“It’s not just the genes we have—
It’s how we use them”

Carsten Friis
1. Microarrays measure the expression levels of genes. But how?
The Central Dogma
Microarrays measure mRNA concentrations

Gene

mRNA

Gene specific DNA probes

Labeled target
2. Fine, probes bind mRNA, But what’s this process called?
Hybridization
3. Ok, hybridization. But how many genes can then hybridize to one array slide?
Microarrays are a high-throughput method

Measure the level of transcript from a complete genome in one go
4. Gee, thousands then? Neat, but what is this sample and control stuff?
Experiment setup – Sample preparation

1. Design experiment
2. Perform experiment
3. Precipitate RNA
4. Label RNA/cDNA

- Eukaryote/prokaryote?
- Amplification? Direct or indirect? Label?
5. Ok, so we search for changes in expression; fine, but which technologies are most popular for this?
Microarrays - The Technologies

Stanford Microarrays

Affymetrix
Microarrays – Test Questions

6. Stan & Affy it is; Now, what characterizes the Stanford technology?
The Stanford cDNA Microarrays
Quantitative monitoring of gene expression patterns with a complementary DNA microarray.

Schena M, Shalon D, Davis RW, Brown PO.

Department of Biochemistry, Beckman Center, Stanford University Medical Center, CA 94305, USA.

A high-capacity system was developed to monitor the expression of many genes in parallel. Microarrays prepared by high-speed robotic printing of complementary DNAs on glass were used for quantitative expression measurements of the corresponding genes. Because of the small format and high density of the arrays, hybridization volumes of 2 microliters could be used that enabled detection of rare transcripts in probe mixtures derived from 2 micrograms of total cellular messenger RNA. Differential expression measurements of 45 Arabidopsis genes were made by means of simultaneous, two-color fluorescence hybridization.

PMID: 7569999 [PubMed - indexed for MEDLINE]
Making Microarrays

1. Produce probes
   - oligos
   - cDNA library
   - PCR products

2. Print (spotting) by the use of a robot
Spotting – Mechanical deposition of probes
16-pin microarray spotter
mRNA -> cDNA

SAMPLE

DESIGN and ORDER PROBES

CONTROL

Cy3-cDNA

Cy5-cDNA

Stanford microarrays

mRNA

cDNA
7. So, I guess that was Stan. What then characterizes the Affy technology?
Affymetrix GeneChip® oligonucleotide array

Pre-fabricated arrays
- On-chip synthesis of 25’mers
- 11 to 20 oligonucleotide probes for each gene
- >50,000 probe sets pr. chip

Automation of routine procedures
- better reproducibility
- lighter workload
- faster scans
Examples of Catalog Arrays

<table>
<thead>
<tr>
<th>Human</th>
<th>E. coli</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mouse</td>
<td>P. aeruginosa</td>
</tr>
<tr>
<td>Rat</td>
<td>Plasmodium/Anopheles</td>
</tr>
<tr>
<td>Arabidopsis</td>
<td>Vitis vinifera (Grape)</td>
</tr>
<tr>
<td>C. elegans</td>
<td>Xenopus laevis</td>
</tr>
<tr>
<td>Canine</td>
<td>Yeast</td>
</tr>
<tr>
<td>Drosophila</td>
<td>Zebrafish</td>
</tr>
</tbody>
</table>

(+ many more...)

NimbleExpress™ Array Program
Spacers bound to surface with photolabile protection groups
Photolithography - Micromirrors

NimbleExpress™ Array Program
Facility setup:
- Stanford Microarrays: < 100,000 USD
- Affymetrix: < 250,000 USD

Cost pr. array
- Stanford Microarrays: 30-50 USD
- Affymetrix: 300-400 USD

+ sample preparation costs

NimbleExpress™ Array Program
- a bit more expensive
Reproducibility of data:
(Pearson’s correlation coefficient)

- Stanford microarrays: 0.80 - 0.95
- Affymetrix: ≈ 0.95
8. And that’s Affy folks; Well, except, what was that about several probes pr. gene? How does that work?
How probe sets bind

Probes bind to different positions on the same gene

Regions not suitable for probes
eg. BLAST hits >75%
& longer than 15bp
9. Ok, then *that* must be the end for Affy, right? Or, what was that again about PM & MM probes?
Affymetrix uses PM & MM probes

- Perfect Match (PM)
- MisMatch (MM)

PM: CGATCAATTGCACACTATGTCATTTTCT
MM: CGATCAATTGCAGTATGTCATTTTCT
Great, and the MM’s don’t work, so Affy have wasted half of the chip. Cool going, dudes.

10. And so we come to the final question, what to do about all that noise (or, why are microarrays such a bother to analyze)?
Sources of variation

Array-specific variation:
- Systematic
  - Similar effect on many measurements
  - Corrections can be estimated from data

Gene-specific variation:
- Stochastic
  - Too random to be explicitly accounted for

- "Noise"
- Statistical testing and/or Error modelling
- Normalization
Facts on your project

We have three data sets for you to choose between
   – Bladder Cancer, HIV, Leukemia

Your report should as a minimum demonstrate that you have understood the basic principles of the microarray technology and data analysis
   – That is, after all, the core of the course

You should preferably also demonstrate some understanding of the biological problems behind the data set you choose
   – Because data are more than just numbers

To get the very highest grades you must demonstrate ability to formulate your data analysis in biological terms
   – i.e. don’t just talk statistics – what does the numbers mean to the cell?
Identify differences between different stages/types of bladder cancer based on DNA chips run on a biopsy.

From the biopsy RNA is extracted and run on a GeneChip. The biopsy is also given to histopathologist, who use a microscope to evaluate and stage the suspicious growth into:

- Superficial Ta
- Intermediate T1
- Invasive T2-T4

The purpose here is to identify differences in gene expression between these stages.

- To learn more about the disease and its progression
- To classify tumors based on a biopsy

(This data has been gathered by Skejby Sygehus and it cannot be used without their permission)
The purpose of this study is to measure the effect of HIV-1 on the transcription of genes in the infected host cell.

The human cell line MT4 was infected \textit{in vitro} with HIV-1 and compared to control cultures grown without HIV-1 infection.

– Thus, we have two classes, sick and healthy

After 7 days of growth of both cultures, cells were harvested and RNA was extracted and run on Affymetrix chips.

– The purpose being to identify genes relevant to the HIV disease

Replicates were performed to assure reproducibility and allow measurement of experimental variation.
Study of Childhood Leukemia

Diagnostic bone marrow samples from leukemia patients
Platform: Affymetrix Focus Array
  – 8793 human genes

Immunophenotype
  – 18 patients with precursor B immunophenotype
  – 17 patients with T immunophenotype

Outcome 5 years from diagnosis
  – 11 patients with relapse
  – 18 patients in complete remission

Paper out in Leukemia:
“Prediction of immunophenotype, treatment response, and relapse in childhood acute lymphoblastic leukemia using DNA microarrays”