DNA Microarrays and Smart Pooling.

Deepthi P, C.H.Renumadhavi

Department of Instrumentation Technology (BMSP&I), RVCE

Abstract: *DNA microarray paves methods in identifying different expression levels of thousands of genes during biological process.Microarray technology helps biologists for monitoringexpression of thousands of genes in a single experiment on a small chip. Microarray is also called as DNA chip, gene chip, orbiochip is used to analyze gene expression. DNA microarraysare rapidly becoming a fundamental tool in genomic research. The problem with microarrays is to measure gene expression from thousands of genes (features) from only tens of hundreds of samples.The compressibility of microarray data to reduce the number of chips used and increase the robustness to noise in measurements can be tested by employing smart pooling.*
Keywords: *microarray,gene expression, smart pooling*

I.   Introduction

Gene microarrays are an emerging technology with the potential to make an enormous impact on medical research and clinical practice. They will generate enormous amounts of information about the activity of genes in development and disease.The most common use for DNA microarrays is to findgenes with changed expression between two samples. Forexample, one might look for genes that are expressed morehighly in cancerous versus healthy lung tissue. Microarray technology monitors thousands of geneexpression levels simultaneously [1].

DNA chips are glass surface bearing thousands of DNA fragments at discrete sites at which the fragments are available for hybridization. Hybridization of RNA and DNA derived samples to DNA chips allows the monitoring of gene expression or occurrence of polymorphisms in genomic DNA.

Microarray experimentsproduce data usually in the form of large matrices ofexpression levels of genes (rows) under different experimentalconditions (columns).In most cases, microarray data sets frequently containsome missing values; typical reasons include insufficientresolution, image corruption, spotting or scratches on theslide, dust or hybridization failures [2]. Therefore missingvalue prediction is essential as a pre-processing step to obtainproper results from microarray data analysis. There are severalapproaches to deal with missing values.

The first approach iselimination, which removes objects containing missing values[3], that loses too much useful information and bias theresults. The second approach substitute the missing values bya global constant such as 0 [3], or by the average of theavailable values for that gene [4], which distorts therelationships among variables. The third approach usesimputation method where the gene with missing value isselected in first step and in second step the missing values arepredicted using the observed values of selected genes.

II.   Gene Microarray Technology

A human being's genetic makeup is determined by billions of DNA base pairs, which are believed to be organised into about 100,000 genes. Each gene has a specific function, which is determined by its particular sequenceof base pairs. Some genes encode the instructions for the manufacture of proteins. Others are believed to regulate the rates at which development occurs.

2.1 Making microarrays.
* Robotic spotting.
* In-situ synthesis.

Robotic spotting: Three steps are involved in forming microarray.
* Making DNA probe.
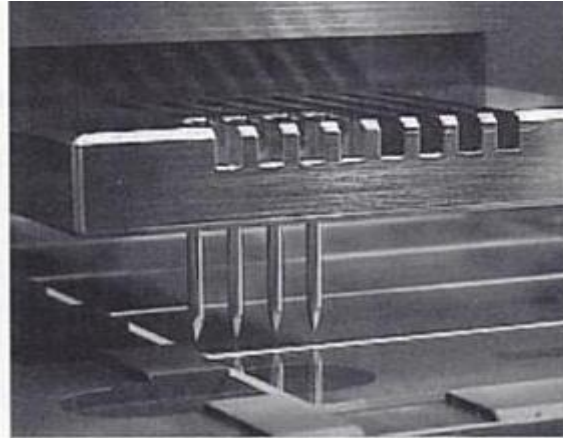* Spotting DNA onto glass surface.
* Post spotting processing of slide.

Fig 1.Spotting Robot.

The spotting robot consists of series of pins arranged in cassette,pins are used to transfer sample from microtier plate to glass slide.Most array making robots have pins with reservoirs of liquid,this enables spotting samples on arrays without returning to sample plate ,increasing throughput.
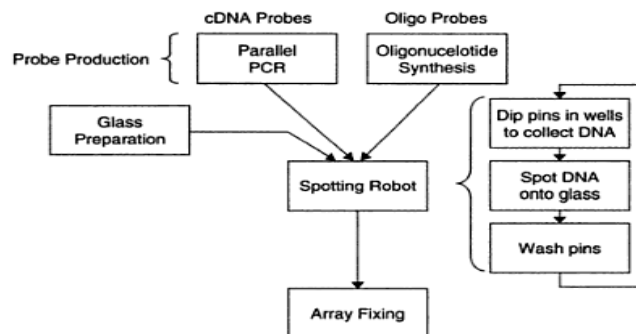

Fig 2.Spotted array synthesis.

Sythesis of microarrays starts with pin depositing samples to glass slide,and wash the pins to remove residual sample.

In-situ Synthesis :Instead of using presynthesizedoligonucleotides,oligos are built base by base,and protective group(PG) is added after adding base to prevent addition of more base.Photolithography techniques are used to convert (PG) to hydroxyl group for attaching to another nucleotide.

Affimetrixtechnology:Photolithography where light is directed towards PG through mask to convert PG to hydroxyl group.

Fibet and Nimblegen:Light is directed towards the target areas without using mask,it is controlled by computer .

Ink jet technology:Base are fired to desired areas as programmed in computer,can generate arrays at faster rate,flexible.

2.2 Using Microarrays.

The base pairs are organised into two strands in normalDNA. These strands can be separated by heating the DNA, exposing it to ultraviolet light, or subjecting it to the activity of an appropriate enzyme. Single strands of DNA have the property that they will spontaneously attach themselves to complementary strands to form the normal double-stranded structure. This behaviour persists even when the DNA has been cut up into short segments.Single strands of DNA can therefore act as templates on which complementary strands can be constructed, or as substrates to which complementary fragments of DNA can attach themselves.This property OF spontaneous attachment is used to advantage in gene microarray technology.

Sample preparation and Labeling: Samples of known base pair sequences are spotted onto glass substrates(specially coated microscope slides) using a robotic printing device. A sample of DNA whose composition is unknown is separated into strands, which are cut into pieces using enzymes. Molecules of a fluorescent compound are attached to these pieces as Labels.

Hybridization: The array is then bathed in R solution of the labelled DNA fragments, which attach themselves to the samples on the substrate as shown in fig 3.
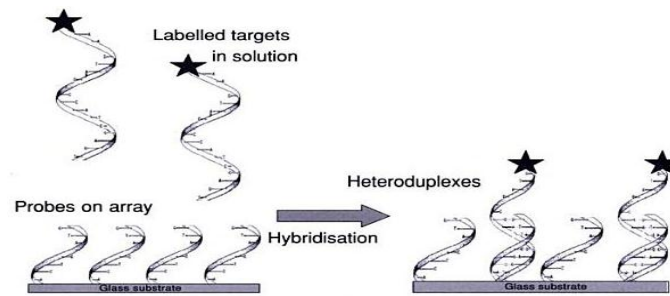
Fig 3.Hybridization

Image Aquisition: The extent of this attachment can bedetermined by exciting the fluorescent labels with ultraviolet or laser light and measuring the amount of light emitted.To create an array, samples of DNA base sequences are printed onto a specially prepared microscope slide as shown in fig 4.
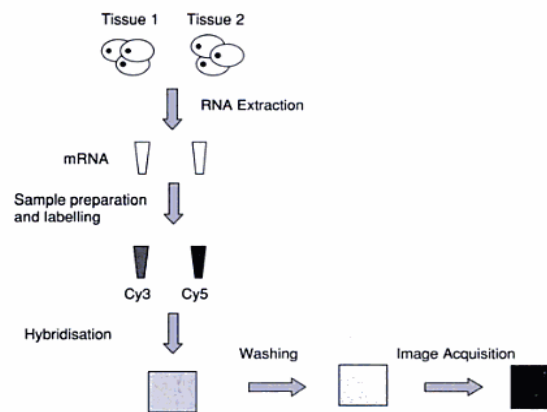

Fig 4.Image acquisition.

Manual processing of this information would be extremely tedious and prone to error. Automation of the analysis of this information is essential. The first step in the automation of this process is the digitisation of the pattern of fluorescence on the slide and the measurement of the intensity of the fluorescence of each spot. This is a non-trivial irnage processing problem.

Introducing Compressive sensing:Compressive sensing microarrays (CSMs) are DNA-based sensors that operate using group testing and compressive sensing (CS) principles. In contrast to conventional DNA microarrays, in which each genetic sensor is designed to respond to a single target, in a CSM, each sensor responds to a set of targets.

III.   Smart Pooling

Pooling of mRNA samples in microarray experiments implies mixing mRNA from several biological-replicate samples before hybridization onto a microarray chip. alternative smart pooling strategy in which different samples, not necessarily biological replicates, are pooled in an information theoretic efficient. Further, each sample is tested on multiple chips, but always in pools made up of different samples. The end goal is exploit the compressibility of microarray data to reduce the number of chips used and increase the robustness to in measurements.The mathematical properties of smart pooling designs ensure that each sample is tested on multiple chips, but always in pools made up of a different set of samples, such that, data from all the chips taken together capture the same information as the standard one-sample-one-chip approach. Because of the convolution step involved in testing pools of samples on multiple chips, the measurements made from the smart pooling strategy must be decoded to obtain the gene expression value of each gene in every sample. To save cost and to accurately transmit information across digital communication channels, where bandwidth is limited and the channel is noisy, a similar compression and recovery strategy is used. Similarly, smart pooling can achieve an overall savings by using fewer microarray chips than samples being tested. The built-in redundancy of testing each sample on multiple microarray chips can also provide robust expression measurements

Significance of microarray: The aim is to identify novel rare alleles, or detect individuals who are carriers of known SNPs in a large population.

3.1 Methodology

Sparsity: A key requirement for any smart pooling strategy is that the data must be sparse to allow compression. A gene's expression profile is said to be sparse if, across several samples, there are only a small number of samples in which the gene's expression is significantly different from its median expression value in these samples.Sparsity can be obtained from group testing.

Smart pooling :The central idea of smart pooling is to exploit sparsity in agene's expression across several samples.The goal of smart pooling is to recover the unique expression value of each gene in each of the samples being tested. To achieve this goal, samples are pooled in a systematic way across several chips while keeping the number of chips used lower than the number of samples tested. The pooling design ensures that each sample is tested on multiple chips, always in pools with different samples.

Decoding: Expression of each gene in all samples is decoded accurately using reconstruction algorithms.

IV.   Conclusion

In this paper, we explore gene microarray technology its application and determine missing value using the different techniques. Hybridization model is to be designed to predict whether given probes mimic the behavior of a binary matrix(pool design), and algorithms are designed, respectively, to find probe sequences satisfying the binding requirements,and decoding the expressions after pooling.

V.   Future Scope

Microarrays are massively parallel biosensors thatcan simultaneously detect and quantify large numberof different genomic particles. A DNA microarrayis a nucleic acid-based microarray that contains probespots testing a multitude of targets in one experiment.Ideas from compressive sensing have been utilized indifferent ways in the analysis of DNA microarrays. Oneof the methods is compressed microarrays,where each spot contains copies of several probes andthe total number of spots is lower, resulting in significantlyreduced costs due to cheaper array manufacturing the measurements that canbe captured by compressed microarrays can be recoveredaccurately using the norm-minimizationmethods.

References

[1]   Peng X, Wood C, Blalock E, Chen K, Landfield P, Stromberg A: Statistical implications of pooling RNA samples for microarray experiments**.** BMCBioinformatics.
[2]   Dai W, Sheikh MA, Milenkovic O, Baraniuk RG: Compressive sensing DNA microarrays**.** EURASIP journal on bioinformatics & systems biology 2009.
[3]   DeVore RA: Deterministic constructions of compressed sensing matrices**.** J Complex 2007, 23(4-6):918-925.
[4]   Bakhtiyar Uddin ,M. EmreCelebi, Hassan Kingravi, Gerald Schaefer "Accurate Genomic Signal Recovery using Compressed Sensing" 21st International Conference on Pattern Recognition (ICPR 2012) November 11-15-2012. Tsukuba, Japan.
[5]   Microarray[online]:http://www.books.google.co.in/books/about/Microarray_Bioinformatics.html.
[6]   Ahmed Fadiel , Frederick Naftolin ,"Microarray applications and challenges: a vast array of possibilities" Int Arch Biosci 2008.
[7]   UgrappaNagalakshmi,KarlWaern,Michael Snyder "RNA-Seq: A Method for Comprehensive Transcriptome Analysis" Curr. Protoc. Mol. Biol. 89:4.11.1-4.11.13.2010.
[8]   K, C.Y. Cheung, Christopher J.S. deSilva "Analysis of Gene Microarray Images" 0-7803-5871-h/991999 IEEE.
[9]   Fred van Ruissen, Frank Baas, "Serial Analysis of Gene Expression (SAGE)"volume 383, 2007, pp 41-66.
[10]   Todd C. Mocklera, Joseph R. Eckera, "Applications of DNA tiling arrays for whole-genome analysis" Genomics 85 (2005).