Review III: DNA Microarrays

Rajan Munshi

BBSI @ Pitt 2006

Department of Computational Biology
University of Pittsburgh School of Medicine

May 25, 2006

Also called

- DNA chips
- biochips
- gene chips
- gene arrays
- genome chips
- genome arrays
What is a microarray?

- An arrangement of DNA sequences on a solid support
- Each microarray contains thousands of genes
- Able to simultaneously monitor gene expression levels in all these genes
- Used for:
  - gene expression studies
  - disease diagnosis
  - pharmacogenetics (drug discovery)
  - toxicogenomics

Types

- Two basic microarray technologies
- cDNA arrays (Stanford)
- High-density oligonucleotide arrays (Affymetrix)
- Each technology has its merits and demerits
Definition

Solid support: glass slides, plastic base

High-density oligonucleotide arrays (1)

- Pioneered by Affymetrix (GeneChip®)
- DNA probe sequences are 25-mer fragments
- Built in situ ("on-chip") by photolithography
- Uses 1 fluorescent dye
High-density oligonucleotide arrays (2)

- Each sequence is represented by a probe set
- 1 probe set = 16 probe pairs
- Each probe pair = 1 Perfect Match (PM) probe cell and 1 MisMatch (MM) probe cell
- PM = perfectly complementary to target
- MM = central base is mismatched to target

Affymetrix Probe Sets

5' // 3'

GATGTTGGATCCGTACTCCATGCCTAGCTAGCTAGTCCGTATGGCTACCAAT

PM

MM
Affymetrix chip

A single probe set
cDNA arrays

- Also known as spotted arrays
- Support can be glass or membrane
- DNA sequences are robotically “imprinted”
- Sequences can range from 30 bp to 2 kb
- Sequences are cDNA clones
- Uses 2 fluorescent dyes (cy3, cy5)
cDNA arrays

Animation
(Courtesy: Dr. A. Malcolm Campbell, Davidson College, NC)
(www.bio.davidson.edu/courses/genomics/chip/chip.html)

Genome-on-a-chip (yeast)
General Steps

<table>
<thead>
<tr>
<th>Probe</th>
<th>Chip Fabrication</th>
<th>Target</th>
<th>Assay</th>
<th>Readout</th>
<th>Informatics</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA or cDNA with known identity</td>
<td>Putting probes on chip (robotic imprinting, photolithography)</td>
<td>Fluorescently labeled cDNA (single channel, dual channel)</td>
<td>Hybridization (Southern Blot)</td>
<td>Fluorescence intensities, fold-change ratios (up- or down-regulated)</td>
<td>Visualization, data mining</td>
</tr>
</tbody>
</table>

Analysis

- Low-level analysis
  - Extraction of signal intensities
  - Normalization of samples

- High-level analysis
  - Unsupervised learning (clustering)
    - Aggregation of a collection of data into clusters based on different features in a data set (e.g. heirarchical clustering, SOM)
  - Supervised learning (class discovery)
    - Incorporates knowledge of class label information to make distinctions of interest by using a training set.
Low-level analysis

Gene Expression Intensity (Signal)

In other words, a numerical value is obtained

Now, these values can be compared because fluorescence intensity is directly proportional to gene expression

High-level analysis

Now what??
High-level analysis (Hierarchical Clustering)

- Algorithm that “pairs” similarly expressed genes
- Uses Pearson’s correlation coefficient ($r$)
- Useful to gain a general understanding of genes involved in pathways

Time course of serum stimulation of human fibroblasts

- Identify clusters of genes that are co-regulated
- Identification of novel genes
- Very widespread method for microarray analysis
High-level analysis (self-organizing maps)

- Algorithm that clusters genes based on similar expression values
- Useful for finding patterns in biological data
- Cocaine study
- 5 regions of the rat brain under treated and untreated conditions
- e.g. cluster 3

Overall Goal

>10,000 genes

<50 genes

Identify potential therapeutic targets

Experimental confirmation
Potential Problems

- Local contamination

Array Contamination
Potential Problems

- Local contamination
- Normalization
- Statistical significance of difference in expression
- cDNA arrays
  - must have the genes cloned
  - need relatively pure product
- Affymetrix arrays
  - need sequence information

Additional Reading

- Affymetrix website: www.affymetrix.com
- Stanford University: genome-www.stanford.edu
- Nature Genetics, vol. 21 supplement, “The Chipping Forecast”
- www.microarray.org
- www.gene-chips.com/
- ihome.cuhk.edu.hk/~b400559/array.html
- www.stat.wisc.edu/~yandell/statgen/reference/array.html