International Journal of Basic and Applied Biology
Print ISSN: 2349 – 5820; Online ISSN: 2349 – 5839 ; Volume. 2 No. 2, November 2014. pp. 1 – 6

http://www.krishisanskriti.org/ijbab.html

Identification of genetic modifiers and its study in modulation of disease severity in cystic fibrosis patients

Amit Kumar Sharma[1], Rambaksh Prajapti[2], Sneha Patil[3]

[1,2,3] *School of Biotechnology & Bioinformatics,*
D. Y. Patil University,
Belapur, Navi Mumbai

Abstract: **Cystic Fibrosis (CF) is a common, phenotypically variable disease caused by mutations in the CFTR gene, which encodes a cAMP-dependent chloride channel found in many cell types. Symptoms are primarily gastrointestinal and pulmonary in nature; with the pulmonary disease accounting for the majority of the mortality in CF. There are reports dedicated toward the characterization of the genetic contribution to the disease variance, with many modifiers gene for Cystic Fibrosis. In this study we have undertaken a novel approach to the identification of cystic fibrosis modifier genes by establishing the Gene Modifier Study (GMS), which utilizes a candidate gene approach to compare the genetic profile of the most severe CF patients to that of the mildest CF patients. We present probable causes for CF due to modifiers prominently variants within HLA-DQA1 and HLA-DQA2 which are associated with CF patient with long-term disease severity and survival.**

1. INTRODUCTION

Cystic Fibrosis (CF) is the most common lethal autosomal recessive disease in the Caucasian population, affecting nearly 1 in 2500 newborns. Cystic Fibrosis is caused by mutations in the gene encoding the Cystic Fibrosis Transmembrane Conductance Regulator (CFTR), a cAMP regulated Cl^- channel[1, 2, 3] expressed in the apical (luminal) membrane of epithelia. The cystic fibrosis gene Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) encodes for a protein of 1480 amino acids CFTR is also known as ABCC7 and belongs to a large family of ATP-binding cassette (ABC) transporters [1, 4]. The proteins consists of two transmembrane domains (TM's), each spanning the membrane six times, two nucleotide binding domains (NBD's) and a highly charged cytoplasmic domain (theR domain) containing consensus sequences for phosphorylation.

The protein CFTR functions as a channel across the membrane of cells that produce mucus, sweat, saliva, tears, and digestive enzymes [5]. Under normal conditions the CFTR protein functions as a chloride channel pumping Cl^- ions out of cell. CFTR also functions as regulator of other channels. The transport of chloride ions helps control the movement of water in tissues, which is necessary for the production of thin, freely flowing mucus. The CFTR protein also regulates the transport of sodium ion across cell membranes. These channels are necessary for the normal function of organs such as the lungs and pancreas[6].

Mutations which result in a loss of channel activity in CFTR can be classified into four major categories, class I-IV. ▲F508 is a class II mutation found in more than 90% of all CF alleles. A single amino acid deletion of a phenylalanine at position 508 (▲F508) causes a defect in CFTR maturation. This results in an incorrect targeting to the apical membrane; the mutant protein is retained and degraded in the endoplasmic reticulum(ER) by the quality control mechanism[7].

Variability in pulmonary disease progression in CF cannot be explained by the CFTR defect alone. The CFTR defect in association with gene defects in other genes involved in modulating the severity of the inflammatory response and response to infection may contribute to more severe and rapid phenotypic progression of pulmonary disease in CF.

Although the gene defective in CF patients has been identified, CF disease pathogenesis is still a not well understood process. In order to explain the CF pathogenesis, a detailed understanding of the CFTR protein is essential.

For this, we have studied the regulation of CFTR, its physiological functions, and the effects of the ▲F508 CFTR mutation. Based on the functions of CFTR and the specific

properties of CFTR mutants, strategies towards a therapy for CF can be developed. With an intent to better treatment to CF patients, present study focusses on the understanding the upregulating and downregulating genes during the CF disease progression. The study addresses whether there are genes other than CFTR gene responsible for variability in frequency of pulmonary infections requiring intravenous therapy seen in CF patients and susceptibility to recurrent pulmonary infection.

2. MATERIALS AND METHODOLOGY

2.1. Gene datasets

The Gene Expression Omnibus (GEO) isa public repository that archives and freely distributes microarray, next-generation sequencing, and other forms of high- throughput functional genomic data submitted by the scientific community. Two gene datasets were obtained from GEO viz. GDS2142 and GDS2143. This database stores curated gene expression Datasets, as well as original series and platform records in the Gene Expression Omnibus (GEO) repository. Dataset records contain additional resources including cluster tools and differential expression queries.

2.2. DNA microarray

DNA microarray known as gene chips, bio- chips or silicon-chips, are solid support usually made up of glass or silicon, holding DNA that represent thousands of genes that act as probes for mRNA.

2.3. Analyzing the datasets

R is a free software environment for statistical computing and graphics. It comprises and runs on a wide variety of UNIX platforms, Windows and MacOS. R provides a wide variety of statistical and graphical techniques, including linear and nonlinear modelling, classical statistical tests, time-series analysis, classification, clustering, and others. It is an environment within which statistical techniques were implemented. Bioconductor is a free, open source and open development software project for the analysis and comprehension of genomic data generated by wet lab experiments in molecular biology. Bioconductor is based primarily on the statistical R programming.

MATLAB is high level language and interactive environment for numerical computation visualization, and programming. Using MATLAB, we can analyze data, develop algorithms, and create models and applications. Normalization was done for resolving the systematic errors and bias introduced by the microarray experimental platform. Normalization is useful for a number of situations including within-slide comparison, multiple-slide comparison, and paired-slide comparison for dye exchange experiments. Normalization removes unwanted systematic variability in hybridization and to adjust the spatial effects from microarray data by adjusting the intensities of hybridization to balance them appropriately to make decisive analysis. It identifies the differences introduced in labelling or detection efficiencies when different fluorescent dyes are used. It also adjusts to bias introduced because of scanner settings. The microarray data generated by the feature extraction software is typically in the form of one or more text files.

Multi Experiment Viewer (MeV) is an application that allows the user to view processed microarray slide representations and identifies genes and expression patterns of interest. This Crude Analysis was used to perform the back validation of genes which were obtained by using SAM test and T-test. In this, the selection of upregulated and down regulated genes was done using MS Excel.

3. RESULTS

3.1. Box plot analysis

Box plot analysis is method for visualizing the distribution of the data values throughout the dataset providing information about the spread and skewness in the dataset. Fig. 3.1 and Fig. 3.2 represent the box plot analysis of the gene datasets GDS2142 and GDS2143 respectively as before and after normalization.

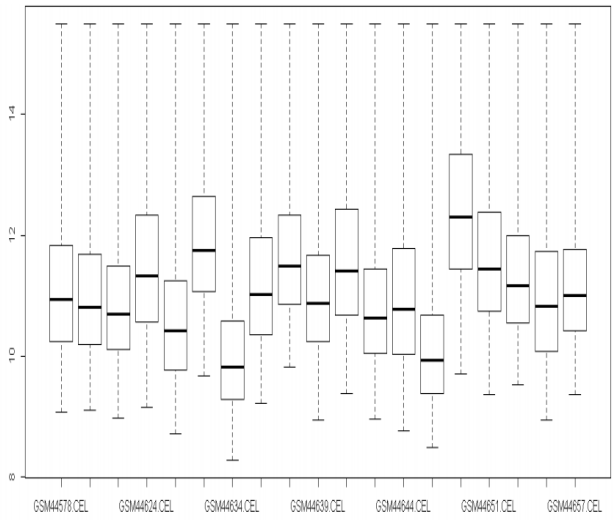

Fig. 3.1: Boxplot of Dataset GDS2142 (GSE2395) before normalization

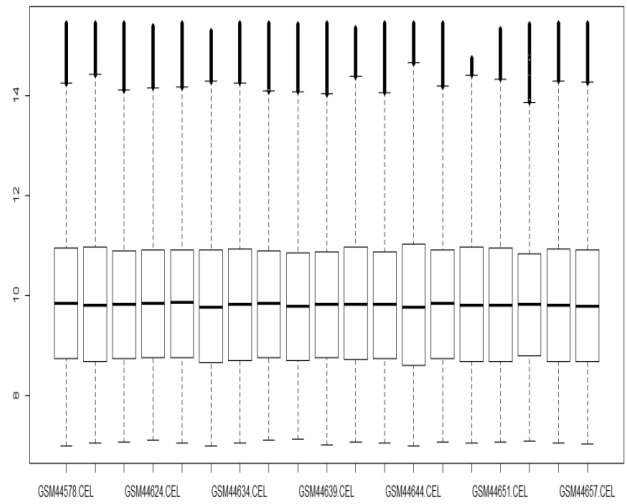

Fig. 3.2: Boxplot of Dataset GDS2142 (GSE2395) after normalization

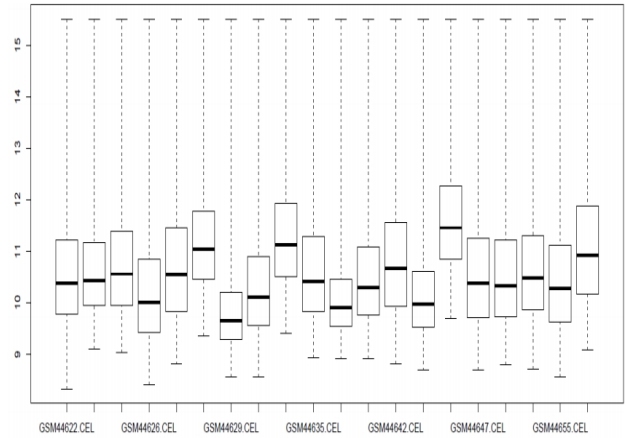

Fig. 3.3: Boxplot of Dataset GDS2143 (GSE2395) before normalization

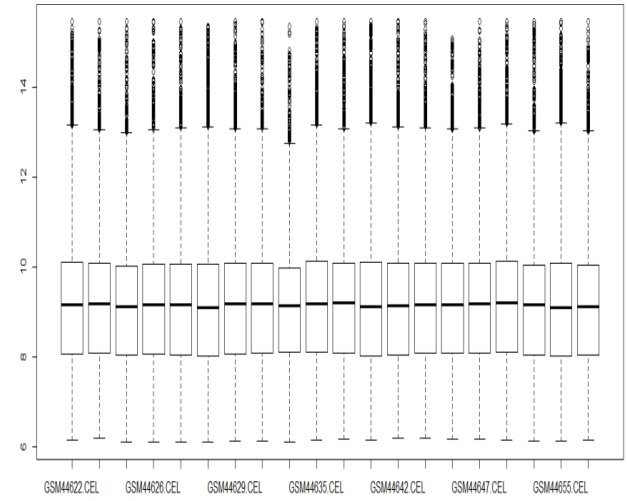

Fig. 3.4: Boxplot of Dataset GDS2143 (GSE2395) after normalization

3.2. Histogram analysis

A histogram is a representation of a statistical graph of a frequency distribution in which vertical rectangles of different heights are proportionate to corresponding frequencies. Fig. 3.3 and Fig. 3.4 represent the histogram curve of the datasets each before and after normalization. The microarray data shows histogram as frequency on y-axis and log transform of pixel intensities along the x-axis.

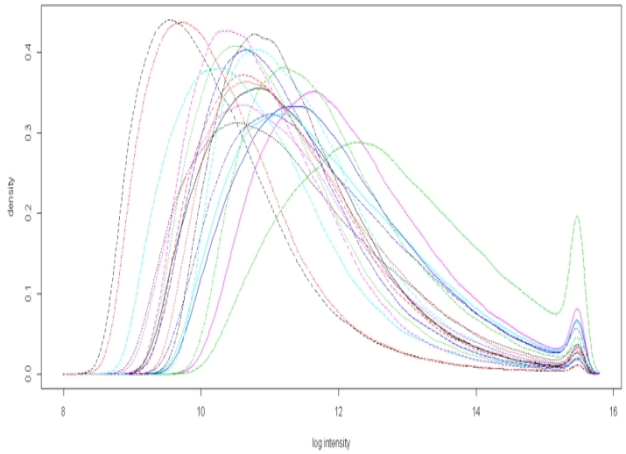

Fig. 3.5: Histogram of GDS2142 (GSE2395) before normalization

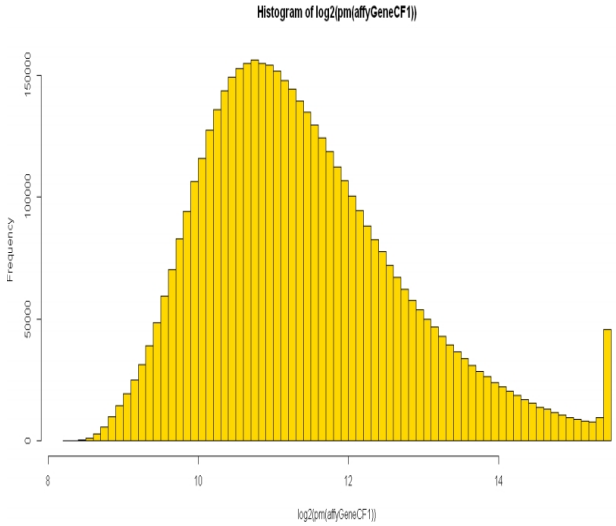

Fig. 3.6: Histogram of GDS2142 (GSE2395) after normalization

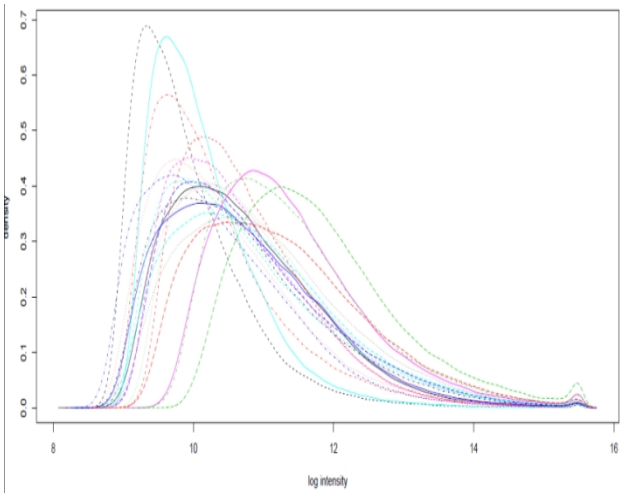

Fig. 3.7: Histogram of GDS2143 (GSE2395) before normalization

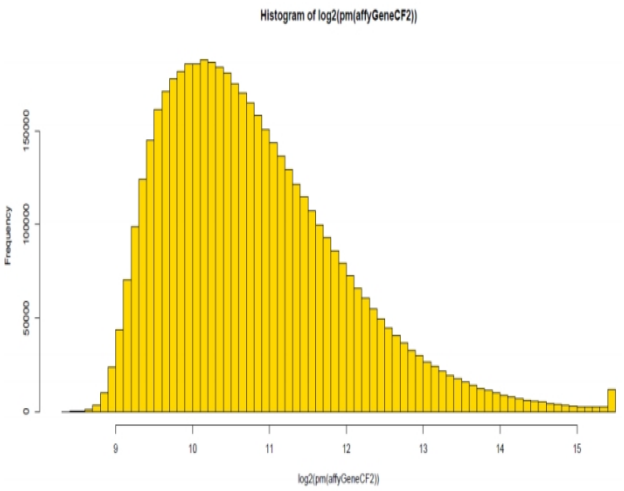

Fig. 3.8: Histogram of GDS2143 (GSE2395) after normalization

3.3. Visualizing the chip data image

The normalized data is viewed in MeV. The respective chip used and organism are selected for all the datasets for automatic annotation loading. When the data is loaded we can visualize all the samples at the top of the image display. Each row of the image is a single gene represented by a Probe ID towards the right**.** Each column represents a sample. When either of the boxes is clicked, the information about the intensity value of that gene in the respective sample along with its gene name, symbol, and Probe IDare displayed. Fig. 3.5 shows the level of gene expression across all the samples of the data set experiment in normal and diseased state.

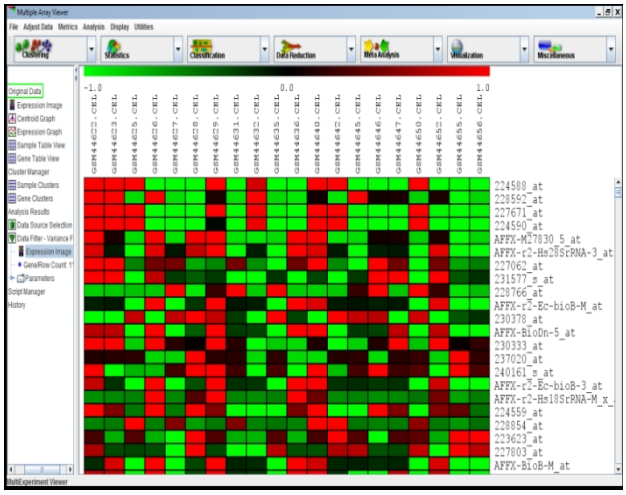

Fig. 3.9: Data image of GDS2143 (GSE2395)

3.4. Sam Results

When SAM test is performed on the normalized data, we obtained a set of positively significant genes and negatively significant genes. The genes were listed along with their respective fold change. Fig. 3.6 represents the positive and negative significant genes inferred by SAM test.

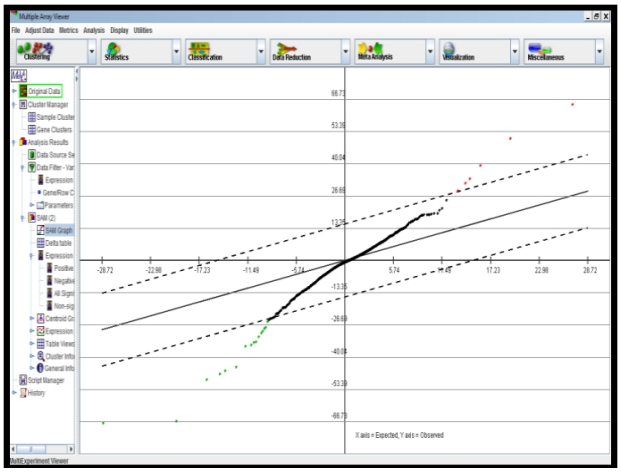

Fig. 3.10: SAM graph showing upregulated genes in red of HG-U133-A chip: GSE39843

3.5. T-test analysis

On performing T-test, we obtained a set of differentially expressed genes characterized by specific p-value for each gene as is observed in Fig. 3.8 and Fig. 3.9. If the p-values are low, the gene is highly significant and vice versa. Volcano plot was also used to visualize the genes differentiated as positively or negatively regulated by considering the fold change as well.

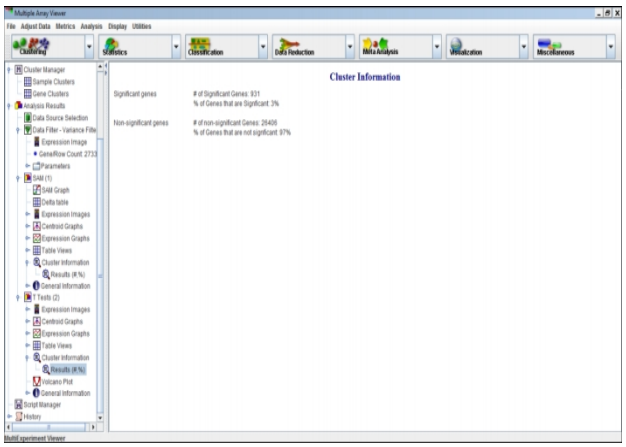

Fig. 3.11: Cluster information of HG-U133-A chip: GSE39843

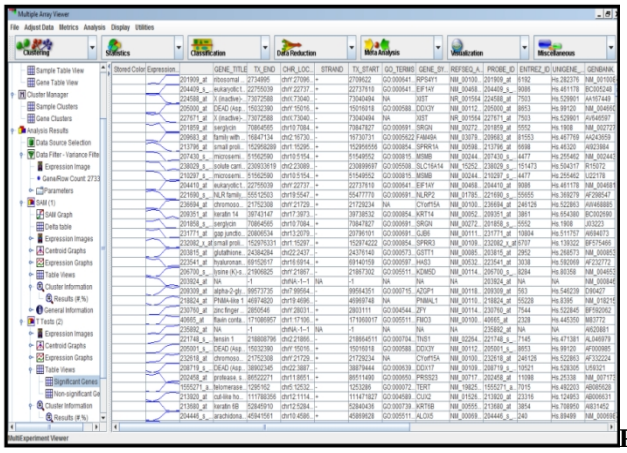

Fig. 3.12: Significant genes from T-TEST of HG-U133-A chip: GSE39843

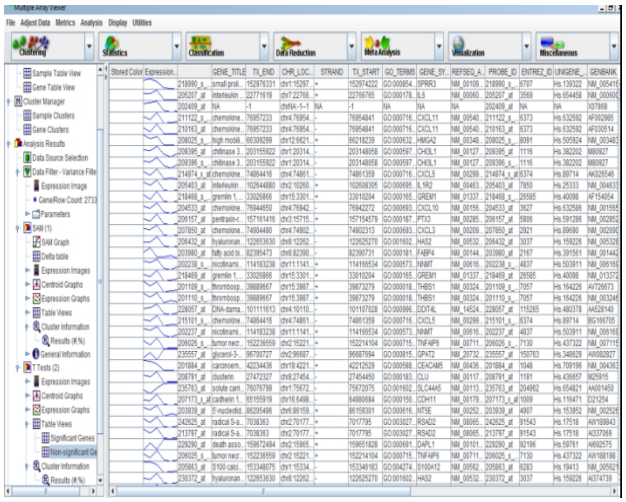

Fig. 3.13: Non-significant genes from T-TEST of HG-U133-A chip: GSE39843

3.6. Back validation analysis

This analysis was used for selection of upregulated and Downregulated genes using Ms.Excel. We segregated the samples of Schizophrenia and Cystic Fibrosis from datasets GSE12649 and GSE12654 and performed their back validation separately. The results were then compared to other datasets in order to identify upregulated and down regulated genes.

Table 1: Downregulated genes associated with cystic fibrosis

Gene symbol	Probe ID	Fold change
HLA-DQA1	212671_s_at	0.4
HLA-DQB1	212998_x_at	0.4
HLA-DQA2	212671_s_at	0.4
IRF1	238725_at	0.5
STAT1	AFFXHUMISGF3A/M 97935_3_at	0.5
FCGBP	203240_at	0.4

4. DISCUSSIONS

In cystic fibrosis, chronic bacterial infection (*Pseudomonas aeruginosa)* and an excessive inflammatory response is the primary cause of chronic respiratory infection which is responsible for most of the morbidity and mortality of these patients. Nitric oxide (NO) having antibacterial property acts in defense of lung epithelial cells by inhibiting the bacterial ribonucleotide reductase. This results into double stranded break in bacterial DNA thereby inhibiting proliferation of bacteria. Synthesis of NO is product of Interferon Regulatory Factor 1 (IRF1) mediated activation of NOS2 gene. Expression of the two genes is activated by Signal transducer for and activator of transcription 1 (STAT1). Study shows that STAT1 level were found to be increased in CF cells but the active phosphorylated form of STAT1 was bound to the protein inhibitor (PIAS1). In CF, the level of STAT1 inhibitor PIAS1(protein inhibitor of activated STAT1) is increased in CF patients which leads to interruption in normal STAT1 signaling. This results in reduction of IRF1 and NOS2 expression level in CF epithelial cell.

Major cause of death in patients with cystic fibrosis (CF) is colonization with *Staphylococcus aureus* and *Pseudomonas aeruginosa*. The wide phenotypic variation in CF patients suggests that genes other than the cystic fibrosis transmembrane conductance regulator (CFTR) gene modify the disease. Another set of downregulated genes

HLA-DQA1, HLA-DQA2 and HLA-DQB1 belonging to the 8.1ancestral Major Histocompatibility Complex (MHC) haplotype modify the disease. The MHC Class II antigens are found on antigen presenting cells (APC). The APC flag the class II antigens to T-cells each with unique T-cell receptor (TCR) variants. A few TCR variants that recognize these DQ/antigen complexes are on CD4+ T-cells (helper T-cells) which can promote the amplification of B-cells. These active B-cells then recognize a different portion on the same antigen.

Downregulated Fc fragment of IgG binding protein (FCGBP) also contributes in modifying the CF disease. More than 90% CF deaths occur due to IgG antibodies inefficient in reacting with the *Pseudomonas* infection. Current studies suggest the severity in Cystic Fibrosis is apparently due to defective opsonophagocytosis. This may be caused either by cleavage of IgG and complement complex mediated by immune system or change from opsonic to monophonic antibody isotope.

Pseudomonas aeruginosa produce a protease Elastase, possessing lactase activity is a major virulence factor in *Pseudomonas aeruginosa* that is believed to cause extensive tissue damage during infection in the human host. Elastase is capable of degrading elastin, collagens, immunoglobulins, complement components, laminin and fibrin. In accordance with substrates it utilizes, elastase of *P. aeruginosa* is also able to inhibit several processes involved in defense mechanisms of hosts. Elastase stimulates the production of proinflammatory mediators such as IL-8, which further induces neutrophil influx. Elastase also impairs mucociliary clearance by direct effects on ciliary function and by stimulating increased mucus production. Elastase produced by *Pseudomonas aeruginosa* cleaves human IgG at hinge region into immune fragments, Fab and Fc that significantly inhibit opsonophagocytosis and uptake of immune complexes by phagocytes. Since macrophages use antibodies to ingest *P. aeruginosa*, opsonophagocytosis is reduced in the presence of excess Elastase. It has been shown that the *Pseudomonas* uptake and killing were improved in absence of cleavage fragments.

The experimental evidence for the hypothesis is reported [8, 9, 10]. This proves the significance of the study.

5. ACKNOWLEDGEMENT

The authors are grateful to the Head of the Department of School of Biotechnology & Bioinformatics, D. Y. Patil University for providing the facilities to carry out the project.

6. REFERENCES

[1] P. M. Quinton, "Physiological Basis of Cystic Fibrosis: A Historical Perspective," *Physiological reviews*, vol. 79, no. 1, pp. S3-S22., 1999.

[2] B. Jeffrey, L. Carolyn and B. Gerald, "Lung Infections Associated with Cystic Fibrosis," *Clinical Microbiology Reviews*, vol. 15, no. 2, pp. 194-215, 2002.

[3] J. Riordan, J. Rommens, B. Kerem, N. Alon, R. Rozmahel, Z. Grzelczak, J. Zielenski and L. Tusi, "Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA," *Science*, vol. 245, no. 4922, pp. 1066-1073, 1989.

[4] M. J. Welsh, "Abnormal regulation of ion channels in cystic fibrosis epithelia," *FASEB*, vol. 4, no. 10, pp. 2718-2725, 1990.

[5] A. MH, "Cystic fibrosis transmembrane conductance regulator. Structure and function of an epithelial chloride channel," *Journal of Biological Chemistry*, vol. 275, no. 6, pp. 3729-32, 2000.

[6] J. Chen, X. Chang, A. Aleksandrov and J. Riordan, "CFTR is a monomer: biochemical and functional evidence," *The Journal of Membrane Biology*, vol. 188, no. 1, pp. 55-71, 2002.

[7] I. Bronsveld, F. Mekus, J. Bijman, M. Ballmann, U. Laabs, D. Halley , G. Mastella, S. Thomas , H. Veeze and B. Tummler, "Chloride conductance and genetic background modulate the cystic fibrosis phenotype of delta F508 homozygous twins and siblings," *Journal of Clinical Investigation*, vol. 108, no. 11, pp. 1705-1715, 2001.

[8] T. J. Kelley and H. L. Elmer, "In vivo alteration of IFN regulatory factor-1 and PIAS1 protein levels in cystic fibrosis epithelium," *The Journal of Clinical Investigation*, vol. 106, no. 3, pp. 403-410, 2000.

[9] S. Zheng,, W. Xu and S. Bose, "Impaired nitric oxide synthase-2 signaling pathway in cystic fibrosis airway epithelium," *American Journal Physiology. Lung Cell and Molecular Physiology*, vol. 287, no. 2, pp. L374-L381, 2004.

[10] J. Laki, I. Laki, R. Ujhelyi, O. Bede, K. Bolbas and A. Halasz, "The 8.1 ancestral MHC haplotype is associated with delayed onset of colonization in cystic fibrosis," *International Immunology*, vol. 18, no. 11, pp. 1585-1590, 2006.