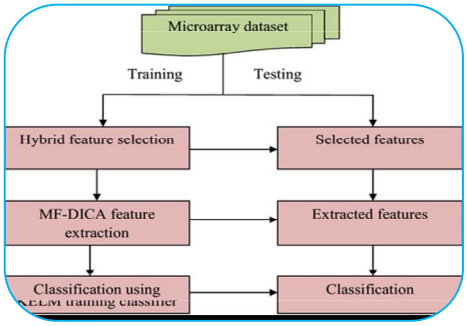
REVIEW OF RESEARCH
ISSN: 2249-894X
IMPACT FACTOR : 5.7631(UIF)
VOLUME - 11 | ISSUE - 6 | MARCH - 2022

SELECTION AND CLASSIFICATION OF GENOME OF EXPRESSION OF MICROARRAY DATA FOR DETECTION OF CANCER

M. D. Acharya[1] and Joshi R. G.[2]
[1] Assistant Professor , Department of Computer Science, Yogeshwari Mahavidyalaya, Ambajogai.
[2] Assistant Professor , Department of Computer Science, Yogeshwari Mahavidyalaya, Ambajogai.

ABSTRACT:

Cancer cells that normally affect the growth of cells, molecules and blood vessels that feed the tumor. For example, cancer affecting the nerves that supply oxygen to the cancer cells and the supplements that the cancer cells need to grow, can remove waste products from the tumor and hide the cells from the immune system. Cancer can spread from one place to another, starting throughout the body, this is called metastatic cancer growth. The process by which cancer cells spread to different parts of the body is called metastasis. Treatment helps prolong the life of the patient. The main goal of treatment is to stop the growth of the cancer or to relieve the symptoms caused by the cancer. Metastatic cancer causes serious death and impairment of body functions in patients.



KEYWORDS: cancer cells , metastatic cancer growth, symptoms.

INTRODUCTION

Molecular biology research is shaped by the innovative techniques used to accomplish it. It is unimaginable to expect to discover numerous qualities using traditional techniques. DNA microarrays are an innovation that enables scientists to diagnose and treat problems that cannot be identified. The outflow of numerous points in a single response can be examined rapidly and efficiently. DNA microarray innovation has enabled established researchers to discover the fundamental components of life development and improvement, as well as the genetic causes of idiosyncrasies in the functioning of the human body. Routine microarray tests involve hybridization of an mRNA particle to the DNA layout from which it originates. A number of DNA tests are used to create a demonstration. How much mRNA is inhibited at each site on the cluster indicates the degree of expression of the various markers. This number can be large. All information is collected and profiled for gene expression in the cell. Microarray innovation is a breakthrough device used to identify the expression of thousands of points in a single response. Microarrays are minute slides of glass that are printed with many small spots in a characteristic position. Each spot contains an unusual deoxyribonucleic acid (DNA) group or duplicate of different quality. These slides are called quality chips or DNA chips. DNA particles attached to a

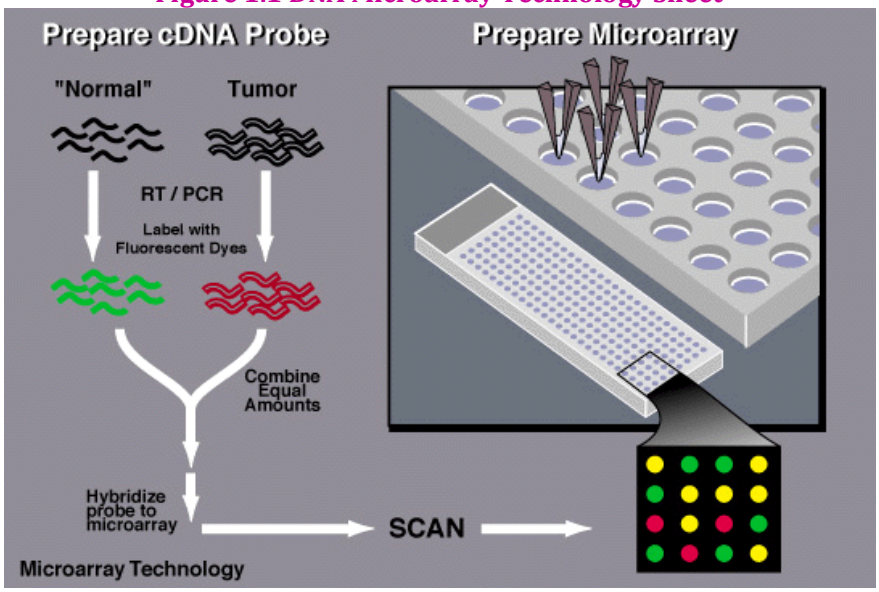
microarray slide test to identify a group of messenger ribonucleic acid (mRNA) records, otherwise called the transcriptome, communicated through a combination of points.

For actual screening of microarray information, mRNA molecules are collected from test and reference organic samples. For example, a reference test is collected from a solid individual and an investigative sample is collected from a person suffering from an infection such as a malignant growth. From that point on, each mRNA test is prepared separately and named an alternative color fluorescent test. The two mRNA probes are then converted into integral DNA (cDNA) around the transcriptase and marked with one of two fluorescent tones, "red" and "green". The test cDNA is labeled with a red fluorescent dye (Cy5) and the reference cDNA is labeled with a green fluorescent dye (Cy3). These two instances are then combined as one and added to the corresponding base match in each spot on the cluster. The cycle by which the cDNA particles are added to the DNA test on the slide is called hybridization. After hybridization, the microarray chip is washed to remove obvious boundaries. The chip is then filtered to measure each quality announcement printed on the slide.

If a particular quality statement is higher in a treatment test than a typical example, it may well be replaced with a ruddy green tone. Alternatively, if the expression in the treatment trial is lower than the normal example, the stain appears green. Finally, assuming that both treatments and typical examples are the same, the space appears yellow. The information collected by the microarray is used to create a gene expression profile, which is designed to examine the expression levels of thousands of markers simultaneously. Such gene expression information can be used to analyze assemblages, discover new subtypes of infection, and identify segregating traits that contribute to disease. The point of this research work is to separate the distinguishing marks associated with mental imbalance problems, malignant breast enlargement and prostate disease.

As shown in the image below, the raw microarray images are converted into a gene expression data matrix, where rows represent scores, sections represent different samples such as tissues or investigational conditions, and numbers in each cell represent the degree of expression of a particular quality. Raw microarray data are images that are transformed into gene expression networks. Figure 1.2 shows the structure of a gene expression grid. Lines in the network compare to scores and sections address examples or test situations. The number in each cell addresses the speech level of a particular quality in a particular instance or condition. The degree of articulation can be direct or relative. Assuming that the two lines are inseparable, this suggests that the distinct traits are co-common and potentially practically related. By looking at examples, altered interactive qualities can be identified.

Figure 1.1 DNA Microarray Technology Sheet



Classification of Microarray:

Based on the types of probes used, microarrays are of twelve different types:

- **DNA Microarray:** DNA microarray is also known as gene chip, DNA chip or biochip. It either measures DNA or uses DNA as part of its detection system. There are four different types of DNA microarrays: cDNA microarrays, oligo DNA microarrays, BAC microarrays, and SNP microarrays.
- **MMCchips:** The MMchip allows cross-platform and inter-laboratory integrated analysis of data. It studies the interaction between DNA and proteins. Two techniques used are ChIP-chip (chromatin immunoprecipitation (ChIP) followed by array hybridization) and ChIP-seq (ChIP followed by massively parallel sequencing).
- **Protein microarrays:** It serves as a platform to characterize hundreds of thousands of proteins in a highly parallelized manner. There are three types of protein microarrays and they are analytical protein microarrays, functional protein microarrays and reverse-phase protein microarrays.
- **Peptide microarrays:** These types of arrays are used for detailed analysis or optimization of protein-protein interactions. It helps identify antibodies by examining proteomes.
- **Tissue Microarrays:** Tissue microarrays are paraffin blocks prepared by isolating cylindrical tissue cores from various tissues and embedding them into a microarray. It is mainly used in pathology.
- **Cellular microarrays:** They are also called transfection microarrays or living-cell-microarrays and are used to screen large-scale chemical and genomic libraries and systematically probe the local cellular microenvironment.
- **Chemical Compound Microarray:** It is used for drug screening and drug discovery. These microarrays have the potential to identify and evaluate small molecules and are therefore more useful than other technologies used in the pharmaceutical industry.
- **Antibody microarrays:** They are also called antibody arrays or antibody chips. These are protein-specific microarrays that consist of a collection of capture antibodies placed on a microscope slide. They are used to detect antigens.
- **Carbohydrate Arrays:** Also called Glycores. Carbohydrate arrays are used in screening proteomes for carbohydrate binding. They can also be used to calculate protein binding affinities and automate solid-support synthesis for glycans.
- **Phenotype Microarrays:** Phenotype microarrays or PMs are primarily used for drug development. They quantitatively measure thousands of cellular phenotypes simultaneously. It is also used in functional genomics and toxicological testing.
- **Reverse Phase Protein Microarrays:** They are microarrays of lysates or serum. Primarily used in clinical trials, especially in the field of cancer, they also have pharmaceutical uses. In some cases, they can also be used in the study of biomarkers.
- **Interferometric Reflectance Imaging Sensor or IRIS:** IRIS is a biosensor used to analyze protein-protein, protein-DNA and DNA-DNA interactions. It does not use fluorescent labels. It is made of Si/SiO2 substrates prepared by robotic spotting.

Applications of Microarray:

- **Gene Expression Analysis:** A major application of DNA microarrays is to measure gene expression levels. In this application, RNA is extracted from cells of interest and either directly labeled, converted to labeled cDNA, or converted to T7 RNA promoter tailed cDNA that is further converted to siRNA by the Eber-wine amplification process. Various methods have been developed for labeling cDNA or siRNA including: incorporation of fluorescently labeled nucleotides during synthesis, incorporation of biotin labeled nucleotides which are then stained with streptavidin, including fluorescently labeled nucleotides. Then is added, and various types of signal amplification methods. The two most frequently used methods are the incorporation of fluorescently labeled nucleotides into the siRNA or cDNA synthesis step or the incorporation of biotin labeled nucleotides into the siRNA synthesis step. The labeled cRNA or cDNA is then hybridized to the microarray, the array is washed, and the signal is detected by measuring the fluorescence at each spot. In the case of biotin

labeled samples, the array is stained after hybridization with fluorescently labeled streptavidin. Laser induced fluorescence is usually measured with a scanning confocal microscope. The signal intensity at each spot is taken as a measure of the expression level of the corresponding gene.

- Transcription Factor Binding Analysis: Microarrays with chromatin immunoprecipitation have also been used to determine binding sites of transcription factors. Briefly, transcription factors (TFs) are bound to DNA with formaldehyde and the DNA is fragmented. The TF(s) of interest are affinity purified either by using an antibody to the TF or by tagging the transcription factor with a peptide suitable for affinity chromatography. After purification, the DNA is released from the TF, amplified, labeled and hybridized to the array. This technique for chromatin immuno-precipitation on a "chip" or microarray is commonly known as "ChIP-chip". Since TFs are often bound at a distance from regulated genes, array structure and fragment length size distributions are correlated. E.g. The array must contain a probe that probes the region of DNA that binds to the transcription factor. For bacteria or yeast, the intergenic regions are much smaller and the same arrays used for gene expression work can be implemented on a ChIP-chip. For mammalian genomes, intergenic regions are large and TFs often bind several kbp away from the gene of interest. Therefore, for mammalian genomes, oligo arrays with evenly spaced oligos throughout the genome are typically used for ChIP-chip experiments.
- **Genotyping:** Microarrays have been widely used as single-nucleotide-polymorphism (SNP) genotyping platforms. Several alternative approaches have been used to detect SNPs but the most commonly used are allele discrimination, allele specific amplification and “bar-code” oligo ligation by hybridization using Affymetrix which hybridizes to the Illumina “Golden Gate Assay” universal array. ” or approaches in which arrayed DNA is extended to SNPs in a single nucleotide extension reaction. In complex genomes, non-specific hybridization counteracts background allelic differentiation by hybridization. To reduce this background, Affymetrix developed a PCR based approach to reduce genomic complexity. Briefly, SNPs for their analysis are selected between restricted sites that are <1kb apart. Genomic DNA is cleaved with restriction enzymes, end repaired and ligated to adapter fragments for PCR. PCR is performed under conditions that selectively amplify products <1kb in size. This method reduces genomic complexity by approximately 50-fold and results in a corresponding increase in signal-to-noise on the array. Both Affymetrix and Illumina methods for SNP genotyping have been highly successful and are widely used. Today SNP arrays capable of detecting >1M different human SNPs are available from both vendors. The call rate and reproducibility of SNP calls exceeded 99.5%. Additionally, the same array, or variations thereof, can also be used to detect copy number variants.

Microarray Data Cancer:

Cancer is a collection of diseases characterized by abnormal cell growth and spread to normal cells. Those abnormal cells grow, divide, and turn into a tumor mass. Cancer causing factors include but are not limited to; Errors during cell division, DNA damage from smoking, ultraviolet rays from the sun, chemicals, obesity, hormones, chronic inflammation. In the development of cancer, genes that control cell growth and differentiation are frequently altered. Microarray-based gene expression profiling enabled scientists to simultaneously quantify the expression levels of thousands of traits and identify traits whose expression is altered in light of malignant growth. Furthermore, cancer begins with damaged cell deoxyribonucleic acid (DNA) in a process known as mutation that occurs when a cell replicates its DNA and makes mistakes before cell division. In a healthy body, cells grow, die, and change in a very controlled manner. Microarray data analysis and classification is becoming one of the most popular research areas in bioinformatics, computational biology, machine learning, pattern classification, and statistics. The main challenge of microarray classification is related to high curve dimensions and small sample size. The importance of research in the field of microarray data analysis, especially in the early stages of cancer, helps to detect the classification of cancer. A treatment plan can be designed to increase the survival rate of cancer patients. Furthermore, domain experts are keen to understand the nature of the traits that contribute to the development of cancer. Increasing the

accuracy of classification problems is another task in the field of microarray data analysis. Medical data classification based on microarray gene expression is one of the most challenging research areas.

Classification of microarray medical datasets plays an important role, especially in identifying which genes contribute most to a particular biological outcome and predicting the outcome when new observations are made. A classification problem requires the creation of a model, which takes input patterns that represent objects and predicts the class or category associated with the objects in question with the aim of delivering accurate predictions on test data.

CONCLUSION:

Cancer gene expression profiles are not normally-distributed, either at the whole-experiment or individual-gene level. Instead, they exhibit a complex, heavy-tailed distribution characterized by statistically-significant sconces and curtsies. The non-Gaussian distribution of these data affects the identification, functional interpretation, and potential molecular classification of differentially expressed genes.

REFERENCES:

1. Blessum C, Jeppsson JO, Aguzzi F, Bernon H, Bienvenu J. Capillary electrophoresis: principles and practice in clinical laboratory. Ann Biol Clin (Paris) 1999;57:643–57.
2. Bienvenu J, Graziani MS, Arpin F, Bernon H, Blessum C, Marchetti C, et al. Multicenter evaluation of the Paragon CZE 2000 capillary zone electrophoresis system for serum protein electrophoresis and monoclonal component typing. Clin Chem 1998;44:599–605.
3. Bossuyt X, Schiettekatte G, Bogaerts A, Blanckaert N. Serum protein electrophoresis by CZE 2000 clinical capillary electrophoresis system. Clin Chem 1998; 44:749–59.
4. Bossuyt X, Bogaerts A, Schiettekatte G, Blanckaert N. Detection and classification of paraproteins by capillary immunofixation/subtraction. Clin Chem 1998; 44:760–4
5. Dias Neto E, Correa RG, Verjovski-Almeida S, Briones MR, Nagai MA, da Silva W Jr. Zago MA, Bordin S, Costa FF, Goldman GH, Carvalho AF, Matsukuma A, Baia GS, Simpson DH, Brunstein A, de Oliveira PS, Bucher P, Jongeneel CV, O'Hare MJ, Soares F, Brentani RR, Reis LF, de Souza SJ, Simpson AJ. Shotgun sequencing of the human transcriptome with ORF expressed sequence tags. Proceedings of the National Academy of Sciences of the United States of America. 2000; 97:3491–3496.
6. Gergen JP, Stern RH, Wensink PC. Filter replicas and permanent collections of recombinant DNA plasmids. Nucleic acids research. 1979; 7:2115–2136.
7. Grunstein M, Hogness DS. Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. Proceedings of the National Academy of Sciences of the United States of America. 1975; 72:3961–3965.