

# Chapter 9

## De-Noising, Clustering, Classification, and Representation of Microarray Data for Disease Diagnostics

**Nitin Baharadwaj**

*Netaji Subhas Institute of Technology, India*

**Sheena Wadhwa**

*Netaji Subhas Institute of Technology, India*

**Pragya Goel**

*Netaji Subhas Institute of Technology, India*

**Isha Sethi**

*Netaji Subhas Institute of Technology, India*

**Chanpreet Singh Arora**

*Netaji Subhas Institute of Technology, India*

**Aviral Goel**

*Netaji Subhas Institute of Technology, India*

**Sonika Bhatnagar**

*Netaji Subhas Institute of Technology, India*

**Harish Parthasarathy**

*Netaji Subhas Institute of Technology, India*

### ABSTRACT

*A microarray works by exploiting the ability of a given mRNA molecule to bind specifically to the DNA template from which it originated under specific high stringency conditions. After this, the amount of mRNA bound to each DNA site on the array is determined, which represents the expression level of each gene. Qualification of the mRNA (probe) bound to each DNA spot (target) can help us to determine which genes are active or responsible for the current state of the cell. The probe target hybridization is usually detected and quantified using dyes/fluorophore/chemiluminescence labels. The microarray data gives a single snapshot of the gene activity profile of a cell at any given time. Microarray data helps to elucidate the various genes involved in the disease and may also be used for diagnosis /prognosis. In spite of its huge potential, microarray data interpretation and use is limited by its error prone nature, the sheer size of the data and the subjectivity of the analysis. Initially, we describe the use of several*

DOI: 10.4018/978-1-4666-4558-5.ch009

*techniques to develop a pre-processing methodology for denoising microarray data using signal process techniques. The noise free data thus obtained is more suitable for classification of the data as well as for mining useful information from the data. Discrete Fourier Transform (DFT) and Autocorrelation were explored for denoising the data. We also used microarray data to develop the use of microarray data as diagnostic tool in cancer using One Dimensional Fourier Transform followed by simple Euclidean Distance Calculations and Two Dimensional Multiple Signal Classification (MUSIC). To improve the accuracy of the diagnostic tool, Volterra series were used to model the nonlinear behavior of the data. Thus, our efforts at denoising, representation, and classification of microarray data with signal processing techniques show that appreciable results could be attained even with the most basic techniques. To develop a method to search for a gene signature, we used a combination of PCA and density based clustering for inferring the gene signature of Parkinson's disease. Using this technique in conjunction with gene ontology data, it was possible to obtain a signature comprising of 21 genes, which were then validated by their involvement in known Parkinson's disease pathways. The methodology described can be further developed to yield future biomarkers for early Parkinson's disease diagnosis, as well as for drug development.*

## **INTRODUCTION**

### **DNA Microarrays**

As a result of the rise in the number of completed genome sequencing projects, the DNA sequence present in a large number of genomes has become available. The central dogma of modern biology takes into account that the information present in the DNA is read to manufacture mRNA or messenger RNA. This, in turn, is read by the cellular machinery to manufacture proteins. The first step of this process is known as transcription (i.e., decoding of the DNA to synthesize mRNA while the second is translation or decoding of mRNA to produce functional proteins). Such a DNA fragment whose information is converted into a mRNA or a protein molecule is termed as a gene. Experimental and computational tools have helped us elucidate the presence of protein coding genes and thus map the vast, unexplored length of DNA. While all cells of the body contain the full set of DNA material (and genes), it is the subset that is transcribed or expressed that confers on the cell its unique properties. These genes are responsible for the response of the cell to its environment in both health and disease. Thus, the gene transcription profile of a cell gives an important insight into

the type of genes that are transcribed under any physiological and pathological condition. The gene transcription profile reflects the response of the cell to its environment, a highly and tightly regulated process. Such a gene transcription profile can be recorded using a microarray.

The microarray capitalizes on the ability of a given mRNA molecule to bind specifically or hybridize to the DNA template from which it originated. Thus, in a microarray experiment, probes consisting of DNA are attached to a solid surface by covalent bonding to a chemical matrix like epoxy-silane, lysine, polyacrylamide, and so forth. The solid surface can be a glass or a silicon chip, commonly known as a gene chip. A gene chip consists of thousands of microscopic spots, containing a small but specific DNA sequence. In order to measure gene expression data, short section of genes are anchored to the solid surface and used as probes to detect the presence of their templates (targets) in a biological sample of interest. The target has the ability to hybridize specifically with its complementary probe mRNA under high stringency experimental conditions. Hybridization can be detected and quantified by labeling the targets with fluorophore or chemiluminescent dye materials.

24 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/de-noising-clustering-classification-and-representation-of-microarray-data-for-disease-diagnostics/79726](http://www.igi-global.com/chapter/de-noising-clustering-classification-and-representation-of-microarray-data-for-disease-diagnostics/79726)

## Related Content

---

### Real Time Motion Detection for Traffic Analysis Using Computer Vision

Ashwin Sai C., Karthik Srinivas K. and Allwyn Raja P. (2020). *International Journal of Computer Vision and Image Processing* (pp. 1-14).

[www.irma-international.org/article/real-time-motion-detection-for-traffic-analysis-using-computer-vision/252230](http://www.irma-international.org/article/real-time-motion-detection-for-traffic-analysis-using-computer-vision/252230)

### Video Summarization by Redundancy Removing and Content Ranking

Tao Wang, Yue Gao, Patricia Wang, Wei Hu, Jianguo Li, Yangzhou Du and Yimin Zhang (2011). *Computer Vision for Multimedia Applications: Methods and Solutions* (pp. 91-101).

[www.irma-international.org/chapter/video-summarization-redundancy-removing-content/48315](http://www.irma-international.org/chapter/video-summarization-redundancy-removing-content/48315)

### Design, Construction, and Programming of a Mobile Robot Controlled by Artificial Vision

Claudio Urrea and Alex Yau (2017). *Advanced Image Processing Techniques and Applications* (pp. 230-250).

[www.irma-international.org/chapter/design-construction-and-programming-of-a-mobile-robot-controlled-by-artificial-vision/177769](http://www.irma-international.org/chapter/design-construction-and-programming-of-a-mobile-robot-controlled-by-artificial-vision/177769)

### A Robust Skeletonization Method for Topological Complex Shapes

Donatella Giuliani (2017). *International Journal of Computer Vision and Image Processing* (pp. 1-18).

[www.irma-international.org/article/a-robust-skeletonization-method-for-topological-complex-shapes/177198](http://www.irma-international.org/article/a-robust-skeletonization-method-for-topological-complex-shapes/177198)

### High-Speed, High-Resolution 3D Imaging Using Projector Defocusing

Song Zhang and Yuanzheng Gong (2012). *Depth Map and 3D Imaging Applications: Algorithms and Technologies* (pp. 121-140).

[www.irma-international.org/chapter/high-speed-high-resolution-imaging/60263](http://www.irma-international.org/chapter/high-speed-high-resolution-imaging/60263)