Terminologies

- Reporter: the nucleotide sequence present in a particular location on the array (a.k.a. probe)
- Feature: the location of a reporter on the array
- Composite sequence: a set of reporters used collectively to measure an expression of a particular gene
- Sample: the biological material from which the nucleic acid has been extracted for subsequent labeling and hybridization (a.k.a. target)
- Experiment: a set of one or more hybridizations that are in some way related

Microarray Process
Clone Selection

Download EST sequences

Gene Assembly

Tentative Singlets (TS) Tentative Contigs (TC)

BLAST (Remove vector, ribosomal, ct and mt DNA) (Remove duplications)

BLAST

Get representative EST

Re-array cDNA Clones

I.D. verification

Non-redundant Clone set

Clone Selection
ARRAY FABRICATION

cDNA Clones
384 well plates

Re-array

Non-Redundant Clones
384 well plates

Clean-up
PCR Products

PCR DNA Amplification

Plasmid DNA isolation

96 well plates

96 well plates

Printing Plate

Microarray

96 well plates

Template preparation: gelqc

Single band/clone
Yield > 2 ug/PCR reaction

TEMPLATE PREPARATION
ARRAY SPOTTER

Omni-grid from GeneMachines

- +/- 2.5 um accuracy
- Can print 100 slides at a time
- Can hold 30 384 printing plates
- 48-pen printhead
- Multiple pin types/brands
- Wash/sonicator
- Humidity control
- HEPA filter
- Blotting capability
- Enclosure
- Sample tracking software

PRINTING PENS

Pin should allow multiple stamps (>100 stamps/dip)
Dips into solution – pick up 0.2-1.0 ul (depends on the size of the pin)
Delivers approximately 0.25-1.0 nl of solution
DNA concentration of 0.5 ug/ul at 0.5 nl/print will give a spot of 0.25 ng DNA
Substrate Coupling Chemistry

1. Non-covalent

2. Covalent
HYBRIDIZATION CHAMBER

MICROARRAY READER

ScanArray 5000XL from PerkinElmer (GSI Lumonics)

- 5-50 um/pixel resolution
- Sequential scanning
- ~5 minutes/scan
- 5 Lasers
- Quantarray software
Data Collection

Control RNA hybridized to probe
Experimental RNA hybridized to probe
Both Control & Experimental RNA hybridized to probe
Neither hybridized to probe
Problems with Microarray analysis

Error can be introduced at every step

Controls

Statistical problems because the data is very high dimensional with very little replication

Normalization within and between arrays

Many assumptions are made that go untested
**Oligonucleotide arrays (DNA Chips)**

- A DNA chip is a solid surface, on which are immobilized a large number of different DNA molecules
- The molecules are produced by step wise synthesis *in situ*
- Two companies have competing methods:
  - Affymetrix using photolithographic techniques
  - Agilent using modified ink-jet deposition technologies
- A predicted open reading frame or gene is represented by a series of oligonucleotides (about 25 mer)
- Each gene is represented by between 10 and 20 different oligonucleotides
- To control for cross-hybridization with similar short sequences, mismatch controls are printed adjacent to each oligonucleotide
The first step in DNA Chip manufacturing is the attachment to a silanized glass surface of linkers carrying photochemically-removable protecting groups.

Light is directed through a photolithographic mask to specific areas of the synthesis surface. This activates specific areas for coupling.

The first of a series of blocked nucleotides (MeNPOC-dT for example) is incubated with the surface. Coupling occurs over the areas that previously had been illuminated.

Light is next directed over other areas of the surface. This activates different sites, including perhaps some occupied by the MeNPOC-dT linked to surface during the previous cycle.

A second blocked nucleotide (for example MeNPOC-dC) is then incubated with the surface. In the case of the areas exposed to light in both cycles, T then C will be attached at that location.

Via appropriate sequencing of masks and chemicals, a defined array of specific oligonucleotides can be assembled on a solid support.

The probes are at most 25 bases long, and ~5% are full-length.

The arrays themselves are 640x640 and the synthesized DNA features are 24µm squares.

Total features are about 410,000

Available arrays:
  E. coli, Human, mouse, rat, murine, C. elegans, Drosophila, P. aeruginosa, B. subtilis, yeast, Arabidopsis, barley
Millions of DNA strands built up in each cell

500,000 cells on each GeneChip® array

Actual strand = 25 base pairs

Actual size of GeneChip®
Eukaryotic Target Labeling for GeneChip® Probe Arrays

Starting RNA samples → Total RNA

1. Primer hybridization
   3' TTTTT - 5' 15 minutes

2. Reverse transcription
   First strand cDNA synthesis
   3' AAAA 5' 1 hours 15 minutes

3. Second strand cDNA synthesis
   3' AAAA - 5' 2 hours 30 minutes

4. Cleanup of double-strand cDNA
   30 minutes

5. Amplification and biotin labeling of antisense cRNA
   Biotinylated Ribonucleotides
   U C
   4 hours

6. Cleanup of biotinylated cRNA
   30 minutes

7. Fragmentation
   45 minutes

8. Hybridization
   Streptavidin-phycoerythrin
   Biotinylated anti-streptavidin antibody
   16 hours

9. Washing/Staining
   75 minutes

10. Scanning
    < 10 minutes

Legend:  RNA  DNA  T7 Primer  • Biotin
**Chip designs**

1. **PM/MM method**
2. **PM only**

![Diagram of PM and MM probes](image)

**PM/MM method**

1. The expression of each gene is monitored by a probe set comprising about twenty nominal 25-mers.
2. These are designed across the gene sequences to be specific for that gene.
3. A second set of 25-mers is designed adjacent to the first, having a single mismatch at the central position of the 25-mers.
4. This is supported to correct for non-specific hybridization.

**PM method**

1. Only the PM set of 25-mers is designed and synthesized.
2. This frees up 50% more space on the GeneChips.

**Technical Reproducibility.**

- Examine intensity values obtained from two biological replications hybridized to two separate microarrays or GeneChips, followed by linear regression analysis.
- Total RNA was isolated from 15-day old plants grown on MS agar plates.
- Two types of Affymetrix chips were used, containing ~8,000 and ~24,000 genes (8K, 24K) respectively.
Long oligonucleotide microarrays

Affymetrix 8K chip
Affymetrix 24K chip

Replication 1

Replication 2

A
Technical replication

B
Biological replication

C

D

Signal intensity (average difference)

Signal intensity (average difference)
<table>
<thead>
<tr>
<th>EXPERIMENT</th>
<th>TRANSFORMATION</th>
<th>N</th>
<th>GROUP</th>
<th>$R^2$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oligo arrays</td>
<td>log</td>
<td>23175</td>
<td>All</td>
<td>93.5</td>
</tr>
<tr>
<td></td>
<td>log</td>
<td>11579</td>
<td>Upper half intensity</td>
<td>85.5</td>
</tr>
<tr>
<td></td>
<td>log</td>
<td>11587</td>
<td>Lower half intensity</td>
<td>82.4</td>
</tr>
<tr>
<td>Affy24</td>
<td>log</td>
<td>22810</td>
<td>All</td>
<td>87.0</td>
</tr>
<tr>
<td></td>
<td>log</td>
<td>11405</td>
<td>Upper half intensity</td>
<td>94.4</td>
</tr>
<tr>
<td></td>
<td>log</td>
<td>11405</td>
<td>Lower half intensity</td>
<td>52.1</td>
</tr>
<tr>
<td>Affy8</td>
<td>log</td>
<td>8297</td>
<td>All</td>
<td>84.7</td>
</tr>
<tr>
<td></td>
<td>log</td>
<td>4148</td>
<td>Upper half intensity</td>
<td>92.4</td>
</tr>
<tr>
<td></td>
<td>log</td>
<td>4149</td>
<td>Lower half intensity</td>
<td>39.2</td>
</tr>
</tbody>
</table>
Problems with Oligonucleotide arrays analysis

- A large proportion of data produced from Affymetrix chips is discarded (tagged as being “absent”)
- Long oligonucleotide microarrays have similar reproducibilities to Affymetrix chips at high signal intensities, and better reproducibilities at low signal intensities.
- Various statistical models have been developed to help resolve Affymetrix chip problems. Some require large numbers of replications and tend to be cost-prohibitive.

General comments

- The requirement for replication addresses both biological and technical variation within expression platform measurements. This cannot be avoided.
- Replication directly impacts total costs of projects, and requires the lowest possible unit cost for measurement.
• Spotted long oligomers provide a cost-effective and accurate platform for expression measurements. But they require good quality input information from the various sequencing projects.
• Long oligomers should be insensitive to SNP variation.

NimbleGen
NimbleGen

- Improved technology and chemistry suggests longer oligos can be produced (up to 70-mers advertised by the company). Slides can contain 195,713 probes or with PM/MM controls 9,615 genes.
- Although the NimbleGen approach might not be cost-effective for routine expression platform production, it will enable rapid identification of oligonucleotides suitable for subsequent mechanical printing.
- In particular, it could permit identification of long oligomers, specific for individual genes, that produce optimal signals during hybridization.

Patently Inefficient

A new industry is thrashed by waves of litigation

Flipping through the quarterly report that Affymetrix issued last November, investors may have noticed a section entitled “The company may lose customers unless it improves its ability to manufacture its products and ensure their proper performance.” Indeed, the firm took almost five years to address frequent complaints from researchers that it delivered chips that sometimes gave spurious results and often arrived months after they had been ordered. Fortunately for Affymetrix, until recently it had no real competitors to lose customers to, thanks largely to a formidable portfolio of issued and pending patents that now number more than 400, according to Stephen P. A. Fodor, its chief executive. “We have license agreements with 20 other companies,” Fodor says. But he acknowledges that the licenses restrict those other firms to making arrays that have only about a tenth as many genetic probes as Affymetrix’s gene chips do.

Other microarray producers responded in two ways: with lawsuits and with patents of their own on different microarray designs. Incyte Genomics, for example, uses robots to deposit up to 10,000 presynthesized genetic probes onto a glass slide. Motorola has prototypes of chips that hold the probes inside a thin slab of gel. But companies’ aggressive patenting has led to a bewildering web of lawsuits (above)—and it may only get worse. “If we want to make a medical diagnostic with 40 genes on it, and 20 companies hold patents on those genes, we may have a big problem,” says Nicholas J. Naderio, head of Motorola’s BioChip division. “It isn’t at all clear how this is going to work out.”

—W.I.G.