Introduction to DNA microarray technologies

Sandrine Dudoit, Robert Gentleman, Rafael Irizarry, and Yee Hwa Yang

© Copyright 2002, all rights reserved
Outline

• Basic principles

• cDNA microarrays

• Affymetrix oligonucleotide chips
DNA microarrays
DNA microarrays rely on the hybridization properties of nucleic acids to monitor DNA or RNA abundance on a genomic scale in different types of cells.

The ancestor of cDNA microarrays: the Northern blot.
Hybridization

- **Hybridization** refers to the annealing of two nucleic acid strands following the base-pairing rules.

- Nucleic acid strands in a duplex can be separated, or **denatured**, by heating to destroy the hydrogen bonds.
Hybridization

Nucleic Acid Hybridization
Hybridization
Gene expression assays

The main types of gene expression assays:

- Serial analysis of gene expression (SAGE);
- **Short oligonucleotide arrays (Affymetrix);**
- Long oligonucleotide arrays (Agilent Inkjet);
- Fibre optic arrays (Illumina);
- cDNA arrays (Brown/Botstein).
Transcriptome

- mRNA or transcript levels sensitively reflect the state of a cell.
- Measuring protein levels (translation) would be more direct but more difficult.
Transcriptome

• The transcriptome reflects
  – Tissue source: cell type, organ.
  – Tissue activity and state:
    • Stage of development, growth, death.
    • Cell cycle.
    • Disease vs. healthy.
    • Response to therapy, stress.
Applications of microarrays

- **Cancer research**: Molecular characterization of tumors on a genomic scale
  → more reliable diagnosis and effective treatment of cancer.
- **Immunology**: Study of host genomic responses to bacterial infections; reversing immunity.
- ...
Applications of microarrays

• Compare mRNA (transcript) levels in different types of cells, i.e., vary
  – Tissue: liver vs. brain;
  – Treatment: drugs A, B, and C;
  – State: tumor vs. non-tumor, development;
  – Organism: different yeast strains;
  – Timepoint;
  – etc.
cDNA microarrays
cDNA microarrays

Prepare cDNA target

"Normal"

Tumor

RT / PCR

Label with fluorescent dyes

Combine equal amounts

Hybridize target to microarray

Microarray Technology

Prepare Microarray

SCAN
cDNA microarrays

- The **relative abundance** of a spotted DNA sequence in two DNA or RNA samples may be assessed by monitoring the **differential hybridization** of these two samples to the sequence on the array.
- **Probes**: DNA sequences spotted on the array, immobile substrate.
- **Targets**: Nucleic acid samples hybridized to the array, mobile substrate.
cDNA microarrays

- The **ratio** of the red and green fluorescence intensities for each spot is indicative of the relative abundance of the corresponding DNA probe in the two nucleic acid target samples.
cDNA microarrays

\[ M = \log_2 \frac{R}{G} = \log_2 R - \log_2 G \]

- **\( M < 0 \)**, gene is over-expressed in green-labeled sample compared to red-labeled sample.
- **\( M = 0 \)**, gene is equally expressed in both samples.
- **\( M > 0 \)**, gene is over-expressed in red-labeled sample compared to green-labeled sample.
Mixture of neuron cells (Control) Label with Green Fluorescent Dye

Certain type of Neuron cell Label with red Fluorescent Dye

cDNA

Denature

Combine equal amounts

Hybridization

Denature cDNA Microarray

RIKEN Mouse EST (sequenced gene)
The process

**Building the microarray:**
- MASSIVE PCR
- PCR PURIFICATION AND PREPARATION
- PREPARING SLIDES
- PRINTING

**RNA preparation:**
- CELL CULTURE AND HARVEST
- RNA ISOLATION
- cDNA PRODUCTION

**Hybing the array:**
- ARRAY HYBRIDIZATION AND SCANNING
- TARGET LABELING
- POST PROCESSING

**Data analysis:**
- DATA ANALYSIS
The arrayer

Ngai Lab arrayer, UC Berkeley

Print-head
Print-tips collect cDNA from wells

96-well plate
Contains cDNA probes

Glass slide
Array of bound cDNA probes
4x4 blocks = 16 print-tip-groups

Print-tip group 1

cDNA clones

Print-tip group 7

Print-tips collect cDNA from wells
Sample preparation
Hybridization

Binding of cDNA target samples to cDNA probes on the slide

Hybridize for 5-12 hours
Hybridization chamber

- Humidity
- Temperature
- Formamide
  (Lowers the Tmp)
Scanning

Detector

PMT

Image

Duplicate spots

Cy5: 635nm
Cy3: 532nm
RGB overlay of Cy3 and Cy5 images
Raw data

E.g. Human cDNA arrays
• ~43K spots;
• 16–bit TIFFs: ~ 20Mb per channel;
• ~ 2,000 x 5,500 pixels per image;
• Spot separation: ~ 136um;
• For a “typical” array, the spot area has
  – mean = 43 pixels,
  – med = 32 pixels,
  – SD = 26 pixels.
Oligonucleotide chips
Probe sets

- Each gene is represented by 16-20 oligonucleotides of 25 base-pairs, i.e., 25-mers.
- **Perfect match probe, PM**: A 25-mer complementary to the reference sequence.
- **Mismatch probe, MM**: same as PM but with a single homomeric base change for the middle (13\(^{\text{th}}\)) base.
- **Probe pair**: A (PM,MM) pair.
- **Probe set**: 16-20 probe pairs.
- The purpose of the MM probe design is to measure non-specific binding and background noise.
Probe sets

Figure 1-3 Expression tiling strategy
Oligonucleotide chips

GeneChip Probe Array

Hybridized Probe Cell

Single stranded, labeled RNA target
Oligonucleotide probe

Millions of copies of a specific oligonucleotide probe

>200,000 different complementary probes

Image of Hybridized Probe Array

Compliments of D. Gerhold
Oligonucleotide chips

• The probes are synthesized \textit{in situ}, using combinatorial chemistry and photolithography.

• \textbf{Probe cells} are square-shaped features on the chip containing millions of copies of a single 25-mer probe. Sides are 18-50 microns.
Oligonucleotide chips

The manufacturing of GeneChip® probe arrays is a combination of photolithography and combinational chemistry.
Image analysis

• About 100 pixels per probe cell.
• These intensities are combined to form one number representing the expression level for the probe cell oligo.
• → CEL file with PM or MM intensity for each cell.
Expression measures

- Most expression measures are based on differences of PM-MM.
- The intention if to correct for background and non-specific binding.
- E.g. MarrayArray Suite® (MAS) v. 4.0 uses Average Difference Intensity (ADI) or $\text{AvDiff} = \text{average of PM-MM}$.
- Problem: MM may also measure signal.
- More on this in lecture *Pre-processing in DNA microarray experiments.*
What is the evidence?

Statistics and Microarrays

1. Biological question
2. Experimental design
3. Microarray experiment
4. Image analysis
5. Normalization
6. Estimation
7. Testing
8. Clustering
9. Discrimination
10. Biological verification and interpretation
Everywhere ...

- for statistical design and analysis:
  - pre-processing, estimation, testing, clustering, prediction, etc.
- for integration with biological information resources (in house and external databases)
  - gene annotation (GenBank, LocusLink);
  - literature (PubMed);
  - graphical (pathways, chromosome maps).
Integration of biological metadata

- Expression, sequence, structure, annotation, literature.
- Integration will depend on our using a common language and will rely on database methodology as well as statistical analyses.
- This area is largely unexplored.
WWW resources

- Complete guide to “microarraying”
  http://cmgm.stanford.edu/pbrown/mguide/
  http://www.microarrays.org
  - Parts and assembly instructions for printer and scanner;
  - Protocols for sample prep;
  - Software;
  - Forum, etc.
- cDNA microarray animation
  http://www.bio.davidson.edu/courses/genomics/chip/chip.html
- Affymetrix
  http://www.affymetrix.com
Next …

Pre-processing in DNA microarray experiments

- cDNA microarrays
  - Image analysis;
  - Normalization.

- Affymetrix oligonucleotide chips
  - Image analysis;
  - Normalization;
  - Expression measures.