ISSN:0975-9646

Microarray Analysis and Gene Expression : A simplified Review

Saad Subair[1] and Hussah AlEisa[2]
[1,2]College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh, KSA

Abstract -Microarray technology has been advancing rapidly during the last decade. The development of powerful robot machines for DNA microarray experiments, new hybridization techniques, and increasing genome-sequencing data sets made it possible to improve the quality and accuracy of microarray experiments. Powerful computers and hardware help in analyzing and classifying microarray huge data. Microarray analysis may provide significant information on diseases mechanisms, resistance to certain drugs, and response to cells interactions. As far as cancer is concerned, microarray analysis may lead to improved early diagnosis and creative treatments to this disease.

Keywords: DNA Microarray analysis, Classification Methods, Gene Expression, Bioinformatics

1. INTRODUCTION

This article has been written to elucidate the microarray analysis and gene expression experiment and interpretation of the results in a simplified way. The article aims balanced and augmented information for both computer scientists and biologists to work in a team to conduct successful microarray experiments. Simply proteins or amino acids are the building blocks of the cells. Genes are specific fragments of the DNA that contain the instructions for creating proteins. The mRNA which is a messenger molecule is used by the genes to direct protein production [1]. Microarrays technologies measure the amounts of mRNA produced by genes. Microarrays are used to compare patterns of gene activity between different cell types or the same cell type under different conditions. Microarrays now are used in distinguishing between specific types of cancers [2, 3, 4, 5].

In an organism, active genes are in certain cells though most cells contain genetic materials. Studying patterns of gene expression allows understanding how cells function normally and when they are abnormal. DNA microarrays measure gene expression within a single sample or compare activity in different cell samples, such as healthy and diseased [2, 4]. The most commonly used microarray approaches today can be divided into two groups, according to the arrayed material: complementary DNA (cDNA) microarrays and oligonucleotide microarrays. Unlike most traditional molecular biology tools, which generally allow the study of a single gene or a small set of genes, microarrays facilitate the discovery of totally novel and unforeseen gene functionalities [2, 3, 4].

2. THE MICROARRAY EXPERIMENT

Microarray technology aims at the measurement of mRNA levels in particular cells or tissues for many genes at once [2]. The principle of a microarray experiment is that mRNA from a given cell or tissue is used to generate a labeled sample, which is hybridized to a large number of DNA sequences, immobilized on a solid surface in an ordered array. Several thousands of transcripts can be detected and quantified simultaneously [2,5]. Figure 1 shows the microarray analysis experiments and its steps.

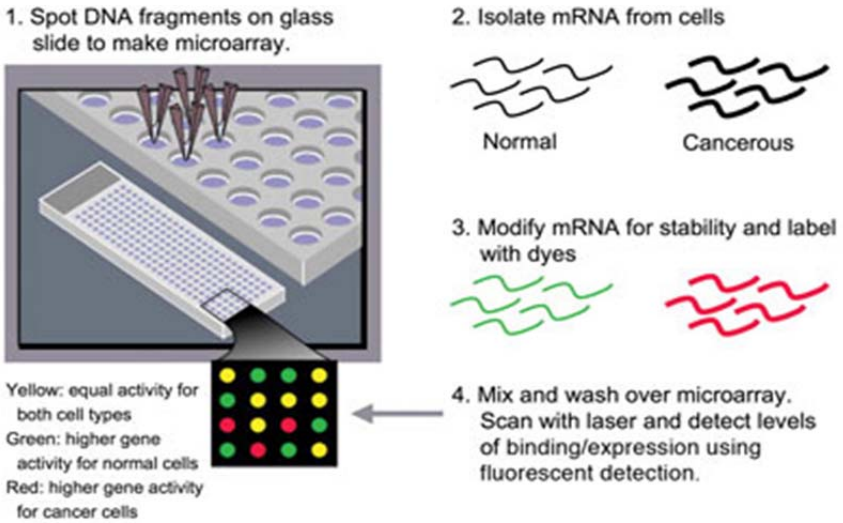

Figure 1: The fundamental Steps in the Microarray Experiment

A microarray may contain thousands of spots and each spot may contain a few million copies of identical DNA molecules that uniquely correspond to a gene. The data of a microarray experiment typically constitute a long list of measurements of spot intensities and intensity ratios, generated either by pairwise comparison of two samples or by comparing several samples to a common control. The challenge is then to investigate through this heap of data to find meaningful results [6]. What microarrays can achieve in analyzing gene expression is a challenge. Microarrays can be used to investigate problems in cell biology in various ways. The experimental approaches can be one of two approaches: Local approach in which the interest is only to finding a single change in gene expression. The other approach is the Global approach in which the interest is to look at the overall patterns of gene expression in order to understand the architecture of genetic regulatory networks [4]. So far, much of the interest in microarrays has been directed towards identifying individual genes, where the regulated expression of these genes can explain particular biological observation. The problem is that microarray experiments, whether based on oligonucleotides or cDNA, are highly capable of generating long lists of genes with altered expression, but they provide few clues as to which of these changes are important in establishing a given phenotype [7,8] . However, careful experimental design of a microarray experiment is critical [9]

3. DESIGN OF THE EXPERIMENT

There are many ways to conduct a DNA microarray experiment and analyze the microarray data. One of the techniques is to assemble the DNA microarray using a machine, mostly a robot to attach small pieces of a gene to a single spot on a glass or silicon slide or any suitable membrane [10, 11]. By arranging other genes elsewhere on the slide in a grid-like pattern (cDNA), thousands of individual spots are organized in a small space then genes can be included. After assembling the microarray, the sample will be isolated and modified by linking it to a fluorescent dye and then the mRNA is washed over the microarray. The mRNA from a specific gene will bind to its corresponding DNA spot on the slide. If the gene is very active, a high level of binding will occur. [10, 11]. Careful experimental design is crucial for a successful microarray experiment. Design issues depend in part on the exact array technology used. However, choosing an array technology may be the first design choice. [10, 11]. Depending mainly on the design of the experiment, the fundamental goal of most microarray experiments is to identify biological processes or pathways that consistently display differential expression between groups of samples. This can be accomplished by examining individual genes or identify groups of functionally related genes to detect differential gene expression [10, 12]. The flow of the microarray experiment is shown in Figure 2.

Hybridization

Hybridization is a phenomenon in which single-stranded deoxyribonucleic acid (DNA) or ribonucleic acid (RNA) molecules anneal or bind to complementary DNA or RNA.[2] Once the samples have been differentially labeled, they are allowed to hybridize onto the same glass slide. Here, any cDNA sequence in the sample will hybridize to specific spots on the glass slide containing its complementary sequence. The amount of cDNA bound to a spot will be directly proportional to the initial number of RNA molecules present for that gene in both samples [13]. Scanning the microarray slide with a laser activates the fluorescent dye attached to the bound mRNA. Measuring the amount of fluorescence at each spot associates with the activity of the genes represented at each position on the microarray. [14].

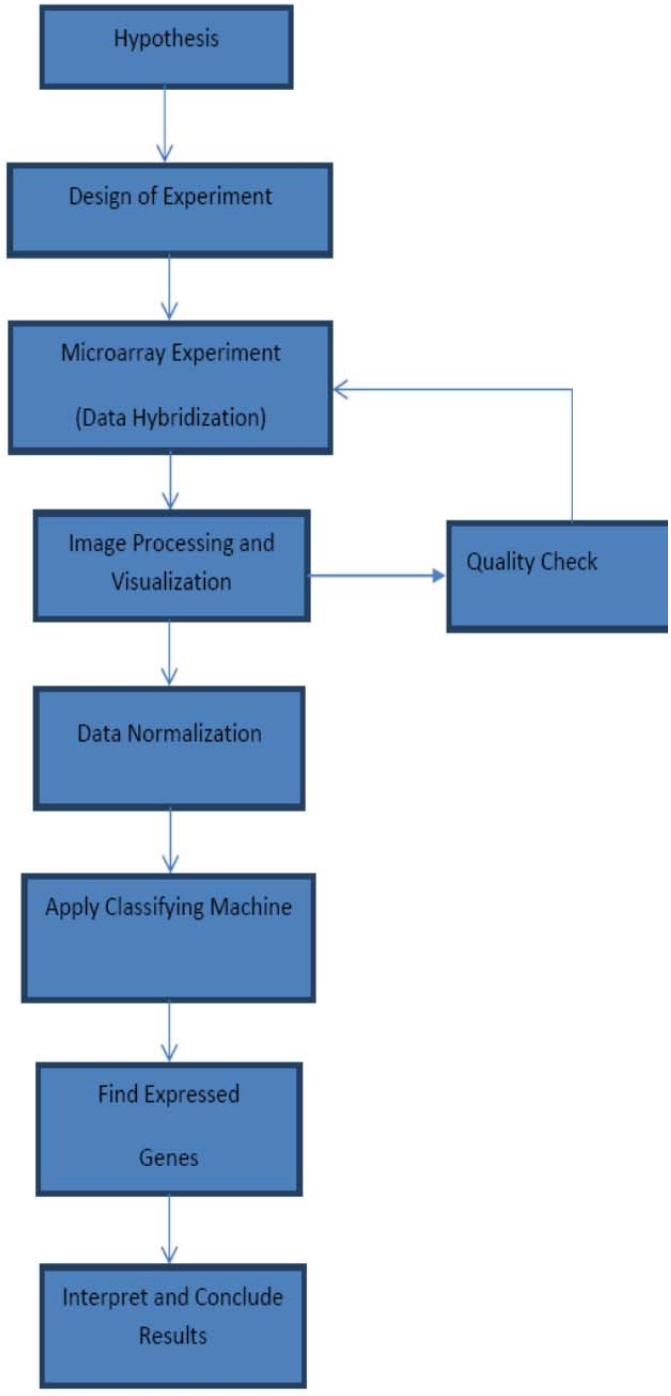

Figure 2: The flow of the microarray experiment

4. IMAGE PROCESSING AND VISUALIZATION

Following the hybridization step, the spots in the hybridized microarray are excited by a laser and scanned at suitable wavelengths to detect the red and green dyes. The amount of fluorescence emitted upon excitation corresponds to the amount of bound nucleic acid [14]. To ensure data quality, visualization of the data is a vital process. Many methods for visualization, quality assessment, and data normalization have been in existence [15, 16]. Clustering has been known as a way of finding and visualizing patterns in the data.[16]. By this way microarray technology has become a brilliant method to classify types of cells. This ability to differentiate expression patterns has been especially useful in many areas especially cancer research. The activities of genes can even lead to the identification of different treatments to different cancers. In general, mRNA from cells or tissue is extracted, converted to DNA and labeled, hybridized to the DNA elements on the array surface, and detected by phospho-imaging or fluorescence scanning [13]. Thus, what is seen at the end of the experimental stage is an image of the microarray, in which each spot that corresponds to a gene has an associated fluorescence value representing the relative expression level of that gene. Microarray image processing uses differential excitation and emission wavelengths. These images are then analyzed to identify the spots, calculate their associated signal intensities, and assess their local noise [16]. Most image acquisition software contains basic filtering tools to process these spots. These results allow intensity to be calculated for every spot on the chip. The products of the image acquisition are the TIFF image pairing and a non normalized data file [13, 14, 15, 16].

5. DATA NORMALIZATION

Normalization is an attempt to remove the technical variation in the microarray experiment while leaving the biological variation intact. The quantity of data generated in a microarray experiment typically requires a dedicated database system to store and organize the microarray data and images. The first role of a local microarray database is the storage and annotation of microarray experiments [15, 17]. Once data have been loaded into the database, it is normalized, and statistics is applied. Normalization is a process that scales spot intensities such that the normalized ratios provide an approximation of the ratio of gene expression. There are many methods of normalizations [15].Table 1 shows some microarray databases, their resources, and description.

6. DATA ANALYSIS AND CLASSIFICATION

Analyzing microarray data takes considerably more time than the laboratory procedures required to generate the data. Assessing the quality of the data and ensuring that all samples are comparable for further analysis is a crucial task.[18]. Clustering analysis is probably the most widely used statistical means in microarray data analysis. In many cases, clustering analysis is considered as a convenient way to display the information present in the data set. The choice of the clustering algorithm has a strong influence on the final result of the microarray data [19, 20, 21, 22]

It is a challenging task to analyze large data sets of gene expression. Generally there are two approaches to classify the gene expression data: supervised and unsupervised classification. The supervised analysis involves assigning predefined classes to expression profiles with a training set and a testing set of the entire data set. Once the classifier has been trained on the training set and tested on the testing set, it can then be applied to data with unknown classification. Supervised methods include k-nearest neighborhood, support vector machines, and neural networks. . [23, 24,25,26,27, 40] On the other hand, unsupervised analysis organizes the data without the benefit of predefined classes' information. Hierarchical clustering, K-means, and self-organizing maps, are examples of unsupervised clustering approaches that have been widely used in microarray analysis [23, 24,25,26,27, 40]. There are many open source tools and commercial packages for microarray analyses. However, commercial tools can be expensive and of steady learning curve that is difficult for novice researchers. One approach is to create a customized data that you can apply on it statistical analysis software packages that allow considerable flexibility such as Matlab or R. there are many packages that are freely available which bundle several R tools for the current gene expression analysis methods. A wonderful tool now available in which you can include R functions is Arlequin [28].

Table 1: Some Microarray databases

No	Source	Description
1	http://genome-www5.stanford.edu/	Stanford Microarray Database (SMD) – contains raw and normalized data from microarray experiments as well as their image fi les. SMD also provides interfaces for data retrieval, analysis and visualization
2	*http://www.ebi.ac.uk/arrayexpress/*	ArrayExpress – public repository for microarray data at the EMBL-EBI.
3	http://info.med.yale.edu/microarray/	Yale Microarray Database (YMD)
4	http://www.ncbi.nlm.nih.gov/geo/	Gene Expression Omnibus at the NCBI, NIH.
5	http://www.mrc-lmb.cam.ac.uk/genomes/madanm/microarray/	Readme file for scripts + datasets PERL scripts: Euclidean distance, Pearson correlation coefficient, Rank correlation coefficient Datasets: Cho et al., Spellman et al.

Table 2: Selected Types of Software used in Microarray Analysis and Their sources

No	Software	Sources	Type of Machine
1	LIBSVM	http://www.csie.ntu.edu.tw/~cjlin/libsvm/	SVM
2	SVMlight	http://svmlight.joachims.org/	SVM
3	SVMTorch	http://www.cs.cmu.edu/afs/cs.cmu.edu/project/learn-43/lib/photoz/.g/mmp/trees/SVM/	SVM
4	Mayday Software	http://www-ps.informatik.uni-tuebingen.de/mayday/wp/?page_id=8	KNN
5	knnGarden	http://cran.r-project.org/web/packages/knnGarden/index.html	KNN
6	Weka-KNN	http://www.ibm.com/developerworks/apps/download/index.jsp?contentid=494038&filename=os-weka3-Example.zip&method=http&locale=	KNN
7	PCP	http://pcp.sourceforge.net/	NN
8	nnet	http://cran.r-project.org/web/packages/nnet/index.html	NN
9	neuralnet	http://cran.r-project.org/web/packages/neuralnet/index.html	NN
10	Iterative Bayesian Model Averaging	http://www.bioconductor.org/packages/2.11/bioc/html/iterativeBMA.html	Bayesian
11	Full Bayesian Network Classifier	http://www.cs.unb.ca/profs/hzhang/FBC.rar	Bayesian
12	Bayesian Trans-dimensional Sampling	http://www2.warwick.ac.uk/fac/sci/statistics/staff/academicresearch/steel/steel_homepage/software/transsup.zip	Bayesian
13	Backward Elimination Random Forest	http://cran.r-project.org/web/packages/varSelRF/index.html	Random Forest
14	Online Random Forest	http://www.ymer.org/research/files/online-forest/OnlineForest-0.11.tar.gz	Random Forest
15	*cforest*	http://cran.r-project.org/web/packages/party/index.html	Random Forest

Most important to microarray analysis methods is to adjust the array for multiple testing. We would like to prove the probability of seeing the observed test score with our null hypothesis that there is no difference in expression related to certain phenotype or disease being studied. In a binary classification of microarray analysis most important is to control the false positive rate [29, 30]. Functional analysis is required for interpreting the results after a list of differentially expressed genes has been assembled. There is several software that can help in this analysis [31]. Table 2 shows a selection of some well known software that are used in microarray analysis with their different approaches of machine types and their sources. Support Vector Machines (SVM) are mainly binary classifiers while the Neural Networks (NN) can classify into multiple classes [32].

However, most of domain experts are interested in classifiers that not only produce high classification accuracy but also reveal important biological information. A huge range of machine learning methods can be applied to the related classification problems. Some researchers reported that simple classification tools often perform as well as and even better than complex ones. [30, 32]

7. EXPRESSED GENES

As mentioned before, the spots in the hybridized microarray are excited by a laser ray and scanned at suitable wavelengths to detect the red and green dyes. The amount of fluorescence produced upon excitation corresponds to the amount of bound nucleic acid. That is, if cDNA from a certain condition for a particular gene is in greater abundance than that from other condition, the spot would be red. If it is not, the spot would be green. If the gene is expressed to the same extent in both conditions, the spot would be yellow, and if the gene is not expressed in both conditions, the spot would be black. Therefore, what is seen at the end of the microarray experiment is an image of the microarray, in which each spot that corresponds to a gene has an associated fluorescence value representing the relative expression level of that gene. So, Gene expression profiles can be linked to external information to explain biological processes and to make new discoveries [3, 33]. Expressed genes and the level of their expression is very crucial. For example determination of genes that are under expressed or over expressed in cancer cells can assists in designing and planning more effective treatments for cancer patients [34 36].

8. INTERPRET RESULTS

Statistical methods like clustering defines groups based on statistical calculations and considerations. What we want in microarray analysis is to interpret these groups in terms of biological functions [26, 34]. To know which genes are differentially expressed is of vital importance for any biological interpretation. The aim of differential analysis is to assess a significant threshold above which a gene or set of genes will be marked differentially expressed [34]. An ordinary way to achieve good interpretation of the results is to try to gather genes having similar expression profiles in a certain set of conditions and or in different tissues into specific clusters. These clusters may then be interpreted as functional groups and the function of an unknown gene or set of genes can be inferred on the basis of the function of a gene or genes belonging to the same cluster [34]. Because of the size of the data sets provided by microarray experiments, the information needs to be grouped or summarized in some way for any interpretation. Clustering techniques of the microarray data are of great importance because they reduce the size of the data sets by gathering genes into a reduced number of groups [23, 24, 25, 26, 27]. Statistical tests constitute the core tool for such analysis and interpretation of results. The identification of appropriate statistics and statistical tests is crucial. We should understand that the statistic should be adapted to the special case of microarrays. Also we have to acknowledge that the hypothesis under test may lead to new definitions that may affect the statistical procedures.

9. DISCUSSION

The microarray analysis field is now well-formed with accessible software and tools to make data analysis manageable by novice researchers. However, the data analysis requires a considerable time and effort that generally exceeds which are devoted to data generation. It is expected that microarray technology will soon be faced by next-generation whole-transcriptome sequencing, in which the transcripts are directly sequenced by low-cost, high-throughput sequencing technologies. Nevertheless, microarrays will remain a desirable alternative for many scientist and researchers [37]. Microarrays has been used effectively in the analysis of diseases especially cancer. The identification of single gene products that are expressed in tumor cells but not in normal ones is of great importance, mainly in the establishment of tumor biomarkers for diagnostic purposes. in addition to the identification of tumor biomarkers, microarray analysis can distinguish between distinct subtypes of leukemia, lymphoma, breast cancer and melanoma [8, 34, 35, 36, 37,38]. Some researcher are talking about the post genomic age, during which the diagnostic, prognostic, and treatment response biomarker genes identified by microarray screening will be integrated to provide personalized management of patients. Medical doctors will be able to use microarrays during early clinical trials to confirm the mechanisms of action of drugs and to assess drug sensitivity and toxicity [39]. It is expected that he costs will decrease in the near future and that the technology will become available and intuitive. The range of applications of microarray technology is enormous. Recent studies in human cancer have demonstrated that microarrays can be used to develop a new molecular taxonomy of cancer, including clustering of cancers according to prognostic groups on the basis of gene expression profiles. The list of potential uses of this technique is not limited to cancer research. For example, the impact on gene expression by drugs, environmental toxins, or oncogenes may be elucidated and regulatory networks and coexpression patterns can then be interpreted [34]. There are several challenges and conventional problems concerning microarray analysis that should be addressed. Some of these problems are:

I. Bias and confounding Problem: This problem occurs during the design phase of microarray and can lead to erroneous conclusion. Poor skills and poor technical abilities could possibly cause bias. Confounding takes place when other factors distort the true relationship between the variables of the experiment.
II. Different standards: cross-platform comparisons of gene expression are difficult to conduct when microarrays were constructed using different standards and setups. Thus, Minimal Information about a Microarray Experiment (MIAME) has been developed to improve reproducibility, sensitivity and robustness in gene expression analysis.
III. High dimensionality: microarray data is a high dimensional data characterized by thousand of genes in few sample sizes, which cause significant problems such as irrelevant and noise genes, complexity in constructing classifiers, and multiple missing gene expression values due to improper scanning. This problem also leads to data over fitting, which requires further validation.
IV. Mislabeled data: mislabeled data of questioned tissues result by experts is another type of drawback that could decrease the accuracy of experimental results and lead to imprecise conclusion about gene expression patterns.
V. Relevance of biological results: this is another integral criterion that should be taken into account in analyzing microarray data rather than only focusing on accuracy of classification.

10. CONCLUSIONS

The use of microarrays to explore gene expression on a global level is a rapidly evolving technology that seems to become more powerful with the achievement of the sequencing of tens of genes including the human genome. Microarray technology has been widely used to investigate tumor classification, cancer progression, and chemotherapy resistance and sensitivity. Expression arrays can be used to gain a better understanding of the basic biology, diagnosis, and treatment of several diseases including cancer. The experimental design of the microarrays is proving very useful and it seems likely that this will improve the interpretation of the data sets generated by the experiments. At present, extracting useful information from these large

quantities of data is a difficult task. This will be improved by advances in experimental design and bioinformatics. Recent work in the field of bioinformatics indicates that microarray technology is set to contribute much to the post-genomic future.

ACKNOWLEDGEMENTS

The authors would like to thank Princess Nourah bint Abdulrahman for facilitating this research. Thanks also go to the colleagues in the college of Computer and Information Sciences in the university.

REFERENCES

[1] Felsenfeld, G; Miles, HT (1967). "The physical and chemical properties of nucleic acids.". Annual review of biochemistry 36: 407–48.

[2] Causton, H., Quackenbush, J., and Brazma, A. (2003). Microarray Gene Expression Data Analysis: A Beginner's Guide. (UK: Blackwell Science)

[3] David J. Duggan, Michael Bittner, Yidong Chen, Paul Meltzer, and Jeffrey M. Trent. Ex-pression profiling using cDNA microarrays. Nature Genetics, 21 (Suppl 1):10–14, 1999.

[4] G.A. Churchill. Fundamentals of experimental design for cDNA microarrays.Nature Ge-netics, 32 Suppl. 2:490–495, 2002.

[5] Slonim DK, Yanai I (2009) Getting Started in Gene Expression Microarray Analysis. PLoS Comput Biol 5(10): e1000543. doi:10.1371/journal.pcbi.1000543

[6] Mary-Huard, T., Robin, S., Daudin, J., Bitton, F., Cabannes, E. and Hilson, P. (2004). Spotting effect in microarray experiments. BMC Bioinformatics. 5(63) 1–9.

[7] Kane, M. D. et al. Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays. Nucleic Acids Res. 28, 4552–4557 (2000).

[8] Hoehndorf R, et al. Analyzing gene expression data in mice with the Neuro Behavior Ontology. Mamm. Genome. 2014:1–9

[9] Shoemaker, D. D. et al. (2001) Experimental annotation of the human genome using microarray technology. Nature 409, 922–927 (2001).

[10] Kerr MK, Churchill GA (2001) Experimental design for gene expression microarrays. Biostatistics 2: 183–201.

[11] Yang YH, Speed T (2002) Design issues for cDNA microarray experiments. Nat Rev Genet 3:579–588.

[12] Ong Huey Fang,Norwati Mustapha and Md. Nasir Sulaiman, "Integrating Biological Information for Feature Selection in Microarray Data Classification" IEEE 2010 Second International Conference on Computer Engineering and Applications

[13] Spellman, P. T. et al. Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. Mol. Biol. Cell 9, 3273–3297 (1998).

[14] Amaral, Telmo, Stephen J. McKenna, Katherine Robertson, and Alastair Thompson. "Classification and Immunohistochemical Scoring of Breast Tissue Microarray Spots." (2013): 1-1.

[15] Quackenbush J (2002) Microarray data normalization and transformation. Nat Genet 32: Suppl496–501.

[16] Duda RO, Hart PE, Stork DG (2001) Pattern Classification. New York: John Wiley & Sons.

[17] Simon RM, McShane LM, Korn EL, Radmacher MD (2003) Design and Analysis of DNA Microarray Investigations. Springer.

[18] Brazma, A. & Vilo, J. Gene expression data analysis. FEBS Lett. 480, 17–24 (2000).

[19] Hartigan JA. Clustering algorithms 1975:351 John Wiley & Sons New York. .

[20] Lin, I.-H., Chen, D.-T., Chang, Y.-F., Lee, Y.-L., Su, C.-H., Cheng, C., … Hsu, M.-T. (2015). Hierarchical Clustering of Breast Cancer Methylomes Revealed Differentially Methylated and Expressed Breast Cancer Genes. PLoS ONE, 10(2), e0118453. http://doi.org/10.1371/journal.pone.0118453

[21] Mukhopadhyay, Anirban, Ujjwal Maulik, and Sanghamitra Bandyopadhyay. "An interactive approach to multiobjective clustering of gene expression patterns."Biomedical Engineering, IEEE Transactions on 60, no. 1 (2013): 35-41.

[22] Jaskowiak, Pablo A., Ricardo JGB Campello, and Ivan G. Costa Filho. "Proximity measures for clustering gene expression microarray data: a validation methodology and a comparative analysis." IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB) 10, no. 4 (2013): 845-857.

[23] Ripley B. Pattern Recognition and Neural Networks. Cambridge, (1996).

[24] K Blekas, Nikolas P. Galatsanos and loannis Georgiou, "an unsupervised artifact correction approach for the analysis of dna microarray images", IEEE, 2003.

[25] Yijuan Lu, Qi Tian, Feng Liu, Maribel Sanchez, and Yufeng Wang, "Interactive Semisupervised Learning for Microarray Analysis", ieee/acm transactions on computational biology and bioinformatics, vol. 4, no. 2, april-june 2007.

[26] Jong Kyoung Kim and Seungjin Choi, Member, IEEE, " Probabilistic Models for Semi-Supervised Discriminative Motif Discovery in DNA Sequences", Ieee/Acm Transactions On Computational Biology And Bioinformatics, February 2, 2010

[27] Maji, Pradipta. "Mutual information-based supervised attribute clustering for microarray sample classification." Knowledge and Data Engineering, IEEE Transactions on 24, no. 1 (2012): 127-140.

[28] http://cmpg.unibe.ch/software/arlequin35/

[29] D. Storey, "The Positive False Discovery Rate: A Bayesian Interpretation and the q-Value," Annals of Statistics, vol. 31, pp. 2013-2035, 2003.

[30] Reiner A, Yekutieli D, Benjamini Y (2003) Identifying differentially expressed genes using false discovery rate controlling procedures. Bioinformatics 19: 368–375.

[31] Siang, TC, Soon, TW, Kasim, S, Mohamad, MS, Howe, CW, Deris, S, Zakaria, Z, Shah, ZA & Ibrahim, Z 2015, 'A review of cancer classification software for gene expression data' International Journal of Bio-Science and Bio-Technology, vol 7, no. 4, pp. 89-108., 10.14257/ijbsbt.2015.7.4.10

[32] Zhu, Y., Shen, X., Pan, W. "Network-based support vector machine for classification of microarray samples". BMC Bioinformatics 2009, 10(Suppl 1):S21doi:10.1186/1471-2105-10-S1-S21

[33] Giannakeas, Nikolaos, Karvelis, Petros S., and Fotiadis, Dimitrios I. "A classification-based segmentation of cDNA microarray images using Support Vector machines. Engineering in Medicine International Journal of Computer Science & Engineering Survey (IJCSES) Vol.2, No.3, August 2011

[34] Dupuy, A., Simon, R.M., 2007. Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. J. Natl. Cancer Inst. (2), 147–157, 17;99.

[35] Wang H., H. Zhang, Z. Dai, M. S. Chen, Z. Yuan (). TSG: A 2013New Algorithm for Binary and Multi-class Cancer Classification and Informative Genes Selection, BMC Medical Genomics, 6(1), 1:53.

[36] Li X., S. Peng, J. Chen, B. Lu, H. Zhang, M. Lai (2012). SVM-T-RFE: A Novel Gene Selection Algorithm for Identifying Metastasis-related Genes in Colorectal Cancer using Gene Expression Profiles, Biochemical and Biophysical Research Communications, 419(2), 148-153.

[37] Yuji Zhang, Jason J. Xuan1, Robert Clarke, Habtom W. Ressom, "Module-Based Biomarker Discovery in Breast Cancer" 2010 IEEE International Conference on Bioinformatics and Biomedicine,978-1-4244-8305-1/10©2010 IEEE

[38] Maulik, Ujjwal, Anirban Mukhopadhyay, and Debasis Chakraborty. "gene-expression-based cancer subtypes prediction through feature selection and transductive SVM." Biomedical Engineering, IEEE Transactions on 60, no. 4 (2013): 1111-1117.

[39] Kim JH, Skates SJ, Uede T, Wong Kk KK, Schorge JO, Feltmate CM, et al. Osteopontin as a potential diagnostic biomarker for ovarian cancer. JAMA 2002;287:1671-1679.

[40] Hanczar, Blaise, and Avner Bar-Hen. "A new measure of classifier performance for gene expression data." IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB) 9, no. 5 (2012): 1379-1386.