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

Jeppe Skytte Spicker
Carsten Friis
Course web page

http://www.cbs.dtu.dk/dtucourse/data.php

Course program
Lecture Slides
Exercises
Project Data Sets
Link to the GenePublisher tool
Course Program – Eight Tuesdays

Tuesday 15/3 (today)
Intro to Microarray

Tuesday 29/3
Normalization and Statistical Analysis

Tuesday 5/4
Clustering, PCA and classification

Tuesday 12/4
Writing on your own

Tuesday 26/4
Writing on your own*

Tuesday 19/4
Q&A session + exercise

Tuesday 3/5
Writing on your own*

Tuesday 10/5
Hand in report before 17:00 to Rasmus Wernersson

*Jeppe and Carsten will be available to answer questions from 13:00-17:00
Introduction to Microarrays
1. Microarrays measure the expression levels of genes. But how?
The Central Dogma

- **Nucleus**: DNA (Information storage)
- **Cytoplasm**: mRNA (Information carrier)
- **Translation**: mRNA → Protein

*Diagram shows the flow from DNA transcription to mRNA, and then to protein synthesis.*
Microarrays measure mRNA concentrations
Microarrays – Test Questions

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 to 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
   - Question?
   - Replicates?
   - Test?

2. Perform experiment
   - Wild type
   - Mutant

3. Precipitate RNA
   - Eukaryote/prokaryote?
   - Cell wall?

4. Label RNA
   - 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
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

1. Touch surface
2. Move pins
3. Repeat

Microarray

4. Deliver drop
5. Move jets
6. Repeat

Microarray
16-pin microarray spotter
mRNA → cDNA → Cy3-cDNA
mRNA → cDNA → Cy5-cDNA

DESIGN and ORDER PROBES

Stanford microarrays

SAMPLE

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

11 to 20 oligonucleotide probes for each gene

On-chip synthesis of 25’mers

~20,000 genes per chip

good quality data – low variance
<table>
<thead>
<tr>
<th>Organism</th>
<th>Organism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>E. coli</td>
</tr>
<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>

NimbleExpress™ Array Program
Fluid station and scanner
Photolithography

_mask #1

_mask #2

Spacers bound to surface with photolabile protection groups
Photolithography - Micromirrors

NimbleExpress™ Array Program
The Technologies - Cost

Facility setup:
Stanford Microarrays < 100,000 USD
Affymetrix < 250,000 USD

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

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?
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: CGATCAATTGCACTATGTCATTTCT
MM: CGATCAATTGCAAGTATGTCATTTTCT
Great, and the MM’s don’t work, so Affy have wasted half of the chip. Cool going, dudes.

- And so we come to the final question, what to do about all that noise (or, why are microarrays such a bother to analyze)?
The DNA Array Analysis Pipeline

Question
Experimental Design

Array design
Probe design

Sample Preparation
Hybridization

Buy Chip/Array

Image analysis

Normalization

Comparable
Gene Expression Data

Expression Index
Calculation

Statistical Analysis
Fit to Model (time series)

Advanced Data Analysis
Clustering
PCA
Classification
Promoter Analysis
Regulatory Network
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?
Study of Bladder Cancer

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 *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”
How do we do the final four days?

There is the Q&A session
  – Where we run through a GenePublisher exercise
  – And you can ask us to elaborate on topics you just didn’t get the first time

And there is the advisory task
  – As best we are able Jeppe and Carsten will be available for questioning
  – But there’s four of you and two of us
# Student Accounts

<table>
<thead>
<tr>
<th>Account</th>
<th>Password:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maja, Cathrine, Søren</td>
<td>msc01</td>
</tr>
<tr>
<td>Sarah, Henrik, Anne, Ingeborg</td>
<td>msc02</td>
</tr>
<tr>
<td>Thomas, Nikolaj, David</td>
<td>msc03</td>
</tr>
<tr>
<td>Simi, Bakhtiar, Henrikas</td>
<td>msc04</td>
</tr>
</tbody>
</table>