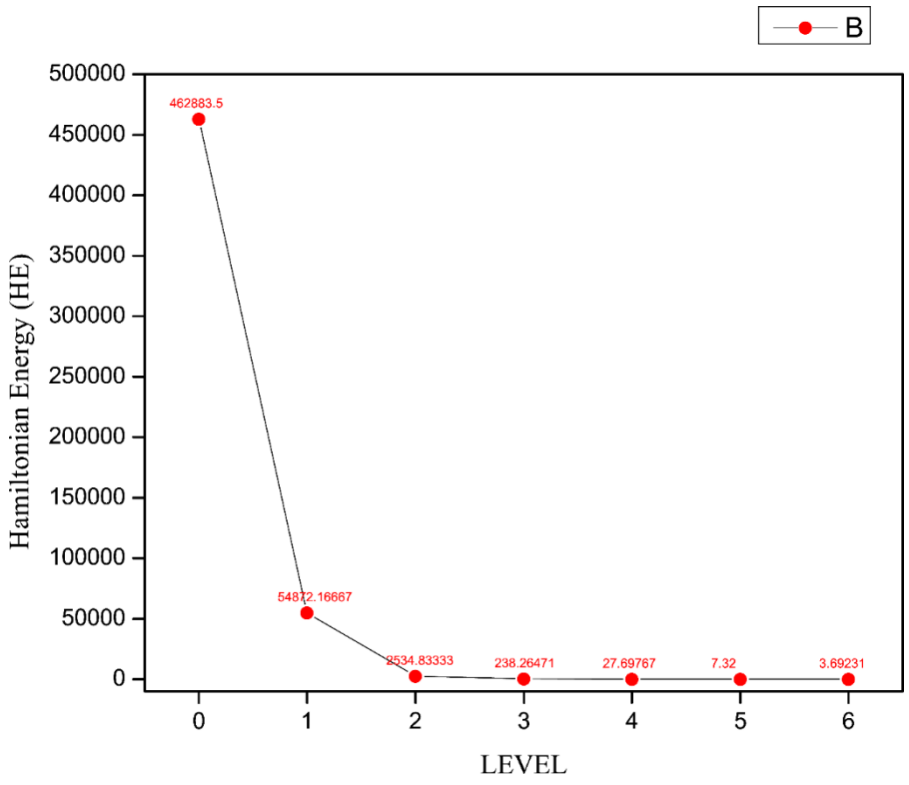

Fig. 9 Plots of distribution of Hamiltonian energy of Down-regulated genes network.

Distribution of energy is beneficial to understand the roles of modules and hubs in network organization (Sporns, 2018). The Hamiltonian energy decrease as one goes down from top to bottom indicating its important regulating activity at complete network level then at a basic level (Haider et al., 2019). Large biological networks are partitioned into clusters or modules where similar or interacting nodes are grouped together. Modules or communities are actually the connected hub networks that are densely connected within themselves and sparsely connected to the rest of the network (Cherifi et al., 2019). Such a grouping of nodes can help us to identify the underlying structure of the network and extract insights from it. In biological networks modularity is defined as a degree to which its components i.e. nodes are relatively connected or separated to each other (Serban, 2020). From this notion we can say that modularity degree is high when many of its nodes belong to dense modules and low if many nodes do not belong to dense modules at all. Modularity (Fig. 10, 11) of a network helps us to understand the structure of the network as well as the mechanism that define the network (Serban, 2020).

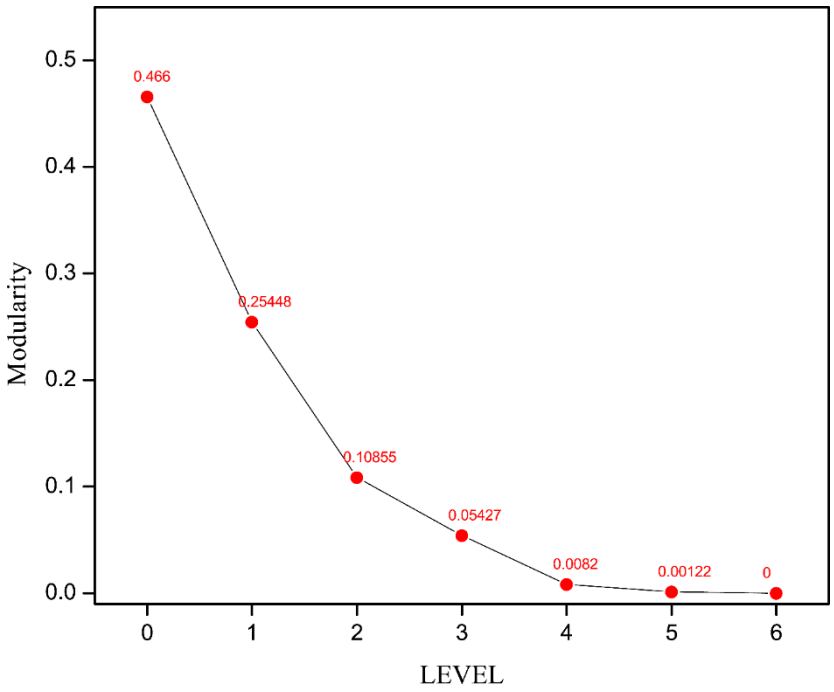

Fig. 10 Modularity distribution plots of Up-regulated genes network.

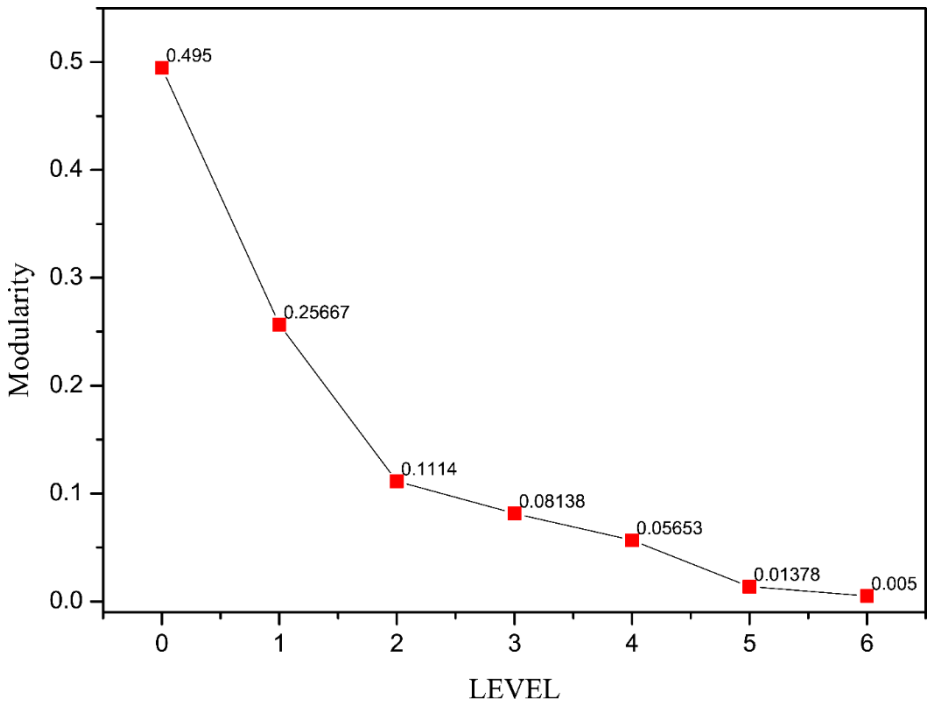

Fig. 11 Modularity distribution plots of Down-regulated genes network.

In hierarchical networks, communication between the different highly clustered neighbourhoods is maintained by a few hubs. The appearance of hierarchical modularity in biological networks supports the assumption that evolution acts on many levels. Modularity shows function of level of organization which is found to be decreasing as one goes from top to bottom organization.

4 Conclusion

The CF is characterized by lung abnormality with inflammation and pancreatic insufficiency. It is not known whether defects in CFTR are responsible or a consequence. The one simple hypothesis with defective CFTR is responsible to decrease airway surface liquid which fails to clear infected secretions from the lung, accelerating the excessive inflammatory response. It is therefore still unknown fact, whether this hyper inflammation is solely the result of chronic infection or is a primary task due to CFTR defects. Despite the role of residues in CFTR's sequence and its variation over lung functions, signature genes are of main importance. Indeed, it is defined that differences in CF genotype are not only responsible for the disease variation. Microarray studies were undertaken to find genetic determinants of phenotypic variation. Nevertheless, further studies are still needed to get a deep insight. Therefore, our aim was to identify new modifier genes associated with CF through microarray data analysis. We found four genes that were obtained from microarray analysis and highlighted their importance for CF, which were also found involved in other diseases (called as Inferred genes). Interestingly, ANKRD7 is one of the highlighted among these four genes that is specifically involved in very basic level of CF network. The ANKRD7somehow modifies the compensation of X-linked gene during spermatogenesis as obtained by RPL10L. We suggest that some other modifier genes may be missing and more studies are required to provide some physiological relevance. Finally, these biomarker/modifier genes reported in our study may contribute to clinical severity in CF, which may have prognostic significance and could prompt to start a more targeted therapy in the CF patients.

Acknowledgments

The authors are grateful to the Centre for Interdisciplinary Research in Basic Sciences (CIRBSc), Jamia Millia Islamia-110025 for providing the research infrastructure. MMA is thankful to Indian Council of Medical Research (ICMR), New Delhi, India. RI duly acknowledges the Indian Council of Medical Research (ICMR).

References

Alansari A, Hajeer A, The LS, Bayat A, Mysercough A, Gül A, Inanc M, Ordi-Ros J, Ollier W. 2002. Transforming growth factor-12 polymorphism and systemic lupus erythematosus. The Journal of Rheumatology, 29(6): 1189-1191

Al Balushi S, Al Balushi Y, Al Busaidi M, Al Mutawa L. 2021. A novel Cystic Fibrosis gene mutation C. 4242+ 1G> C in an Omani patient: A Case Report. Oman Medical Journal, 36(2): e243

Aris RM, Merkel PA, Bachrach LK, Borowitz DS, Boyle MP, Elkin SL, Guise TA, Hardin DS, Haworth CS, Holick MF, Joseph PM. 2005. Guide to bone health and disease in cystic fibrosis. The Journal of Clinical Endocrinology and Metabolism, 90(3): 1888-1896

Assenov Y, Ramírez F, Schelhorn SE, Lengauer T, Albrecht M. 2008. Computing topological parameters of biological networks. Bioinformatics, 24(2): 282-284

Augarten A, Tov AB, Madgar I, Barak A, Akons H, Laufer J, Efrati O, Aviram M, Bentur L, Blau H, Paret, G, 2008. The changing face of the exocrine pancreas in cystic fibrosis: the correlation between pancreatic status, pancreatitis and cystic fibrosis genotype. European Journal of Gastroenterology and Hepatology, 20(3): 164-168

Berger SI, Iyengar R. 2009. Network analyses in systems pharmacology. Bioinformatics, 25(19): 2466-2472

Bruzzese E, Raia V, Gaudiello G, Polito G, Buccigrossi V, Formicola V, Guarino A, 2004. Intestinal inflammation is a frequent feature of cystic fibrosis and is reduced by probiotic administration. Alimentary Pharmacology and Therapeutics, 20(7): 813-819

Cheng K, Smyth RL, Motley J, O'Hea U, Ashby D, 2000. Randomized controlled trials in cystic fibrosis (1966–1997) categorized by time, design, and intervention. Pediatric Pulmonology, 29(1): 1-7

Cherifi H, Palla G, Szymanski BK, Lu X. 2019. On community structure in complex networks: challenges and opportunities. Applied Network Science, 4(1): 1-35

Clarke LA, Sousa L, Barreto C, Amaral MD. 2013. Changes in transcriptome of native nasal epithelium expressing F508del-CFTR and intersecting data from comparable studies. Respiratory Research, 14(1): 1-19.

Clauset A, Shalizi CR, Newman ME. 2009. Power-law distributions in empirical data. SIAM Review, 51(4): 661-703

Cohn JA, Strong TV, Picciotto MR, Nairn AC, Collins FS, Fitz JG. 1993. Localization of the cystic fibrosis transmembrane conductance regulator in human bile duct epithelial cells. Gastroenterology, 105(6): 1857-1864

Colombo C, Battezzati PM, 2004. Liver involvement in cystic fibrosis: primary organ damage or innocent bystander? Journal of Hepatology, 41(6): 1041-1044

Cui X, Churchill GA. 2003. Statistical tests for differential expression in cDNA microarray experiments. Genome biology, 4(4): 1-10

Cui X, Wu X, Li Q, Jing, X. 2020. Mutations of the cystic fibrosis transmembrane conductance regulator gene in males with congenital bilateral absence of the vas deferens: Reproductive implications and genetic counseling. Molecular Medicine Reports, 22(5): 3587-3596

Dahabreh IJ, Giannouli S, Gota V, Voulgarelis M. 2010. GSTT1 and GSTM1 polymorphisms and myelodysplastic syndrome risk: A systematic review and meta-analysis. International Journal of Cancer, 126(7): 1716-1723

De Boeck K. 2020. Cystic fibrosis in the year 2020: A disease with a new face. Acta Paediatrica, 109(5): 893-899

Declercq D, Robberecht E. 2016. Diagnosis of pancreatic insufficiency in cystic fibrosis: a practical approach. International Journal of Gastroenterology Hepatology Transplant and Nutrition, 1(1): 7-10

Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, Lane,HC, Lempicki RA. 2003. DAVID: database for annotation, visualization, and integrated discovery. Genome Biology, 4(9): 1-11

Elkins MR, Robinson M, Rose BR, Harbour C, Moriarty CP, Marks GB, Belousova EG, Xuan W, Bye PT. 2006. A controlled trial of long-term inhaled hypertonic saline in patients with cystic fibrosis. New England Journal of Medicine, 354(3): 229-240

Fagerberg L, Hallström BM, Oksvold P, Kampf C, Djureinovic D, Odeberg J, Habuka M, Tahmasebpoor S, Danielsson A, Edlund K, Asplund A. 2014. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. Molecular and Cellular Proteomics, 13(2): 397-406

Farooqui A, Tazyeen S, Ahmed MM, Alam A, Ali S, Malik MZ, Ali S, Ishrat R. 2018. Assessment of the key regulatory genes and their Interologs for Turner Syndrome employing network approach. Scientific Reports, 8(1): 1-13

Freshour SL, Kiwala S, Cotto KC, Coffman AC, McMichael JF, Song JJ, Griffith M, Griffith OL, Wagner AH. 2021. Integration of the Drug–Gene Interaction Database (DGIdb 4.0) with open crowd source efforts. Nucleic Acids Research, 49(D1): D1144-D1151

Fuchs HJ, Borowitz DS, Christiansen DH, Morris EM, Nash ML, Ramsey BW, Rosenstein B.J, Smith AL, Wohl ME. 1994. Effect of aerosolized recombinant human DNAs on exacerbations of respiratory symptoms and on pulmonary function in patients with cystic fibrosis. New England Journal of Medicine, 331(10): 637-642

Gadsby DC, Vergani P, Csanády L. 2006. The ABC protein turned chloride channel whose failure causes cystic fibrosis. Nature, 440(7083): 477-483

Gary MR, Johnson DS. 1979. Computers and Intractability: A Guide to the Theory of NP-completeness. WH Freman and Co, USA

Gibson-Corley KN, Meyerholz DK, Engelhardt JF. 2016. Pancreatic pathophysiology in cystic fibrosis. The Journal of Pathology, 238(2): 311-320

Gibson RL, Burns JL, Ramsey BW. 2003. Pathophysiology and management of pulmonary infections in cystic fibrosis. American Journal of Respiratory and Critical Care Medicine, 168(8): 918-951

Gorter RR, Karimi A, Sleeboom C, Kneepkens CMF, Heij HA. 2010. Clinical and genetic characteristics of meconium ileus in newborns with and without cystic fibrosis. Journal of Pediatric Gastroenterology and Nutrition, 50(5): 569-572

Haeusler G, Frisch H, Waldhör T, Götz M. 1994. Perspectives of longitudinal growth in cystic fibrosis from birth to adult age. European Journal of Pediatrics, 153(3): 158-163

Haider S, Ponnusamy K, Singh RB, Chakraborti A, Bamezai RN. 2019. Hamiltonian energy as an efficient approach to identify the significant key regulators in biological networks. Plos one, 14(8): e0221463

He L, et al. 2008. Multiple membrane-cytoplasmic domain contacts in the cystic fibrosis transmembrane conductance regulator (CFTR) mediate regulation of channel gating. Journal of Biological Chemistry, 283(39): 26383-26390

Herrmann U, Dockter G, Lammert F. 2010. Cystic fibrosis-associated liver disease. Best Practice and Research Clinical Gastroenterology, 24(5): 585-592

Hwang TC, Kirk KL. 2013. The CFTR ion channel: gating, regulation, and anion permeation. Cold Spring Harbor perspectives in medicine, 3(1): a009498

Ideozu JE, Zhang X, McColley S, Levy H. 2019. Transcriptome profiling and molecular therapeutic advances in cystic fibrosis: recent insights. Genes, 10(3):180

Ideozu JE, Zhang X, Rangaraj V, McColley S, Levy H. 2019. Microarray profiling identifies extracellular circulating miRNAs dysregulated in Cystic Fibrosis. Scientific Reports, 9(1): 1-11

Kelly A, Moran A. 2013. Update on cystic fibrosis-related diabetes. Journal of Cystic Fibrosis, 12(4): 318-331

Kubiczkova L, Sedlarikova L, Hajek R, Sevcikova S. 2012. TGF-β–an excellent servant but a bad master. Journal of Translational Medicine, 10(1): 1-24

Kulkarni H, Kansra S, Karande S. 2019. Cystic fibrosis revisited. Journal of Postgraduate Medicine, 65(4): 193-196

Lopes-Pacheco M. 2019. CFTR Modulators: Advances, Barriers and Future Directions in Precision Medicine for Cystic Fibrosis. Frontiers in Pharmacology, 10:1662

Lyon A, Bilton D. 2002. Fertility issues in cystic fibrosis. Paediatric Respiratory Reviews, 3(3): 236-240

Luke DA, Harris JK. 2007 Network analysis in public health: history, methods, and applications. Annual Review of Public Health, 28.021406.144132

Lebrun JJ. 2012. The dual role of TGF in human cancer: from tumor suppression to cancer metastasis. ISRN Molecular Biology, 2012: 28

Lee LYH, Loscalzo J. 2019. Network medicine in pathobiology. The American Journal of Pathology, 189(7): 1311-1326

Madácsy T, Pallagi P, Maleth J. 2018. Cystic fibrosis of the pancreas: the role of CFTR channel in the regulation of intracellular Ca2+ signaling and mitochondrial function in the exocrine pancreas. Frontiers in Physiology, 9:1585

McCarthy VA, Ott CJ, Phylactides M, Harris A. 2009. Interaction of intestinal and pancreatic transcription factors in the regulation of CFTR gene expression. Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms, 1789(11-12): 709-718

Newman ME. 2006. Finding community structure in networks using the eigenvectors of matrices. Physical Review E, 74(3): 036104

Ooi CY, Durie PR. 2012. Cystic fibrosis transmembrane conductance regulator (CFTR) gene mutations in pancreatitis. Journal of Cystic Fibrosis, 11(5): 355-362

Patel SD, Bono TR, Rowe SM, Solomon GM. 2020. CFTR targeted therapies: recent advances in cystic fibrosis and possibilities in other diseases of the airways. European Respiratory Review, 29(156)

Ramjeesingh M, Kidd JF, Huan LJ, Wang Y, Bear CE. 2003. Dimeric cystic fibrosis transmembrane conductance regulator exists in the plasma membrane. Biochemical Journal, 374(3): 793-797

Reiman EM, Langbaum JB, Tariot PN, Lopera F, Bateman RJ, Morris JC, Sperling RA, Aisen PS, Roses AD, Welsh-Bohmer KA, Carrillo MC. 2016. CAP—advancing the evaluation of preclinical Alzheimer disease treatments. Nature Reviews Neurology, 12(1): 56

Riordan JR, Rommens JM, Kerem BS, Alon N, Rozmahel R, Grzelczak Z, Zielenski J, Lok S, Plavsic N, Chou JL. 1989. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science, 245(4922): 1066-1073

Rose RL. 2013. Cystic Fibrosis: CFTR, Complications, and Prospective Therapies. Chancellor's Honors Thesis Project April 2013. University of Tennessee, USA

Rowe SM, Heltshe SL, Gonska T, Donaldson SH, Borowitz D, Gelfond D, Sagel SD, Khan U, Mayer-Hamblett N, Van Dalfsen JM, Joseloff E. 2014. Clinical mechanism of the cystic fibrosis transmembrane conductance regulator potentiator ivacaftor in G551D-mediated cystic fibrosis. American Journal of Respiratory and Critical Care Medicine, 190(2): 175-184

Sonawane AR, Weiss ST, Glass K, Sharma A. 2019. Network medicine in the age of biomedical big data. Frontiers in Genetics, 10: 294

Sanders DB, Fink, AK. 2016. Background and epidemiology. Pediatric Clinics, 63(4): 567-584

Serban M. 2020. Exploring modularity in biological networks. Philosophical Transactions of the Royal Society B, 375(1796): 20190316

Singh SS, Khundrakpam B, Reid AT, Lewis JD, Evans AC, Ishrat R, Sharma BI, Singh RB. 2016. Scaling in topological properties of brain networks. Scientific Reports, 6(1): 1-19

Sporns O. 2018. Graph theory methods: applications in brain networks. Dialogues in clinical Neuroscience, 20(2): 111

Starling S. 2019. A new metabolic role for TGFβ2. Nature Reviews Endocrinology, 15(4): 191-191

Su G, Morris JH, Demchak B, Bader GD. 2014. Biological network exploration with Cytoscape 3. Current Protocols in Bioinformatics, 47(1): 8-13

Sultan A, Ali R, Sultan T, Ali S, Khan NJ, Parganiha A. 2021. Circadian clock modulating small molecules repurposing as inhibitors of SARS-CoV-2 Mpro for pharmacological interventions in COVID-19 pandemic. Chronobiology International, 1-15

Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M. 2015. STRING v10: protein–protein interaction networks, integrated over the tree of life. Nucleic Acids Research, 43(D1): D447-D452

Traag VA, Van Dooren P, Nesterov Y. 2011. Narrow scope for resolution-limit-free community detection. Physical Review E, 84(1): 016114

Trouvé P, Génin E, Férec C, 2017. In silico search for modifier genes associated with pancreatic and liver disease in Cystic Fibrosis. PloS one, 12(3): e0173822

Xie J, Szymanski BK. 2011, June. Community detection using a neighborhood strength driven label propagation algorithm. In: 2011 IEEE Network Science Workshop. 188-195, IEEE

Yang Z, Algesheimer R, Tessone CJ. 2016. A comparative analysis of community detection algorithms on artificial networks. Scientific Reports, 6(1): 1-18

Zhang WJ. 2016. Selforganizology: The Science of Self-Organization. World Scientific, Singapore

Zhang WJ. 2018. Fundamentals of Network Biology. World Scientific Europe, London, UK

Zheng Q, Wang XJ. 2008. GOEAST: a web-based software toolkit for Gene Ontology enrichment analysis. Nucleic Acids Research, 36(Suppl-2): W358-W363

Zhu QN, Renaud H, Guo Y, 2018. Bioinformatics-based identification of miR-542-5p as a predictive biomarker in breast cancer therapy. Hereditas, 155(1): 1-11