Journal of Advanced Research in Technology and Management Sciences
Volume: 01 Issue: 03 ISSN: 2582-3078 November 2019
Available online at: http://www.jartms.org

Framework of Classification with Microarray data Prediction Cancer Using Gene Express

Sanku Rajendra Kumar, Dr. Anil Kumar

Research Scholar, Associate Professor
Department of Computer Science and Engineering
Sri Satya Sai University of Technology & Medical Sciences, Bhopal Indore Highway, Pachama, Sehore (M.P.)-466001
Email : rajendhar.sanku@gmail.com

Abstract :-Classification of cancer based on gene expression has provided insight into possible treatment strategies. Analysis of gene expression is important in many fields of biological research in order to retrieve the required information. It provides accurate prediction providing better treatment to the patients. One of major challenges is to discover how to extract useful information from huge dataset. A high-capacity system was developed to monitor the expression of many genes in parallel. Microarrays prepared by high-speed robotic printing of complementary DNAs on glass were used for quantitative expression measurements of the corresponding genes. Because of the small format and high density of the arrays.

Cancer classification based on microarray has become a popular research topic in bioinformatics, which can be used to detect subtypes of cancers and produce therapies. Gene selection may provide insights into understanding the underlying mechanism of a specific biological phenomenon. Also, such information can be useful for designing less expensive experiments by targeting only a handful of genes. Finally, incorporating diverse types of genomic data (e.g., protein-protein interaction data and gene expression) increase the prediction accuracy as compared to using gene expression.

Keywords: Classification, gena expression, cDNA array

I. INTRODUCTION

Cancer classification based on gene expression data contains a large number of features, which requires a relatively large training set to learn a classifier (e.g., model) with a low error rate. Over-fitting a classification model can be avoided by choosing a subset of features (or genes) to learn a model. Feature selection methods can address the challenges arising from high data dimensionality and small sample size. Feature selection decreases the dimensionality of the feature space, and mitigates the challenge of small sample size prevalent in most microarray.

Describing gene expression starts necessarily with deoxyribonucleic acid (DNA), the very stuff genes are made from and ribonucleic acid (RNA). Both DNA and RNA are polymers, that is, molecules that are constructed by sequentially binding members of a small set of subunits called nucleotides into a linear strand or sequence. The RNA occurring as intermediate structure is referred to as messenger RNA (mRNA). The term transcription is commonly used to describe Step 1 and the term translation for Step 2. Collectively, the overall process consisting of transcription and translation is known as gene expression. Notice, in most organisms only a small subset of genomic DNA is capable of being transcribed to mRNA or expressed as proteins.

cDNA micro-arraying is currently widely used to asses differential gene expression:

1. Simultaneous hybridization of two samples labeled with different fluorescent dyes providing an intensity ration that reflects the relative mRNA level.
2. The cDNA microarrays consisting of a mix of all probes spotted on the array and hybridizations with both amplified and non-amplified targets.

In addition, a set of filters can be repeatedly closed to identify differentially expd genes in different experiments by striping and re-hybridization. Furthermore, cDNA array can be used to investigate gene induction and suppression at the same time without yield mg fault positive clones as in DDRT-PCR. Finally, there is no need to sub clone the differential expressed genes, because cDNA clones are directly arrayed in filters. It is obvious that cDNA array can be applied to studying gene expression and identifying new functional genes in any area or treatment, such as development, metabolism, hormone, signal transduction, human disease and stress response. However, caution should be taken to identify or monitor those genes with extremely low expression. With improvement in making arrayed cDNA filters and in quantifying expression patterns, together with other technology in reverse genetics such as gene targeting,

cDNA array will find its wide application in post genom. (e.g., genes) and information increase, it becomes more challenging to integrate the disparate data into a reliable classification model. The use of gene expression data to develop classification models presents The small number of cancer samples typically available to train the model compared with the number of features present (e.g., genes) can degrade the performance of the classifier and increase the risk.

Classification based on an integrative approach combining gene expression with other genomic information (e.g. protein-protein interaction data) improved the classification performance over using gene expression data alone. Classification methods applied to cancer data but have focused predominantly on the performance of the classification models from a computational perspective without an in-depth exploration of the biologically relevant information that could be extracted Using this reference in the analysis of cDNA microarray data enables reliable computations of expression levels, which is especially useful when comparing more than two hybridizations. This is because RNA-RNA hybrids are more stable than RNA-DNA or DNA-DNA hybrids. It single stranded RNA generated by in vitro transcription avoids the competitive hybridization resulting from the two strands in DNA probes. Microarray hybridization experiments. By pooling a small part of all the probes before they are spotted on the array and performing subsequent amplication, a batch of common reference is produced that shows a reproducible hybridization

The aim of the present study was to develop a common reference that enables comparison of expression levels in multiple microarray hybridizations. An important prerequisite for a common reference is to obtain a signal for each spot. We accomplished this by mixing all arrayed cDNAs and labeling them in one color. This approaches the ideal situation, but cell culturing is very time and space consuming. In addition, gene expression in the pooled cell lines may not represent all genes present on the array and it may change over time under even slightly different growth conditions and other variables like passage number. Furthermore, it is difficult to repeatedly quantify and pool large amounts of RNAs from multiple sources in a reliable and reproducible way.

II. LITERATURE SURVEY

Next generation sequencing methods (NGS) enable DNA and RNA-based analyses of uncultured species and, with exploiting the data cumulating in the data banks, biodiversity evaluation of phytoplankton can be renewed. Strong positive correlation between rRNA gene copy numbers and genome size (Prokopowich et al., 2003)

One challenge involves DNA/RNA extraction from the cells, as many comparative studies have described differences inisolation efficiencies. Although several reference databases exist for rRNA genes of prokaryotes (Decelleet al., 2015)

Environmental samples contain cells with diverse cell properties, varying in cell size and firmness of cell walls, which may favor certain cells when using particular extraction procedures. Various physical, chemical and enzymatic cell lysis protocols. Yuan et al. (2015)

III. RESEARCH METHODLOGY

Every organism has a unique DNA sequence, r barcode, associated with it. This DNA barcode is a highly variable region interspersed between conserved genomic regions. Gene expression analysis has become routine microarrays and RNA sequencing, which is helping researches discover novel RNA forms and variants. n methods applied to cancer data but have focused predominantly on the performance of the classification models from a computational perspective without an in-depth exploration of the biologically relevant information.

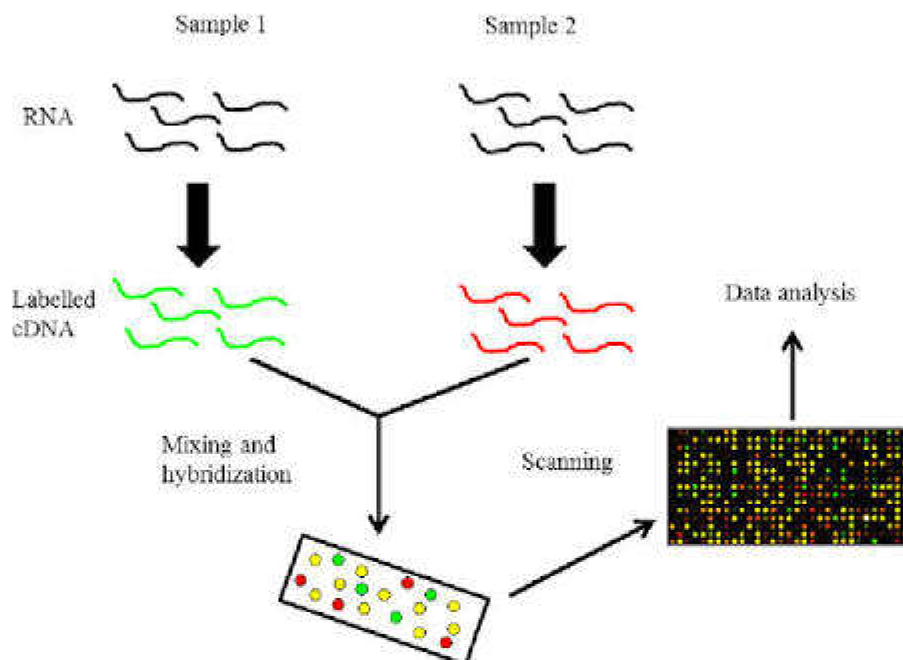

Figure 1: DNA Micro Array for Gene Expression

The synthesis of DNA from an RNA template, via reverse transcription, results in complementary DNA (cDNA). cDNA can then sever as template in a variety of downstream applications for RNA studies such as gene experession, therefore cDNA synthesis is the critical steps to help you ensure your cDNA synthesis results in highest efficiency.

- **Prepare Sample:** RNA servers as the template in cDNA synthesis. RNA is routinely used in cDNA synthesis for downstream applications. Maintaining RNA integrity is critical and requires special precautions during extraction, processing, storage and experimental. Best practices to prevent degradation of RNA include wearing gloves, pipetting and decontamination of work area.

- **Remove Genomic DNA**: Trace amounts of genomic DNA (gDNA) may be co-purified RNA. The traditional method of gDNA removal to preparations of isolated RNA . DNase must removed prior to cDNA synthesis

- **Perform cDNA Synthesis**: the critical step is during DNA polymerization. In this step, reaction temperature and duration may vary according to the primer choice.

- **Mixing and Hybridization:** In this process of combining two complementary single stranded DNA or RNA molecules and allowing them to form a single double stranded molecule through base paring. DNA mixuture interpreatation protocols is provided for application of the CPI/CPE method in the analysis. Labeled reference cDNA and labeled target cDNA pooled, purified and hybridized

- **Scanning:** cDNA scan accepts sequencing data and optionally variant files. The pipeline firstly performs an alignement step and followed by a customizable data analysis protocols. Finally, results are annotated and a user friendly report is genearated.

- **Statistical Analysis:** Interpreting a DNA analysis requires a valid scientific method for estimating the probability that a random might chance matches the forensic of DNA. Calculate the frequency of the complete multilocus genotype. The frequency of a complete genotype is calculated by multiplying assumption no of correlations.

IV. CONCLUSION

The cancer diseases are present in proportional to benevolent tumors, which may not be expand or distributed. Finding uniqueness between oncogenic tumors assumes a critical job in disease analysis and treatment. The group of sicknesses that consists of development of irregular cell will leads a cancer.

It can have impact on various parts of body by means of attack or spread the parts of body. The cancer was originated from features of physiochemical. cDNA microarray data enables reliable computations of expression levels which is especially useful when comparing more than two hybridizations. The sensitivity of the hybridization technique can be increased through the use of single-stranded RNA probes.

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2015. CA Cancer J Clin 2015; 65: 5–29.

2. Foreman D. Bray F. Brewster DH, et al., eds. Cancer Incidence in Five Continents. Volume X. Lyon, International Agency for Research on Cancer, 2014.

3. Brown P.O and Bostein D (1999) Exploring the new world of the genome with DNA microarrays, Nature Genet., Genet., 21, 33-37

4. Ferlay J, Soerjomataram I, Ervik Metal. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11. Lyon, IARC, 2013

5. Schena M. , Shalon, D. , Heller , R. et a1 . , Parallel human genome analysis: microarray-based expression monitoring of1 000 genes, Pm . Natl . Acad. Sci . USA, 1996, 93: 10614.

6. Berrar D., Dubitzky W., Granzow M., Eils R. (2001). Analysis of gene expression and drugactivity data by knowledge-based association mining. Proceedings of Critical Assessment of Microarray Data Analysis (CAMDA 2001), pp. 23-28.

7. Alizadeh,A., Eisen,M., Davis,R.E., Ma,C., Sabet,H., Tran,T., Powell,J.I.,Yang,L., Marti,G.E., Moore,D.T. et al. (1999) The lymphochip: a specialized cDNA microarray for the genomic-scale analysis of geneexpression in normal and malignant lymphocytes. Cold Spring Harbor Symp. Quant. Biol., 64, 71±78.

8. Schena, M., Shalon.D.,Schena,M., Shalon,D., Davis, R.W. and Brown,P.O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science, 270, 467±470.3.

9. Ross ena, M., Shalon.D.,Schena,M., Shalon,D., Davis, R.W. and Brown,P.O. (1995) Quantitative monitor Ross,D.T., Scherf,U., Eisen,M.B., Perou,C.M., Rees,C., Spellman,P. (2000) Systematic variation in gene expression patterns in human cancer celllines. Nature Genet., 24, 227±235.13.

10. Zehetner,G., Pack,M. and Schafer,K. (2001) Preparation and screening of high-density cDNA arrays with genomic clones. Methods Mol. Biol., 175,169±188.

11. Granjeaud S., Bertucci F., Jordan B.R. (1999). Expression profiling: DNA arrays in many guises. BioEssays 21(9):781-790.