An Introduction to DNA microarrays
Rebecca Fry, Ph.D.

What is a DNA Microarray?

- genes or gene fragments attached to a substrate (glass)
- Tens of thousands of spots
- Hybridized slide
- Two dyes
- Image analyzed
The Beginnings of Microarray Technology

Lockhart et al., 1996
Nature Biotechnology
“Expression monitoring by hybridisation to high-density oligonucleotide arrays”

Schena et al., 1995
Science
“Quantitative monitoring of gene expression patterns with a complementary DNA microarray”

8 years later
4162 references
Uses and Applications

- Pathway mapping
- Target identification
- Gene screening
- Disease characterization
- Mechanism of Action Studies
- Molecular Diagnosis of Disease
- Personalized medicine
- Developmental Biology
- Toxicology
- Prediction of Drug Efficacy/Toxicity

A model experiment: Two samples of interest

EGFP expressing cells

![EGFP expression](http://www.oardc.ohio-state.edu/plantranslab/gfp-a.gif)

Q: Which genes are differentially expressed in EGFP cells versus EGFP KD cells?

EGFP KD
Inside every cell: DNA Serves as a genetic blueprint

Nucleic acids must be translated into amino acids that make up proteins

Process: “TRANSLATION”

Process of transcribing deoxyribonucleic acid to ribonucleic acid is “TRANSCRIPTION”

Relating Gene Expression

- DNA
  - High throughput protein assays complicated
  - We measure transcript level
  - Is a gene expressed?
  - Is protein produced
  - Ideally measure protein levels

- RNA

- Protein
Introduction

Two Popular Microarraying Platforms

Spotted microarrays

- cDNA: PCR products (500-2,000bp)
- synthesized oligos
- >10,000 probes

Commercially available microarray

- Affymetrix “Gene Chip”
- 500,000 probes
- 25 mer (represents a fragment of a gene)

Designing Oligos

- 70 mer oligo specific to gene of interest

1 846

ATTCTGGAGTTGACGGTGACCCGCTGGGCGGGATCCACCGGCTGCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACC GGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACCCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTGGTACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTCAAGAAGCTTAGCCATGGCTTCCCGCCGGCGGTGGCGGCGCAGGATGATGGCACGCTGCCCATGTCTTGTGCCCAGGAGAGCGGGATGGACCGTCACCCTGCAGCCTGTGCTTCTGCTAGGATCAATGTGTAGGCGGCCGCGACTCTAGATCAAATATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGA
Introduction
Overview of fabrication of spotted microarrays

Robotics: Liquid Handling
Resuspension of oligos

16 FEBRUARY 2001 VOL 291 SCIENCE www.sciencemag.org

Human DNA Repair Genes
Richard D. Wood,1,6 Michael Mitchell,2 John Sgouros,2 Tomas Lindahl3

Base excision repair (BER)
Nucleotide excision repair (NER)
Direct reversal of slippage
Homologous recombination
Editing and processing nucleases
Mismatch repair (MMR)
Nonhomologous end-joining
DNA polymerases (catalytic subunits)

Growth arrest and DNA-damage-inducible

Gadd45α
Gadd45β
Gadd45γ

Excision repair

Hap1
Rad23A
Rad23B
Rad50
Rad51
Rad51C
Rad51L1
Rad51L3
Rad52
Rad54

alkB homolog
Breast cancer 1

ABH
ABH2
ABH3
ADPERT
ADPERTL2
ADPERTL3
APEX
APEXLI2
ATM
ATR
BID
BLM
BRCA1
BRCA2

CASP2
CASP3
CASP8
CASP9
CCNH
CDK2
CDK4
CDK6
CDK7
CDK8
CETN2
DDB1

G23P1
DMC1
DUT
EGFP
ENDO
ERCC1
ERCC2
ERCC3
ERCC4
ERCC5
EXO1
FANCA
FANCC
FANC
FANC
FANC
FANC
FANC
FANC
FANC
MAD2L2

PRKDC
PRSS25
RAD1
RAD17
RAD18
RAD23A
RAD23B
RAD50
RAD51
RAD51C
RAD51L1
RAD51L3
RAD52
RAD54

Breast cancer 1

Growth arrest and DNA-damage-inducible

Gadd45α
Gadd45β
Gadd45γ

Excision repair

Hap1
Rad23A
Rad23B
Rad50
Rad51
Rad51C
Rad51L1
Rad51L3
Rad52
Rad54

alkB homolog
Breast cancer 1
Introduction

A closer look at Spotted microarrays
Some nomenclature

each spot represents a gene or gene fragment

gene

“probe”

RNA

“target”

Introduction

cy3 and cy5: Commonly used dyes

Differential dye incorporation
cy5 less well than cy3
Light sensitivity: cy5 more easily degraded
**Introduction**

Spotted microarray target preparation
Direct labeling

**Target preparation**

- EGFP
- EGFP KD
- cy3
- Reverse transcription Flourescent dyes
- cDNA
- Combined in equal amounts
- Co-hybridized to array

**Spotted Microarray**

- RNA
- cy5
- red cy5>cy3
- yellow cy3=cy5
- green cy3>cy5

**Direct Labeling (Spotted Arrays)**

**RNA**

- Reverse transcription RNA
- Enzyme (Superscript RT)
- Dye
- Oligo d(T) nucleotides

**DNA**

- EGFP
- EGFP KD

**RT**
RNA quality control

Pre-labeling quality control:
Determine RNA Quality
Agilent Bioanalyzer: 50-500 ng
No more formaldehyde gels!!

Gel Image (in silico)
Sharp, Clear Bands

Electropherogram (28S/18S Ratio~2)

Microarray Measurements

Image Analysis: Spotted arrays
### Microarray Measurements

**Signal: Spotted arrays**

Spotted microarrays

![Image of spotted microarrays](www.molgen.mpg.de)

- Signal is average of pixel intensities of spot
- 2 numbers per spot

### Image Analysis: Spotted arrays

**What information do we see?**

<table>
<thead>
<tr>
<th>Name</th>
<th>Clone ID</th>
<th>Spot Mean Intensity (w595)</th>
<th>Spot Mean Intensity (w685)</th>
<th>Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>control1</td>
<td>control1</td>
<td>6642.5</td>
<td>2371.163</td>
<td>Gene name</td>
</tr>
<tr>
<td>control1</td>
<td>control1</td>
<td>5786.72</td>
<td>2068.51</td>
<td>Oligo id</td>
</tr>
<tr>
<td>control1</td>
<td>control1</td>
<td>6434.93</td>
<td>2370.575</td>
<td>Cy3 intensity</td>
</tr>
<tr>
<td>control1</td>
<td>control1</td>
<td>6242.5</td>
<td>2351.697</td>
<td>Cy5 intensity</td>
</tr>
<tr>
<td>PRKDC</td>
<td>H200014101</td>
<td>348.522</td>
<td>122</td>
<td></td>
</tr>
<tr>
<td>PRKDC</td>
<td>H200014101</td>
<td>340.933</td>
<td>120.5333</td>
<td></td>
</tr>
<tr>
<td>PRKDC</td>
<td>H200014101</td>
<td>306.8235</td>
<td>117.8676</td>
<td></td>
</tr>
<tr>
<td>PRKDC</td>
<td>H200014101</td>
<td>296.5224</td>
<td>115.6119</td>
<td></td>
</tr>
<tr>
<td>ADPRT</td>
<td>H200008289</td>
<td>324.886</td>
<td>131.2286</td>
<td></td>
</tr>
<tr>
<td>ADPRT</td>
<td>H200008289</td>
<td>265.7391</td>
<td>108.5797</td>
<td></td>
</tr>
<tr>
<td>ADPRT</td>
<td>H200008289</td>
<td>259.5362</td>
<td>108.9855</td>
<td></td>
</tr>
<tr>
<td>ADPRT</td>
<td>H200008289</td>
<td>272.5676</td>
<td>117.8919</td>
<td></td>
</tr>
</tbody>
</table>

**Steps:**

- Normalize globally
- Calculate average of cy3 and cy5
- Bring cy5 numbers to cy3 by multiplying by common factor
- Ratios (take ratio of wt/KD...look for 2 fold cutoff)
- Log2 (calculate log 2 of ratio to differentiate increase or decrease)
- Reproducibility (how did the four replicates perform?)
Plots

Data Analysis
Requires software: Spotfire

Requires ability to search for Patterns and Trends
The DNA damage response: putting checkpoints in perspective
Bin-Bing S. Zhou* & Stephen J. Elledge²
NATURE | VOL 408 | 23 NOVEMBER 2000 |