Outline

• Basic principles

• Spotted DNA 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

(a) DNA is denatured by heating

(b) Renaturation on cooling

(c) RNA

(d) DNA/RNA hybrid

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);
- Spotted cDNA arrays (Brown/Botstein).
Applications of microarrays

- Measuring transcript abundance (cDNA arrays);
- Genotyping;
- Estimating DNA copy number (CGH);
- Determining identity by descent (GMS);
- Measuring mRNA decay rates;
- Identifying protein binding sites;
- Determining sub-cellular localization of gene products;
- …
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.
- ...
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

• 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.
Spotted DNA microarrays
Spotted DNA microarrays

**Prepare cDNA target**

- "Normal"
- Tumor
- RT / PCR
- Label with Fluorescent Dyes

**Combine Equal Amounts**

**Hybridize target to microarray**

**Microarray Technology**

**Prepare Microarray**

**SCAN**
Spotted DNA 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.
Spotted DNA 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.
Spotted DNA 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.
The process

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

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

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

Print-tip group 7

cDNA clones
Sample preparation

human sample collection
tissue banks

→

model systems

→

pathology

→

microdissection

→

sources of RNA

→

cell lines
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

- Pairs of 16-bit TIFFs, one for each dye.
- E.g. Human cDNA arrays:
  - ~43K spots;
  - ~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.
Animation

http://www.bio.davidson.edu/courses/genomics/chip/chip.html
Oligonucleotide chips
Terminology

- Each gene or portion of a gene is represented by 16 to 20 oligonucleotides of 25 base-pairs.

- **Probe**: an oligonucleotide of 25 base-pairs, i.e., a 25-mer.
- **Perfect match (PM)**: A 25-mer complementary to a reference sequence of interest (e.g., part of a gene).
- **Mismatch (MM)**: same as PM but with a single homomeric base change for the middle (13th) base (transversion purine <-> pyrimidine, G <->C, A <->T).
- **Probe-pair**: a (PM,MM) pair.
- **Probe-pair set**: a collection of probe-pairs (16 to 20) related to a common gene or fraction of a gene.
- **Affy ID**: an identifier for a probe-pair set.
- The purpose of the MM probe design is to measure non-specific binding and background noise.
Figure 1-3 Expression tiling strategy
## Spotted vs. Affymetrix arrays

<table>
<thead>
<tr>
<th>Spotted arrays</th>
<th>Affymetrix arrays</th>
</tr>
</thead>
<tbody>
<tr>
<td>One probe per gene</td>
<td>16 – 20 probe-pairs per gene</td>
</tr>
<tr>
<td>Probes of varying length</td>
<td>Probes are 25-mers</td>
</tr>
<tr>
<td>Two target samples per array</td>
<td>One target sample per array</td>
</tr>
</tbody>
</table>
Oligonucleotide chips

GeneChip Probe Array

Hybridized Probe Cell

- Single stranded, labeled RNA target
- Oligonucleotide probe

- Image of Hybridized Probe Array

- Millions of copies of a specific oligonucleotide probe
- >200,000 different complementary probes

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.
• 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 is 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 DNA Microarray Data*.
What is the evidence?

Biological question

Experimental design

Microarray experiment

Image analysis

Expression quantification

Normalization

Pre-processing

Analysis

Estimation

Testing

Clustering

Prediction

Biological verification and interpretation
Everywhere …

• Statistical design and analysis:
  – image analysis, normalization, estimation, testing, clustering, prediction, etc.

• Integration of experimental metadata with biological metadata from WWW-resources
  – gene annotation (GenBank, LocusLink);
  – literature (PubMed);
  – graphical (pathways, chromosome maps).
Integration of experimental and 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 DNA Microarray Data

• Spotted DNA microarrays
  – Image analysis;
  – Normalization.

• Affymetrix oligonucleotide chips
  – Image analysis;
  – Normalization;
  – Expression measures.