Introduction to DNA microarray technology

Johan Vallon-Christersson
Dept. Oncology, Lund University
What is DNA microarray and how to use it?
Genomic tool

• Gene Expression

• Array of Probes
Gene Expression

Cell

DNA

Protein

mRNA
Gene
↓
Transcription
RNA processing
↓
mRNA
nuclear transport
↓
degradation
5´capping
endonuclease cleavage
polyadenylation
splicing
Translation
Modification
glycosylation
lipid-anchoring
metal binding
phosphorylation
...
Sorting
Degradation
(enzymatic) activity
A “probe” can be any nucleic acid

- cDNA clones of genes or expressed sequence tags (ESTs)
- Synthesized DNA oligomers of 15 to 80 bases
  - In situ
  - Longmers
- Genomic DNA sequences
Detection of mRNA – array lingo

Target

Target represents your sample

The probe is immobilized and is used to capture your target.

Probe
Detection of mRNA

Sample
RNA → RT → cDNA → labeled target → Cy detector

Cy3 nucleotides → cDNA probe

high → low
DNA microarray slide

Gene A
Gene B
Gene C
Human genome

3,180,000,000 bp

20,000 genes?
80,000 genes?
30,000 genes?
Oligo Sets™ (Qiagen)

Array-Ready Oligo Sets™ (Qiagen)

- Human 34k
- Mouse 32k
- Rat 16k
Slide vs. Blot – principal differences

Northern Blot
- Immobilize target on membrane
- Probe with labeled gene
- Detect: 1 gene at a time

Microarray
- Immobilize probes on glass slide
- Apply labeled targets
- Detect: 1000’s of genes in one experiment
• What is a DNA Microarray slide?
Getting the probe on the slide

- In situ synthesized probes
- Spotted probes
In situ synthesized probes

- Photo regulated (Affy)
- Spatial separation (Rosetta)
Spotted probes

- cDNA
- Oligo Longmers
- BAC (genomic DNA)

RNA → cDNA

Cloned into bacteria

Purify / Amplify insert

cDNA Clone Library 45K clones
(cDNA clones, sequenced and identified)
Clone preparation schematic

**cDNA platform**

- Library of Expressed Sequences (cDNA clones, sequenced and identified)
- Grow bacteria
- Isolate plasmid or use lysate directly
- PCR to amplify insert
- Reaction clean up
- Reformatting
- Dissolve in buffer
- QC
- Print plate setup

**PRINT READY**

**Oligonucleotide platform**

- Library of Oligonucleotide probes (designed using bioinformatics)
- Dissolve in buffer
- Print plate setup

**PRINT READY**
Printing DNA / oligo microarrays
Printing DNA / oligo microarrays

Dia. \(~50\text{um}\)

Vol. \(~60\text{pl}\)
Printing DNA / oligo microarrays
Array Robot
Printing set-up

Arrayer:
- BioRobotics MG2
- Contact printer
- Split-pin system
- BioRobotics MicroSpot2500
  - 100um
  - 50um

Print conditions:
- 384 well source plates
- 50% DMSO
- 22degC and 45% rel Hum

Spots:
- ~145um

Pin wash:
- Between samples
- Recirculating baths
- Vacuum wash
- Between runs:
- Sonication
10,000 spots, dia. ~130um
Array design

Landing lights

Replicates spatially

Controls, e.g. negative

~ 10 – 30 µm between spot boundaries, 130 µm between centers allowing up to 55000 features

Replicates spatially separated
• How do you produce the slide?
Label your target

Total RNA, mRNA, amplified RNA
- TRIZOL (Gibco-BRL), RNeasy (Qiagen)
- aRNA T7-based in vitro transcription (Ambion)

Fluorescent nucleotides
- Cy3/Cy5-dUTP and –aminoallyl (Amersham)

RT Enzymes
- SuperScript II (LifeTech) vs. CyScribe (Amersham)

Labeling strategies
- Direct vs. indirect incorporation
Two-colour microarray

Sample
- RNA
  - RT
  - cDNA
  - Cy3 nucleotides

Reference
- RNA
  - RT
  - cDNA
  - Cy5 nucleotides

Cy detector
- high
- low

cDNA probe
Reference

- Tumor 1
- Tumor 2
- Tumor 3
- Tumor 4
- Tumor 5
- Tumor 6
- Tumor 7
- Tumor 8

- Normalization
- Compare results
Hybridization

Tumor sample

RNA → cDNA

Reference sample

RNA → cDNA

- incubation
- wash

Hybridize
Scan Slide

Scan Slide

Laser scan

Cy5 Image

Cy3 Image

Pseudocolored Merged Image

Image & Data Analysis
• What is a DNA Microarray slide?
• How do you produce the slide?
• How do you use the slide?
Introduction to DNA microarray technology

Johan Vallon-Christersson
Dept. Oncology, Lund University
• Materials - Methods
• Analysis – Examples
Using DNA Microarrays

Sample

Extract

Labeled extract

Hybridization

Wash

Scan

Image 1

Image 2

Ratio image

Sample1 / Sample2

Sample1 / Sample2
CloneID: 244628
Green (Cy3) Intensity: 4537
Red (Cy5) Intensity: 26520
Ratio (Cy3/Cy5): 0.17

CloneID: 47291
Green (Cy3) Intensity: 26520
Red (Cy5) Intensity: 4537
Ratio (Cy3/Cy5): 5.85

CloneID: 35835
Green (Cy3) Intensity: 36331
Red (Cy5) Intensity: 37234
Ratio (Cy3/Cy5): 0.98
## Array Raw Data

### Quantified Data Values

<table>
<thead>
<tr>
<th>Spots/Probes</th>
<th>Int 1</th>
<th>Int 2</th>
<th>Int 1 BG</th>
<th>Int 2 BG</th>
<th>Int1 SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1342</td>
<td>214</td>
<td>34</td>
<td>67</td>
<td>345</td>
</tr>
<tr>
<td>B</td>
<td>452</td>
<td>7313</td>
<td>32</td>
<td>68</td>
<td>42</td>
</tr>
<tr>
<td>C</td>
<td>51728</td>
<td>52168</td>
<td>34</td>
<td>59</td>
<td>536</td>
</tr>
<tr>
<td>D</td>
<td>321</td>
<td>213</td>
<td>36</td>
<td>87</td>
<td>53</td>
</tr>
<tr>
<td>E</td>
<td>6725</td>
<td>324</td>
<td>41</td>
<td>56</td>
<td>312</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

One hybridization < 55000 Spots/Probes

- Massive amounts of data
- Data matrices 55000x48 > 2,5 million Data Points
• What is image analysis?
M vs. A - plot

Log2Ratio vs. Log Int.

High

Low
Normalization

M (log2ratio)

High

Low

A (log int)
Normalization

M (log2ratio)

A (log int)

High

Low
QC filter

M (log2ratio)

High

Low

A (log int)
### Gene Expression Data

<table>
<thead>
<tr>
<th>Genes</th>
<th>Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>sample1</td>
</tr>
<tr>
<td>A</td>
<td>0.46</td>
</tr>
<tr>
<td>B</td>
<td>-0.10</td>
</tr>
<tr>
<td>C</td>
<td>0.15</td>
</tr>
<tr>
<td>D</td>
<td>-0.45</td>
</tr>
<tr>
<td>E</td>
<td>-0.06</td>
</tr>
</tbody>
</table>

Relative Gene Expression (Sample / Reference)
• What is normalization?
Copy Number Profile of a Breast Tumor

45 kbp array-CGH

Lund 32k BAC
2 Arrays
CGH – BAC library

Library of Genomic Sequences (end-sequenced and identified)

BAC library
32 K clones

Genomic sequence
~100 kb

Cloned into bacteria

BAC clones
Copy Number Profile
Copy Number Profile in MCF7
Copy Number Profile in HCC-1937

Chromosome 10

Homozygous deletion of the PTEN tumor suppressor gene

BAC clone RP11-79A15 including PTEN

10q23.31
CGH vs mRNA plot
• What is CGH-array analysis?
Microarray Experimentation

- Biological Questions
- Biological Samples
- Sample Extracts
- Labeled Extract
- Hybridization
- Image Acquisition
- Image Analysis
- Data Management
- Data Analysis
- Biological Answers

Clone Set
Preparation of Probes
Array Printing
Arrays

Results only make sense when viewed in a context......

Preprocessing
Normalization
Additional Analysis
Analysis

Integrate:

Backend Production
Analysis

Integrate:

Backend Production + Experimental Data
Analysis

Integrate:

- Backend Production
- Experimental Data
- Biological / Sample Data
Data Storage and Access

Web Interface

DNA Resource Center

Bioinformatics Facility

BASE Server

Users
• Analysis methods
## Gene Expression Data

### Samples

| Genes | tumor1 | tumor2 | tumor3 | tumor4 | tumor5 | ...
|-------|--------|--------|--------|--------|--------|------
| A     |  0.46  |  0.30  |  0.80  |  2.51  |  0.90  | ...
| B     | -0.10  |  0.49  |  0.24  |  0.06  |  0.46  | ...
| C     |  0.15  |  0.74  |  0.04  |  0.10  |  0.20  | ...
| D     | -0.45  | -1.03  | -0.79  | -0.56  | -0.32  | ...
| E     | -0.06  |  1.06  |  1.35  |  1.09  | -1.09  | ...
| ...   | ...    | ...    | ...    | ...    | ...    | ...  

### Relative Gene Expression (Sample / Reference)
Methodology - Data Analysis

“Traditional” techniques:

- Multi-Dimensional Scaling (MDS)
  Calculate Pearson correlation coefficients for all pairwise comparisons of tumor log(ratios)
  
  \[(1 - \text{coefficient}) = \text{distance}\]

  Plot tumors in 2-3 dimensions, forming a best-fit of the tumors for all distance measurements

Gruvberger et al.
Multi Dimensional Scaling
Methodology - Data Analysis

“Traditional” techniques:

• Multi-Dimensional Scaling (MDS)

• Weighted Gene Analysis (WGA)
  Iterative process to discover genes that maximize MDS inter-cluster distance & minimize intra-cluster distance
Inter-cluster distance

Gruvberger et al.
Intra-cluster distance

Gruvberger et al.
Methodology - Data Analysis

“Traditional” techniques:

- Multi-Dimensional Scaling (MDS)
- Weighted Gene Analysis (WGA)

Iterative process to discover genes that maximize MDS inter-cluster distance & minimize intra-cluster distance

More influential genes have higher rank
Methodology - Data Analysis

“Traditional” techniques:

- Multi-Dimensional Scaling (MDS)
- Weighted Gene Analysis (WGA)
- Hierarchical Dendrograms

Use Pearson correlation coefficients for tumors and genes to group the two most correlated tumors and genes, and so forth to extend the branching
Arrays

Gene clusters

Gruvberger et al.
• Materials - Methods
• Analysis – Examples
Site visit: DNA microarray resource center

BMC B10 – kl 13.15..

Clone preparation
Array printing
Hybridization
Image analysis
Introduction to DNA microarray technology

Johan Vallon-Christersson
Dept. Oncology, Lund University
Molecular genetic characterization of inherited breast cancer
Hereditary breast cancer

- Breast cancer
  - >6000 new BRCA cases in Sweden yearly
  - 1 in 10 women affected

Hereditary breast cancer
- high penetrance genes
- dominant segregation

5-10%

'Sporadic BRCA'

- low penetrance genes
- recessive genes
- environmental effect
- chance clustering

< 10%

 major susceptibility genes

- BRCA1 30%
- BRCA2 20%
- BRCAx ....? ~ 50%

rare multi-cancer syndrome genes

- p53 (Li Fraumeni)
- PTEN (Cowden)
- CDKN2A (melanoma)
- CHK2 (Li Fraumeni-like)
Inherited breast cancer

BRCA1

Breast-ovarian cancer family
Inherited breast cancer

BRCA2
Breast-ovarian cancer family including a male breast cancer
The Hunt for Breast Cancer Genes

Linkage to 17q21 = BRCA1
Cloning of BRCA1

Linkage to 13q12 = BRCA2
Cloning of BRCA2

1990 1994 1995

BRCA3 - BRCAx
Multidimensional Scaling (MDS) analysis

Hedenfalk et al., NEJM 2001
Hierarchical Clustering Dendrograms – Weighted Gene Analysis

Hedenfalk et al., NEJM 2001
Hierarchical Clustering Dendrograms – Weighted Gene Analysis

Hedenfalk et al., NEJM 2001
Cancer genes

**Oncogenes**
– activated proto-oncogenes – accelerate cell division

Mechanisms of activation:
– point mutation
– chromosomal translocation $\rightarrow$ increased expression
– gene amplification $\rightarrow$ increased expression

**Tumor suppressor genes**
– the cell’s brake on cell division and tumor growth

Mechanisms of inactivation:
– mutation $\rightarrow$ decreased expression
– deletion $\rightarrow$ decreased expression
– DNA tumor virus proteins
– promoter methylation $\rightarrow$ decreased expression
New approaches to find BRCAx
Introduction to DNA microarray technology

Johan Vallon-Christersson
Dept. Oncology, Lund University