Journal of Advanced Bioinformatics Applications and Research.
Vol 1, Issue 1, June 2010, pp 59-68

A REVIEW OF BIOINFORMATICS APPLICATION IN BREAST CANCER RESEARCH

Vidya Vaidya, Shriram Dawkhar
Department of Bioinformatics, M.I.T. Group of institutions, MACS College. S. No. 124
Paud Road, M.I.T. Campus, Pune 411038.

ABSTRACT

Now a day's cancer is a worldwide common problem. Among these the breast cancer is rapidly growing in woman. So the study of breast cancer is important to predict the prognostic factors affecting for breast cancer, the resistance of drug for individual patient, and to predict the ratio/percentage of transmitting the cancer in next generation.

So far various studies are carried out in the study of breast cancer by applications in bioinformatics.

In this paper various methods adopted for the comprehensive analysis of microarray data are discussed. This paper highlights the clinical and biological aspects of human breast cancer using molecular profiling techniques.

Such study will be useful to develop a model which fulfils the preclinical setting for cancer patients.

Keywords: breast cancer, resistance of drug , clinical and biological aspects

Introduction

There are many different types of breast cancer, with different stages (spread), aggressiveness, and genetic makeup; survival varies greatly depending on those factors.[1] Computerized models are available to predict survival.[2]

Now a day's cancer is a worldwide common problem. Among these the breast cancer is rapidly growing in woman. So the study of breast cancer is important to predict the prognostic factors affecting for breast cancer, the resistance of drug for individual patient, and to predict the ratio/percentage of transmitting the cancer in next generation.

The bioinformatics has following applications

1. Systematic collections and Organization of biological data to help researcher.
2. Develop most appropriate tools and resources that aid in analysis of data such as FASTA and BLAST [C].
3. Use these tools to analyze data and interpret the result in biological meaningful manner.
4. Different application areas of bioinformatics are gene comparison, genomics, microarrays, proteomics, etc.

So far various studies are carried out in the study of breast cancer by applications in bioinformatics.

We observed that Sofia K Gruvberger-Saal et al, gives the platform for gene profiling by microarray technique. Breast cancer derived by BRCA1 and BRCA2 mutation carriers. The extended knowledge of defects causing the development of intervention approaches for affected individuals. PE Lonning et al, gives the issue of

prognostication and prediction. For relevance of cancer he also focuses on how our present knowledge helps in improving therapy.

Worldwide, breast cancer comprises 10.4% of all cancer incidences among women, making it the most common type of non-skin cancer in women and the fifth most common cause of cancer death.[3] In 2004, breast cancer caused 519,000 deaths worldwide (7% of cancer deaths; almost 1% of all deaths).[4] Breast cancer is about 100 times more common in women than in men, although males tend to have poorer outcomes due to delays in diagnosis.[5][6][7][8]

In this paper various methods adopted for the comprehensive analysis of microarray data are discussed. This paper highlights the clinical and biological aspects of human breast cancer using molecular profiling techniques.

Such study will be useful to develop a model which fulfils the preclinical setting for cancer patients.

Bioinformatics is "the mathematical, statistical, and computing methods that aim to solve biological problems using DNA and amino acid sequences and related information". Bioinformatics can be applied in the field of medical sciences to know the pathways to know which genomic changes could give raise to each known inherited disease i.e. identification of the gene causing disease, and also genetic therapies that can reverse disease phenotype.

Colin S Cooper suggest about applications of microarray in the field of breast cancer. Approximately there are 50000 genes are present and 10000 genes functions have been established. Microarrays utilize the preferential binding of complementary single-stranded nucleic acid (cDNA) sequences. Microarray is build using glass slide, on to which cDNA molecules are attached at fixed locations. There are very large numbers of spots on an array, each containing a huge number of identical DNA molecules, of lengths from twenty to hundreds of nucleotides. Gene expression monitoring and Single Nucleotide Polymorphisms (SNP) detection are two important application of microarray technology.

Sofia K Gruvberger-Saal et al gives platform for illustration of clinical issues by using gene expression profile by using molecular signature microarray have the potential to be direct benefit to the patient in the near future. Microarrays provide a tremendous opportunity to explore the complex biology of breast cancer and to potentially identify new target for therapy.

PE Lonning et al. gives the platform for microarray technology used for drug resistance. Here the main target is to control the cell growth on the tumor. There are various technologies are used from them microarray is used. By merging picture of gene and then linked it with drug resistance the factors affecting on drug resistance can be affected.

Microarray Technology

The technology that revolutionalized Molecular Biology, first time developed and demonstrated by Pat Brown at Standford University, U.S.A.

The actual design of microarray is suggested by Colin S Cooper. According to him microarray is an orderly arrangement of known and unknown DNA sample at glass support. Each DNA on microarray is called probe. Several distinct chemistries may be used to attach the probe for hybridization. The target sequences are either hybridized or fluorescently labeled.

From the study it comes to know that there are following types of microarray.

Filter arrays:

Filter arrays are produced by the high-density application of DNA samples, usually cDNAs, to nylon hybridization membranes such as those used for Southern analysis.

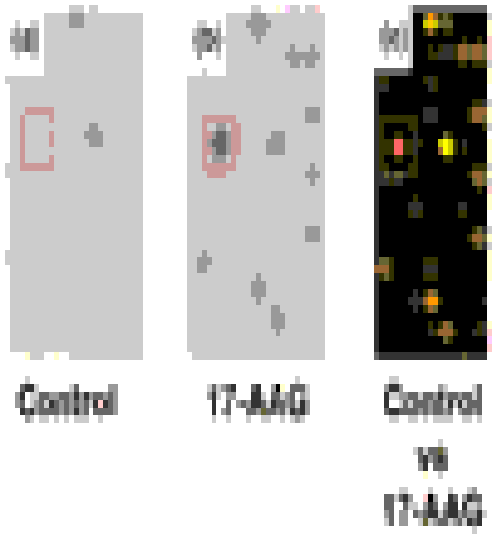
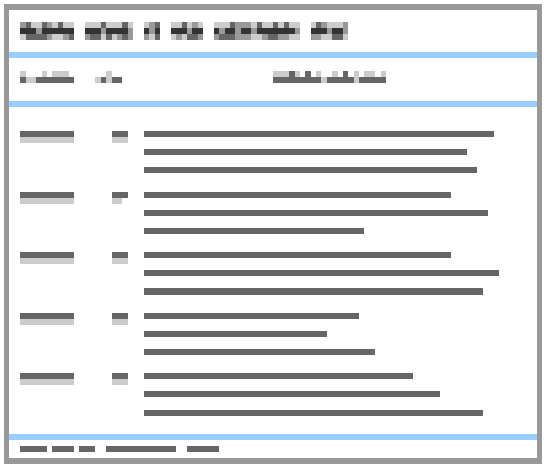


Fig-1: Filter microarrays hybridized to [α-^{33}P]-labeled cDNA probes prepared from 50-100 ng of poly $(A)^+$ mRNA from a human cell line before **(a)** and after **(b)** treatment with 17-allylamino-17-demethoxygeldanamycin (17-AAG). Hybridization signals were detected.

Glass DNA Microarray

Glass DNA microarrays are produced by the robotic application of DNA to a glass surface; cDNAs or genomic clones are usually applied, although oligonucleotide may also be used. This approach has the advantage that the probe can be covalently attached to an appropriately treated glass surface. Glass is extremely durable and its non-porous properties mean that both the access of the hybridization target to attached probes and the washing steps are rapid and highly reproducible. In addition, the transparency and rigidity of glass improves image acquisition, and its low-fluorescence background results in superior sensitivity. As the target or test samples are fluorescently labeled, several targets can be hybridized simultaneously in a single reaction, allowing direct comparison between samples. A disadvantage of glass DNA microarrays is that large target samples are required: probes would typically be prepared from 1-5 µg poly$(A)^+$ RNA or 50-200 µg total RNA by reverse transcription under conditions that allow incorporation of one of two fluorescently labeled dyes designated Cy3 (red) and Cy5 (green).



High density oligonucleotide arrays:

In this format, arrays of oligonucleotide (20-25 mers) are synthesized *in situ* on a flat solid support. Affymetrix achieves this by first attaching linkers containing a photochemically removable protecting group to a glass surface. Selected areas of the surface are then activated by illumination through a photolithographic mask, yielding reactive hydroxyl groups. A given nucleotide (A, G, C or T) is incubated with the surface under conditions that allow attachment of a single nucleotide only to the areas illuminated in the previous step. Repeating this process using a series of photolithographic masks that each direct light to different areas of the surface allows the synthesis of a large number of distinct oligonucleotides.

Microarray platform for gene expression analysis:

The microarray technology is the combination of wet-lab technique, statistical analysis and application of informatics to the data. There are four basic steps for gene expression study using microarray:

1. Print and cross-link DNA Clones (probes) onto a glass side.
2. Reverse transcribe mRNA from sample tissue into cDNA (target) and label with different fluorescence dyes (Cy3 and Cy5).
3. Hybridize targets to probes. The amount of target hybridized to a probe is measured by the intensity of fluorescence emission from that probe.
4. Images of fluorescence emission are compared to fine out differentially expressed genes. Those genes annotated and further studied.

The types of microarray are spotted array and high density oligonucleotide array. Scanners used red and grebe lasers operating at appropriate wavelength to excite Cy3 and Cy5 dyes to create a separate 16-bitTIFF image for each Channel. For image analysis the spot is identified. After this the background subtracted fluorescent intensity of each spot is calculated. The ratios of measured Cy3 to Cy5 intensities are used to identify differentially expressed gene. The mechanism of genechip preparation for microarray and overview of microarray experiment can be well understood from following figure.

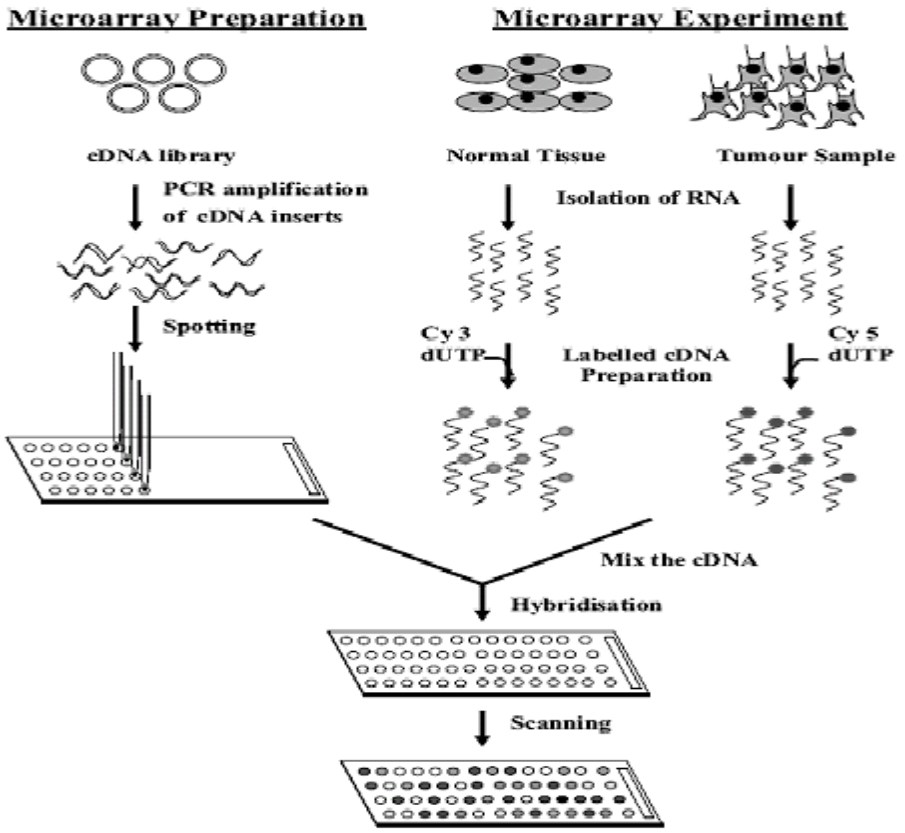


Fig-2: Microarray preparation and Microarray experiment

Data analysis:

Microarray data is characterized by a large number (typically >20.000) of measurement values per patient sample. It is difficult to be interpreted, because the number of genes is much higher than the number of samples and the data correlation structure is not well understood. Therefore the challenge is to distinguish between random and significant patterns of gene expression. Clinical data is complicated, too. There are many different items, both quantitative and categorical, often with complex coding schemes. Clinical data is multi-dimensional, because each clinical data source (e.g. a specific lab technique) provides a different view on the same patient. For this reason a systematic approach in data management with detailed planning and documentation is strongly recommended to keep an overview with microarray analysis in a clinical setting. Microarray studies generate vast quantities of data.

A program called UniGene has been used to identify clusters of overlapping Expressed Sequence Tag (EST) contigs (continuous runs of sequence) that represent a unique gene set, and representatives of each cluster are commercially available as the UniGene set from selected companies including Genomic Solutions and Research Genetics.

The technique of comparative genomic hybridization (CGH) was developed to monitor genome-wide variations in

DNA copy number. In this technique, differentially labelled test DNAs from a tumour and from normal tissue are co-hybridised to metaphase chromosomes. The relative levels of hybridization of test and control probes can indicate loss, gain or amplification of specific genomic regions. The resolution of this technique is poor, probably in the region of 10-20 megabases (Mb).

Amplification of genomic regions is believed to result in the overexpression of genes that drive tumour development and/or are involved in conferring drug resistance. In breast cancer, an amplification event at 17q results in overexpression of the *CERBB2* gene; amplification at 12q13-15 can cause overexpression of the *MDM2* and *CDK4* genes; and several different genes including *AIB1, ZNF217* and *BTAK* have been proposed to correspond to the gene driving an amplification event found at 20q13.

For identifying the actual region or spot related to can DNA some methods can be used. Mostly suitable methods are supervised and unsupervised. After the image processing of scanned image the unknown and known sequence data is collected. For supervise analyses require prior knowledge i.e. biological or clinical to discriminate genes distributed between sample groups within statistical probability.

Unsupervised methods are based on pattern-recognition algorithm to define groups of sample with similar global molecular profiles.

The microarray technology can be used to explore the biology of tumor types using hierarchical clustering. The breast cancer may be classified on the basis of global gene 'signatures'. These classifications have been found robust. The identification of 'basal' or 'triple-negative' class, revealing tumors lacking expression of the ER as well as HER-2 may be by specific gene profile.

Hierarchical as well as supervised clustering was able to reveal strong prognostic signature. But developing signature predicting drug resistance has not been successful.

As the Human Genome Project nears completion and alterations in more genes become implicated in human pathology, there will be a growing demand for techniques allowing high-throughput resequencing of genomic DNA. In the breast cancer field, there is already a requirement for large-scale population-based detection of inherited mutations within genes such as *BRCA1*, *BRCA2*, *P53* and *CHK2*, which predispose towards the development of this disease

Distinguishing tumors on the basis of their gene expression profile

After detecting differential gene expression it is often necessary to accurately classify samples into known groups. There are quite a few methods to classify samples for instance support vector machines

(SVM), PAM classification and regression trees (CART), k-Nearest-Neighbor (k-NN) and others. There is evidence that SVM-based prediction slightly outperforms other classification techniques. To estimate diagnostic accuracy, the prediction model is built using a training set of patient samples. Accuracy is determined based on predictions in an independent test set. Apparent accuracy usually is determined by 10-fold crossvalidation (10-fold CV): The data set is divided into 10 approximately equally sized subsets, the prediction model is trained for 9 subsets and predictions are generated for the remaining subset. This training / prediction process is repeated 10 times to include predictions for each subset. Apparent accuracy is the overall rate of correct predictions. Random sets of training and test data can be generated iteratively to assess robustness of diagnostic accuracy. Since most software packages still rely on strong informatics/statistic knowledge and programming skills robust, intuitive and user-friendly software is needed. Bioconductor, for example, is an open source and open development software project to provide tools for the analysis and comprehension of genomic data.

Early microarray studies were focused on molecular definition of already known subtypes of cancer. The Breast cancers derived from BRCA1 and BRCA2 mutation carriers. The gene expression profiles associated with ER-protein expression, HER-2/neu, p53 mutational status etc.

Breast cancer is morphologically and genetically highly heterogeneous. The traditional methods for sub grouping breast cancer relying on pathological and clinical data. Data can only partially reflect the clinical diversity of the disease. Molecular profiling has well suited to phenotypic characterization of Breast cancers. Several microarray studies demonstrate that one can build a molecular taxonomy of Breast tumors using this technology.

Microarrays used to identify distinct sub groups among familial Breast cancers. Germaline mutations in BRCA1 and BRCA2 genes have been shown to leave a characteristic imprint on the panel of genes express by the tumors. Mutation screening techniques use in oncogenetic clinics are time consuming and expensive. Thus, the realization that tumors derived from individuals with BRCA1/ BRCA2 mutations each displays characteristic gene expression profile opens the possibilities for novel screening strategies.

The extended knowledge of defect causing the development of Breast cancer improves treatment strategies and intervention approaches for affected individuals in high risk families.

Predicting Sensitivity to Therapy in Breast Cancer

The factors found correlated to resistance or drug sensitivity have a potential biological explanation as specific target molecules for therapy applied , as gene products modifying therapy response through DNA

metabolism , as gene products participating in the execution of apoptosis and growth arrest.

Treatments, DNA repair and phenotype of progeny

This review highlights the prediction of available treatment & curable status for breast cancer. It also specifies the DNA repair and identifies the phenotype of progeny.

Microarray technology can, in principle, provide many types of information to help drug discovery and development. In their simplest application, microarrays can be used to screen for changes in gene expression following exposure of tumour cells to drugs either in culture or in patients following treatment. Studies of this type may provide information on the precise mechanism of action of the drug or could lead to the identification of early markers of drug response. For predicting the longer-term clinical response to drug treatment, microarray information could also yield information about possible side effects, and identify markers that, in a clinical context, could be used to predict possible adverse events. This technology can also provide information for identifying and validating new therapeutic targets.

Microarray technology can, in principle, provide many types of information to help drug discovery and development. In their simplest application, microarrays can be used to screen for changes in gene expression following exposure of tumor cells to drugs either in culture or in patients following treatment. Studies of this type may provide information on the precise mechanism of action of the drug or could lead to the identification of early markers of drug response. For predicting the longer-term clinical response to drug treatment, microarray information could also yield information about possible side effects, and identify markers that, in a clinical context, could be used to predict possible adverse events. This technology can also provide information for identifying and validating new therapeutic targets.

The DNA repair suggest the exclusion of mutate gene, also termed as domain, by using restriction enzyme. DNA repair involves several mechanisms such as base excision repair, nucleoside excision repair, mismatch repair as well as O^6 – methylguanine DNA mythyltransferase.

The progeny specifies the mode of inheritance for breast cancer. This inheritance has two common forms; Mendelian forms and non Mendelian forms.

Conclusion

Microarrays provide a means for the comprehensive and simultaneous analysis of the expression status of thousands of genes. The resulting signature allows the identification of a distinct molecular phenotype. It is anticipated that the application of microarrays will significantly improve molecular diagnostics and will provide deep insights into the pathogenetic alterations of malignant and non-malignant diseases which will allow the development of novel treatment

approaches. To realize these efforts it is important to cover well-defined questions in well-classified tumor samples. Furthermore, it is expected that the results of microarray experiments will allow the identification of disease-specific target structures and the design of novel and specific drugs.

The comparisons of different gene expression analyses will provide answers for both diagnostic and biological questions. Besides the above mentioned reports on disease-specific gene expression profiles in a variety of malignancies even the potential of a primary tumor to progress into the metastatic status may be detected and predicted by microarray analyses. Gene expression analyses are limited, however, to the transcriptional level and thus cannot be used to analyze alterations at the protein level. These aspects will be covered by novel techniques like proteomics.

With regard to leukemia the microarray analysis represents a novel promising method which may be used as a diagnostic tool in the near future. Today the diagnosis and subclassification of leukemia is based on the application of various techniques like cytomorphology, cytogenetics, fluorescence in situ hybridization, multiparameter flow cytometry, and PCR-based methods which are time-consuming and cost-intensive and require expert knowledge in central reference laboratories.

So we conclude that the proposed model which will gives the gene expression for responsible gene, identifies the type of tumor. The model will also give the drug resistance and response to gene therapy by suggesting the restriction enzyme and will identify the phenotype of progeny. This method is useful in marathwada region. Because this type of work yet not be done for this region. Hence it has tremendous scope.

References

1. Merck Manual Online, Breast Cancer". http://www.merck.com/mmpe/print/sec18/ch253/ch253e.html.
2. CancerMath.net Calculates survival with breast cancer based on prognostic factors and treatment. From the Laboratory for Quantitative Medicine, Massachusetts General Hospital.
3. World Cancer Report". International Agency for Research on Cancer. June 2003. http://www.iarc.fr/en/Publications/PDFs-online/World-Cancer-Report/World-Cancer-Report. Retrieved 2009-03-26.
4. "Fact sheet No. 297: Cancer". World Health Organization. February 2006. http://www.who.int/mediacentre/factsheets/fs297/en/index.html. Retrieved 2009-03-26.
5. "Male Breast Cancer Treatment". National Cancer Institute. 2006. http://www.cancer.gov/cancertopics/pdq/treatment/malebreast/healthprofessional. Retrieved 2006-10-16.
6. "Breast Cancer in Men". Cancer Research UK. 2007. http://www.cancerhelp.org.uk/help/default.asp?page=5075. Retrieved 2007-11-06.
7. "What Are the Key Statistics About Breast Cancer in Men?". American Cancer Society. September 27, 2007. http://www.cancer.org/docroot/CRI/content/CRI_2_4_1X_What_are_the_key_statistics_for_male_breast_cancer_28.asp?sitearea=. Retrieved 2008-02-03.
8. "ACS :: What Are the Key Statistics About Breast Cancer in Men?".

Cancer.org. http://www.cancer.org/docroot/CRI/content/CRI_2_4_1X_What_are_the_key_statistics_for_male_breast_cancer_28.asp?sitearea=. Retrieved 2010-05-08.

Websites cited:

A) www.pubmed.com.
B)http://www.atlasgeneticsoncology.org/Deep/MicroarraysID20045.html
C) www.ncbi.nlm.nih.gov.in