

MICROARRAY TECHNOLOGIES

Vivian Cheung, MD
Department of Pediatrics
vcheung@mail.med.upenn.edu
(215) 590-4950

GENES & PHENOTYPES

Phenotype: Characteristics of an individual as determined by his/her genotype and the environment.

Traditionally: short stature, dementia, elevated sweat chloride

Post genome era: expression level of genes, proteins

Tools for Global Analysis of the Genome

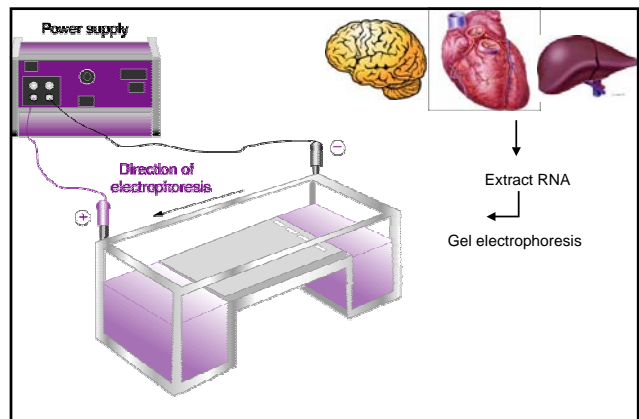
- Analyze many genes simultaneously
- Identify all the genes that are involved in a pathway
- Gene interactions
- DNA, non-coding regions of the genome

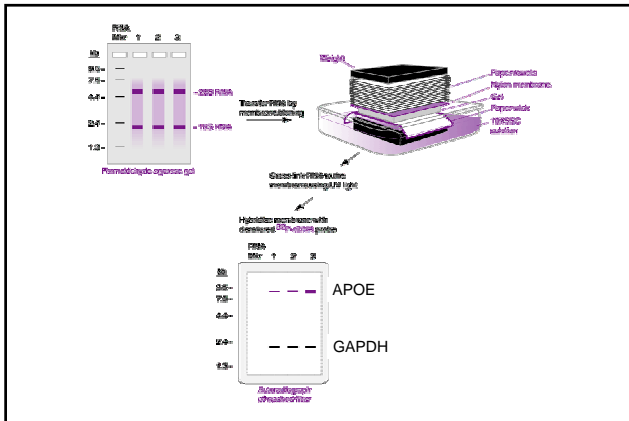
OUTLINE - MICROARRAY

1. Introduction
2. Designing a microarray experiment
3. Microarray data
4. Application
 - gene expression profiling

NORTHERN BLOTS

- Determine the expression level of one or more genes
- RNA samples are run on gels by electrophoresis and transferred onto a membrane.
- Gene(s) of interest are labeled and hybridized onto the membrane.
- Level of gene expression is proportional to the intensity of hybridization.





NORTHERN BLOT

Identified a new gene – what tissue is it expressed in?

32p probe

1. Heart
2. Liver
3. Liver tumor
4. Breast tumor
5. NI. Breast
6. Kidney tumor
7. Pancreatic tumor

Your new gene is expressed at relatively higher level in various tumor samples than “normal” tissues.

INCREASED THROUGHPUT

- Need to analyze more than one gene at a time
- Northern blot analysis is quite laborious and require a large amount of RNA
- Microarray = Reverse-Northern

NORTHERN BLOT to MICROARRAY

32p gene

1. Heart
2. Liver
3. Liver tumor
4. Breast tumor
5. NI. Breast
6. Kidney tumor
7. Pancreatic tumor

**GOAL: Study the expression levels of 1000 genes in these tissues. You need roughly:
1000 ug RNA per sample (10 ug for ~ 10 genes)
1000 radiolabeled genes**

MICROARRAY

1. Heart
2. Liver
3. Liver tumor
4. Breast tumor
5. NI. Breast
6. Kidney tumor
7. Pancreatic tumor

- 7 microarrays each containing 30,000 genes
- RNA (<<10 ug / sample) from the 7 tissues of interest

TERMINOLOGY

- Probes – DNA immobilized on the solid surface
- Targets – DNA or RNA mixture that is being analyzed

Targets

probes

Solid surface

DNA MICROARRAYS

What is a microarray?

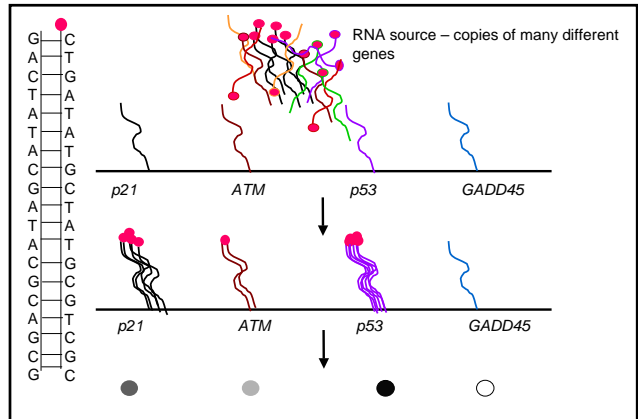
- hundreds of thousands of genes, genomic regions (cDNA clones, BACs or oligonucleotides) on a solid support (usually glass)

2 main types of microarrays:

- Spotted microarrays
- Photolithographic microarrays (commercial)

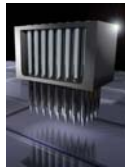
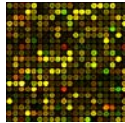
Technical Foundation:

- mRNA/ DNA molecules bind (hybridize) specifically to molecules with complementary sequences



SPOTTED MICROARRAYS – developed by P. O. Brown

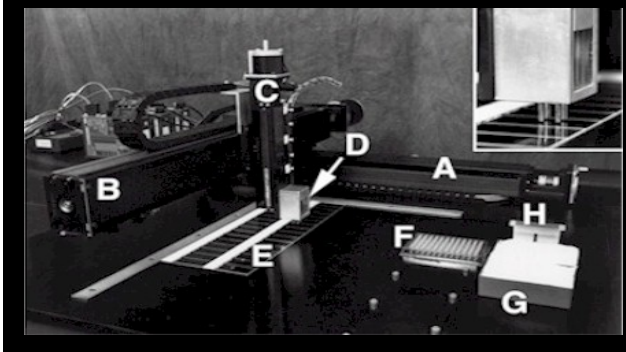
1. Synthesize DNA – PCR of cDNA clones or chemically synthesize oligonucleotides
2. Robotically spot the synthesized DNA onto solid support – glass microscope slides
3. Denature DNA on the “microarray” so that there will be single stranded DNA for hybridization
4. Label mRNA from cells of interest, label them with green fluorescent dye
5. Label reference mRNA with red fluorescent dye
6. Cohybridize the red and green-labeled mRNA onto the microarray.



PROBES

- Material spotted on the solid surface
- PCR products – from cDNA clones, BAC clones
 - gene of interest
 - cDNA clone
 - vector
 - 96 – or 384 well formats
- Oligonucleotides – 40 to 60 “mers” representing unique sequences

FABRICATION OF SPOTTED ARRAYS



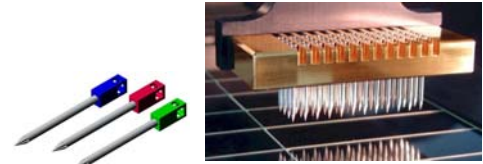
SPOTTING TECHNIQUES

A robot picks up samples from 96- or 384 well plates and deposits the aliquots sequentially onto solid support (microscope slides).

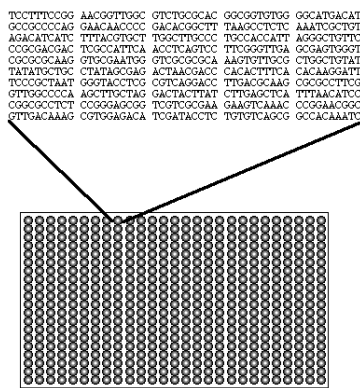
Robots usually have "print-heads" that contain 8 – 24 pins that pick up samples simultaneously. Depending on the speed of the robots, it usually takes about 2 days to make 100 arrays each containing about 5000 clones.

Different types of print tips:

- Quill
- Pin-ring
- Ink-jet

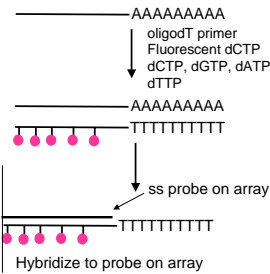


Inkjet technology
- Agilent (spin-off
Of Hewlett-Packard)

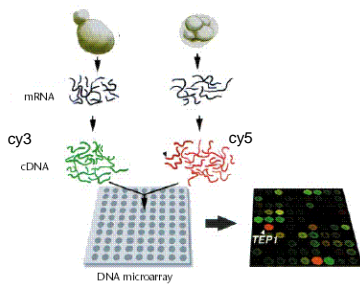
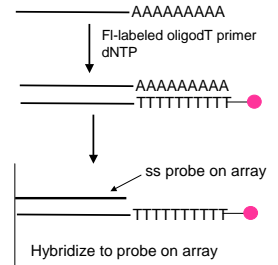


Labeling and Hybridization of cDNA / DNA

Direct Labeling



Indirect Labeling



Brown & Botstein, 1999

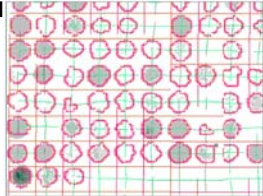
SCANNING HYBRIDIZED ARRAYS



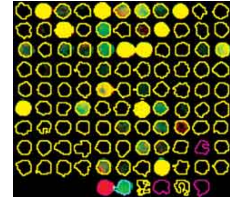
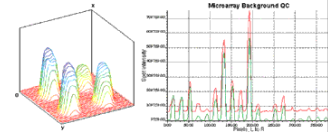
IMAGE PROCESSING

Hybridized arrays are scanned using fluorescent scanners. The fluorescence must then be converted into a measure of abundance.

1. Image acquisition – 16-bit tiff (range 0 to 65,536 or 2^{16}) – one channel
2. Localize spot
3. Measure intensities
4. Data reporting



1. TIFF image of experimental and control samples
2. Superimpose the two images
3. Position a quantitation grid on the images and measure the intensities in the two images
4. Process image – raw intensity, background subtraction, intensity thresholding and define confidence limit
5. Pseudocolor display of processed image and output experimental and control intensity to file.



SAMPLE DATA

Normal Breast Tissue

Gene	Cy3	Cy5	Ratio	Log ₂ Ratio
A	12500	10000	1.25	0.32
B	5000	5000	1	0
C	950	850	1.1	0.14
D	600	700	0.87	-0.20

Breast Tumor

Gene	Cy3	Cy5	Ratio	Log ₂ Ratio
A	7500	6300	1.19	0.25
B	4329	4300	1.01	0.01
C	1800	900	2.0	1.00
D	10000	700	14.29	3.84

Cy3 = experimental sample ; Cy5 = reference sample

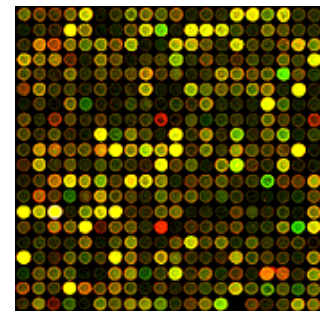
Expression levels of Genes A and B are very similar.
Expression level of Gene C is higher in breast tumor than normal breast tissue.
Expression level of Gene D is also higher in breast tumor than normal breast tissue.

Green = test sample
Red = reference sample

Yellow : test = reference

Green spot: test sample contains more copy of that gene than the reference sample.

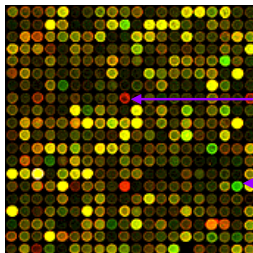
Red spot: test < reference



Each spot represents one gene.

Green = test sample, breast cancer cell

Red = reference sample, normal breast cell



B-actin:
Equal copies in cancer and normal cells

MSH2:
More copies in normal cells than in cancer cells.

CDKN2D:
More copies in cancer cells than normal cells

Screen many genes in parallel – in order to obtain a gene expression profile of a cell/ tissue of interest.

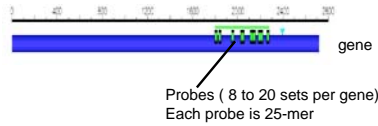
SPOTTED MICROARRAYS

- Need clones or oligonucleotides presenting genes/ regions of interest
- 2 sample hybridization – to normalize the differences in DNA spotted for each gene
- Flexibility in design/ content – organisms, genes, regions of interest

PHOTOLITHOGRAPHIC MICROARRAYS

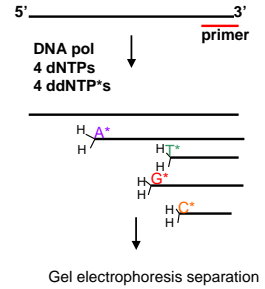


- Commercially available arrays
- Affymetrix
- Multiple oligonucleotides representing a gene



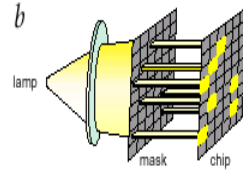
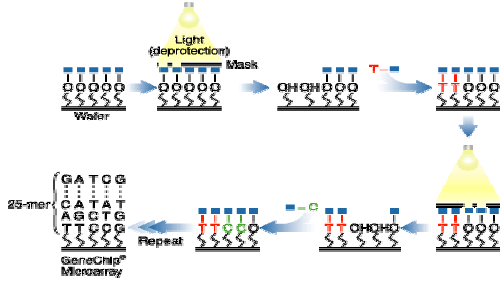
SANGER DNA SEQUENCING METHOD = dideoxy sequencing method

- Primer is used to initiate DNA synthesis.
- Addition of 4 dideoxynucleotides randomly arrests synthesis - lacks 3' OH group.
- Resulting fragments are separated by electrophoresis.



PHOTOLITHOGRAPHIC MICROARRAYS

- Photolithography & DNA Synthesis



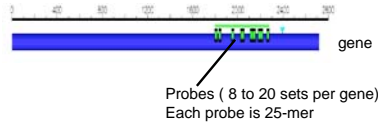
DNA that has length, N can be synthesized in 4 x N cycles.

You need the sequences of the genes of interest.

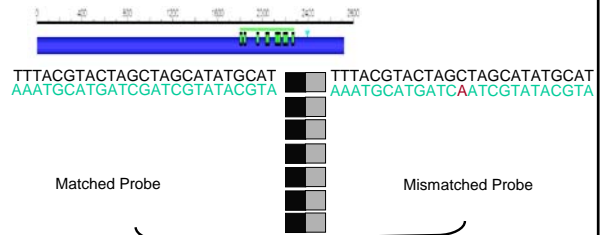
PHOTOLITHOGRAPHIC MICROARRAYS



- Commercially available arrays
- Affymetrix
- Multiple oligonucleotides representing a gene

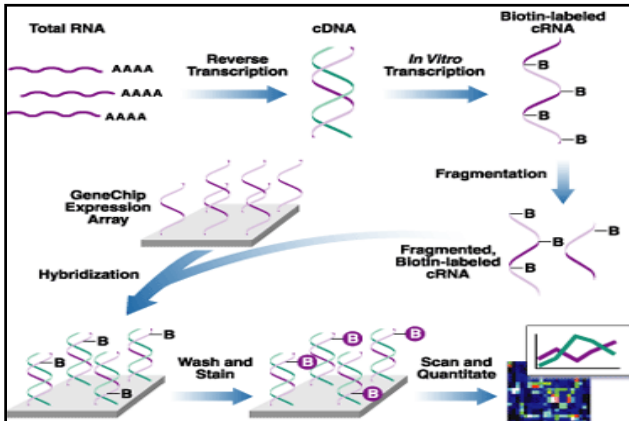


PROBE DESIGN IN PHOTOLITHOGRAPHIC ARRAYS



Intensity of the gene is determined by "averaging" the matched and mismatched probes.

Absent/ Present calls: relative intensities from matched vs. mismatched probes.



~12,000 gene
Only one sample is hybridized to it.
Equal amount of oligos for each gene.

mRNA abundance – reflected by signal / hybridization intensities (not a ratio).
False color scale (less to more) : blue – green – orange – yellow – white

SAMPLE DATA

Normal Breast Tissue			Breast Tumor		
Gene	Intensity	Log ₂ Intensity	Gene	Intensity	Log ₂ Intensity
A	10000	13.29	A	9800	13.26
B	590	9.20	B	600	9.23
C	250	7.97	C	900	9.81
D	900	9.81	D	10000	13.29

Expression levels of Genes A and B are very similar.
Expression level of Gene C is higher in breast tumor than normal breast tissue.
Expression level of Gene D is also higher in breast tumor than normal breast tissue.

	Spotted	Photolithographic
Cost	Less	Chemistry is very expensive
Co-hybridization	Yes – variable amount of DNA on each spot	No – uniform amount of DNA
Reagents	cDNA clones for PCR or sequence	Special equipment Sequence of genes
Sensitivity	Maybe slightly less in spotted arrays, range from 1 molecule in 100,000 to over 1,000,000 depending on hybridization conditions	

APPLICATIONS

Gene expression profiling

- biomarkers of cancer
- target genes of transcription factor, p53

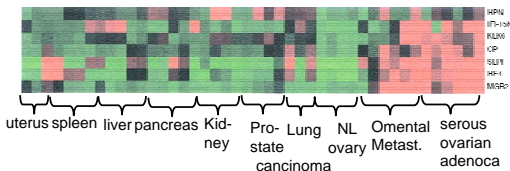
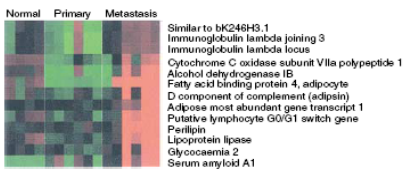
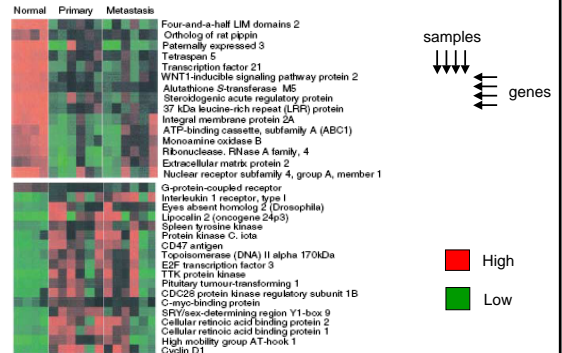
OVARIAN CANCER

- Leading cause of gynecological cancer in the US – about 25,000 new cases per year
- Early stage is usually asymptomatic
- At diagnosis, metastasis is common and 5-year survival is less than 20%
- No reliable screening assay – 90% of patients with advanced disease have elevated serum CA-125 but this marker is not very sensitive or specific.

GENE EXPRESSION PROFILING

- Compare the expression levels of ~12,000 genes among: normal, primary and metastatic diseased tissues obtained during surgical treatment of ovarian cancer
- Abid et al. British J. of Cancer 2004, p686-692.

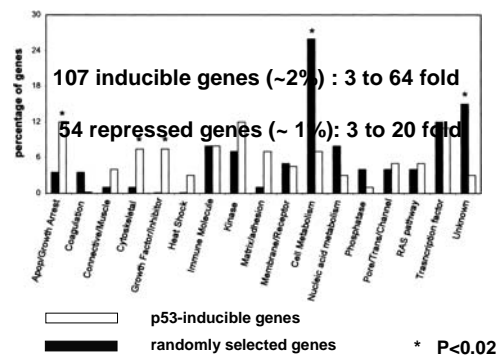
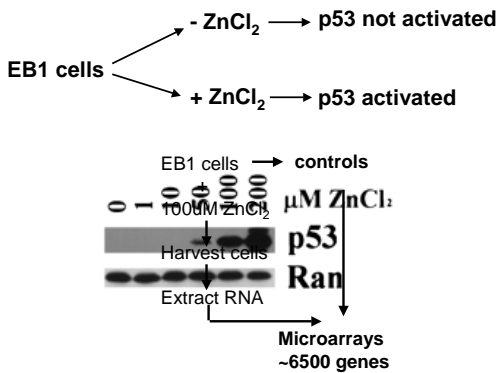
GENE EXPRESSION PROFILING- OVARIAN CANCER

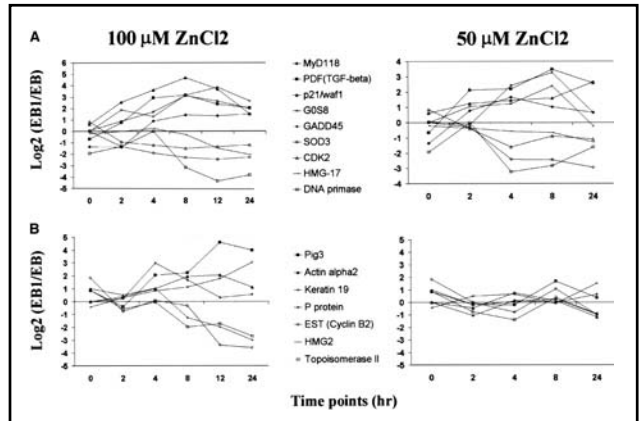
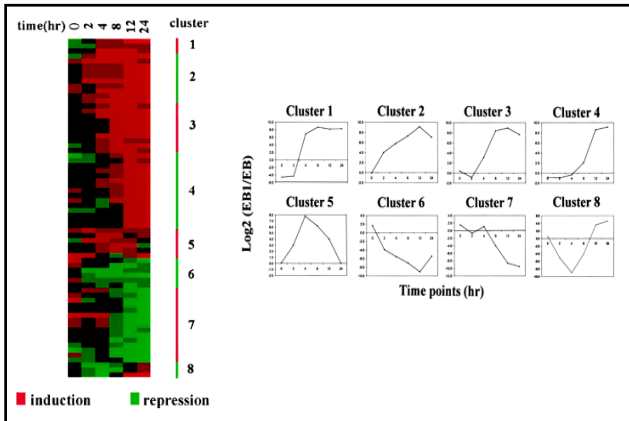


Analysis of p53-regulated gene expression patterns using oligonucleotide arrays

Renbin Zhao,¹ Kurt Gish,² Maureen Murphy,^{2,6} Yixin Yin,^{1,6} Daniel Notterman,^{1,6} William H. Hofman,² Edward Tom,² David H. Mack,³ and Arnold J. Levine^{5,7}
Genes & Dev 14:981 (2000)

- p53 – transcription factor involves in response to cellular stress – cell cycle arrest, apoptosis
- Mechanism: identify target genes
- EB1 – human colon cancer cell line with a stably transfected p53 cDNA driven by metallothionein promoter
- Activation of p53 in EB1 cells leads to apoptosis





Please email me if you have any questions
vcchung@mail.med.upenn.edu