Microarray Based Pharmacogenomics Study Of Celecoxib On Breast Cancer Samples

Vijay Laxmi Saxena[1], Pragya Chaturvedi[2], Neha Agrahari[3]

[1](Coordinator, Bioinformatics infrastructure facility center, Department of zoology, D.G.P.G. College, Kanpur, India)
[2](Research scholar, School of Biotechnology, KIIT University, Bhubaneswar, India
&
Bioinformatics infrastructure facility center, Department of zoology, D.G.P.G. College, Kanpur, India)
[3](Department of Biotechnology, D.G. P.G. College, Kanpur, India)

Abstract : *Microarrays are a powerful technique to monitor the expression level of thousands of genes in a single experiment under various conditions e.g. normal versus disease and control versus drug treated which offers a promising means to understand how cell react to environmental perturbation. Microarray technology has been phenomenal in the past few years. Cancer is a group of disease caused by uncontrolled growth and spread of abnormal cells. The celecoxib is usually used to treat patients who are suffering with osteoarthritis, rheumatoid arthritis, acute pain, painful menstruation but in this study the effect of celecoxib is seen on the patients who were affected with breast cancer and the novel genes which was expressed under certain experimental condition were identified using a microarray technique. Breast cancer is the most common invasive cancer in females worldwide. Breast cancer is the most frequently diagnosed cancer and the leading cause of cancer death among females, which accounts for 23% of the total cancer cases and 14% of the cancer deaths. Current study aims to identify genes involved in effect of celecoxib a drug over the breast cancer cell lines. Different methods of BRB Array Tool have been used to fulfill this approach. SAM and Hierarchical Clustering identified different genes. In SAM 944 genes were significant out of 4743 genes. So we can conclude that celecoxib affects breast cancer gene expression so it can be used for treatment of breast cancer.*
Keywords: *Breast Cancer, Celecoxib, Gene expression analysis, Microarray, Pharmacogenomics*

I. INTRODUCTION

Microarray technology is a high-throughput experimental technique that can measure expression levels of hundreds of thousands of genes simultaneously. Microarrays are gaining popularity in biological laboratories by the day. In their standard application, they measure the gene expression status of a particular sample, by quantifying those RNA levels of all genes in highly parallel fashion [1]. The goal of most microarray experiments is to survey patterns of gene expression by assaying the expression levels of thousand genes in a single assay. Microarray experiments investigate relationships between related biological samples based on patterns of expression, and the simplest approach looks for genes that are differentially expressed [2].Microarrays in their various instances have become a fundamental experimental technique for biomedical research. Gene-expression microarrays have become almost as widely used as measurement tools in biological research [3].

Cancer is now appreciated as not only a highly heterogeneous pathology with respect to cell type and tissue origin but also as a disease involving dysregulation of multiple pathways governing fundamental cell processes such as death, proliferation, differentiation and migration[4].Cancer is a condition where cells in a specific part of the body grow and reproduce uncontrollably. The cancerous cells can invade and destroy surrounding healthy tissue, including organs. Cancer sometimes begins in one part of the body before spreading to other areas. This process is known as metastasis. Cancer is a common condition. More than one in three people will develop some form of cancer during their lifetime [5].

Breast cancer is the most common cancer in women, the second leading cause of cancer death in women after lung cancer in the United States and the leading cause of cancer death in women worldwide. The risk of breast cancer increases with age, and the majority of women are diagnosed in their postmenopausal years. Half of all female breast cancer patients, however, receive their breast cancer diagnosis by age 61, and approximately 12 per cent are diagnosed at ages younger than 45. There are at least five different breast cancer subtypes each with distinct biologic features, clinical outcomes, and responses to therapy. Breast cancer is a global problem. Worldwide, breast cancer is the most commonly diagnosed malignancy and the leading cause of cancer deaths in women, accounting for 14 per cent of cancer deaths in women [6].Breast cancer occurs when a malignant (cancerous) tumour originates in the breast. As breast cancer tumours mature, they may metastasize

(spread) to other parts of the body. Breast cancer is not just a woman's disease. It is quite possible for men to get breast cancer, although it occurs less frequently in men than in women [7].

Celecoxib is a non-steroidal anti-inflammatory drug (NSAID) used in the treatment of osteoarthritis, rheumatoid arthritis, acute pain, painful menstruation and menstrual symptoms, and to reduce numbers of colon and rectum polyps in patients with familial adenomatous polyposis. It is marketed by Pfizer under the brand name Celebrex. In some countries, it is branded Celebra[8].

II.   TOOLS AND TECHNIQUES

2.1 The ARRAYEXPRESS (Database)
It is a public database for storing & providing access to high throughput functional genomic data. It is a public database of microarray gene expression data at EBI, which is a generic gene expression database designed to hold data from all microarray. It uses annotation standard MIAME & associated XML data exchange format Microarray Gene Expression Mark-Up Language (MAGE-ML) & is designed to store well annotated data in a structured way [10].

The data has been taken from Array Express Database. The information of the data is given below
E-MTAB-566 - Transcription profiling by array of breast cancer samples treated and non-treated with celecoxib
Status:-Released on 28 March 2013, last updated on 24 May 2013Organism:-Homo sapiens
Samples (74)
Description:-Biopsies and surgical specimens were obtained from breast tumours treated and non-treated with celecoxib 400mg twice daily for 2 weeks.

2.2 BRB ARRAY TOOL
 BRB stands for "BIOMETRIC RESEARCH BRANCH". BRB-Array Tools is an integrated software package for the analysis of DNA microarray data. It was developed by the Biometric Research Branch of the Division of Cancer Treatment & Diagnosis of the National Cancer Institute under the direction of Dr Richard Simon. [9]
VERSION OF BRB TOOL USED: - 4.3.1 version

2.3 SYSTEM REQUIREMENT

Processor: - Intel(R) Core(TM) i3-2100 CPU @ 3.10 GHZ
RAM: - 2.00 GB
System Type: - 32-bit Operating System

III.   METHODOLOGY
The data set which is used in this study was downloaded from Array Express database had Array Express I.D. E-MTAB-566. It contains 74 samples in which half of the patients were treated with celecoxib and half of them were remain untreated. The data is then analyzed using BRB Array Tool. BRB-Array Tools is an integrated software package for the analysis of DNA microarray data. The data is being re-filtered and normalized. The genes are being filtered using 2 fold changes. The results which have been analyzed are represented through heat map, hierarchical clustering and significance analysis of microarray (SAM).

IV.   RESULTS AND DISCUSSION:-
4.1 Cluster Analysis of Samples (Hierarchical)
In hierarchical clustering, genes with similar expression patterns are grouped together. The samples are being divided into two clusters i.e. cluster A and cluster B. These cluster A and Cluster B are further subdivided into sub-clusters 1,2,3,4. Here green colour represents those genes which are not treated with celecoxib, red colour represents those genes which are treated with celecoxib whereas black colour represents those genes which are neither treated nor untreated with celecoxib i.e. it is totally unexpressed.   In sub cluster 1 it is observed that mostly the patients who were not treated with celecoxib their genes are being expressed but some genes are expressing which were treated with celecoxib whereas some are unexpressed. In sub cluster 2 it is observed that most of genes were expressed which were treated with celecoxib and other genes were expressed which was untreated with celecoxib whereas some are not being expressed. In sub cluster 3 it is observed that mainly the patients not treated with celecoxib their genes are being expressed but some genes are expressing of patients which were treated with celecoxib whereas some are not expressing. In sub cluster 4 there are genes which are treated, untreated and unexpressed genes. Hence it is being concluded that in cluster A mainly the genes of celecoxib treated patients are being expressed whereas in cluster B mainly the genes of patients who were not treated with celecoxib are being expressed.
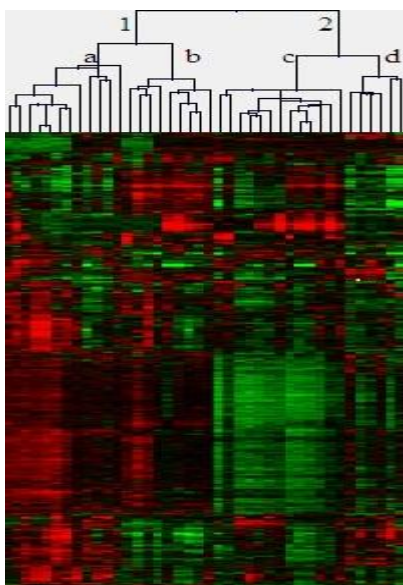
Fig 1- Hierarchical clustering of samples
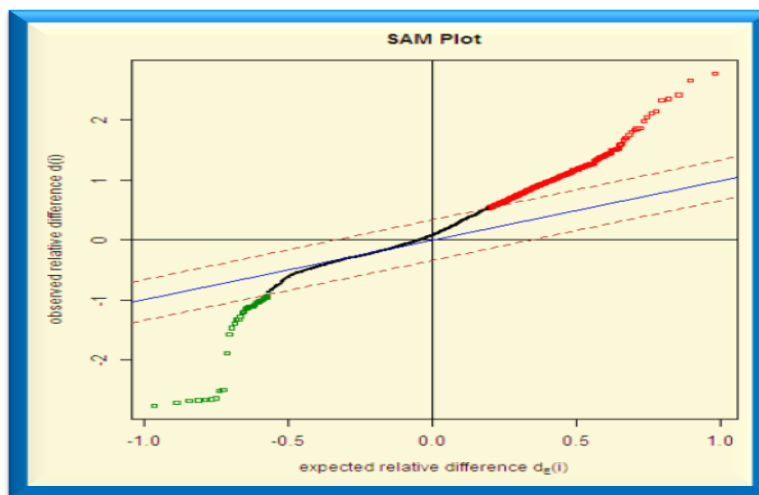
4.2 Significance analysis of microarray (SAM)

Fig. 2 Representation of SAM plot

The SAM plot is a scatter plot of observed relative difference $d(i)$ versus the expected relative difference $dE(i)$. In this SAM plot the significant genes are represented by square symbols. Positive significant genes are represented by red colour whereas Negative significant genes are represented by green colour.

This table represents the two number of classes i.e. class 1 and class 2. Here class 1 represents the breast cancer patients who were treated with celecoxib whereas class 2 represents the breast cancer patients who were not treated with celecoxib. The genes were then put under the filtering criteria so the total number of genes which have passed the filtering criteria is 4743. The target proportion of false discoveries is 0.1 and the percentile for determining called genes that are false is 90. The total number of genes that are significant by SAM is 944. In which there are 904 genes which are negatively significant whereas 40 genes are positively significant. Hence

the estimated false discovery rate among the 944 significant genes is 0.09983 and delta value used to identify the significant genes is 0.33789. The fudge factor for standard deviation is computed as 0.82597. The d(i) represents the observed relative difference for two-class.

V. CONCLUSION

Microarray expression analysis offers an opportunity to generate functional data on a genome-wide scale and consequently, should provide much-needed data for the biological interpretation of genes and their functions. It has also shown promise for classifying physiological and disease states. As the discussion presented here should show, the careful handling and interpretation of microarray expression data is not yet an exact science. The hypothesis behind using clustering techniques is that genes in a cluster must share some common function or regulatory elements. The use of microarrays in basic and applied research in drug discovery is only going to increase, but as these data sets grow in size, it is important to recognize that untapped information and potential discoveries might still be present in existing data sets. The field of microarray analysis has moved beyond the initial simple approaches and is becoming increasingly sophisticated as more powerful algorithms are being used to increase sensitivity and specificity. Microarray based experiments are frequently seen as the stronghold of hypothesis-free genome research. While debatable in itself, this assertion simply shifts the responsibility to the computational scientist analysing the data. The hierarchical clustering have been performed in which there are two clusters i.e. cluster A and cluster B. In cluster A mainly the genes of celecoxib treated patients are being expressed whereas in cluster B mainly the genes of patients who were not treated with celecoxib are being expressed. Significance analysis of microarray (SAM) have also been performed in which there are two number of classes i.e. class 1 and class 2. Here class 1 represents the breast cancer patients who were treated with celecoxib whereas class 2 represents the breast cancer patients who were not treated with celecoxib. In which 4743 genes passed the filtering criteria among which 944 genes were significant.

Acknowledgements

REFERENCES

[1] Adi L. Tarca, Roberto Romero, Sorin Draghici, Analysis of microarray experiments of gene expression profiling (2006) American Journal of Obstetrics and Gynecology, Volume 195.
[2] Alvis Brazma, Helen Parkinson, UgisSarkans, Mohammadreza Shojatalab, Jaak Vilo,Niran Abeygunawardena, Ele Holloway, Misha Kapushesky, Patrick Kemmeren, Gonzalo Garcia Lara, Ahmet Oezcimen, Philippe Rocca-Serra and Susanna-Assunta Sansone, ArrayExpress—a public repository for microarray gene expression data at the EBI (2003), Nucleic Acids Research, Vol. 31.
[3] American Cancer Society, Global Cancer Facts & Figures 2nd Edition. Atlanta: American Cancer Society; 2011.
[4] American Cancer Society. Breast Cancer Facts and Figures 2011-2012. Atlanta,American Cancer Society; 2011.
[5] Andre Koschmieder et al, Tools for managing and analyzing microarray data(2011), BRIEFINGS IN BIOINFORMATICS,VOL 13,10.1093/bib/bbr010.
[6] AtulButte,The Use And Analysis of Microarray Data,(2002) ,Nature Publishing Group, VOLUME 1.
[7] Breast Cancer and the Environment: Prioritizing Prevention Report of the Interagency Breast Cancer and Environmental Research Coordinating Committee
[8] Catherine Olivier and Bryn Williams-Jones,Pharmacogenomic technologies: a necessary "luxury" for better global public health(2011),BioMed Central, Volume 7.
[9] Dr. Richard Simon, Amy Lam BRB-Array Tools Version 3.0.
[10] Richard S. Segall Data Mining of Microarray Databases for the Analysis of Environmental Factors on Plants Using Cluster Analysis and Predictive Regression (2004) Arkansas State University, Department of Computer & Information Technology.