Introduction to Microarray Technology

Presenter: Wayne Xu, PhD
Computational Genomics Consultant
Supercomputing Institute

Email: wxu@msi.umn.edu

Phone: (612) 624-1447

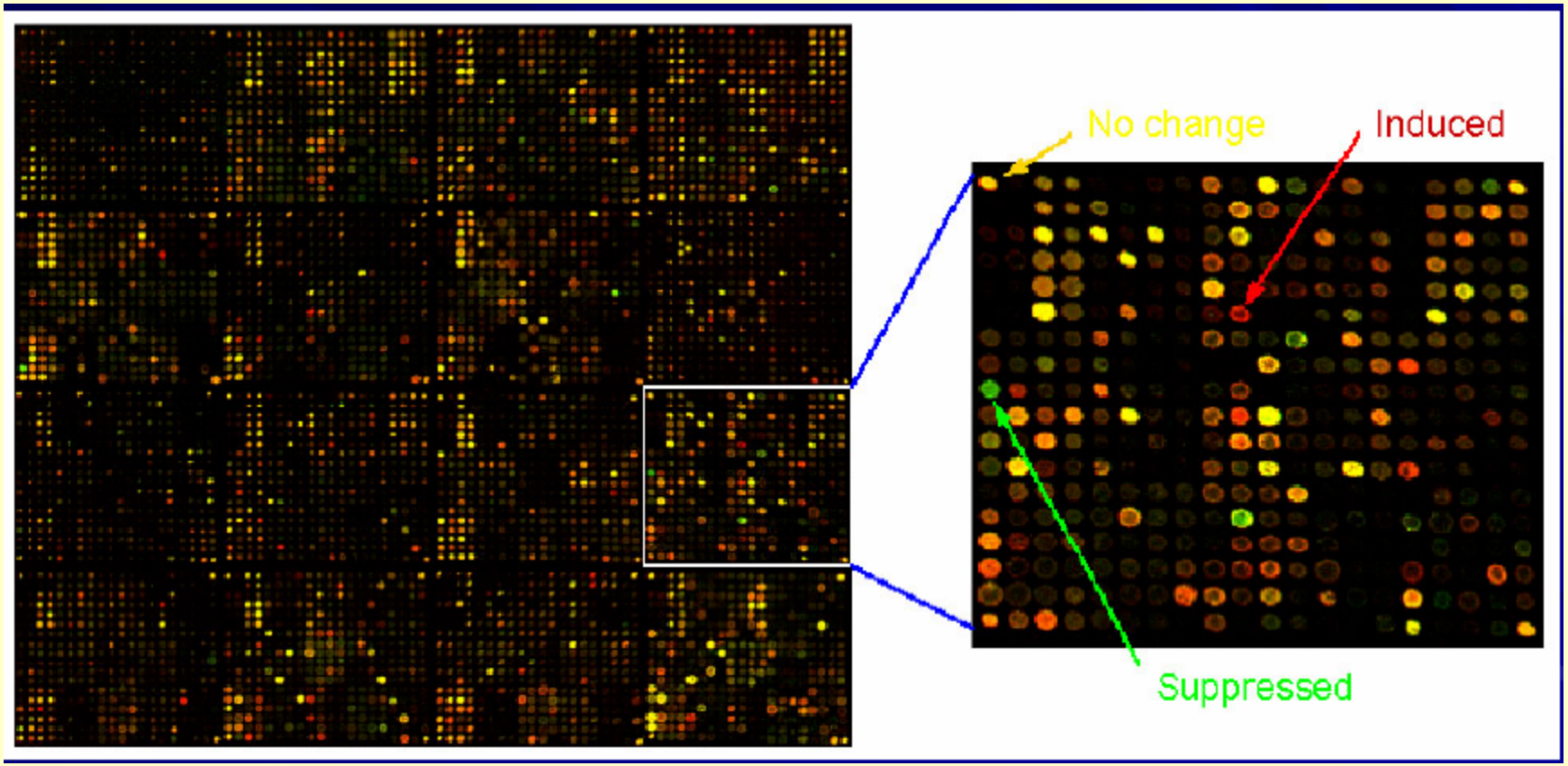
Help: help@msi.umn.edu
(612) 626-0802

Outline

- Introduction
- Array chips
  - cDNA array
  - Affymetrix array
- Microarray experiment and data acquisition
- Data analysis

Introduction

What is microarray?



Microarray

- **A high throughput technology that allows detection of thousands of genes simultaneously**
- Principle: base-pairing hybridization
- Much rely on computer aids
- Central platform for functional genomics

Features

- **Parallelism**
  - **Thousands of genes simultaneously**
- **Miniaturization**
  - **Small chip size**
- **Multiplexing**
  - **Multiple samples at the same time**
- **Automation**
  - **Chip manufacturing**
  - **Reagents**

What circumstances brought out this technology?

- Fact: biological processes are complicated with many molecules working together. Biologists are eager to obtain the “whole picture”
- Genome sequences availability
- Computer aids

What problems can it solve?

- Differing expression of genes over time, between tissues, and disease states
- Identification of complex genetic diseases
- Drug discovery and toxicology studies
- Mutation/polymorphism detection (SNP's)
- Pathogen analysis

What is its pitfall?

- Detect transcription mRNA level, not translation protein level
- Many factors (variations) can affect the result
  - Chip and probe design
  - Experiment design
  - Sample preparation
  - Image acquisition
  - Data normalization
  - Data analysis
  - ….
- Success crucial:
  - You know both the biology problem and the computer aids (software, statistics).

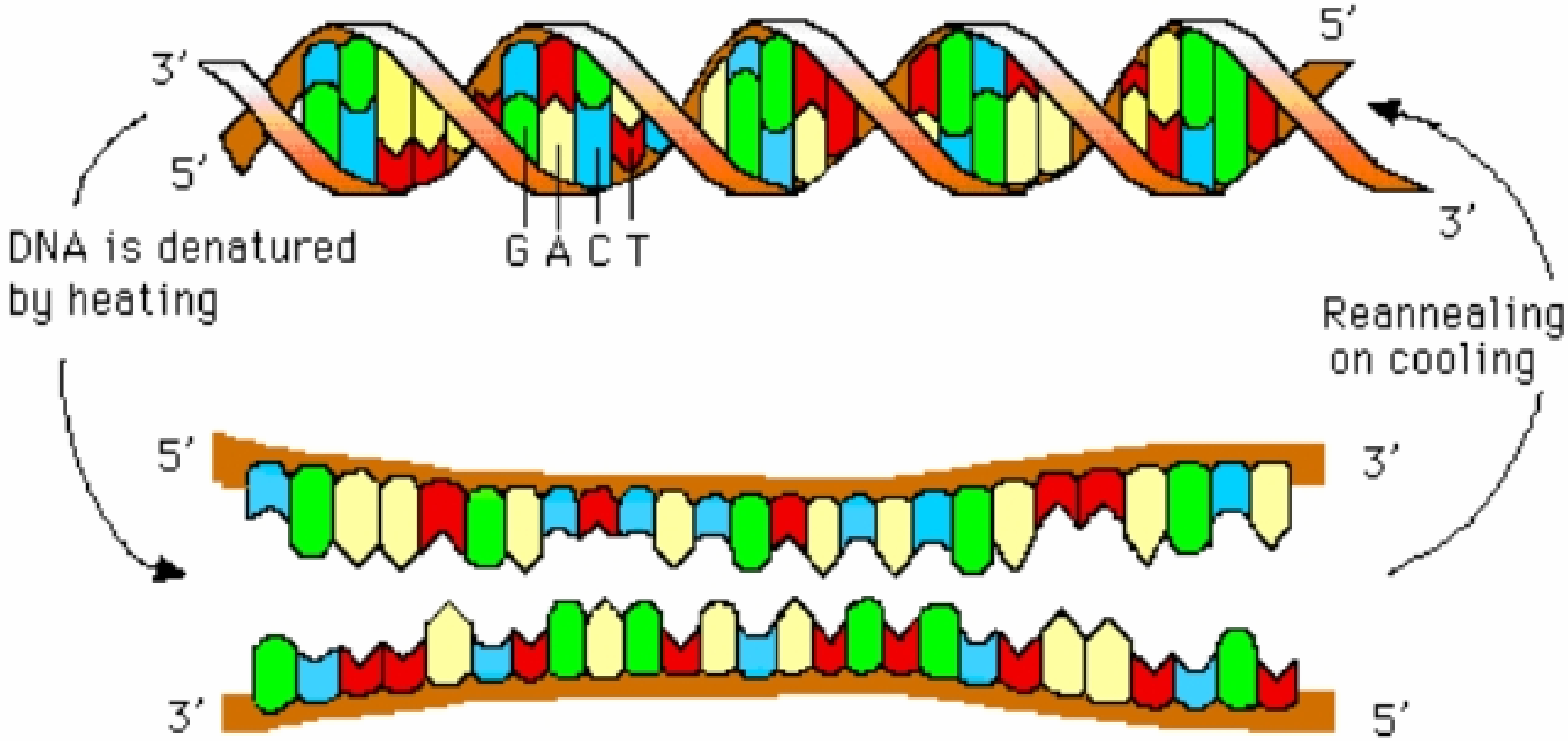
Principle

- **Similar to Northern**
  - **Base-Pairing, hybridization between nucleic acids**
- **Major differences from Northern**
  - **Detects thousands of genes simultaneously /individual**
  - **Probes fixation on glass slide / nylon membrane**
  - **Target samples labeling with fluorescent/radioactive dNTP**

Principle

- **Base-pairing**
  - DNA: A-T and G-C
  - RNA: A-U and G-C

Principle

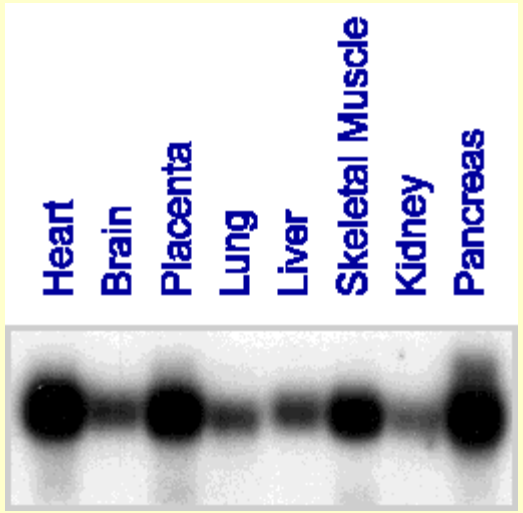

3′
5′
5′
3′
G A C T
DNA is denatured
by heating
Reannealing
on cooling
5′
3′
3′
5′

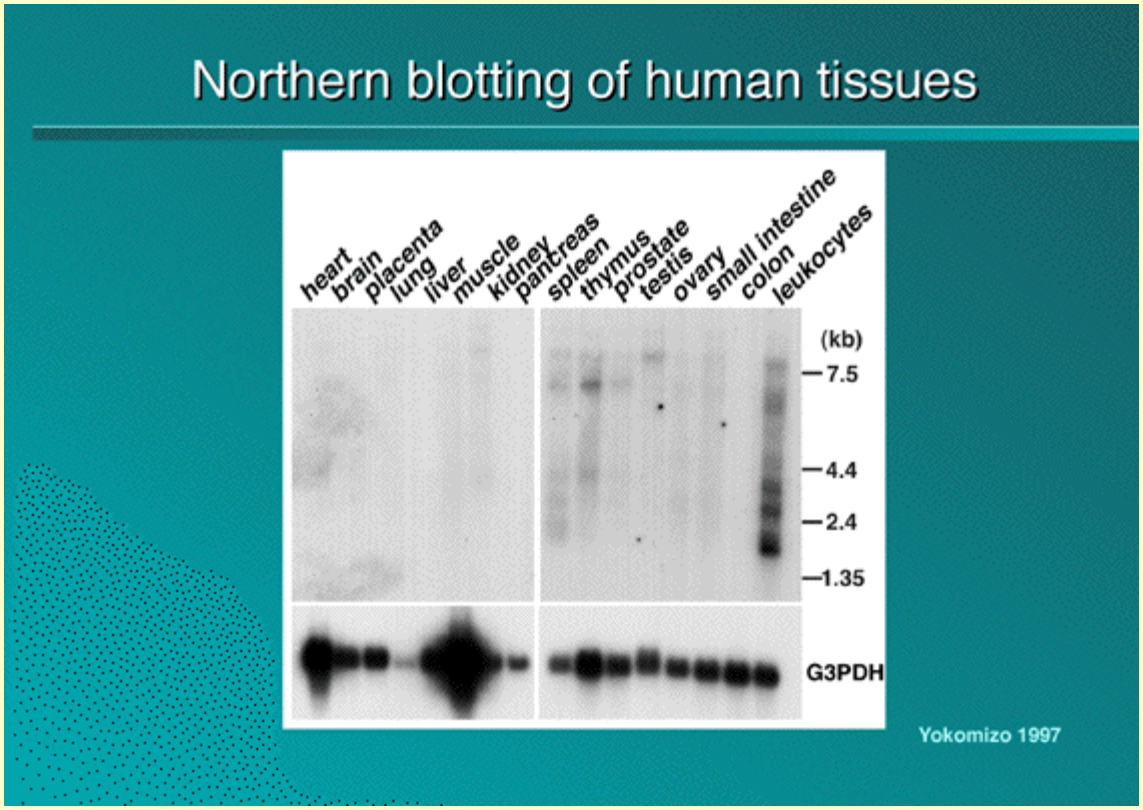
Northern Blotting

Northern measures relative expression levels of mRNA

- mRNA isolation and purification
- electrophoreses on a gel
- The gel is probed by hybridizing with a labeled clone for the gene under study.

Northern Blotting





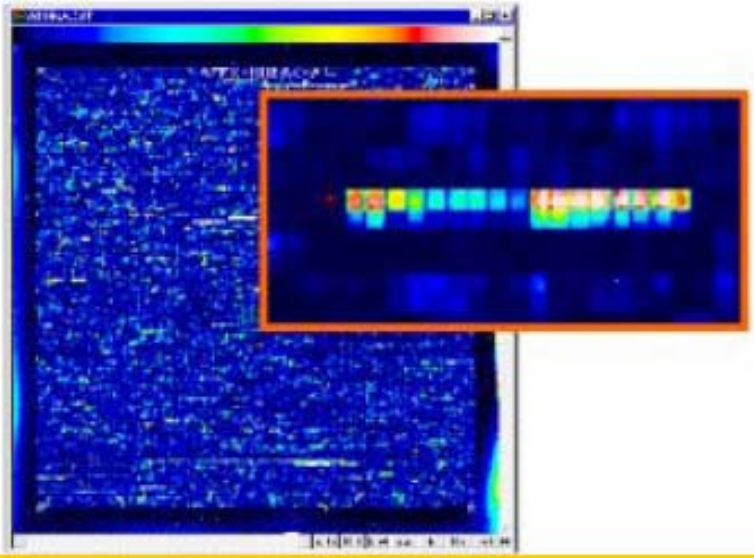
Microarray Steps

- **Experiment and Data Acquisition**
  - **Chip manufacturing**
  - **Sampling and labeling**
  - **Hybridization**
  - **Image scaling**
  - **Data acquisition**
- **Data normalization**
- **Data analysis**
- **Biological interpretation**

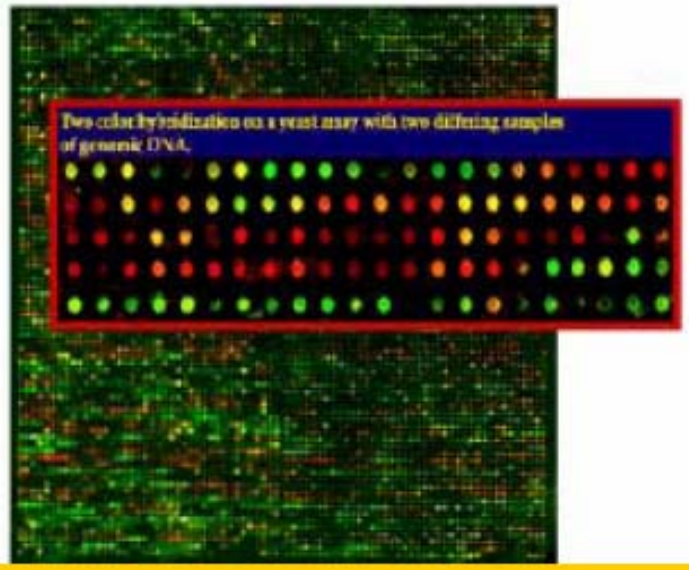
Array Chips

DNA Array Technologies

(A) Affymetrix



(B) "Spotting"



Array Chip Types

1. cDNA chip (DNA microarray, two-channel array):
   - Probe cDNA (500~5,000 bases long) is immobilized to a solid surface such as glass
   - Using robot spotting
   - Traditionally called DNA microarray
   - Firstly developed at Stanford University.

Array Chips

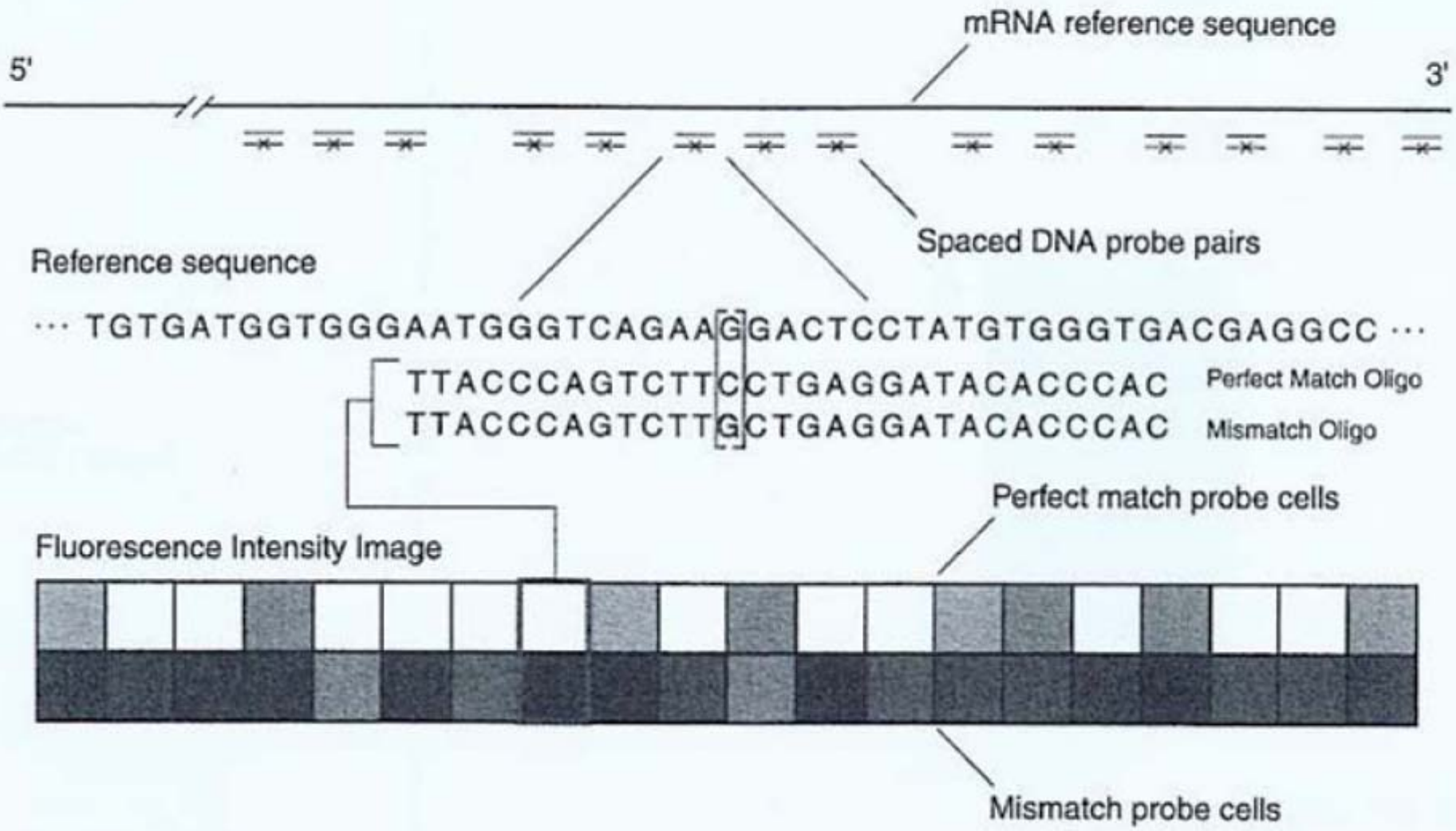
2. Gene chip (DNA chip, Affymetrix chip):

- Oligonucleotide (20~80-mer oligos) is synthesized either *in situ* (on-chip) or by conventional synthesis followed by on-chip immobilization
- Historically called DNA chips
- Developed at Affymetrix, Inc. , under the *GeneChip*® trademark
- Many companies are manufacturing oligonucleotide based chips using alternative technologies

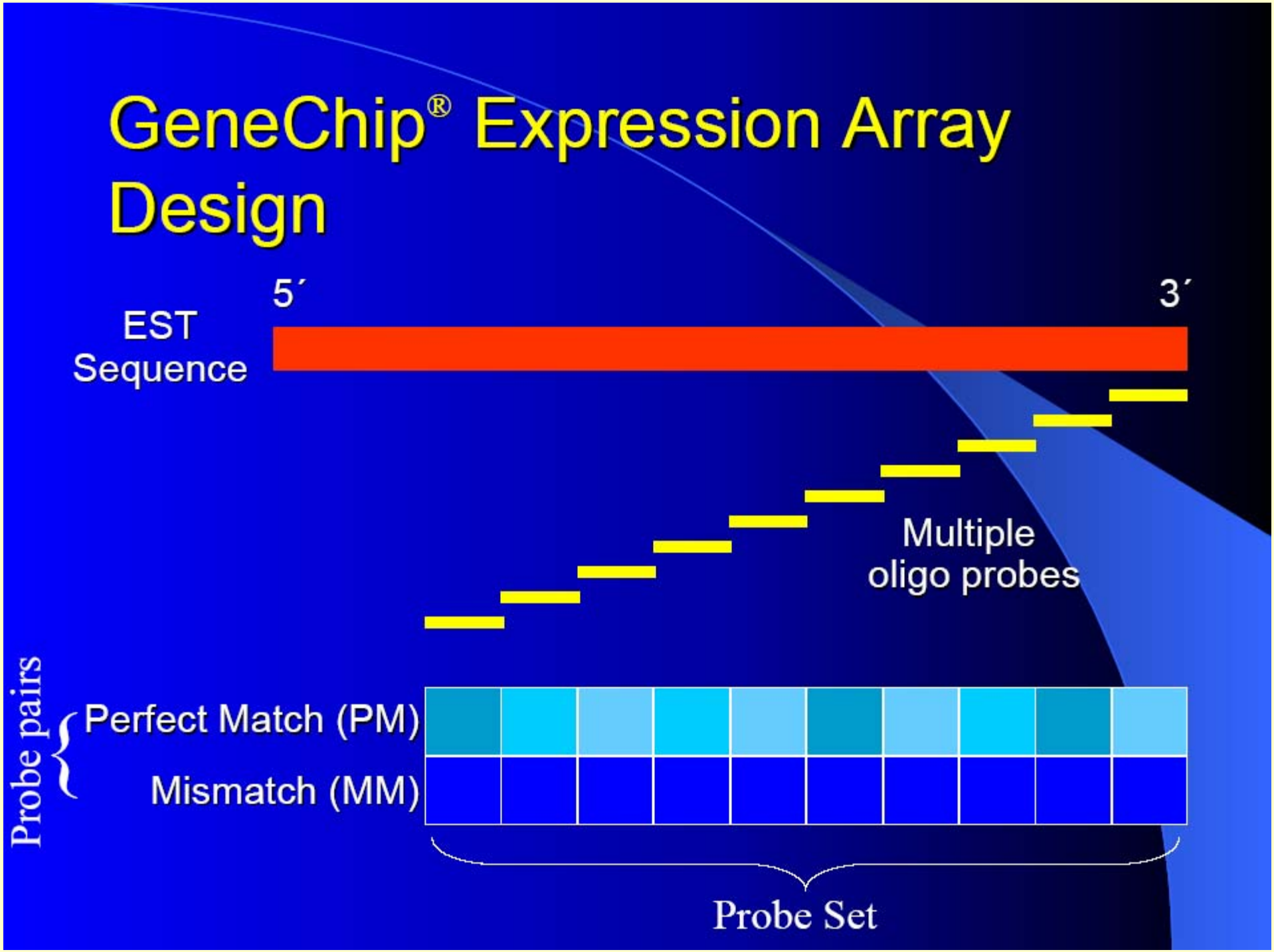
Affymetrix Chip

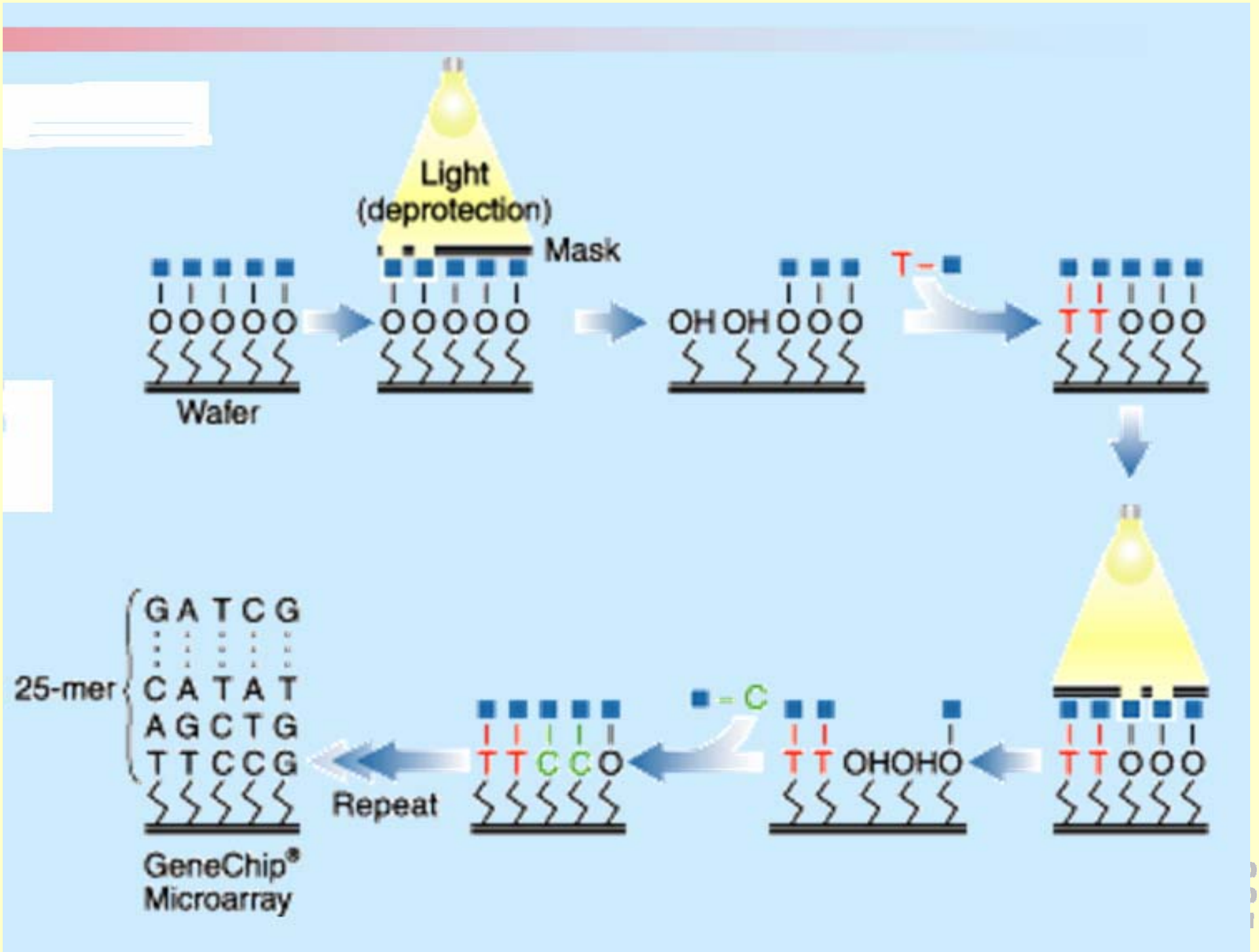
- **Each gene has 16 – 20 pairs of probes synthesized on the chip**
- **Each pairs of probes have two oligonucleotides**
  - **Perfect match (PM, reference seq) ATG…C…TGC (20-25 bases)**
  - **Mismatch (MM, one base change) ATG…T…TGC**
- **The scanned result for a given gene is the average differences between PM and MM signals, over probes**
  - **(MAS5 algorithm)**

GeneChip® Expression Array Design



GeneChip® Expression Array Design

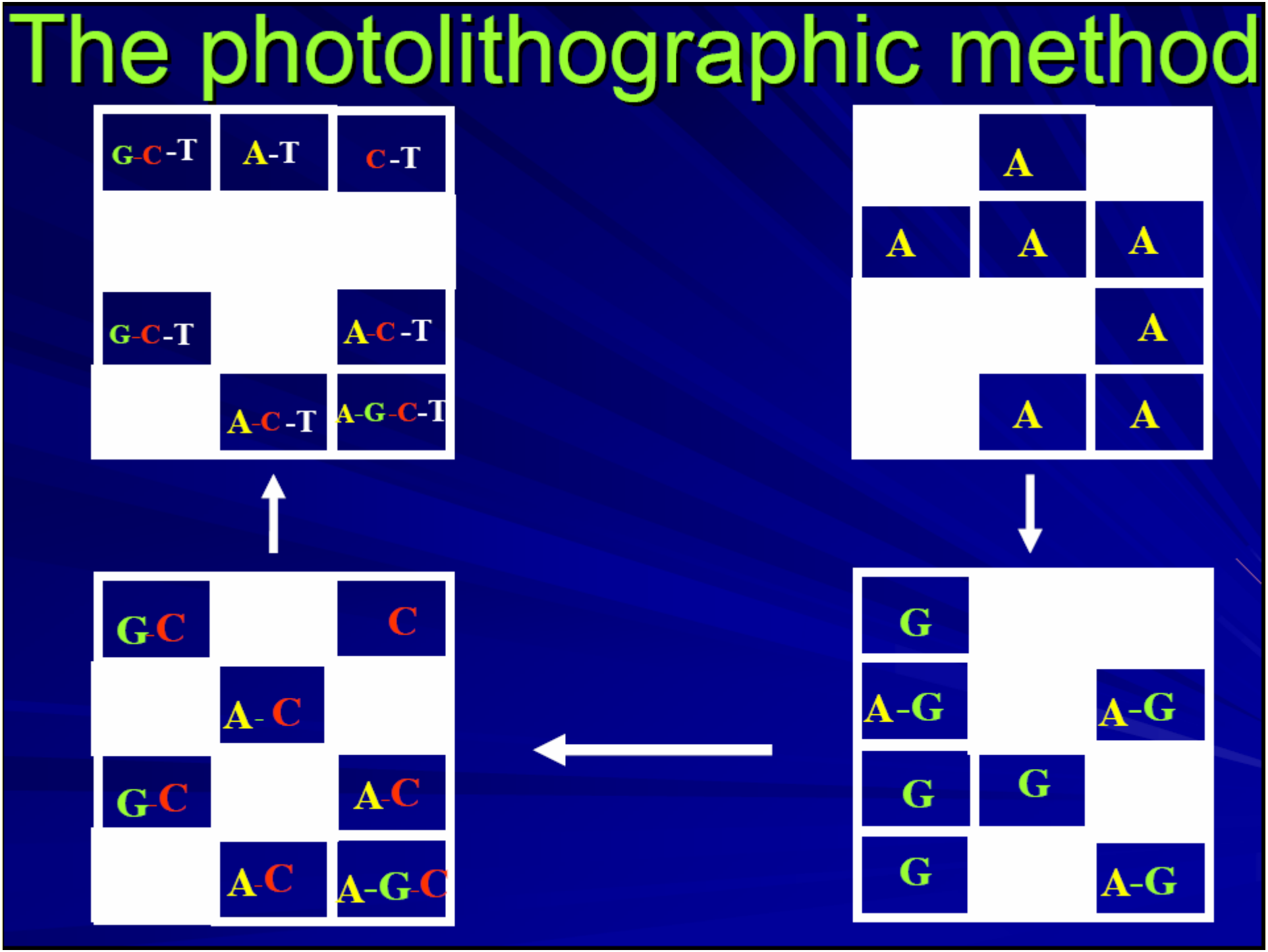



Light
(deprotection)
Mask
Wafer
T –
– C
Repeat
25-mer
G A T C G
C A T A T
A G C T G
T T C C G
GeneChip®
Microarray

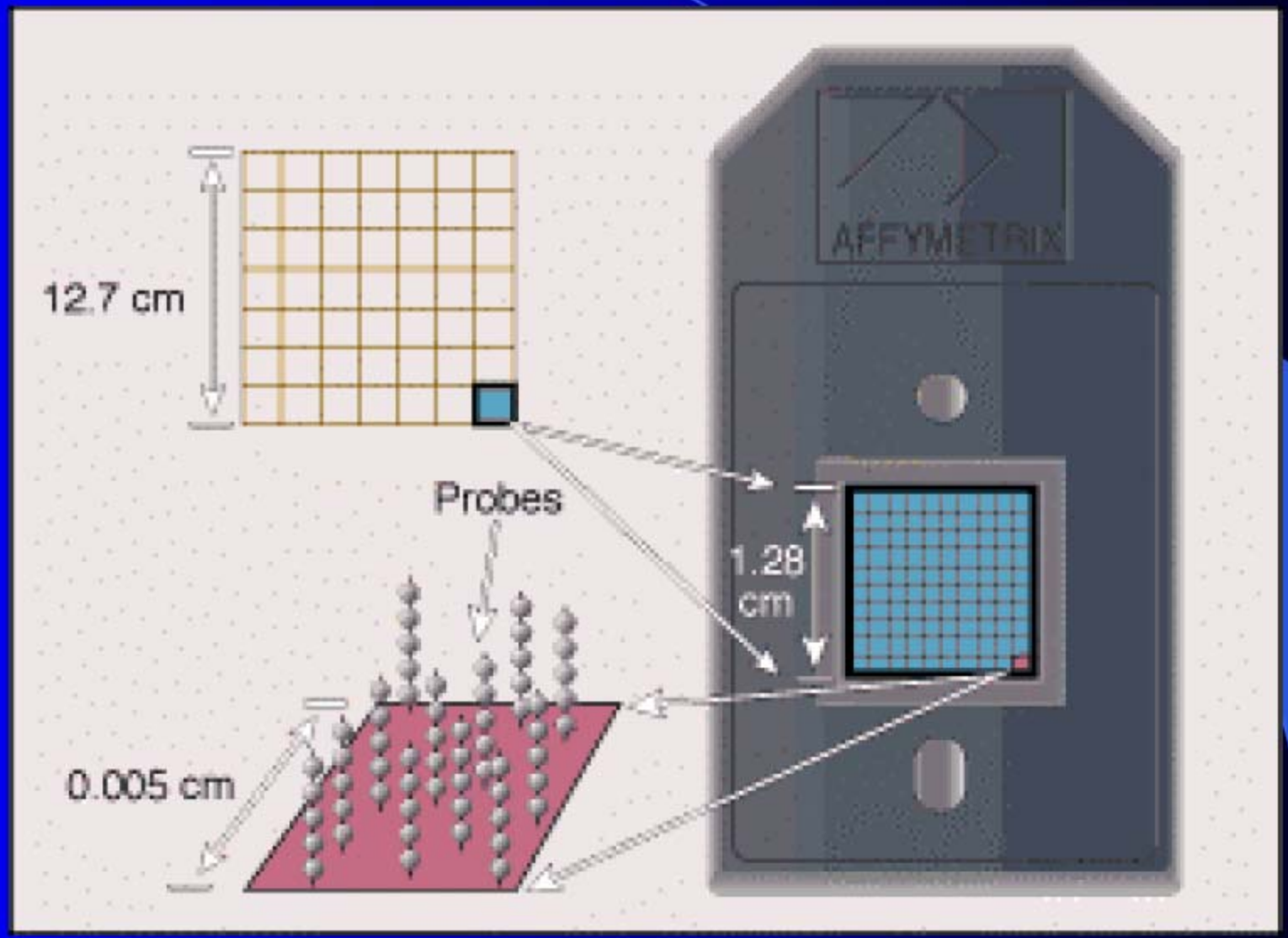
The photolithgraphic method

1. **Treat substrate with chemically protected "linker" molecules, creating rectangular array**
   - **Site size = appro. 10x10 um**
2. **Selectively expose array sites to light**
   - **Light deprotects exposed molecules, activating further synthesis**
3. **Flush chip surface with solution of protected A,C,G,T**
   - **Binding occurs at previously deprotected sites**
4. **Repeat steps 2&3 until desired probes are synthesized**

The photolithographic method

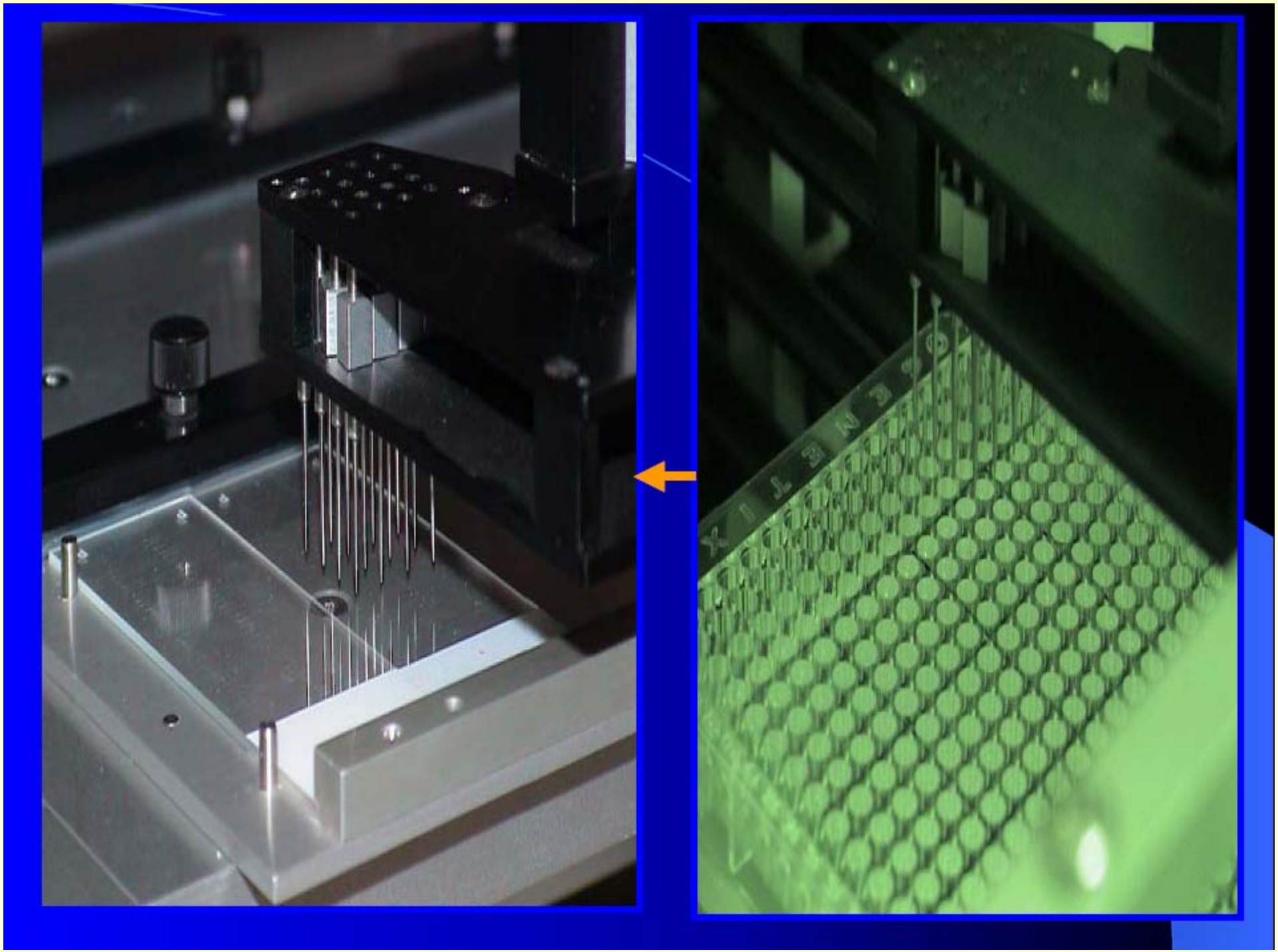


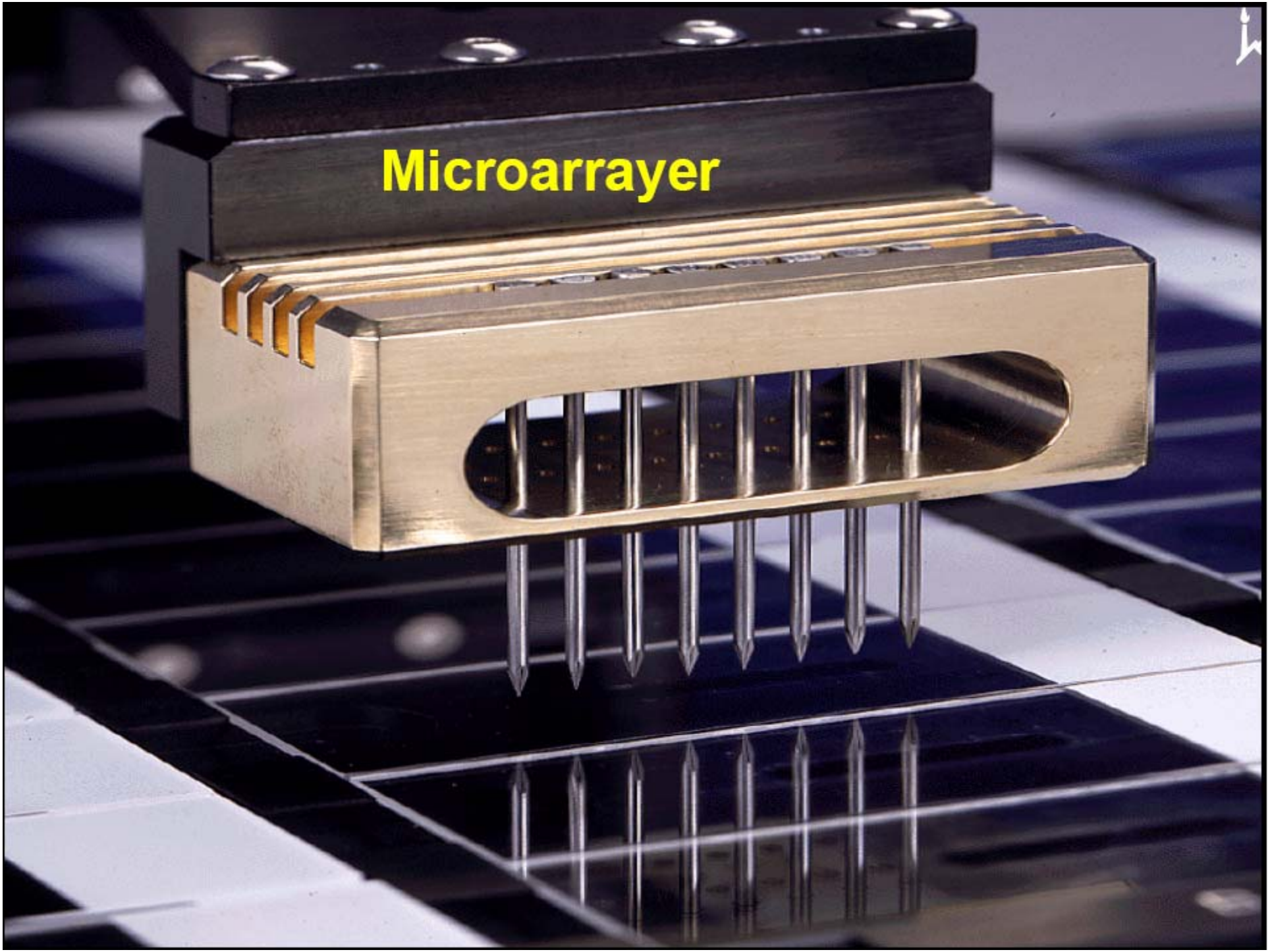
Affymetrix Chip



Manufacture cDNA Array

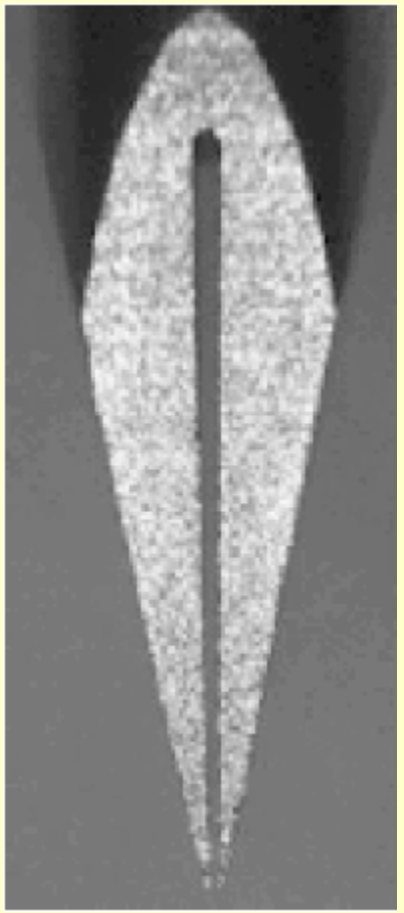
- Start with individual genes, e.g. the ~6,200 genes of the yeast genome
- Amplify all of them using polymerase chain reaction (PCR)
- “Spot” them on a medium, e.g. an ordinary glass microscope slide
- Each spot is about 100 µm in diameter
- Spotting is done by a robot




Microarrayer

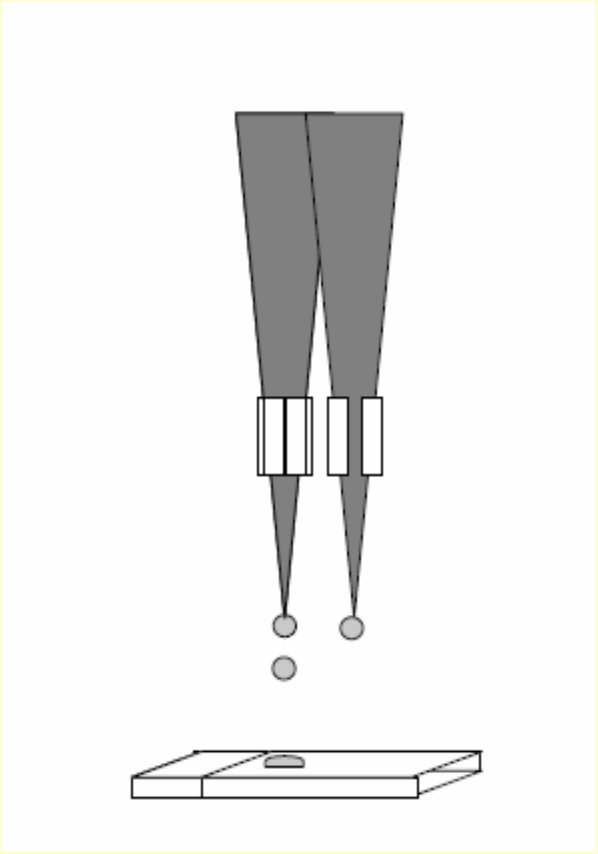
Contact Printing

- **pins**
  - Uptake 0.25 ul
  - Dispense 0.6 nl
  - (approximately 1-10ng per spot)
  - 100 um feature size



Non Contact Printing

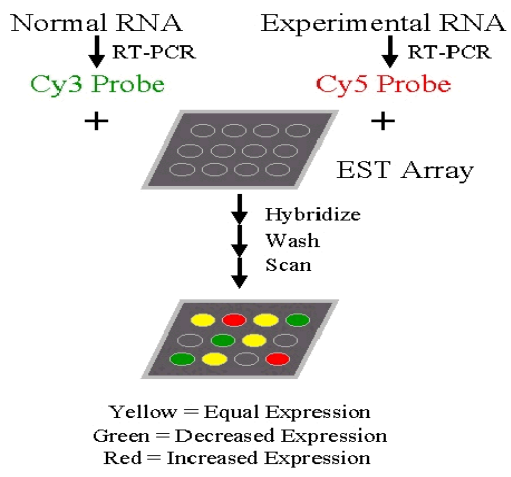
- **Piezoelectric**
- **Ink jet**
- **Higher reproductivity**
- **1 drop = 100 picolitres**

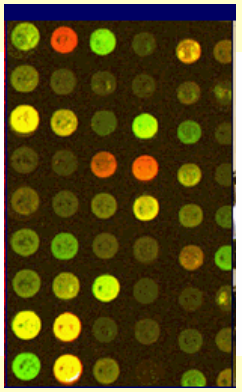


Experiment & Data Acquisition

Example

- Extract mRNA
- Convert mRNA into colored cDNA (fluorescently labeled)
- Mix labeled cDNA together
- Hybridize cDNA with array
- Each cDNA sequence hybridizes specifically with the corresponding gene sequence in the array
- Wash unhybridized cDNA off
- Read array with laser
- Analyze images

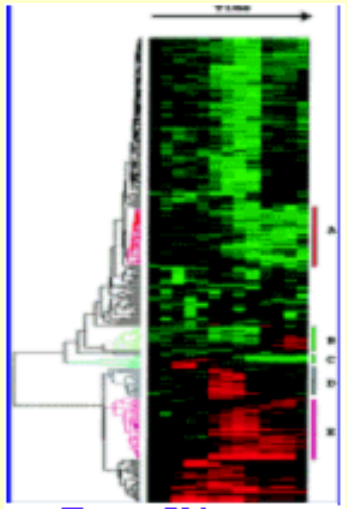




Microarray Scanner

	A	B	C	D	E	F	G
1	YORF	NAME	GWEIGHT	spo0	spo30	spo2	spo5
2	EWEIGHT			1	1	1	1
3	YAL003W	EFB1	1	0.23	-1.79	-1.29	-1.56
4	YAL004W	YAL004W	1	0.41	-0.38	-0.89	-1.06
5	YAL005C	SSA1	1	0.61	-0.07	-1.29	-1.29
6	YAL010C	MDM10	1	0.16	-0.15	-0.76	-1.25

Cluster



TreeView

Affymetrix Microarray Experiment

1. Sample RNA labeling

- First-strand cDNA synthesis
  - Reverse transcriptase
- Second-strand synthesis
  - DNA polymerase
- cDNA purification
- In Vitro transcription to synthesize biotin-labeled RNA
  - T7 enzyme

2. Fragmentation

- **Use heat and Mg++**
- **Reduce RNA to 25-200 bp fragment**
- **Facilitate efficient and reproducible hybridization**

3. Hybridization

- Preheat hybrid mix solution (99 C)
- Affy chip in hybrid solution 5min
- Add probe and hybridization for 16 hours

4. Wash and Stain

- **Wash buffer**
- **Stain with a fluorescent molecule (streptavidin-phycoerythrin) that binds to biotin**
- **A signal amplification step that employs anti-Streptavidin antibody (goat) and biotinylated goat IgG antibody**

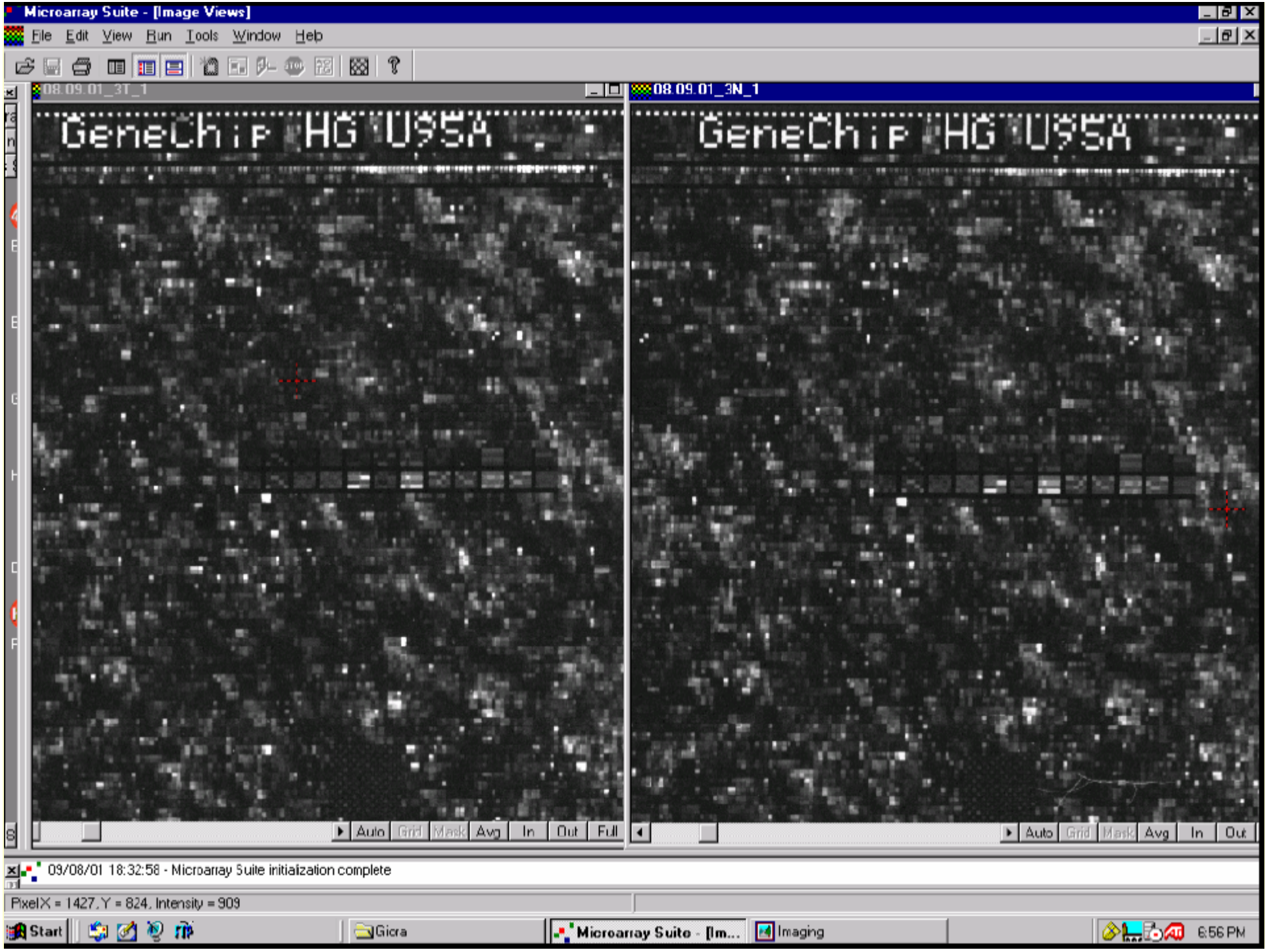
5. Scan

- Affymetrix scanner and follow the menu
- .dat image file
- .cel tab delimited file
- .CHP data file

Data Acquisition

- **Affymetrix Microarray Suite**
- **GCOS** (Genechip Operating System)

- Need chip description file (CDF)
  - For probe location


Microarray Suite - [Image Views]
File Edit View Run Tools Window Help
08.09.01_3T_1
GeneChip HG U95A
Auto Grid Mask Avg In Out Full
08.09.01_3N_1
GeneChip HG U95A
Auto Grid Mask Avg In Out
09/08/01 18:32:58 - Microarray Suite initialization complete
PixelX = 1427, Y = 824, Intensity = 909
Start
Gicra
Microarray Suite - [Im...
Imaging
6:56 PM

.Cel file

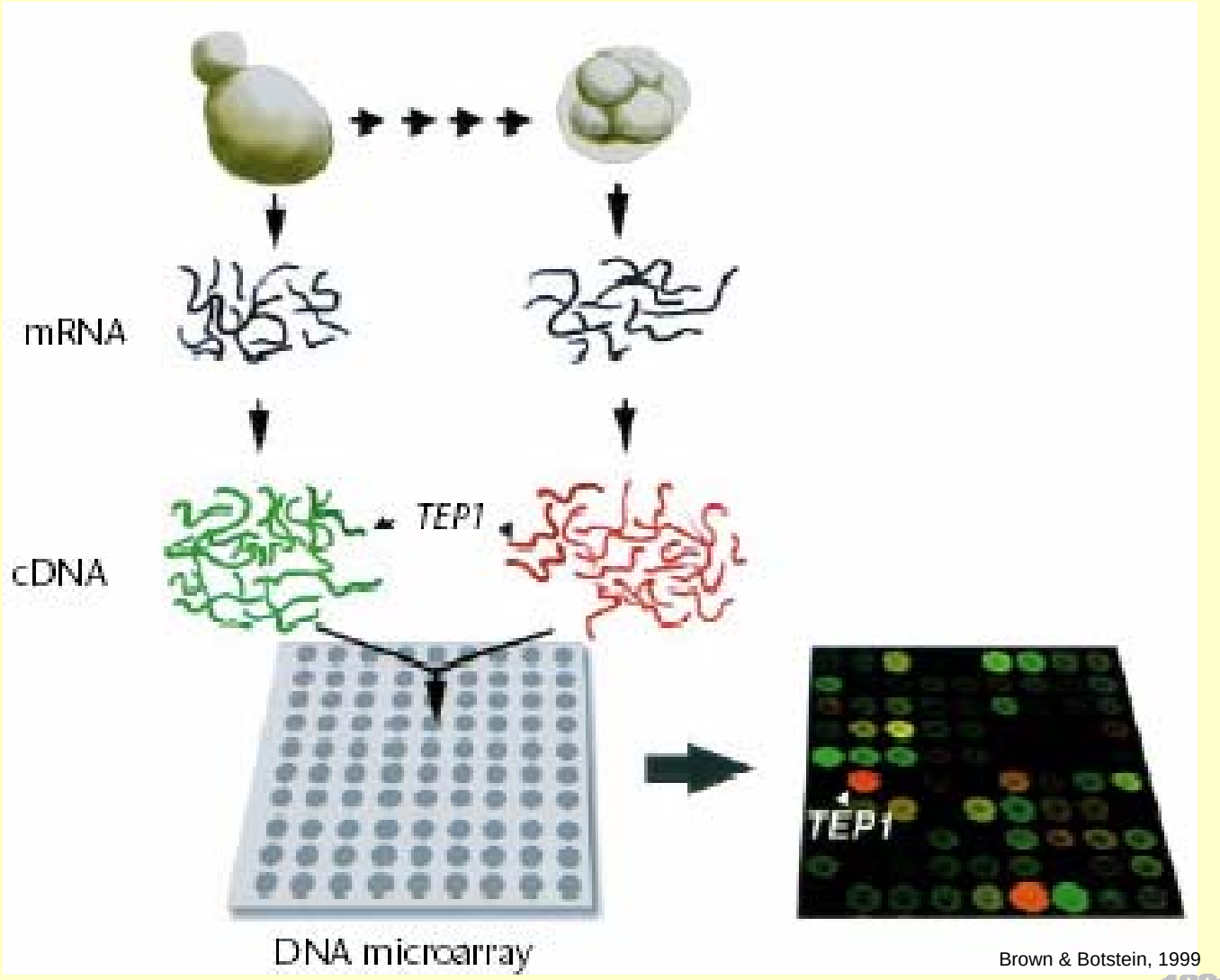
X	Y	Mean	STDV	NPixels
0	0	166	30.8	16
1	0	13135	1216.2	16
2	0	165.3	25.5 16	
3	0	13706	1305.2	16
4	0	95	24.9	16
5	0	155.8	21.8	16
6	0	11675.8	1296.9	16
7	0	184	24.3	16
8	0	11465.5	1533.1	16
..	..	.. ..	..	

.CHP text file

ID	Signal	Det	P-value	Desc
AFFX-CreX-5_at	1200.5	P	0.0007	X03453
AFFX-CreX-3_at	235.8	P	0.0005	L38424
AFFX-CreX-9_at	15	A	0.5	K01391
..	..	..	..	..

cDNA Microarray Experiment

1. **Array fabrication**
   - **DNA clones**
     - **Unigene**
     - **EST clustering**
   - **PCR amplification of clones**
   - **Array printing**



Brown & Botstein, 1999

2. Probe preparation

- RNA extraction (control, test)
- RNA labeling
  - Incorporate fluorescently labeled deoxyribonucleotides
  - First strand cDNA
  - Cyanine5 labels Test sample RNA
  - Cyanine3 labels Control sample RNA
- Mix the labeled two RNAs

3. Hybridization

- Prehybridize slide 42C 45 min
- Hybridize preheated probes 16-20 hours

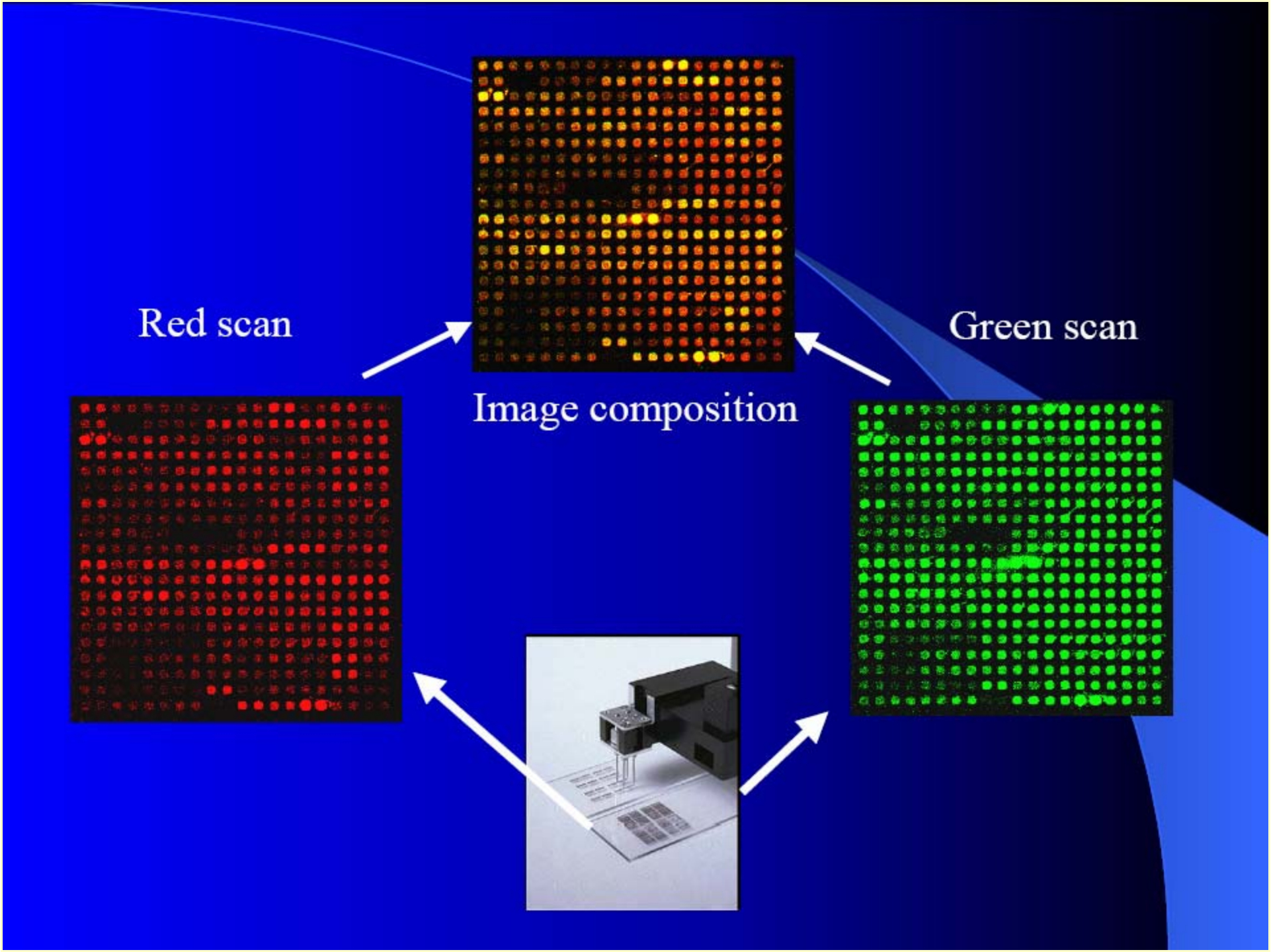
4. Slide scaning

- C3 16-bit TIFF image file
- C5 16-bit TIFF image file

Data Acquisition

- **GenePix**
- **Quantarray**

- Need chip description file (CDF)
  - For probe location


Red scan
Green scan
Image composition

Raw Data

Name	ch2/ch1 Ratio
Control_M12	2.953803
PPSL_25A09	1.206626
PPSL_25C09	2.389387
PPSL_25E09	2.24675
.....	....

Data Normalization and Data Analysis

Data Normalization

- **Why?**
  - **Reliability**
    - **Remove non-biological variation**
  - **Comparability**
    - **Scale (multiplicative factor)**

Data Analysis and Visualization

- **Use microarray software**
- **Address biological questions**

Address Biological Questions

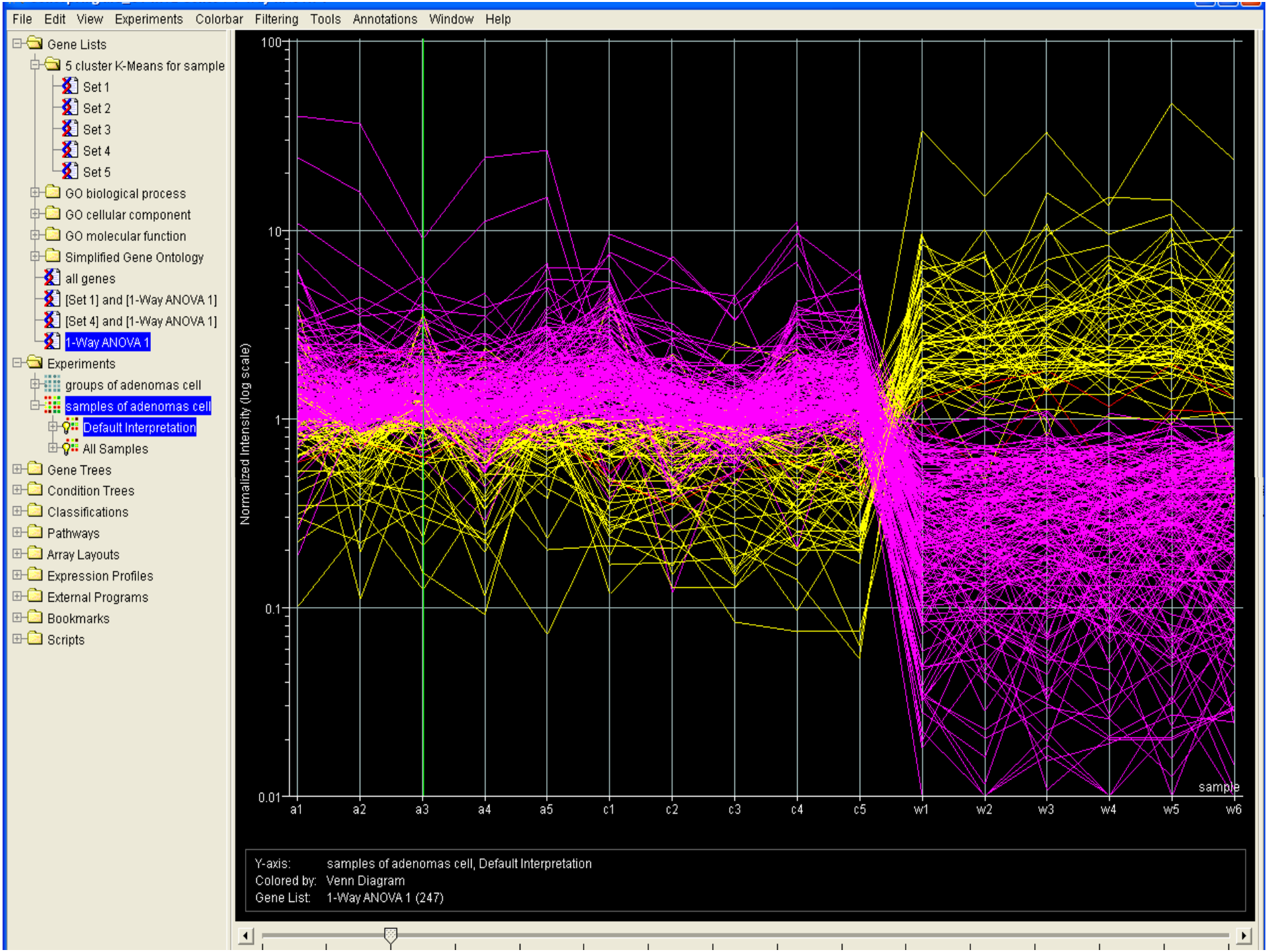
- What genes are involved in a particular biological process?
- What genes are turned-on?
- What genes are turned-off?
- What genes are the key elements in a biological process?
- Similar clinic samples share similar gene expression profile?
  - Sample classification

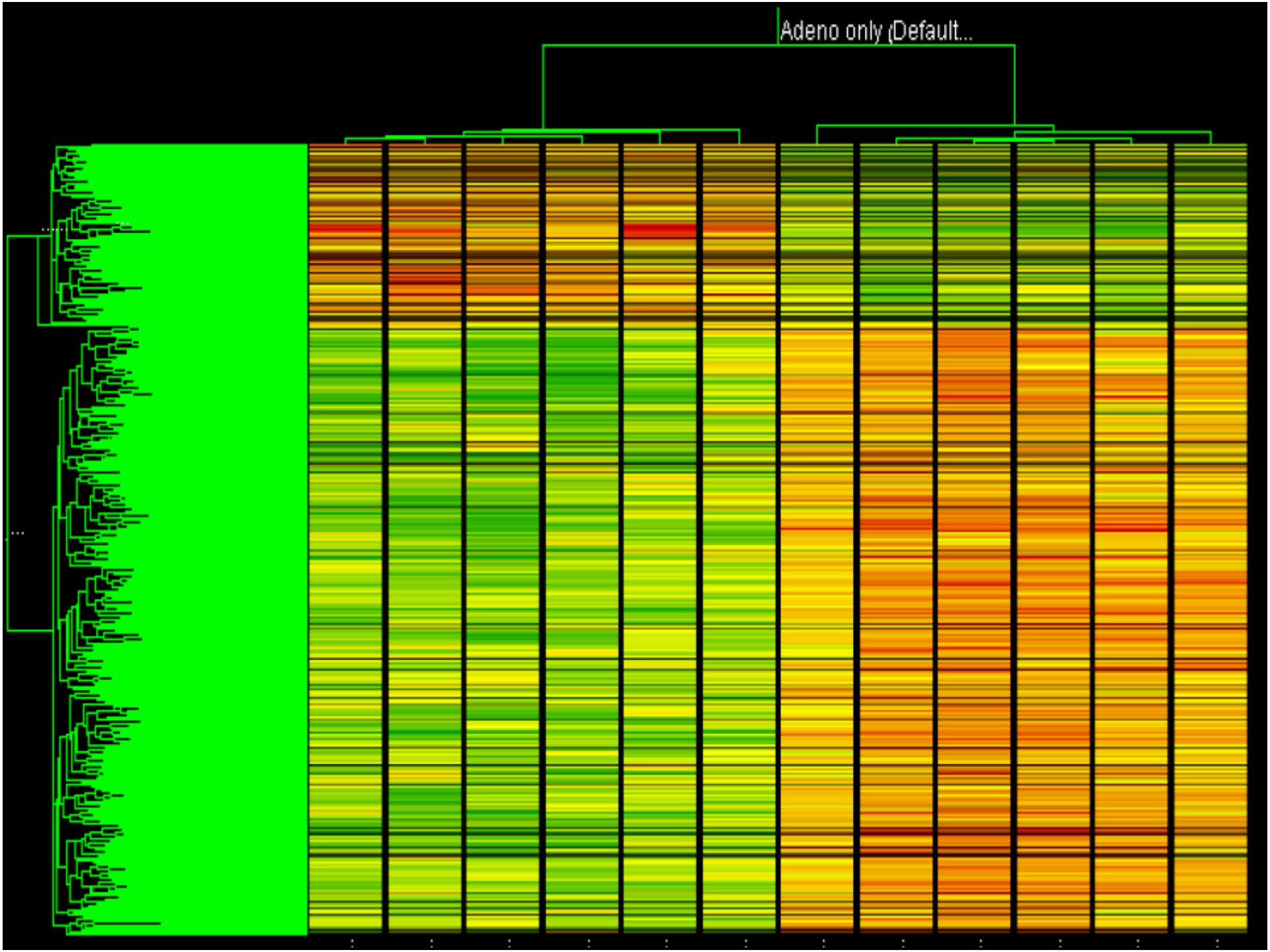
Address Biological Questions

- What genes have similar profile?
- What are the features for the similar profile genes?
  - Gene classification
    - Functional annotation
    - Pathways
- What is the functional behavior of a particular gene?
  - Functional screening

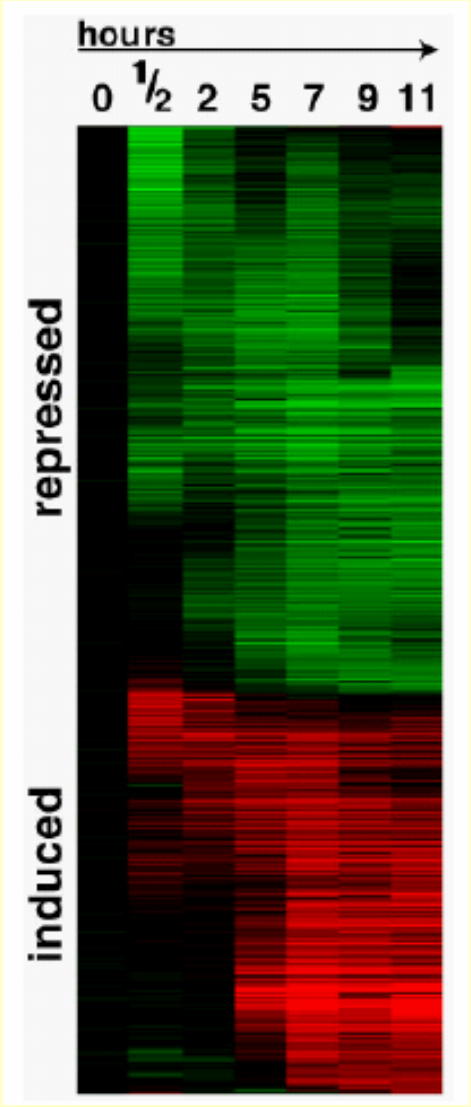
Software Tools

- GeneSpring (SiliconGenetics)
- Expressionist (GeneData)
- GeneTraffic (Iobion)
- Spotfire (Spotfire)
- Cluster and TreeView (free)
- …..


File Edit View Experiments Colorbar Filtering Tools Annotations Window Help
Gene Lists
5 cluster K-Means for sample
Set 1
Set 2
Set 3
Set 4
Set 5
GO biological process
GO cellular component
GO molecular function
Simplified Gene Ontology
all genes
[Set 1] and [1-Way ANOVA 1]
[Set 4] and [1-Way ANOVA 1]
1-Way ANOVA 1
Experiments
groups of adenomas cell
samples of adenomas cell
Default Interpretation
All Samples
Gene Trees
Condition Trees
Classifications
Pathways
Array Layouts
Expression Profiles
External Programs
Bookmarks
Scripts
Normalized Intensity (log scale)
100
10
1
0.1
0.01
a1
a2
a3
a4
a5
c1
c2
c3
c4
c5
w1
w2
w3
w4
w5
w6
sample
Y-axis: samples of adenomas cell, Default Interpretation
Colored by: Venn Diagram
Gene List: 1-Way ANOVA 1 (247)


Adeno only (Default...

Clustering of entire yeast genome



Campbell & Heyer, 2003