IN- SILICO ANALYSIS OF MICRO ARRAY DATA FOR PROSTATE CANCER

Singh Satendra*[1], Lall Rohit[2], Jain Prashant. A.[1]
[1]Department of Computational Biology & Bioinformatics, JSBB, SHIATS, Allahabad (India) – 211007
[2]Department of Molecular & Cellular Engineering, JSBB, SHIATS, Allahabad (India) – 211007
*Email: satendralike@gmail.com

ABSTRACT

The micro array data analysis for prostate cancer was carried out by clustering algorithms SOM and K-mean. The Genes were clustered into nine different clusters in both techniques based on the expression profile of those genes in prostate cancer. The expression of genes in some of clusters was found to be similar and genes were having both expressions i.e. under and over. Some of genes were found to be highly co-expressed in relation to other genes. Clustering results obtained by two techniques were approximately same and accurate. Sixteen identified genes were co-expressed in different clusters.

Keywords: Expression profile, K-mean, microarray, Self-Organizing Map, Prostate cancer.

INTRODUCTION

Prostate cancer is a disease in which cancer develops in the prostate gland of the male reproductive system. It occurs when cells of the prostate mutate and begin to multiply out of control. These mutations can be caused by radiation, chemicals and carcinogens or by certain viruses that can insert their DNA into the human genome. Mutations occur spontaneously and may be passed down from one cell generation to the next as a result of mutations within lines[1].

These cells may metastasize from the prostate to other parts of the body, especially the bones and lymph nodes. Prostate cancer may cause pain, difficulty in urinating, problems during sexual intercourse, erectile dysfunction and other symptoms. However these symptoms are present only in an advanced stage of the disease. It is the most common cancer in men older than age 50. Gene expression profiling by DNA microarrays has become an important tool for studying the transcriptome of cancer cells and has been successfully used in many studies of tumor classification and of identification of marker genes associated with cancer[2]. With an increasing number of available microarray data, the comparison of studies with similar research goals e.g. to identify genes being differentially expressed in normal versus tumor tissue, has gained high importance [3]. The microarray data for prostate cancer can be studied for the analysis of gene expression profile with the help of clustering algorithms. The present work was carried out with the objective to identify and retrieve microarray data for prostate tumor cell and to carry out the clustering for identification of co-expressed genes.

MATERIALS AND METHODS

A progress in micro array data generation for prostate cancer provides considerable resources for the in-silico analysis of its gene expression and defines genes that are important in the development of prostate cancer. For the purpose of the present study the microarray raw data of five thousand genes of eight different experiments conducted on different dates was retrieved from SMD (Stanford Microarray Database) in the Excel format. This raw data was then normalized because there are many sources of systematic variation in microarray experiments that affect the measured gene expression levels. [3]

After removing the bad quality data we were left with reliable data in the excel file. Clustering of final dataset was carried out for further analysis by using Genesis, which is a versatile and transparent software suite for large-scale gene expression cluster analysis. The software enables data import and visualization, normalization and clustering using K-mean and Self Organizing Maps. After uploading the dataset normalization was carried out as an attempt to remove the non-biological influences on biological data. The normalized data from excel file containing the expression values of the genes from different experiment was converted into a separate excel file for the final analysis. Clustering of the normalized data by using K-mean and Self Organizing Maps was carried out and in both the cases entire dataset was clustered into nine different clusters based on the expression profile of genes in the prostate cancer. [4]

By using K-mean nine clusters K1, K2, K3, K4, K5, K6, K7, K8, and K9 were generated. Similarly nine clusters S1, S2, S3, S4, S5, S6, S7, S8 and S9 were generated by using Self Organizing Maps. For the identification of co-expressed genes the expression profile of K-mean and Self Organizing Maps were compared and genes which showed similar expression were then compared for identification of co-expressed genes.

RESULTS AND DISCUSSION

The Genes were clustered into 9 different clusters in case of both techniques, based on the expression profile of those genes in prostate cancer.

Table 1: Distribution of Genes in different cluster by using SOM and K-mean clustering algorithm.

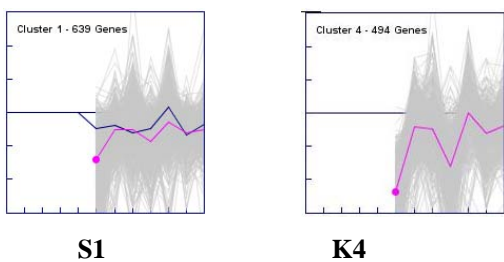| Self Organizing Map			K-mean		
Cluster	Genes	Genes (%)	Cluster	Genes	Genes (%)
S1	639	17	K1	606	11
S2	351	7	K2	761	15
S3	131	3	K3	652	13
S4	73	1	K4	494	10
S5	131	3	K5	202	4
S6	1668	34	K6	566	11
S7	1432	29	K7	531	11
S8	418	8	K8	366	7
S9	117	2	K9	782	16

S1   **K4**

Figure 1.1: Comparison of Cluster no.1 of SOM & Cluster no.4 of K-mean based on their expression profile

The expression of genes in cluster S1 and K4 of Figure 1.1 was found to be similar and under express as they have similar pattern of peaks, which represents the status of gene expressions.
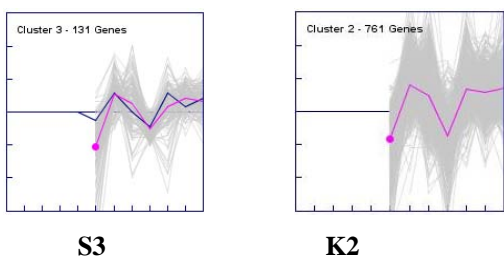

S3   **K2**

Figure 1.2: Comparison of Cluster no.3 of SOM & Cluster no.2 of K-mean based on their expression profile.

The expression of genes in cluster S3 and K2 of Figure 1.2 was found to be similar and some genes are showing under express while others are over expressed.
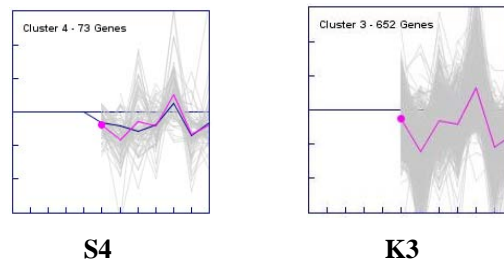

S4   **K3**

Figure 1.3: Comparison of Cluster no.4 of SOM & Cluster no.3 of K-mean based on their expression profile.

The expression of genes in cluster S4 and K3 of Figure 1.3 was found to be similar and most of the genes were under expressed.
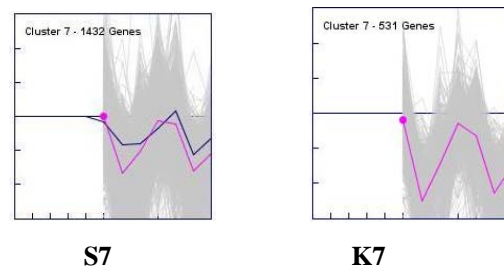

S7   **K7**

Figure 1.4: Comparison of Cluster no.7 of SOM & Cluster no.7 of K-mean based on their expression profile.

The expression of genes in cluster S7 and K7 of Figure 1.4 was found to be similar and under expressed.

Table 2: Genes which were present in clusters of similar expression profile and co- expressed.

S. no.	Gene no.	Gene Name
1	73	Hypothetical gene supported by AK026189
2	81	Growth arrest-specific 2
3	99	Polymerase (RNA) II (DNA directed) polypeptide
4	105	Adaptor-related protein complex 4, sigma 1 subunit
5	114	Hematological and neurological expressed 1
6	121	Rho/rac guanine nucleotide exchange factor (GEF) 2
7	145	Neuron navigator 2
8	149	Chromosome 7 open reading frame 11
9	170	Lymphotoxin alpha (TNF super family, member 1)
10	201	Transcribed locus Hs.561661
11	203	Transcribed locus Hs.163555
12	205	Synapsin II Hs.445503
13	230	Asparagine-linked glycosylation 8 homology (yeast, alpha-1,3-glucosyltransferase)
14	261	Transcribed locus Hs.282800
15	274	Transcribed locus Hs.156048
16	491	AA778627

By observing the results of SOM and K-mean it was observed that the expression patterns of most of the clusters were almost same and these clusters were also having almost same genes. Table 2 gives the list of co-

expressed genes, which was selected after the comparative analysis of clusters showing similar expression profile.

SUMMARY AND CONCLUSION

The micro array data analysis for prostate cancer was carried out by using Genesis. The data set was retrieved from SMD. Genesis was used to carry out for normalization of raw data retrieved from SMD and to carry out clustering of the genes by clustering algorithms. The results were generated with the help of SOM and K-mean techniques. The Genes were clustered into 9 different clusters in case of both techniques, based on the expression profile of those genes in prostate cancer.

In case of SOM; clusters S6, S7 and S1 had maximum number of genes. Genes that were clustered in S6, S7 and S1 were highly co-expressed. Similarly, In case of K-mean Cluster number K9, K2, K3 and K1 has maximum number of genes. Genes, that were clustered in cluster number K9, K2, K3 and K1, are highly co-expressed. The expression of genes in some of clusters has found to be similar and some genes were under express while others were over expressed.

Thus it can be concluded that clustering results obtained by two techniques were same and approximately accurate. Sixteen genes have been identified which were co-expressed in different clusters[5]. In future work promoter analysis can be carried out to analyze the regulatory systems of these sixteen genes. Drug target can be identified with the help of this regulatory system analysis. Functions of these sixteen genes are unknown and can be predicted on the bases of the known genes of similar cluster.

REFERENCES

1. DeMarzo AM, Nelson WG, Isaacs WB, Epstein JI. Pathological and molecular aspects of prostate cancer, (2003) Lancet; 361:955–64.

2. Lapointe J, Li C, Higgins JP, Gene expression profiling identifies clinically relevant subtypes of prostate cancer, (2004) Proc Natl Acad Sci; 101:811–6.

3. Jacques Lapointe, Chunde Li, , Craig P. Giacomini, Keyan Salari, Stephanie Huang, Pei Wang,Michelle Ferrari, Tina Hernandez-Boussard, James D. Brooks, and Jonathan R. Pollack ,Genomic Profiling Reveals Alternative Genetic Pathways of Prostate Tumorigenesis, (2007) Cancer Res; 67:18.

4. Tomlins SA, Rhodes DR, Perner S, Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer, (2005) Science; 310:644–8.

5. Pollack JR, Perou CM, Alizadeh AA, Genomewide analysis of DNA copy-number changes using cDNA microarrays, (1999) Nat Genet; 23:41–6.

Concern Websites:

1. http://www.helsinki.fi/biochipcenter

2. http://www.microarrays.btk.utu.fi

3. http://www.med.uio.no/dnr/microarray/

4. http://www.genome.tugraz.at/genesisclient/1.7.2/install.htm

5. http://www.smd.Stanford.EDU/

6. http://www.ncbi.nlm.nih.gov.in

7. http://www.aacr.gov

****