International Journal of PharmTech Research
CODEN (USA): IJPRIF ISSN : 0974-4304
Vol.4, No.4, pp 1432-1438, Oct-Dec 2012

In Silico Microarray Data Analysis of over-expressed LYN Gene responsible for Adult T-cell leukemia/lymphoma (ATL)

Pallavi Gangwar*, Santosh Kumar, Ankur Mohan

BCS-InSilico Biology, Lucknow (U.P.), India

Corres. author: g.pallavi2007@gmail.com
Contact Number: +91-8756093333

Abstract: Adult T-cell leukemia/lymphoma (ATL) is a rare cancer of the immune system's own T-cells. Human T cell leukemia/lymphotropic virus type 1 (HTLV-1) is believed to be responsible for it. ATL is usually a highly aggressive non-Hodgkin's lymphoma with no characteristic histologic appearance except for a diffuse pattern and a mature T-cell phenotype. Circulating lymphocytes with an irregular nuclear contour (leukemic cells) are frequently seen. ATL is frequently accompanied by visceral involvement, hypercalcemia, lytic bone lesions, and skin lesions. Most patients die within one year of diagnosis. Co-regulated genes may share similar expression profiles, may be involved in related functions or regulated by common regulatory elements. There are different approaches to analyse the large-scale gene expression data in which the essence is to identify gene clusters. This approach has allowed us to (i) determine expression profiles of previously described developmentally regulated genes: in this work raw data of patients infected by HTLV-1 is used as a sample where three over-expressed genes such as LYN, were found. (ii) identify novel developmentally regulated genes: here Co-expressed genes of these three over-expressed genes were analyzed by using Insilico approaches such as Clustering (HCL, KMC) and phylogenetic analysis.
Keywords: Leukemia, histologic, T-cell phenotype, skin lesions, hypercalcemia, gene expression, HTLV-1, phylogenetic analysis.

Introduction:

Adult T-cell leukemia-lymphoma (ATLL) is an HTLV-1-associated lymphoproliferative malignancy that is frequently fatal [1]. ATLL is a peripheral T-cell malignancy associated with human T-cell lymphotropic virus type I (HTLV-1) infection that develops after a very long latency period. Clinically, ATLL is classified into four subtypes: acute, lymphoma, chronic and smoldering type. Although the prognosis of chronic and smoldering-type ATLL is relatively good, that of patients with acute- or lymphoma-type ATL still remains extremely poor [2]. Leukemia is preceded by oligoclonal expansions arising from a polyclonal background of activated HTLV-1-infected T cells as a result of the expression of the viral transactivator protein Tax, which activates various cellular genes [3,4] and creates an autocrine loop involving IL-2, IL-15 and their cognate receptors. Patients with aggressive ATL, either acute or lymphoma type, generally have a poor prognosis because of intrinsic chemoresistance of the malignant cells, a large tumor burden frequently associated with multiorgan failure, hypercalcemia and/or frequent infectious complications due to profound T-cell immune deficiency [5]. It is a unique T-cell cancer first described in Japan. Chemoresistance is considered to be due to multiple factors, including

overexpression of the multidrug resistance protein, *TP53* mutations and dysregulation of various cellular oncogenes in ATL cells [6]. On the other hand, patients with indolent ATL, either chronic or smoldering type, have a better prognosis.

Family studies showed that the routes of natural infection of HTLV-I are from mother to child and also from husband to wife. The third route is blood transfusion. The borderline between the healthy carrier state and smoldering ATL remains unclear. In the endemic areas smoldering ATL is frequently diagnosed in patients with fungus infection of the skin, chronic lymphadenopathy, interstitial pneumonitis, chronic renal failure, and stronglyloidiasis [7].

A microarray is an array of DNA molecules that permit many hybridization experiments to be performed in parallel. It can monitor expression levels of thousands of genes simultaneously. Microarray emerged in late 90s as a high-throughput technology for gene expression analysis. It has become a powerful tool for biomedical research [8]. In just a few years, microarrays have gone from obscurity to being almost ubiquitous in biological research. Microarray Data Analysis is one of the best and most widely used techniques in bioinformatics to study gene expression, disease diagnosis, target identification, gene screening, marker mapping and other developmental biology. Raw data of this microarray work was available in Stanford Microarray Database. At the same time, the statistical methodology for microarray analysis has progressed from simple visual assessments of results to a weekly deluge of papers that describe purportedly novel algorithms for analysing changes in gene expression [9].Various microarray gene clustering algorithms like hierarchial clustering, self organizing maps and k-means are found useful for discovering groups of correlated or co-expressed genes potentially co-regulated or associated to the disease or conditions under investigation.

Microarray Experiment:

Microarrays are a novel technology that facilitates the simultaneous measurement of thousands of gene expression levels [10]. The DNA microarray is made out of a glass, plastic or silicon chip. This chip has many microscopic DNA spots on its surface which forms the array. The DNA spots are known as probes, because they are probing the sample which is hybridized to the chip. The sample which contains cDNA is called the target since the probes are looking to match these targets. Microarray experiments are typically made to compare two or more samples which represent two or more conditions, one being for example a cell that has mutated into a cancer cell and the other a normal cell [11].

We identified several genes anomalously over-expressed in the ATLL leukemic cells at the mRNA level, including LYN, CSPG2, and LMO2, and confirmed LMO2 expression in ATLL cells at the protein level. In this report we performed gene clustering on the raw microarray data sets provided by Stanford Microarray Database for LYN expression in various scenarios of T-cell leukemia-lymphoma. Gene ontology gave a further insight into cellular and molecular processes of LYN. Using GENESIS's HCL (hierarchial clustering) and K-means analysis, we found 70 common genes in clusters, out of which 10 genes were considered to be harmful after reading their literature in GeneCard and GENE (NCBI). Further a phylogenetic analysis was done on the 10 harmful genes using MEGA 5.05 which gave us the 10 most closely related genes to our target gene LYN. Thus, these closely related 10 genes can also be considered as potential novel targets for designing drugs for T-cell Leukemia-Lymphoma in near future of biomedical research.

Materials and Methodology

Data retrieval:

The Stanford Microarray Database (SMD; http://genome-www.stanford.edu/microarray/) serves as a microarray research database for Stanford investigators and their collaborators. In addition, SMD functions as a resource for the entire scientific community, by making freely available all of its source code and providing full public access to data published by SMD users, along with many tools to explore and analyze those data [12]. The SMD was used to obtain the raw data for microarray data analysis. Under the publications section of SMD, organism *Homo sapiens* was selected. The data was taken from List data for Publication for *Homo sapiens*. **Citation:** Alizadeh AA, et al. (2010) Leuk Lymphom**.** **Topic:** Expression profiles of adult T-cell leukemia-lymphoma and associations with clinical responses to zidovudine and interferon alpha. Raw data was downloaded and saved. The data was present in the form of excel sheets. Total of Eighteen excel sheets were present. These contained data from eighteen different experiments.

Export data in Genesis

On SMD this raw data are provided in 18 excel file sets each file contain expression data of different time and different pH. The raw data files are sorted and scaled by taking logarithm at base 2 of R/G normalized (mean) ratio. All excel files were merged in one excel file. Then data is normalzed by the rule if missing value in a row more than 80% then delete that row. After normalization we got 5222 genes in excel file. Finally this file is imported in the genesis [13].

Genesis:

A versatile, platform independent and easy to use Java suite for large-scale gene expression analysis was developed. Genesis integrates various tools for microarray data analysis such as filters, normalization and visualization tools, distance measures as well as common clustering algorithms including hierarchical clustering, self-organizing maps, k-means, principal component analysis, and support vector machines. The results of the clustering are transparent across all implemented methods and enable the analysis of the outcome of different algorithms and parameters. Additionally, mapping of gene expression data onto chromosomal sequences was implemented to enhance promoter analysis and investigation of transcriptional control mechanisms [14].

Clustering for data:

Here Hierarchical cluster (HCL) is obtained as output by using microarray gene expression data in cluster which can visualize in tree-view as a hierarchical tree. It has been found that this tree contain all given data in a hierarchical form. According to gene expression value, closely related (co-express) gene would in same cluster. By using different correlation type it has been found that the centered correlation is better and suitable for hierarchical clustering and gives more appropriate output for further process [15]. Then through manual subclustering we got 34 clusters. For the k means clustering numbers of clusters of hcl is used (Fig 1 & Fig: 2). After that k means clustering was done, parameter number of cluster was 34 and maximum iteration of 2000 was selected.

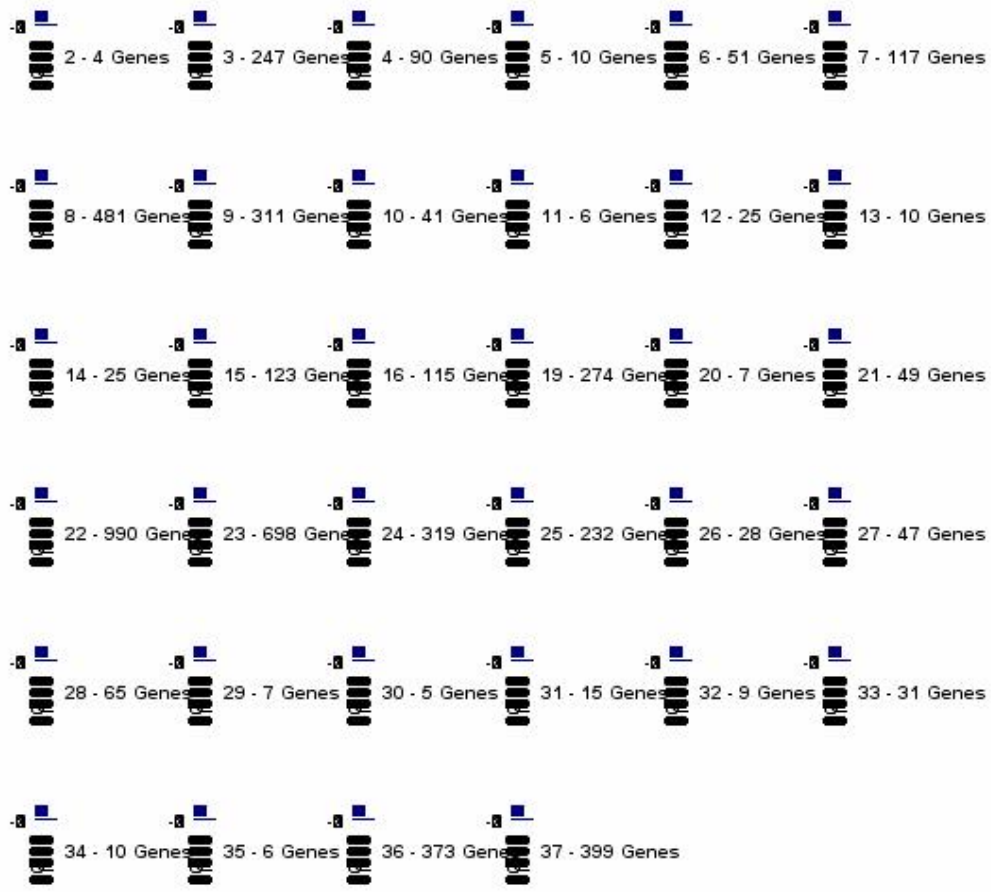

Fig: 1 Clusters obtained by HCL method.

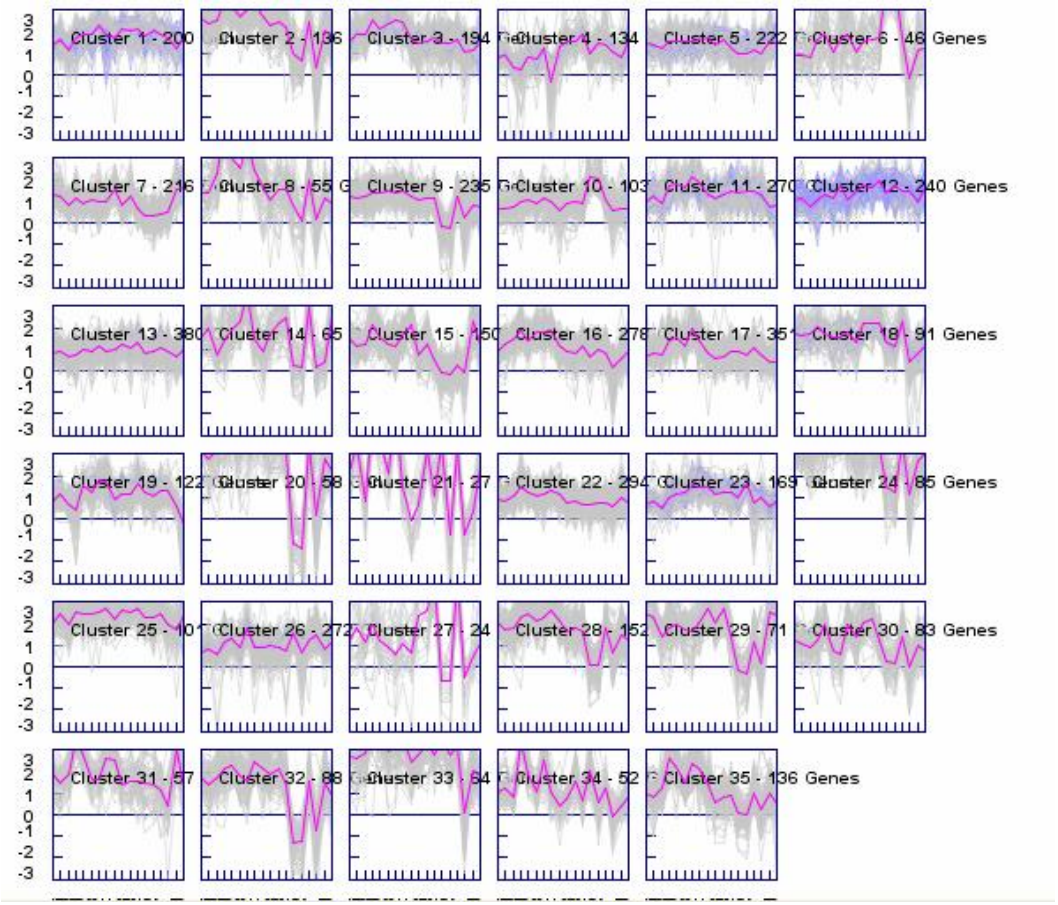

Fig: 2 Clusters obtained by KMC method

Comparision of clusters

Each and every cluster of hcl is compared with the k means clusters (34 clusters). It was done manually by studying each and every gene independently. For clusters which were having same genes (by comparing hcl and k means clusters), their gene list were obtained.

.

Phylogenetic analysis of our ten important genes

Phylogenetic analysis is done to find the evolutionary relationship between these genes. Genes which are evolutionary related can have similar sequences and similar promoters. So we can exploit this property for preparing potential drug targets against the proteins coded by those genes.

Result and Discussion

Prediction of coexpressed genes

From the manual analysis of k means clusters and HCl clusters, 2 clusters were found which were showing same gene contained, which is shown in table1.

Table1: cluster of HCl matching with K means cluster and their expression pattern

Cluster no. of HCl	Cluster no of k means	Shown in figure
HCL 4	K24	Fig: 3
HCL 21	K22	Fig: 4

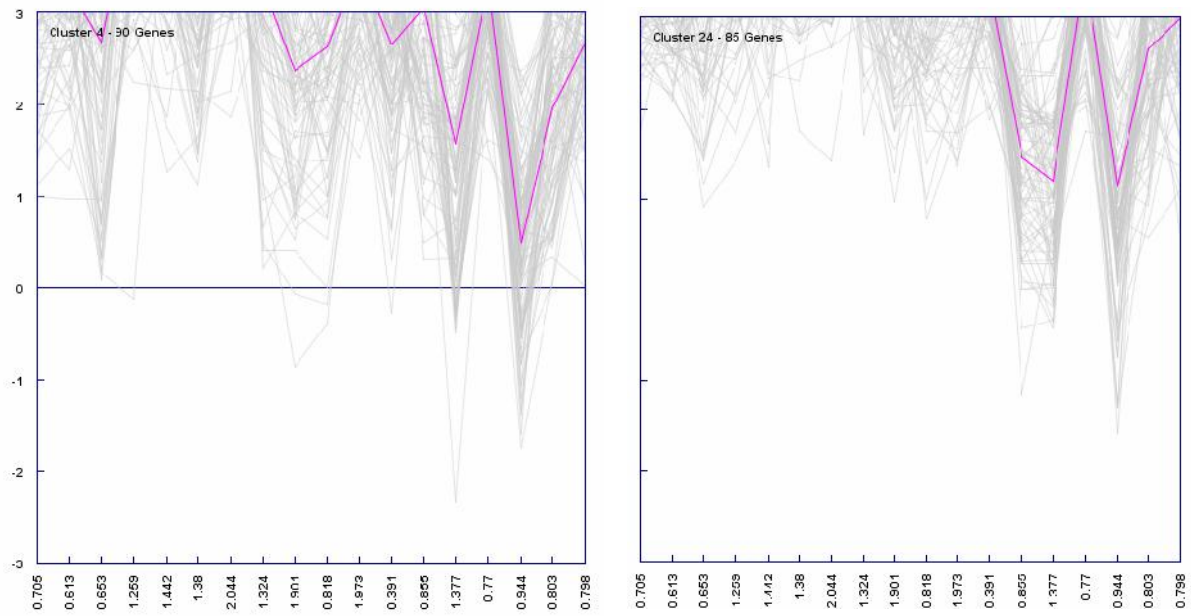

Fig 3: Expression Pattern of HCL4 & K24

Common genes in clusters

The genes of these selected clusters 'seed cluster' can be used for further analysis. This both cluster contains different number of gene which shows same gene expression. By applying clustered gene in genesis and some other plotting option for gene expression pattern analysis, plot by them shows that according to time period and provided environment condition, the expression of gene become changed, and this change is observable. This fluctuation seems same for most of genes which are in same cluster Figure 4 & 5. Gene expression pattern shows that, when we going to calculate the variance for all gene at every given condition, we found that at some point it is very high for some gene. It is shows that after clustering there is a chance of getting some false positive gene in cluster. For those genes whose function is tilled not known but are came in these 2 clusters we can conclude that they may responsible for ATLL. By phylogenetic analysis we have seen that 1KBKE gene is the closest to our target gene (LYN) and a breast cancer oncogene (Fig: 5).

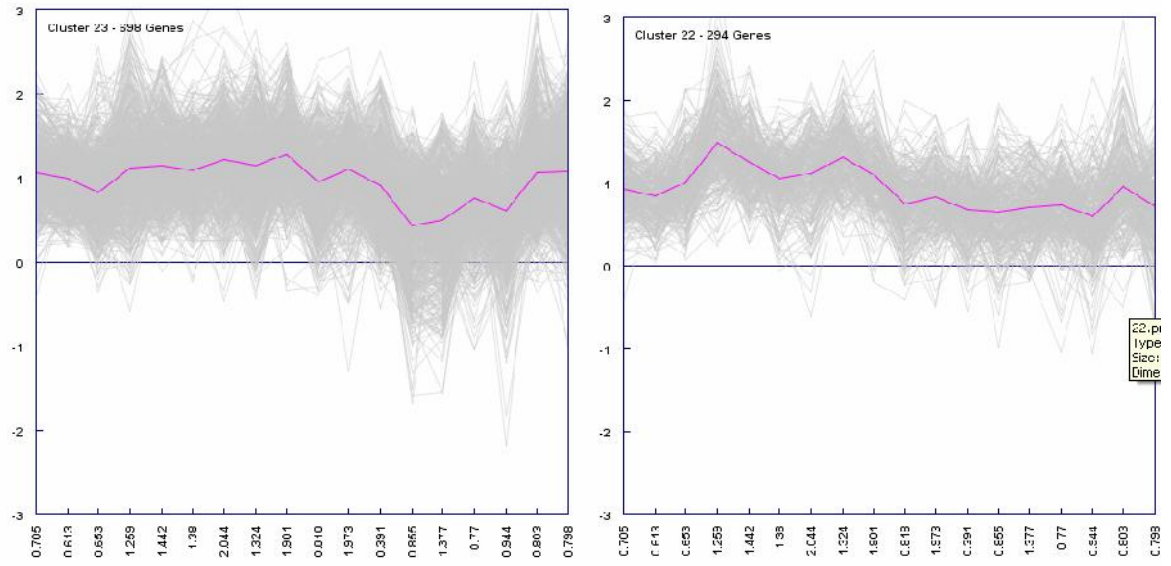

Fig 4: Expression Pattern of HCL 21 & K22

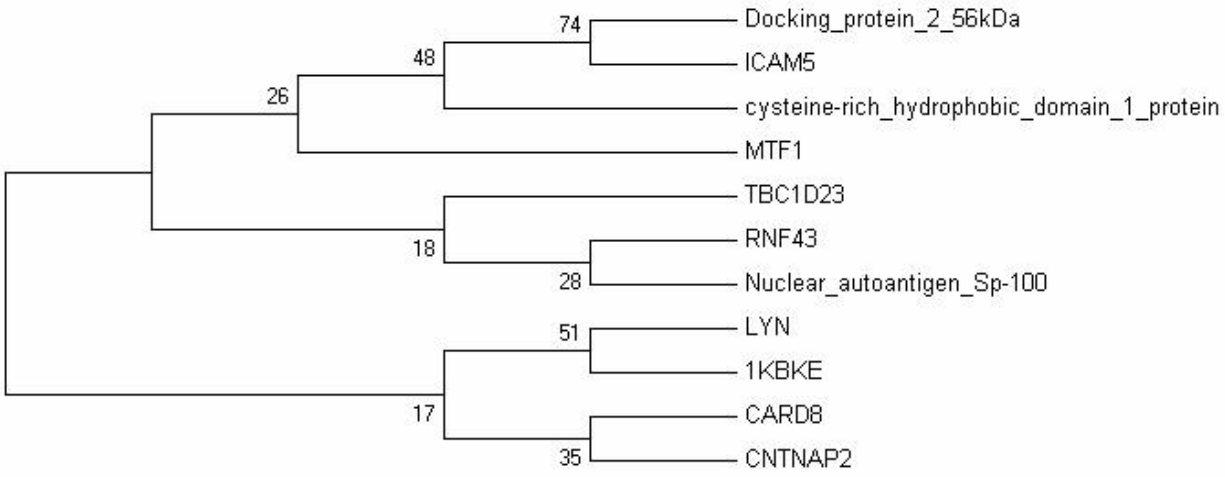

Fig 5: Phylogenetic analysis of 10 harmful genes

Conclusion

Clustering result from both the methods (HCL clustering and K-means clustering) shows that genes which are common in specific clusters of Hierarchical Clustering and cluster of k-means clustering(k24 HCl4 , k22 HC21) have similar expressions pattern of the respective clusters (which are present in both type of clustering) are also same. Finally it was concluded that common genes of both clustering methods, viz-aviz the different clusters, obtained by comparing the genes of clusters, k24 hcl4, k22 hcl21differentially coexpressed. Thus on the basis of comparative analysis this can be concluded that the coexpression is present within the genes of the same clusters.

.

References:

1) Alizadeh AA, Bohen SP, Lossos C, Martinez-Climent JA, Ramos JC, Cubedo-Gil E, Harrington WJ Jr, Lossos IS. (2010): Expression profiles of adult T-cell leukemia-lymphoma and associations with clinical responses to zidovudine and interferon alpha. 51(7):1200-16. (Pubmed)
2) Tobinai K, Watanabe T, Jaffe ES. Human T-cell leukemia virus type-I-associated adult T-cell leukemia-lymphoma. In: *Non-Hodgkin Lymphoma, Second Edition (Chapter 27)*. Mauch PM, Armitage JO, Coiffier B, Dalla-Favera R, Harris NL (Eds.), Lippincott Williams & Wilkins, PA, USA, 404–414 (2010).
3) Franchini G. Molecular mechanisms of human T-cell leukemia/lymphotropic virus type I infection. Blood 86(10), 3619–3639 (1995).
4) Matsuoka M, Jeang KT. Human T-cell leukaemia virus type 1 (HTLV-1) infectivity and cellular transformation. Nat. Rev. Cancer 7(4), 270–280 (2007).
5) Shimoyama M. Chemotherapy of ATL. In: *Adult T-cell Leukaemia*. Takatsuki K (Ed.). Oxford University Press, Oxford, UK, 221–237 (1994).
6) Taylor GP, Matsuoka M. Natural history of adult T-cell leukemia/lymphoma and approaches to therapy.*Oncogene* 24(39), 6047–6057 (2005).
7) Kiyoshi Takatsuki, Kazunari Yamaguchi, Fumio Kawano, Toshio Hattori,Hiromichi Nishimura, Hiroyuki Tsuda, Isao Sanada, Kiyonobu Nakada, and Yayeko Itai (1985): Clinical Diversity in Adult T-Cell Leukemia-Lymphoma. J. Cancer Research, September 1985 45;4644s
8) Module 10: Microarray Data Analysis I Yan Cui (ycui2@utmem.edu http://compbio. utmem.edu/MSCI814/Module10.htm

9) David B. Allison, Xiangqin Cui, Grier P. Page & Mahyar Sabripour Nature Reviews Genetics 7, 55-65 (January 2006)
10) Bradley Efron, Robert Tibshirani, John D Storey andVirginia Tusher, Empirical Bayes Analysis of a Microarray Experiment,Journal of the American Statistical Association, Volume 96, Issue 456, 2001 pages 1151-1160
11) David J. Lockhart & Elizabeth A. Winzeler, Genomics, gene expression and DNA arrays, NATURE | VOL 405 | 15 JUNE 2000 |
12) Jeremy Gollub, Catherine A. Ball, Gail Binkley1, Janos Demeter, David B. Finkel stein, Joan M. Hebert3, Tina Hernandez - Boussard, Heng Jin,Miroslava Kaloper, John C.Matese, Mark Schroeder, Patrick O. Brown, David Botstein and Gavin Sherlock, Oxford Journals Life Sciences Nucleic Acids Research Volume 31, Issue 1 Pp. 94-96.
13) Alexander Sturn, John Quackenbush, Zlatko Trajanoski. (2002). Genesis: cluster analysis of microarray data, 18.1. 207-208
14) Alexander Sturn, John Quackenbush and Zlatko Trajanoski1, 2000. Genesis: cluster analysis of microarray data, Oxford Journals Life Sciences & Mathematics & Physical Sciences Bioinformatics Volume 18, Issue 1Pp. 20
15) Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci 95: 14863-8.

*****