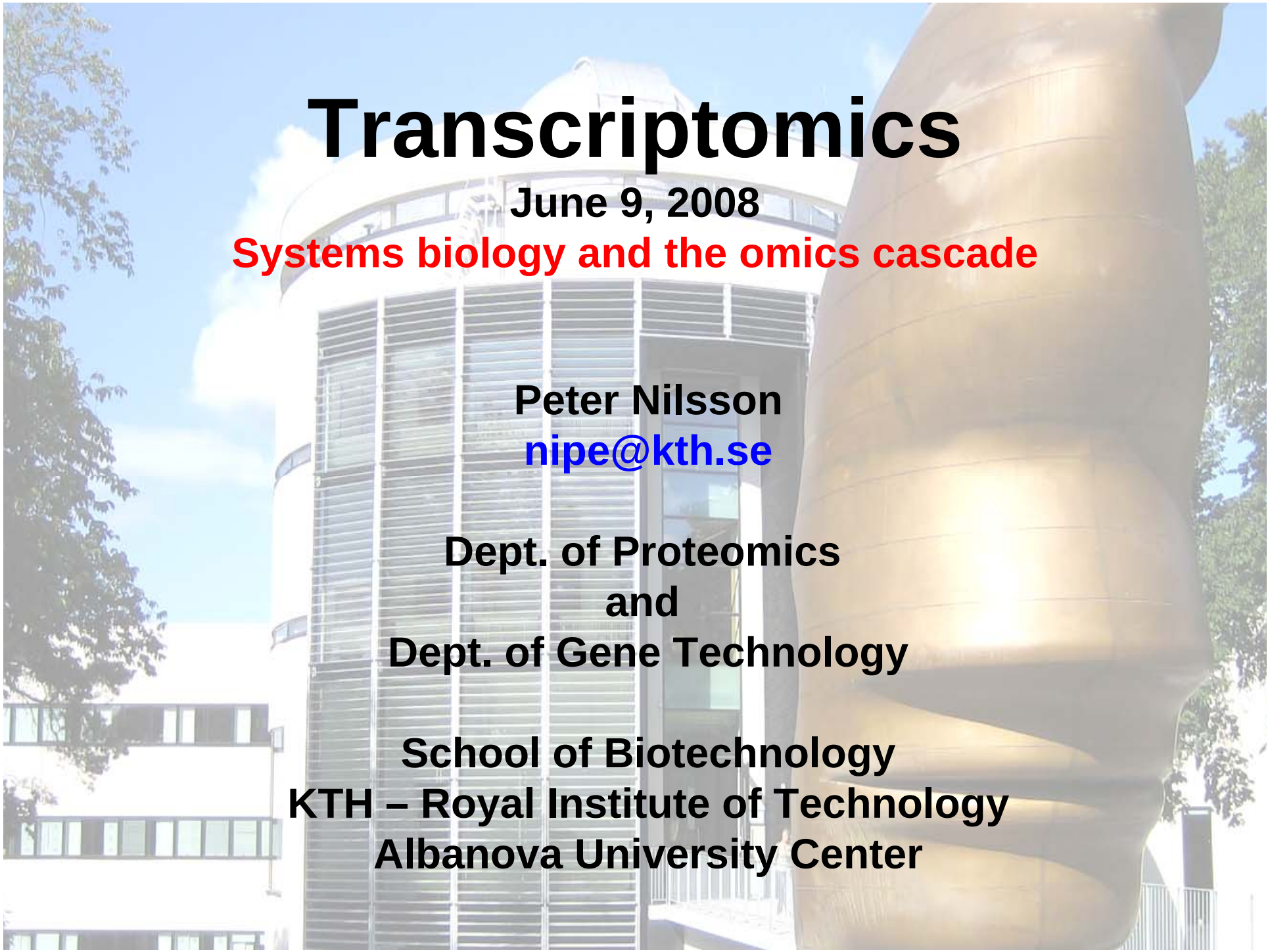


Transcriptomics

June 9, 2008

Systems biology and the omics cascade

Peter Nilsson
nipe@kth.se

Dept. of Proteomics
and
Dept. of Gene Technology

School of Biotechnology
KTH – Royal Institute of Technology
Albanova University Center

Transcriptome

”The complete set of transcripts and their relative levels of expression in a particular cell or tissue type under defined conditions.”

“The full complement of all activated genes, mRNAs, or transcripts in a particular cell at a particular time”

“A transcriptome is a collection of all the gene transcripts present in a given cell. ”

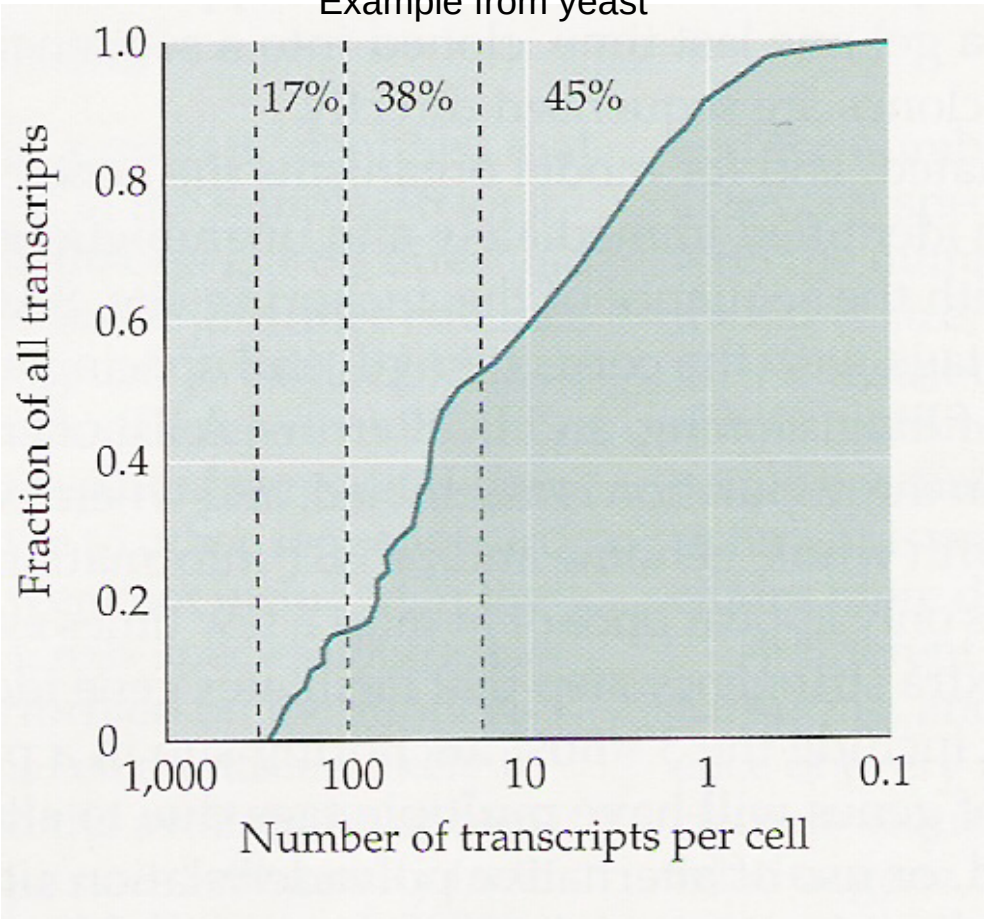
mRNA abundance

<u>High:</u>
<100 different transcripts
100-1000/cell
~20% of the transcriptome

<u>Intermediate:</u>
several hundred different transcripts
10-100/cell
~30% of the transcriptome

<u>Low</u>
>10.000 different transcripts
0.1-10/cell
~50% of the transcriptome

Example from yeast



Noncoding RNAs

- Recently very much focus, interesting research and new findings

2001

NON-CODING RNA GENES AND THE MODERN RNA WORLD

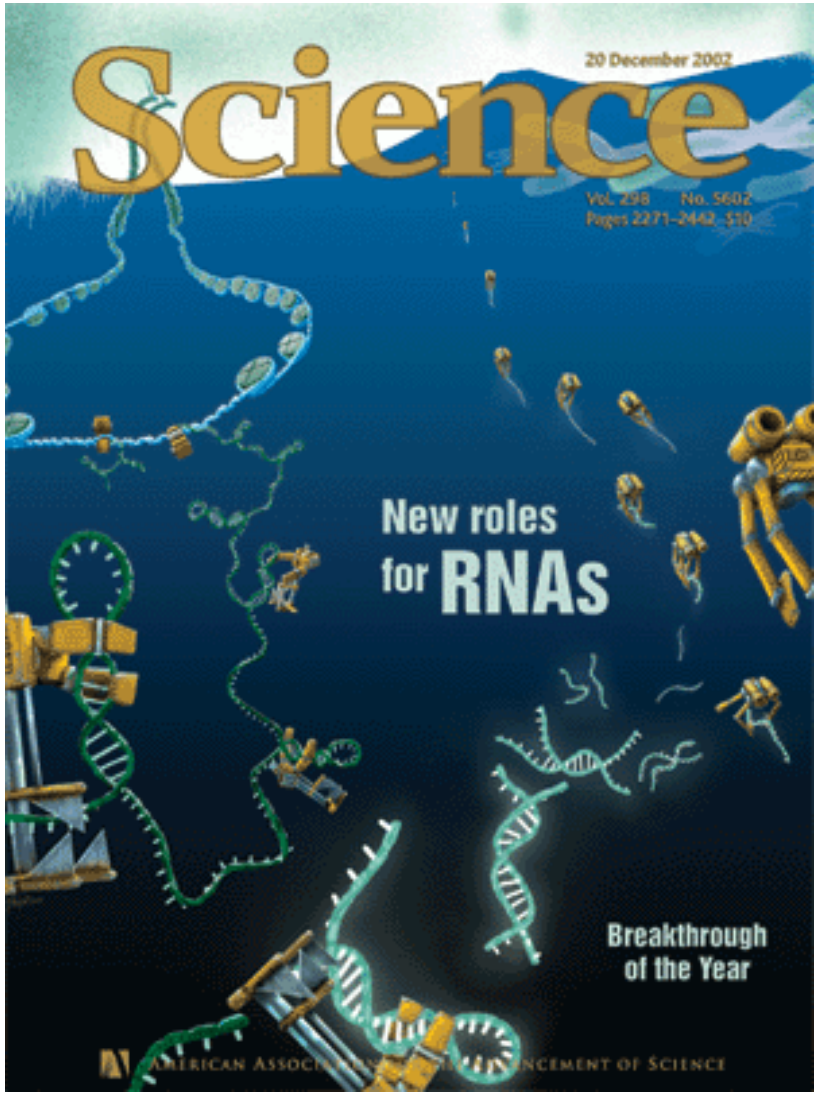
Sean R. Eddy

Non-coding RNA (ncRNA) genes produce functional RNA molecules rather than encoding proteins. However, almost all means of gene identification assume that genes encode proteins, so even in the era of complete genome sequences, ncRNA genes have been effectively invisible. Recently, several different systematic screens have identified a surprisingly large number of new ncRNA genes. Non-coding RNAs seem to be particularly abundant in roles that require highly specific nucleic acid recognition without complex catalysis, such as in directing post-transcriptional regulation of gene expression or in guiding RNA modifications.

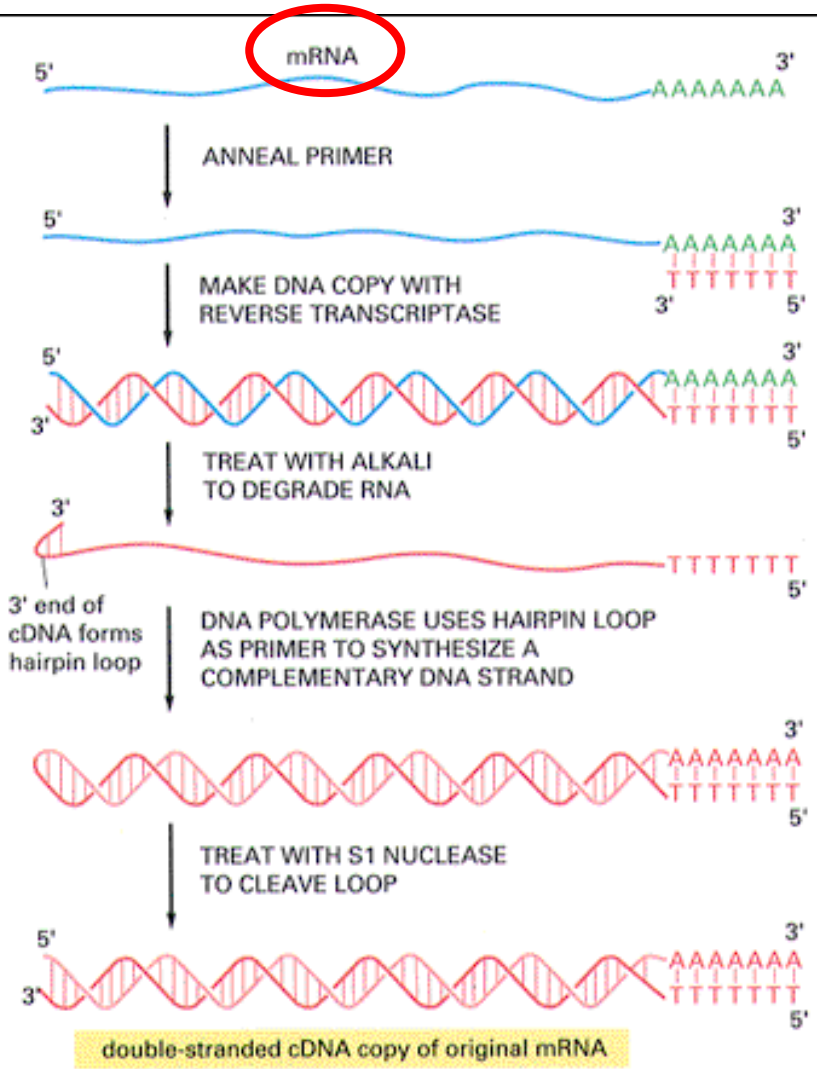
Box 1 | **Abbreviations for different classes of non-coding RNA**

- fRNA
  Functional RNA — essentially synonymous with non-coding RNA[104]
- miRNA
  MicroRNA — putative translational regulatory gene family
- ncRNA
  Non-coding RNA — all RNAs other than mRNA[13]
- rRNA
  Ribosomal RNA
- siRNA
  Small interfering RNA — active molecules in RNA interference
- snRNA
  Small nuclear RNA — includes spliceosomal RNAs
- snmRNA
  Small non-mRNA — essentially synonymous with small ncRNAs[14]
- snoRNA
  Small nucleolar RNA — most known snoRNAs are involved in rRNA modification
- stRNA
  Small temporal RNA — for example, *lin-4* and *let-7* in *Caenorhabditis elegans*
- tRNA
  Transfer RNA

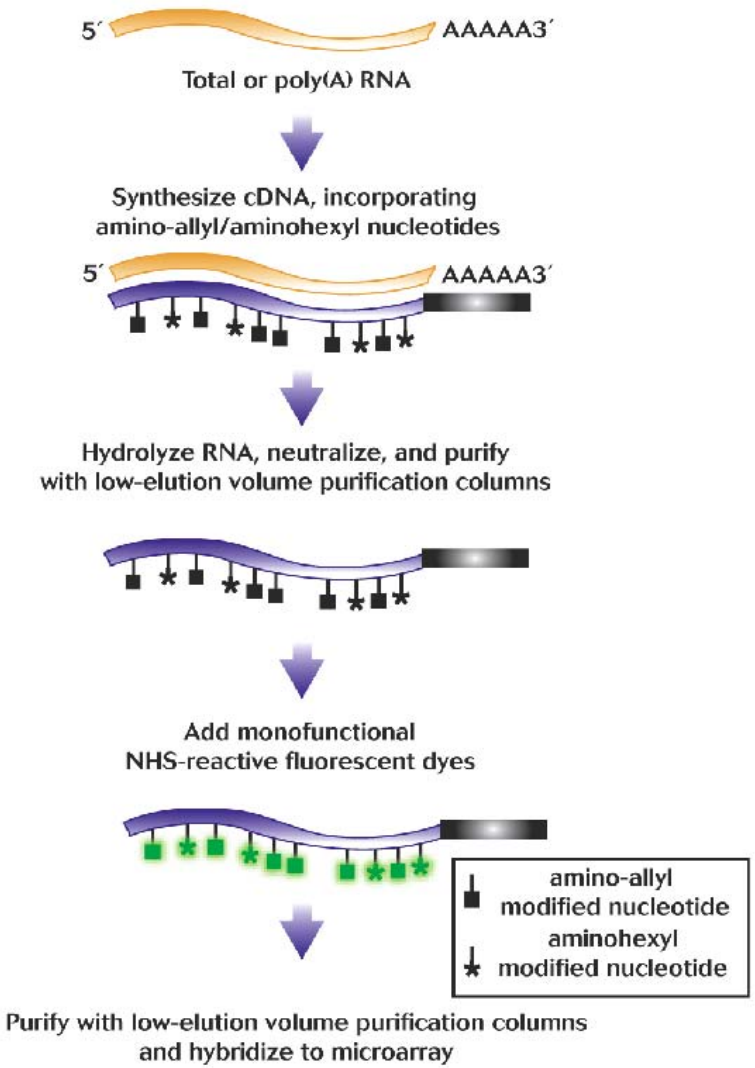
2002



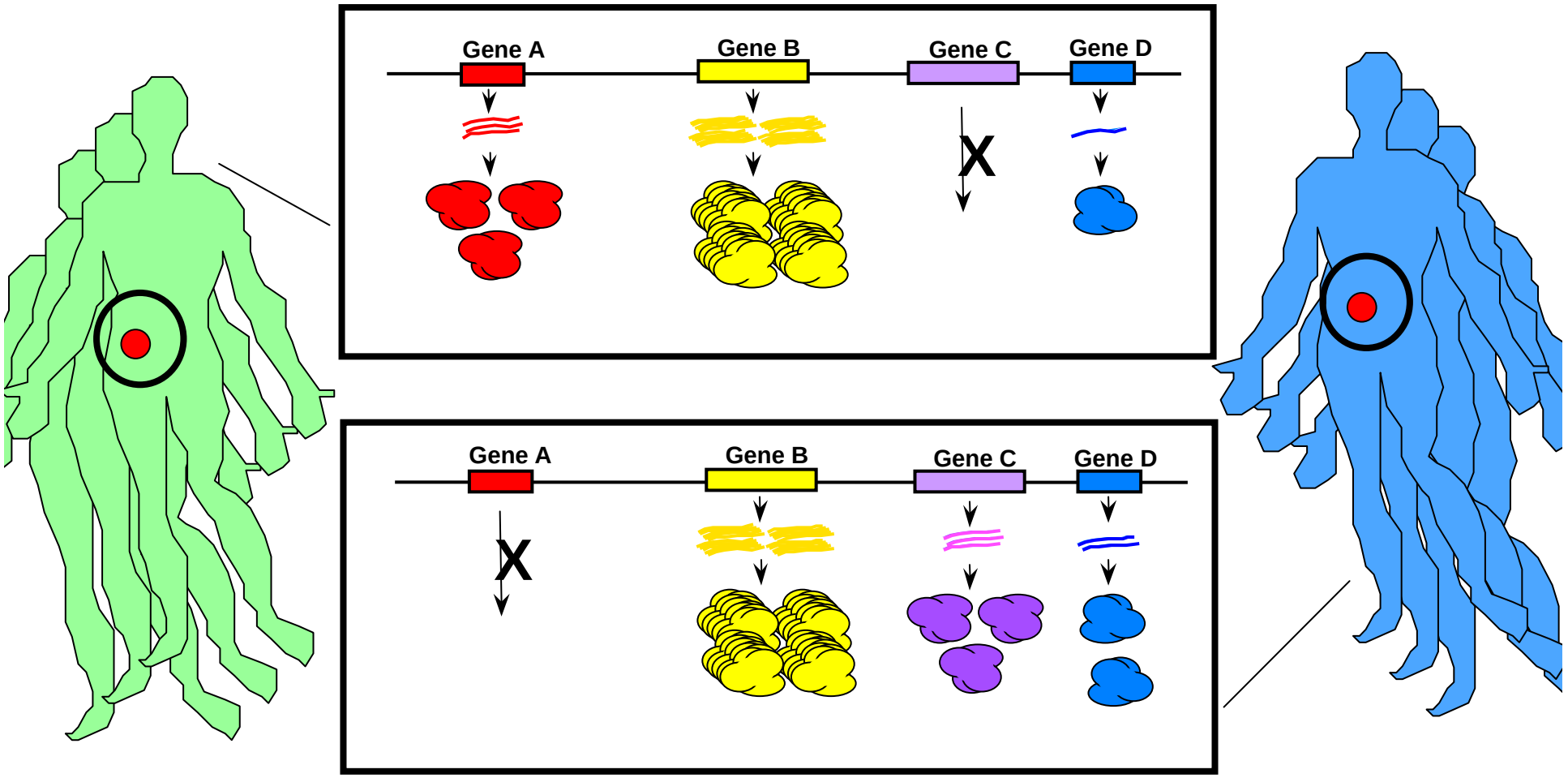
cDNA synthesis



Labelled cDNA



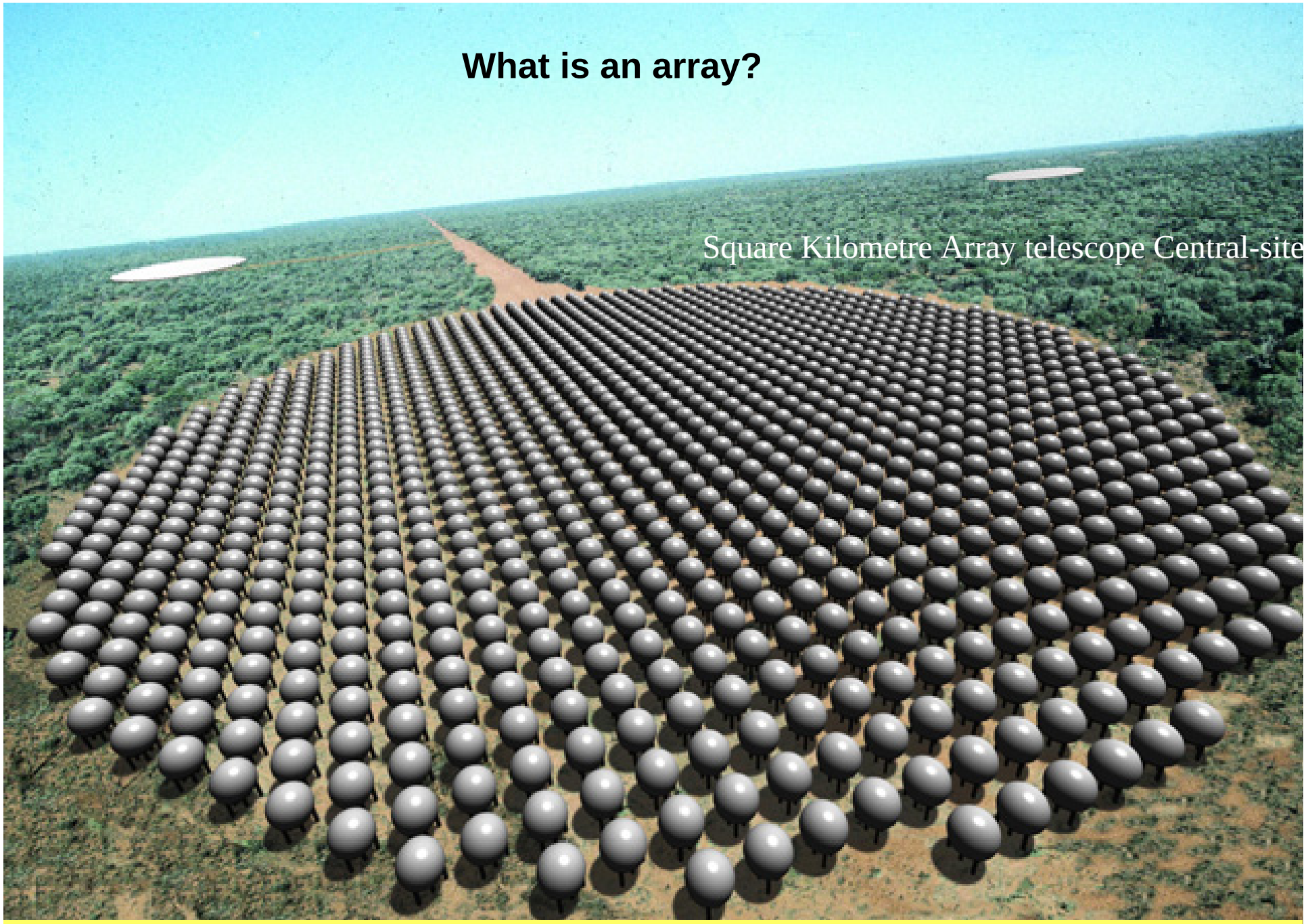
Disease/tissue/sample specific gene expression



Identify relative differences in expression of gene products

Analyse the relative amount of genes that are transcribed at a certain timepoint or condition
Comparison of gene activity

What is an array?

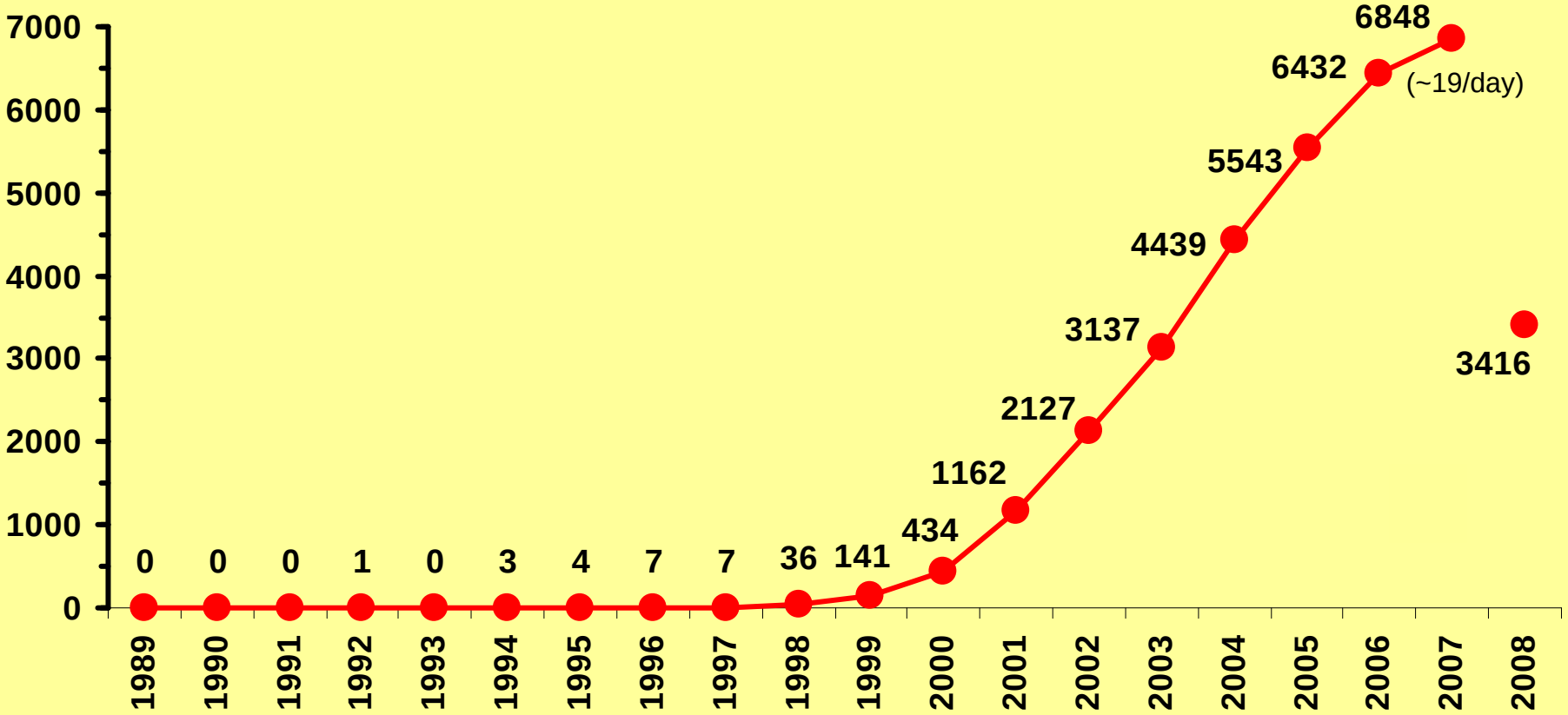


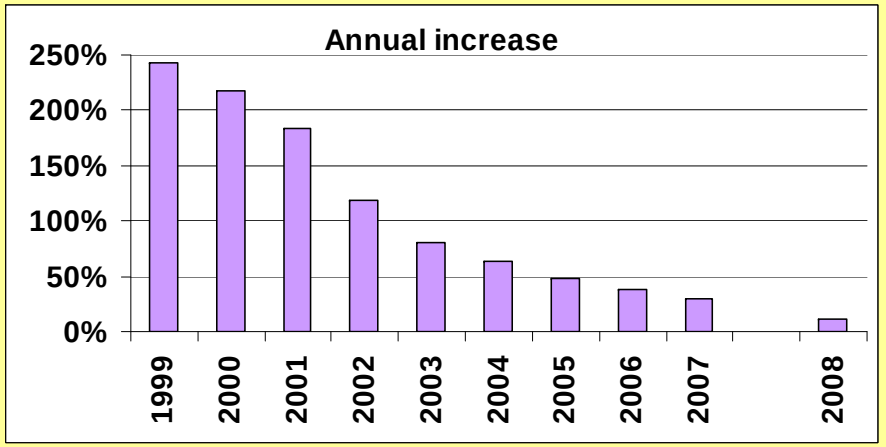
Square Kilometre Array telescope Central-site

Array = impressive ordered collection

Microarray = “Mikromatris” in Swedish

>31.000 microarray articles in PubMed





Science, Vol 270, Issue 5235, 467-470 , 20 October 1995

Reports

Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray

Mark Schena (1), Dari Shalon (1), Ronald W. Davis (2), Patrick O. Brown (3)

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.

M. Schena and R. W. Davis, Department of Biochemistry, Beckman Center, Stanford University Medical Center, Stanford, CA 94305, USA.
D. Shalon and P. O. Brown, Department of Biochemistry and Howard Hughes Medical Institute, Beckman Center, Stanford University Medical Center, Stanford, CA 94305, USA.
(1) These authors contributed equally to this work.
(2) Present address: Synteni, Palo Alto, CA 94303, USA.
(3) To whom correspondence should be addressed. E-mail: pbrown@cmgm.stanford.edu

Microarray publications from 1996

Schena M.
Genome analysis with gene expression microarrays.
Bioessays. 1996 May;18(5):427-31. Review.

Shalon D, Smith SJ, Brown PO.
A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization.
Genome Res. 1996 Jul;6(7):639-45.

Schena M, Shalon D, Heller R, Chai A, Brown PO, Davis RW.
Parallel human genome analysis: microarray-based expression monitoring of 1000 genes.
Proc Natl Acad Sci U S A. 1996 Oct 1;93(20):10614-9.

DeRisi J, Penland L, Brown PO, Bittner ML, Meltzer PS, Ray M, Chen Y, Su YA, Trent JM.
Use of a cDNA microarray to analyse gene expression patterns in human cancer.
Nat Genet. 1996 Dec;14(4):457-60.

Nelson N.
Microarrays pave the way to 21st century medicine.
J Natl Cancer Inst. 1996 Dec 18;88(24):1803-5.

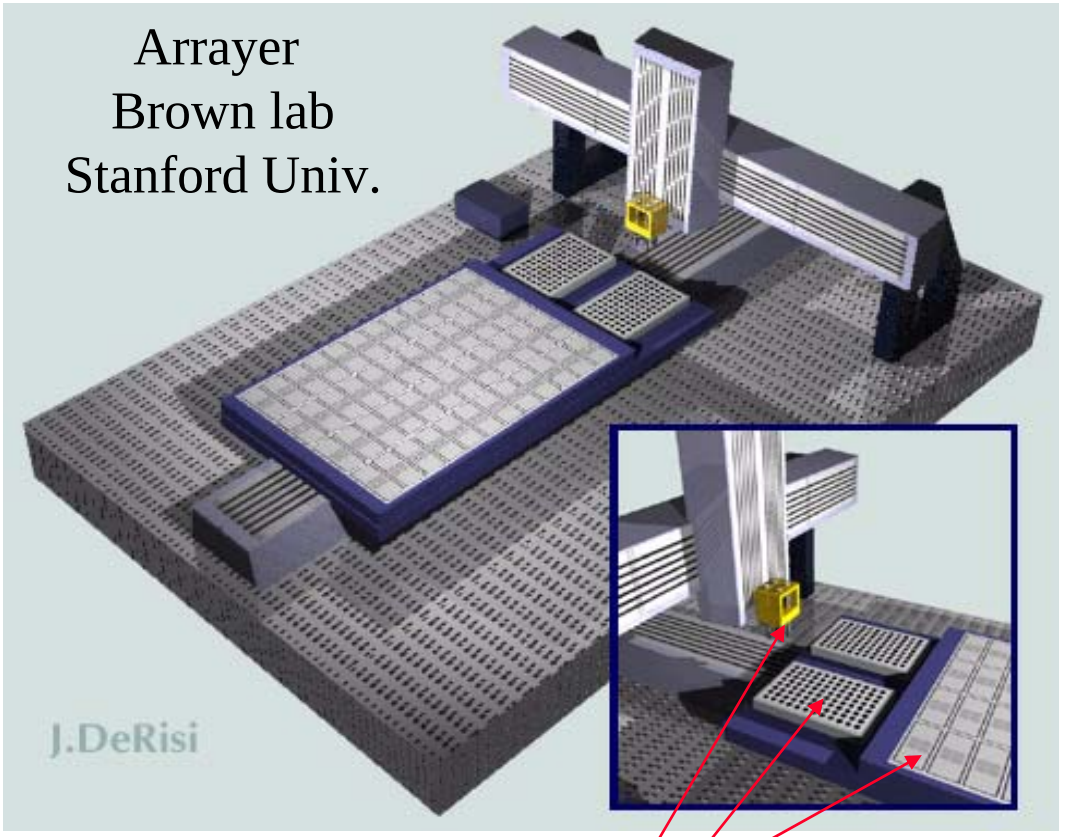
Fragment-based DNA printing
Array of PCR products

http://cmgm.stanford.edu/pbrown/mguide/index.html

Purified PCR products are spotted by an arrayer, on a glass slide, with a positively charged coating.

The DNA is snap-dried and UV-crosslinked to the surface.

Pat Brown



Schena (1995), Science, 270: 467

Methods for creating a DNA array

I) In situ (on-chip) synthesis of oligonucleotides

Photolitography (Affymetrix)

- Light-directed oligonucleotide synthesis on chip.
- Adapted from semiconductor industry

Inkjet technology (Agilent)

- Adapted from the technique used in ink-jet printers.
- Piezoelectric effect: a charge on a narrow tube containing nucleotides, forces out a small drop onto a coated glass slide.
- The oligonucleotides are synthesized drop-by-drop.

II) Spotting of long DNA fragments

Fragment based cDNA printing (Stanford University)

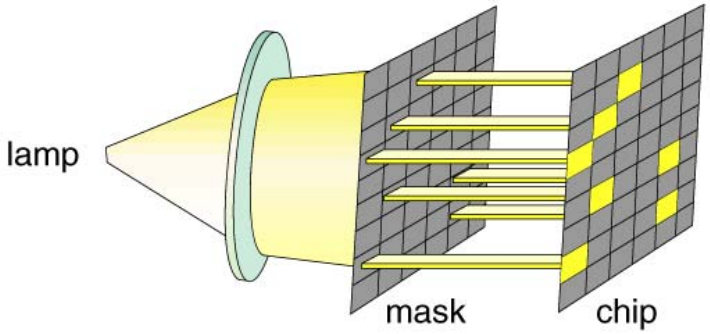
- Array of spotted PCR products.
- For gene expression analysis.

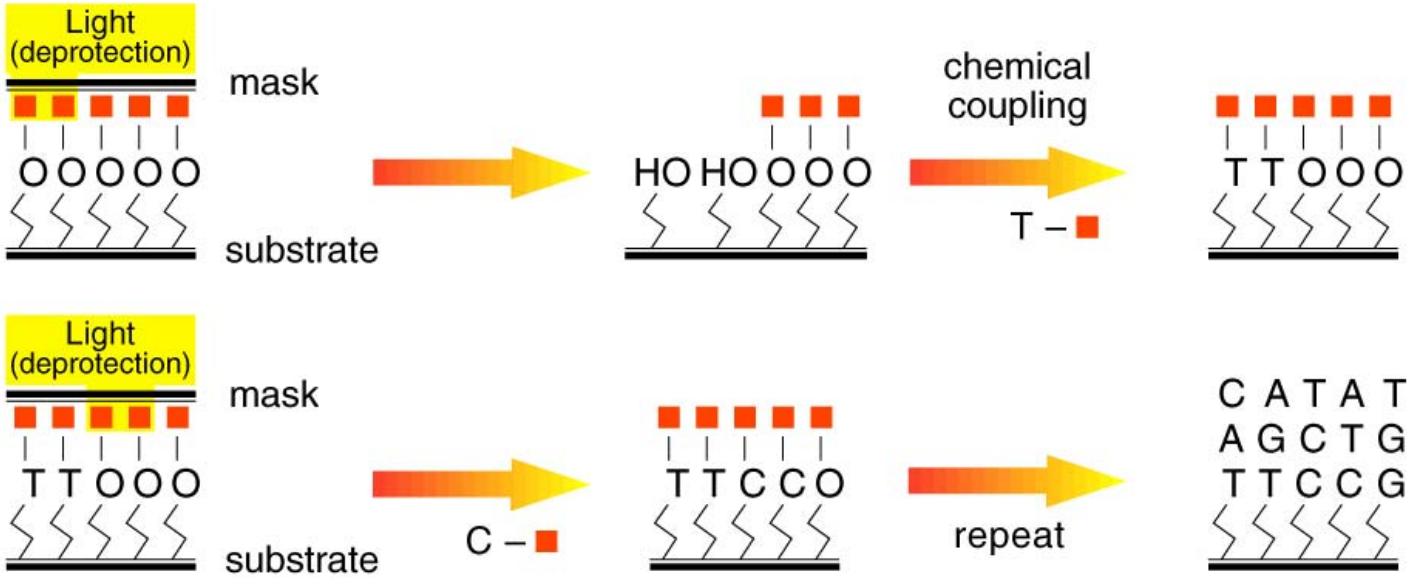
Spotting of prefabricated oligonucleotides

- Usually 50-70 nt longmers

Contact or non-contact printing

Photolithograpic *in situ* synthesis of high density oligonucleotide arrays



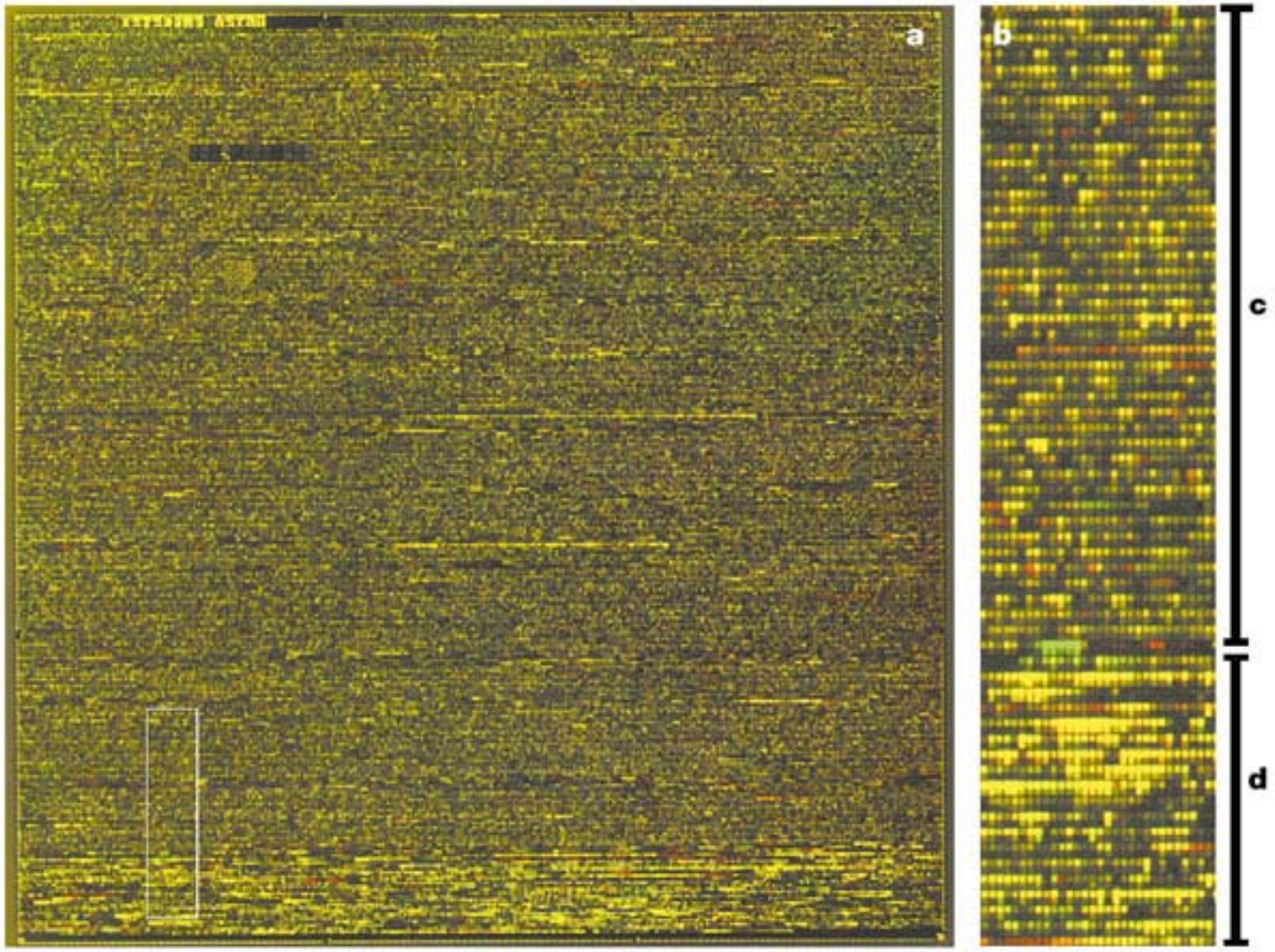


4 x n cycles

n = length of oligonucleotides, n ≤ 20

Affymetrix GeneChip

Capacity of >400.000 oligos / cm^2



Spotted probes on chip

cDNA

buy clones

plasmid preparation

create clone libraries

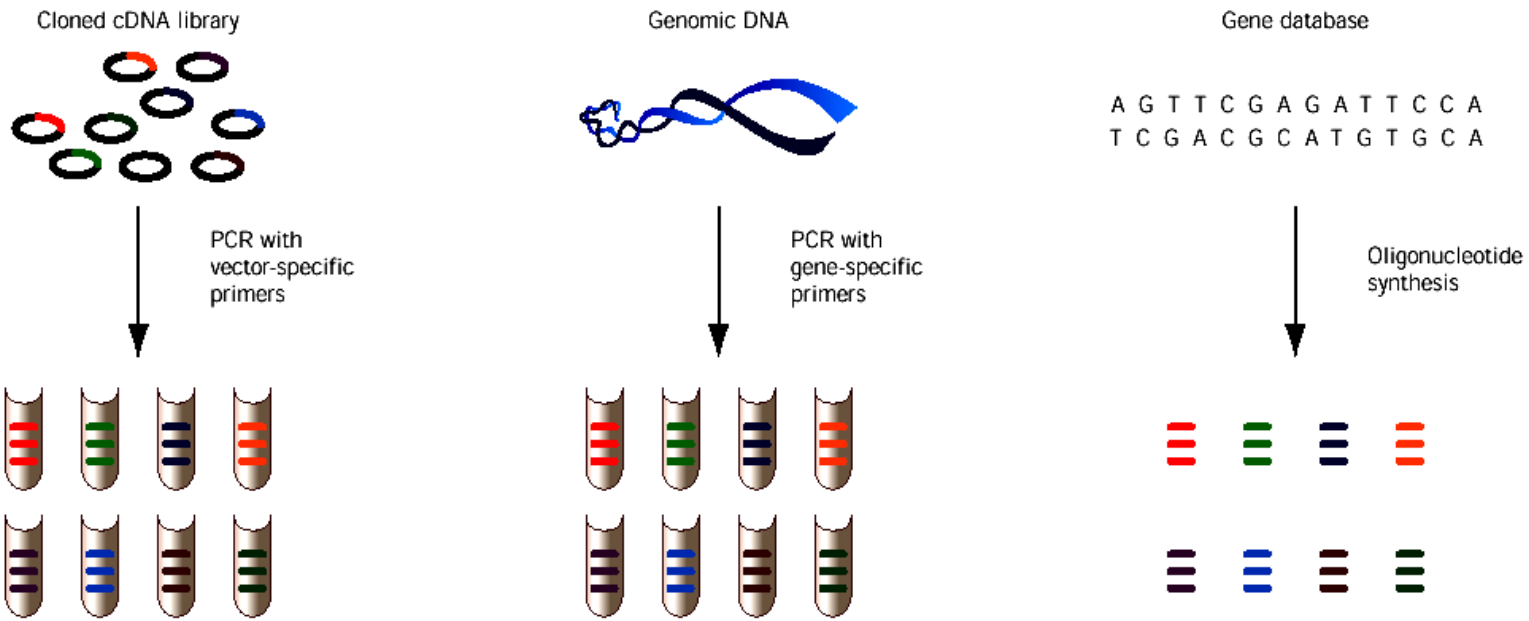
plasmid preparation

sequence

create unigene collection

PCR
purify

Oligonucleotides
(50-70 nt)
(design)
buy



Example of Oligo Sets for Microarray Applications

operon
molecules for life

Mammalian
Homo sapiens (human)
Mus musculus (mouse)
Rattus norvegicus (rat)
Bos taurus (bovine)
Canis familiaris (dog)
Sus scrofa (pig)

Animal
Caenorhabditis elegans (nematode)
Danio rerio (zebrafish)
Drosophila melanogaster (fruit fly)
Gallus gallus (chicken)
Xenopus tropicalis (Pepid frog)

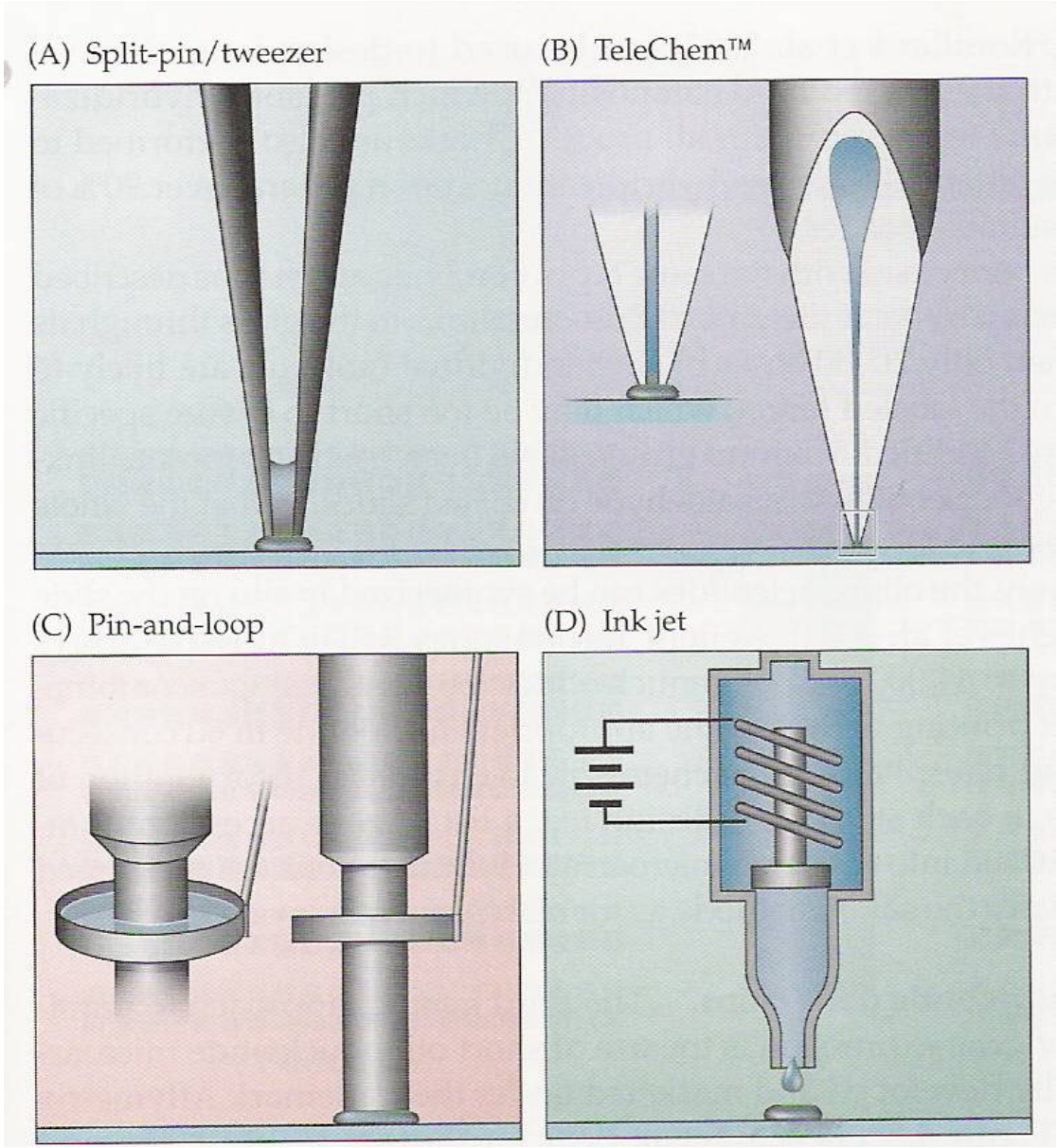
Plant
Arabidopsis thaliana
Lycopersicon esculentum (tomato)
Zea mays (maize)
Medicago truncatula (barrel medic)
Oryza sativa (rice)
Prunus persica (peach)
Pisum sativum (garden pea)
Vitis vinifera (grape)

Bacterial
Bacillus Genus (anthrax)
Bordetella pertussis
Campylobacter coli
Campylobacter jejuni
Chlamydophila pneumoniae
Escherichia coli
Haemophilus influenzae
Helicobacter pylori
Lactobacillus sakei
Listeria monocytogenes
Mycobacterium tuberculosis
Rhizobium leguminosarum
Salmonella Genus
Sinorhizobium meliloti
Streptococcus mitis
Streptococcus pneumoniae

Yeast and Fungi
Candida albicans
Magnaporthe grisea
Saccharomyces cerevisiae

Additional Genomes
Plasmodium falciparum (malaria)
Rhodopirellula baltica (pirellula)

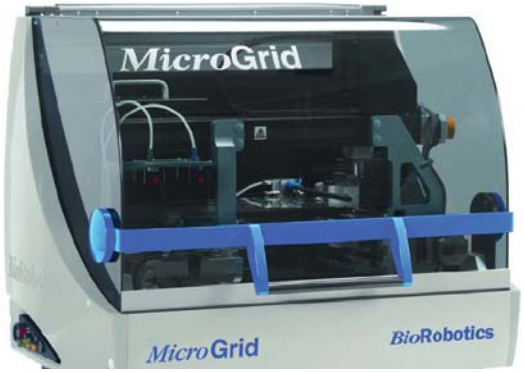
Different microarray printers



~400 pl is deposited

Robotics for microarray production

GeSIM Nanoplotter 2, 8-pins, non-contact

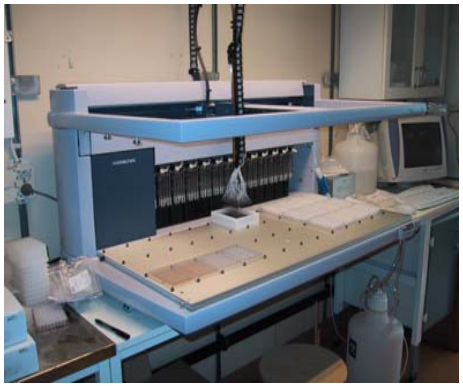


BioRobotics MicroGrid II, 64-pins

Genetix Qarray2, 48-pins

Genetix Qarray, 24-pins

GMS 417, 4-pins

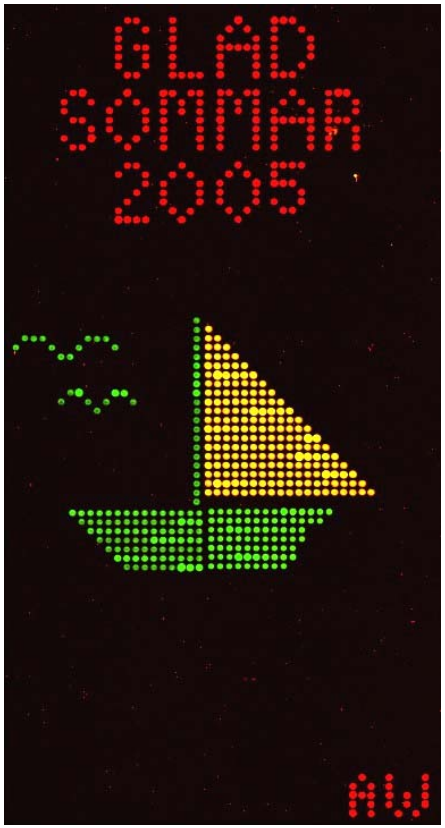


Robotic workstations

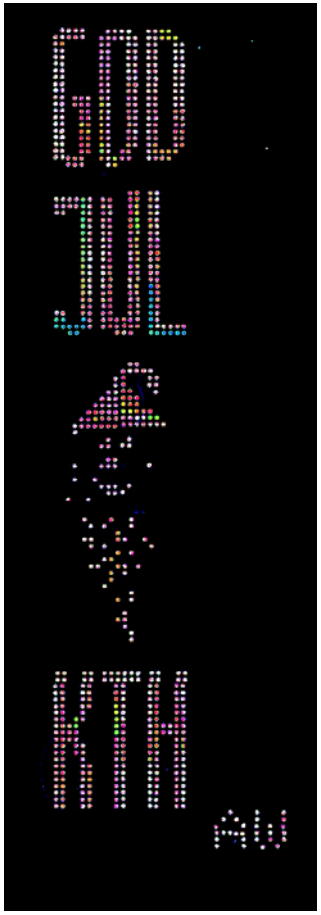
Scanner

KTH Microarray Center – spotting projects

DNA microarray production for internal projects and academic collaborations



Human	cDNA oligo genomic (CGH)	46 k 34 k 11 k
Poplar	cDNA	25 k
Mouse	cDNA (stem cells) oligo	20 k 35 k
Rat	oligo	27 k
Chicken	cDNA	14 k
Arabidopsis	gene spec. PCR gene spec. PCR (promoters) oligo	22 k 10 k 29 k
Drosophila	oligo	15 k
C. elegans	oligo	20 k
Yeast	oligo	6,4 k
E. coli	oligo	6,0 k
Plasmodium falciparum	oligo	7,1 k
Sulfolobus solfataricus	gene spec. PCR	3,0 k
Sulfolobus acidocaldarius	gene spec. PCR	2,8 k
Helicobacter pylori	oligo	1,9 k
Campylobacter jejuni	oligo	1,6 k
Streptococcus pneumoniae	oligo	2,0 k
Bartonella henselae	gene spec. PCR	1,9 k
Bartonella grahamii	oligo	4,4 k



KTH SCHOOL OF BIOTECHNOLOGY STOCKHOLM SWEDEN WWW.KTHARRAY.SE

www.ktharray.se

cDNA microarray technology

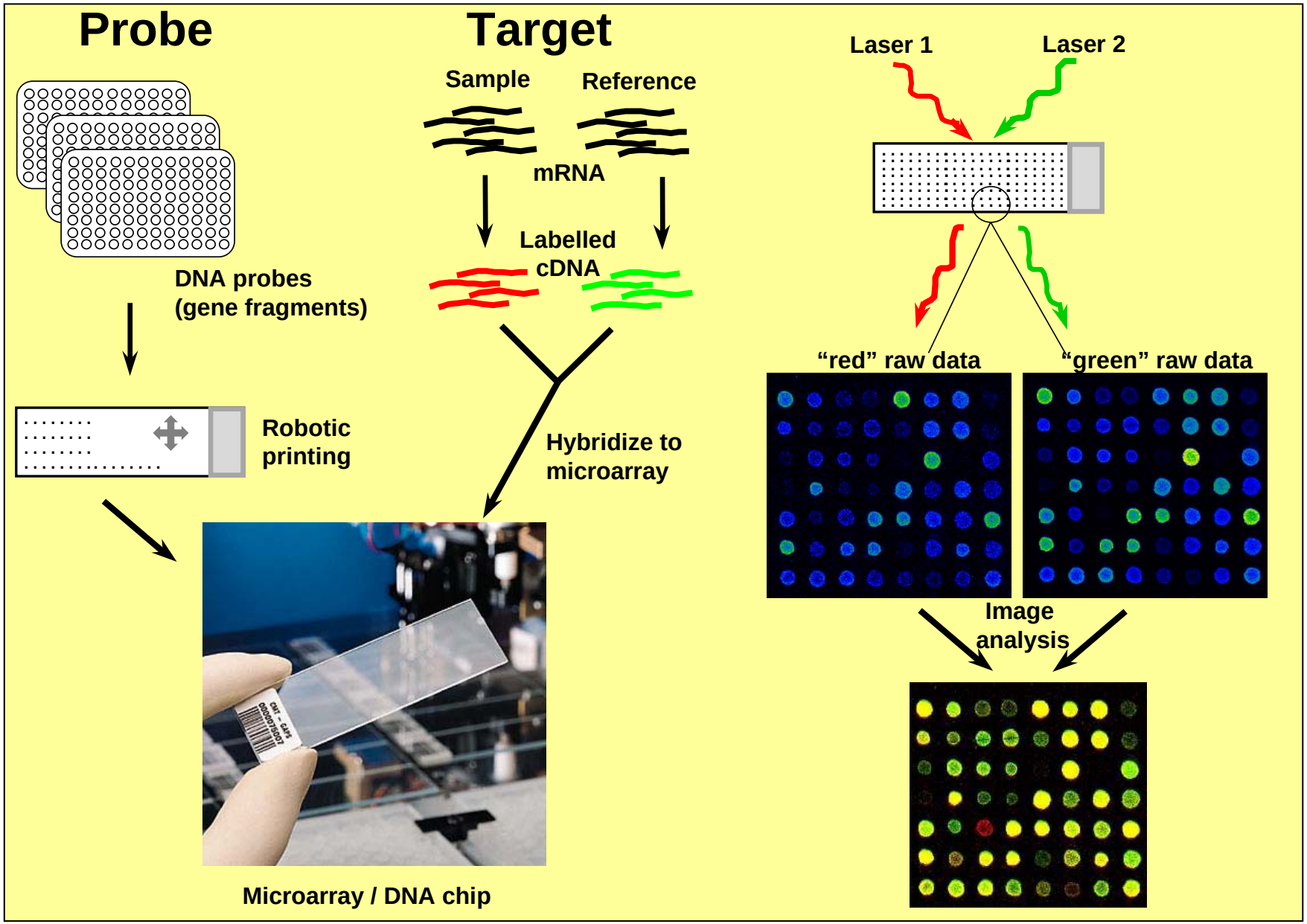
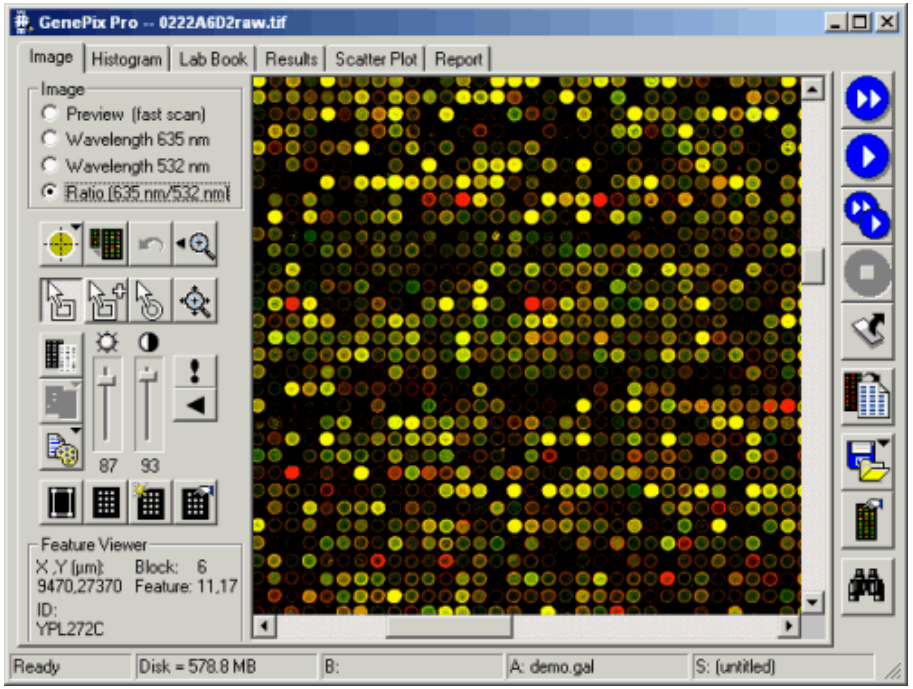
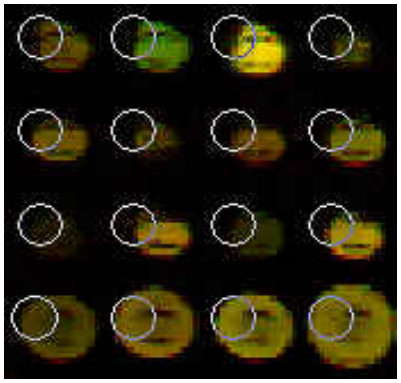


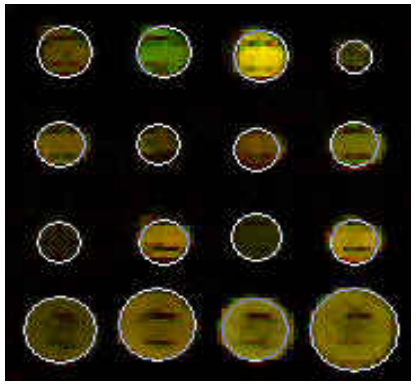
Image analysis



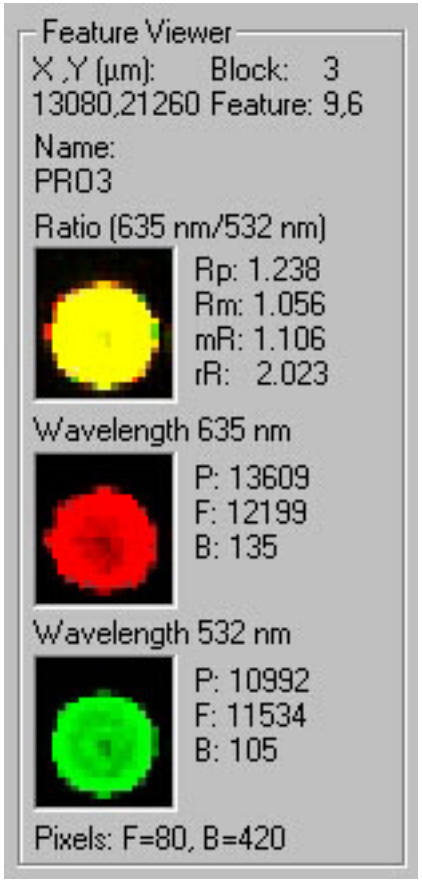
Before auto align

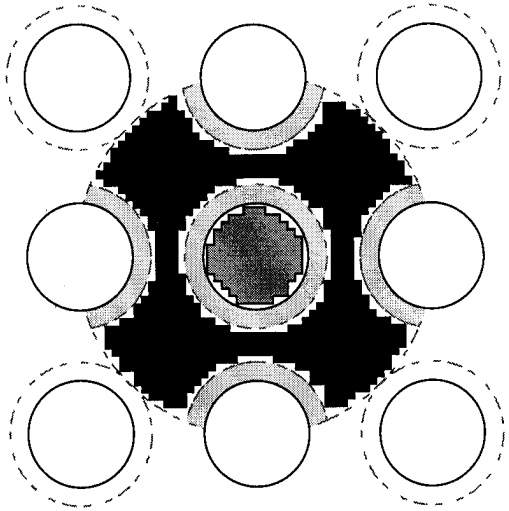


After auto align



Calculate ratios



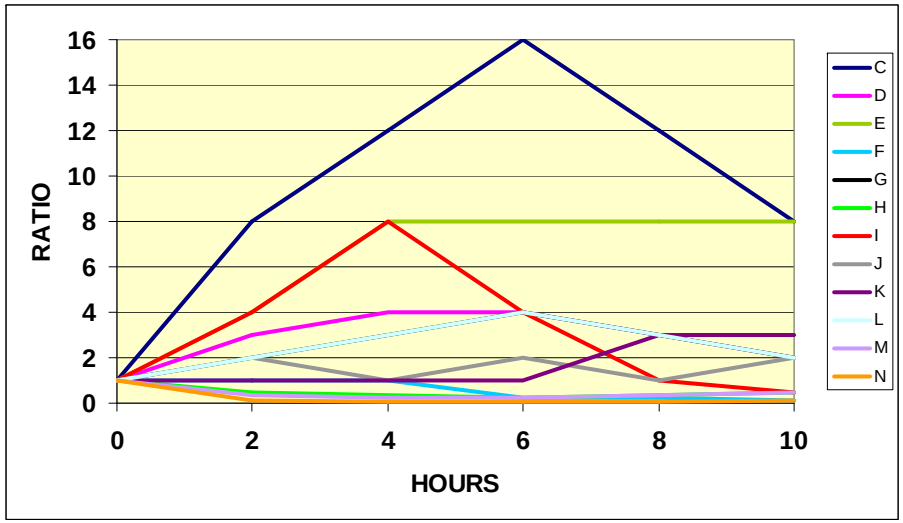


Background subtraction

Fold change vs $\log_2$ ratios

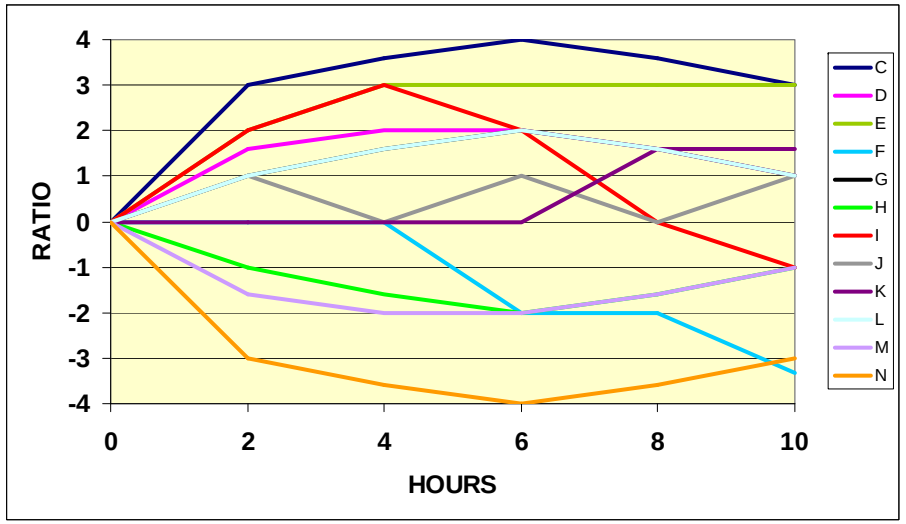
Fold change
sample intensity/ref intensity

Name	0	2	4	6	8	10
C	1	8	12	16	12	8
D	1	3	4	4	3	2
E	1	4	8	8	8	8
F	1	1	1	0,25	0,25	0,1
G	1	2	3	4	3	2
H	1	0,5	0,33	0,25	0,33	0,5
I	1	4	8	4	1	0,5
J	1	2	1	2	1	2
K	1	1	1	1	3	3
L	1	2	3	4	3	2
M	1	0,33	0,25	0,25	0,33	0,5
N	1	0,125	0,0833	0,0625	0,0833	0,125



$\log_2$ ratios
$\log_2$ (sample intensity/ref intensity)

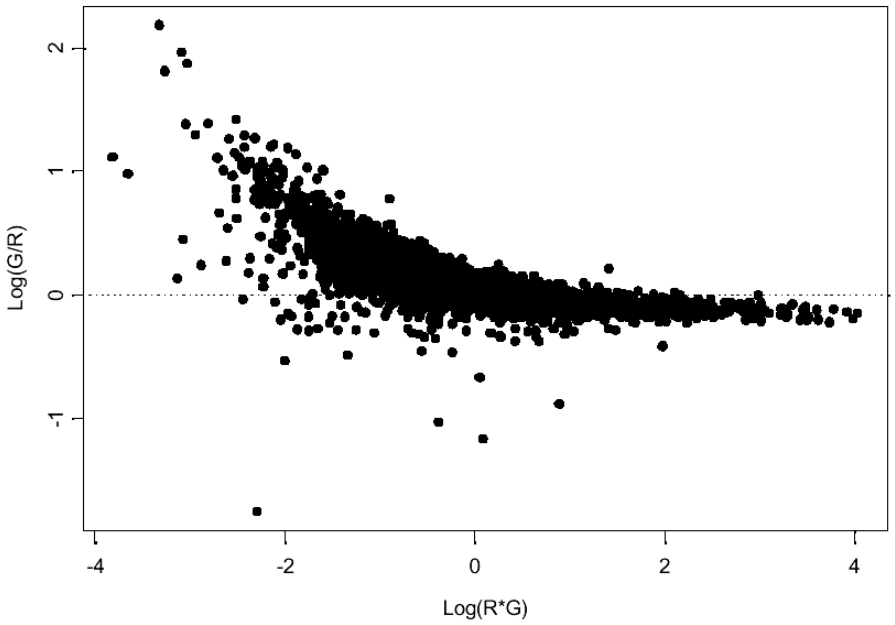
Name	0	2	4	6	8	10
C	0	3	3,58	4	3,58	3
D	0	1,58	2	2	1,58	1
E	0	2	3	3	3	3
F	0	0	0	-2	-2	-3,32
G	0	1	1,58	2	1,58	1
H	0	-1	-1,60	-2	-1,60	-1
I	0	2	3	2	0	-1
J	0	1	0	1	0	1
K	0	0	0	0	1,58	1,58
L	0	1	1,58	2	1,58	1
M	0	-1,60	-2	-2	-1,60	-1
N	0	-3	-3,59	-4	-3,59	-3



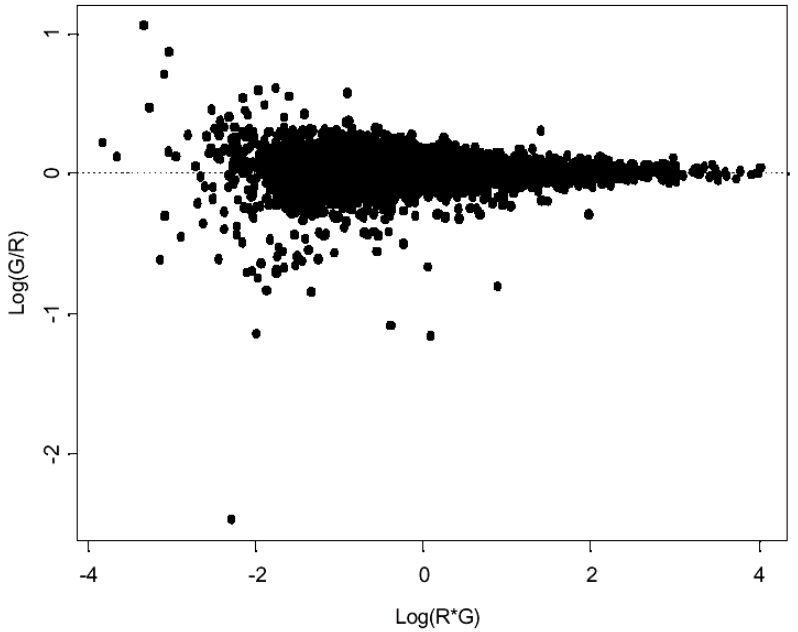
Normalization

X: intensities
Y: ratio

2-color data before applying Lowess

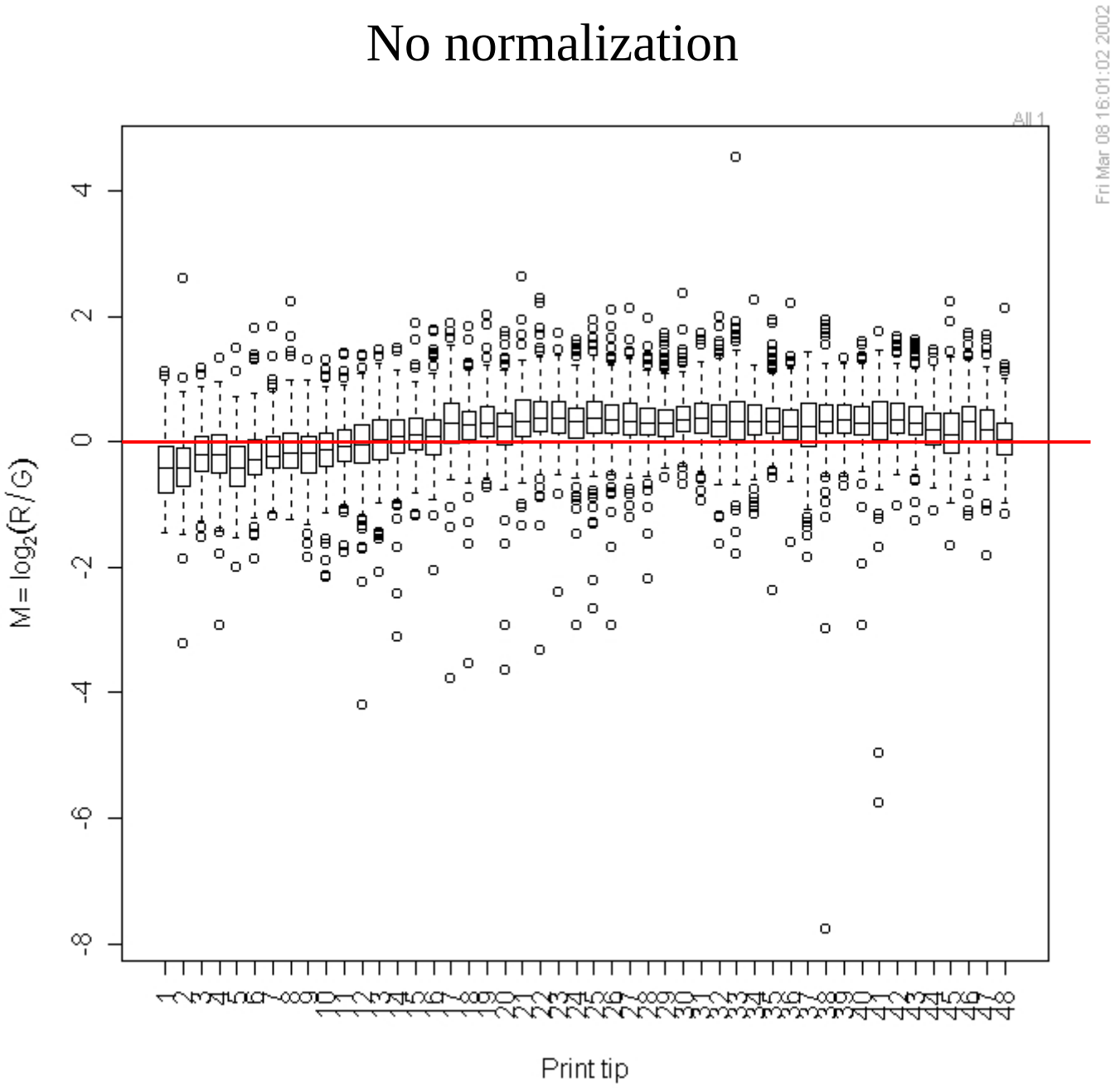


2-color data after applying Lowess



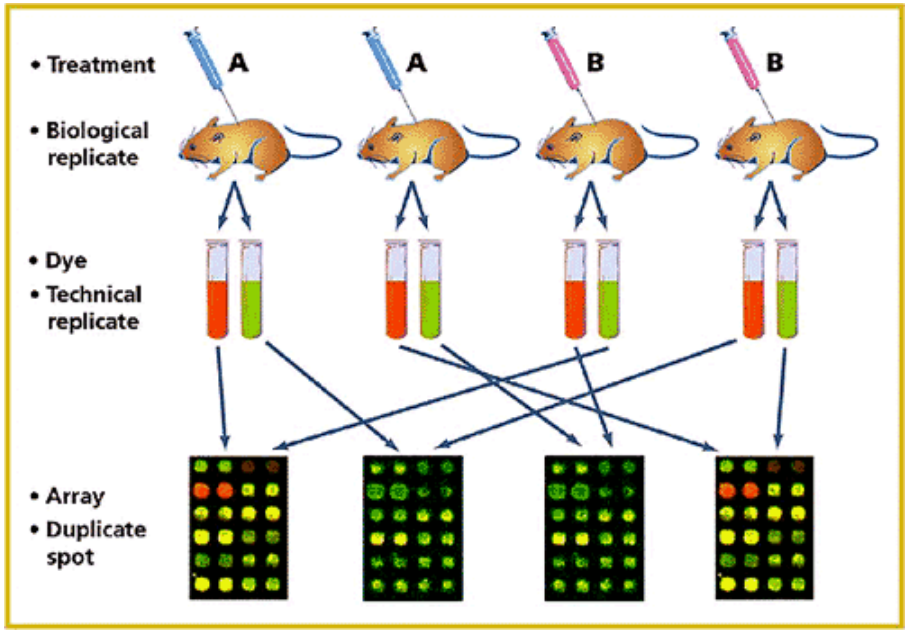
objective of normalization is to scale the spot intensity measurement so that the ratio of intensities is approximately equal to the ratios of gene expression

No normalization



Experimental design

How many replicates are needed and how should they be used?



Different levels of replicates

Individual animals/cell-lines/etc

Repeated sampling of the same animal

Repeated RNA extraction

Repeated amplification of the mRNA

Repeated labelling&hybridisation

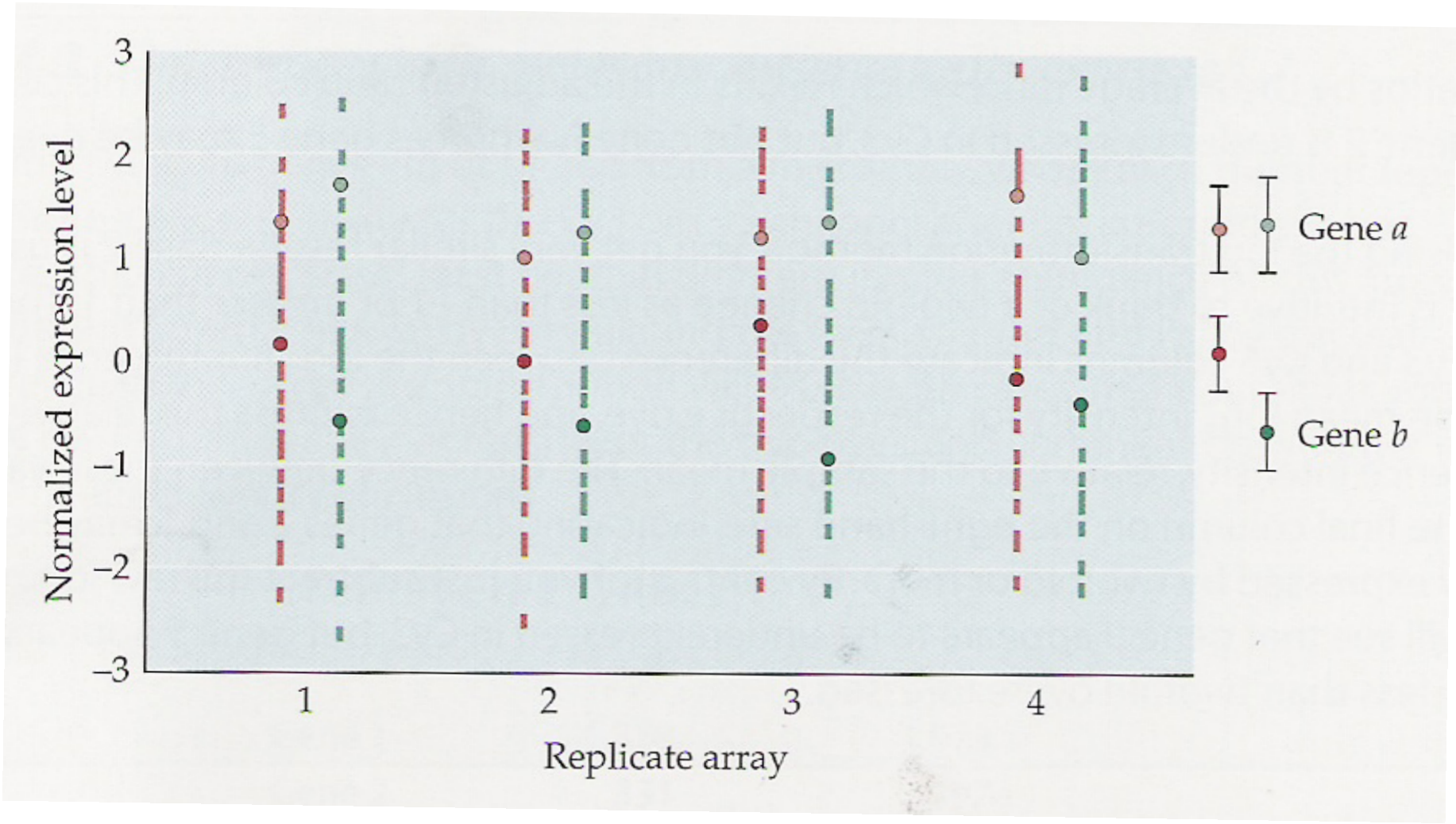
Repeated measurement of each gene on each slide (duplicate spots)

<u>Controls...</u>

...biological variation

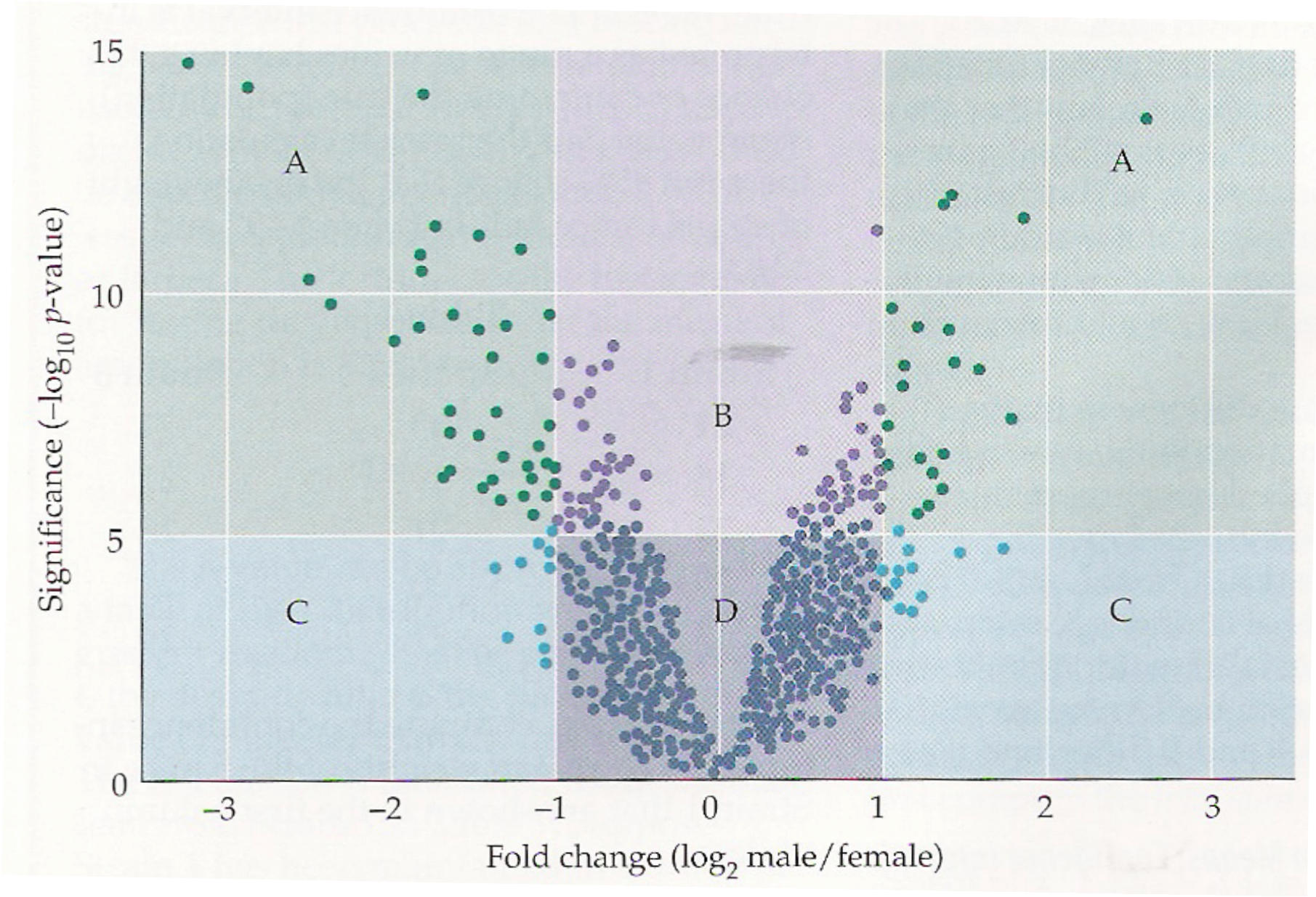
...technical variation

Analysis of variance (ANOVA) for gene expression data



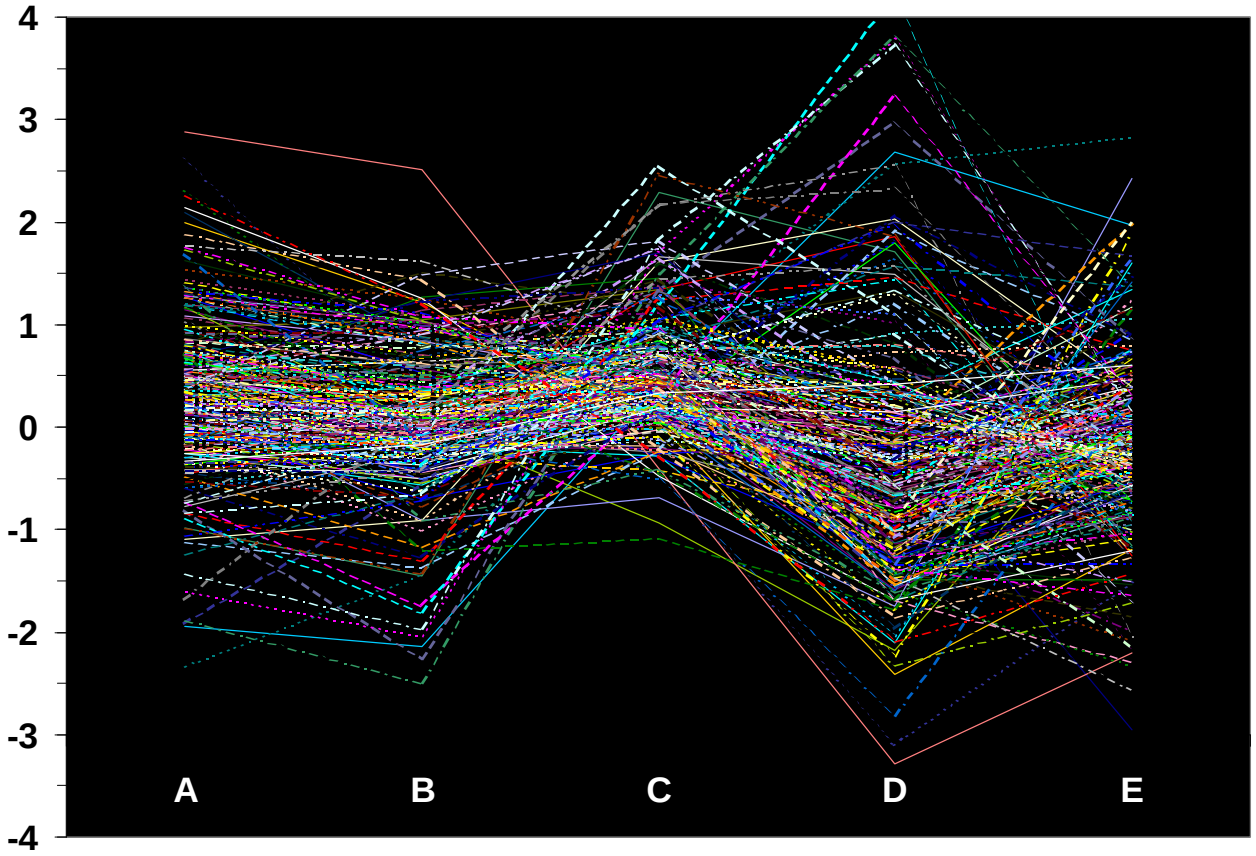
Significance testing

Don’t look only on fold change, take also variance into account



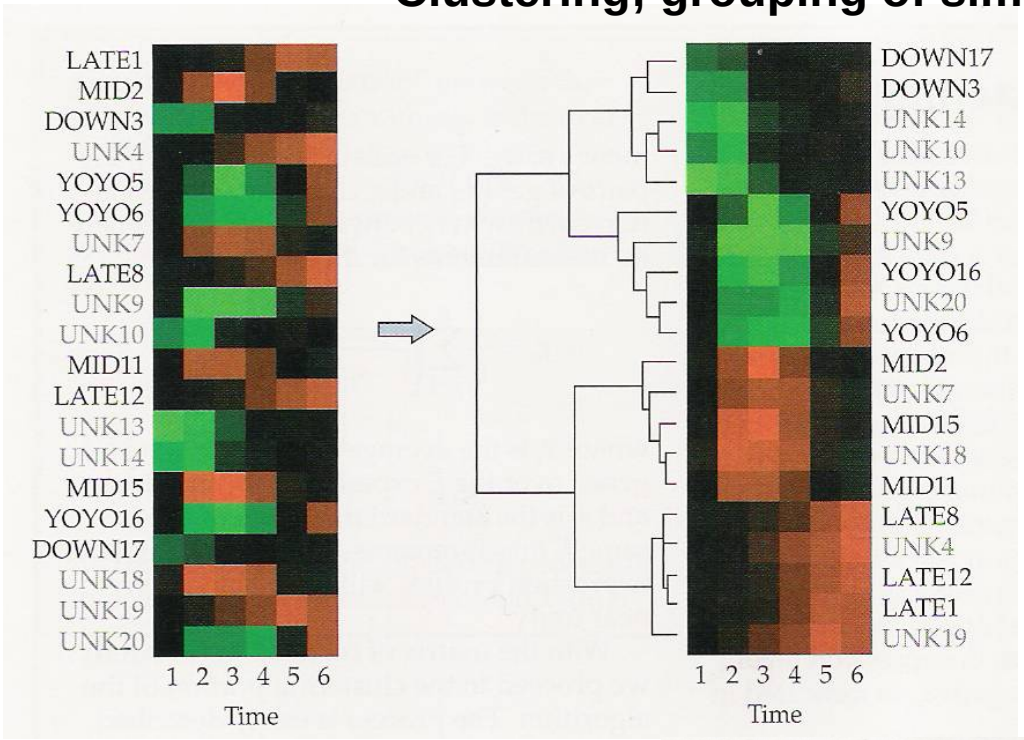
Microarray data mining

How to visualize data

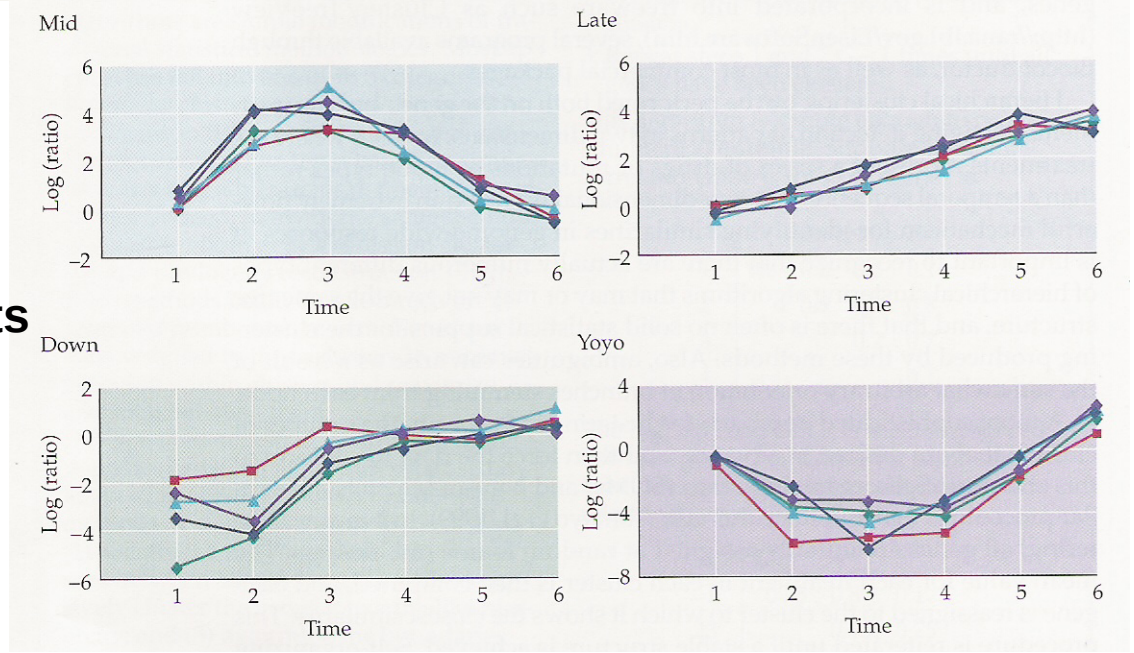


- **Hierarchical clustering**
- First, find the two genes, with the most similar behaviour.
- Join these together into a cluster. Next, join the next two most similar objects (gene or cluster), forming a new cluster.
- Continue until there is only one object left, a single cluster containing all genes.

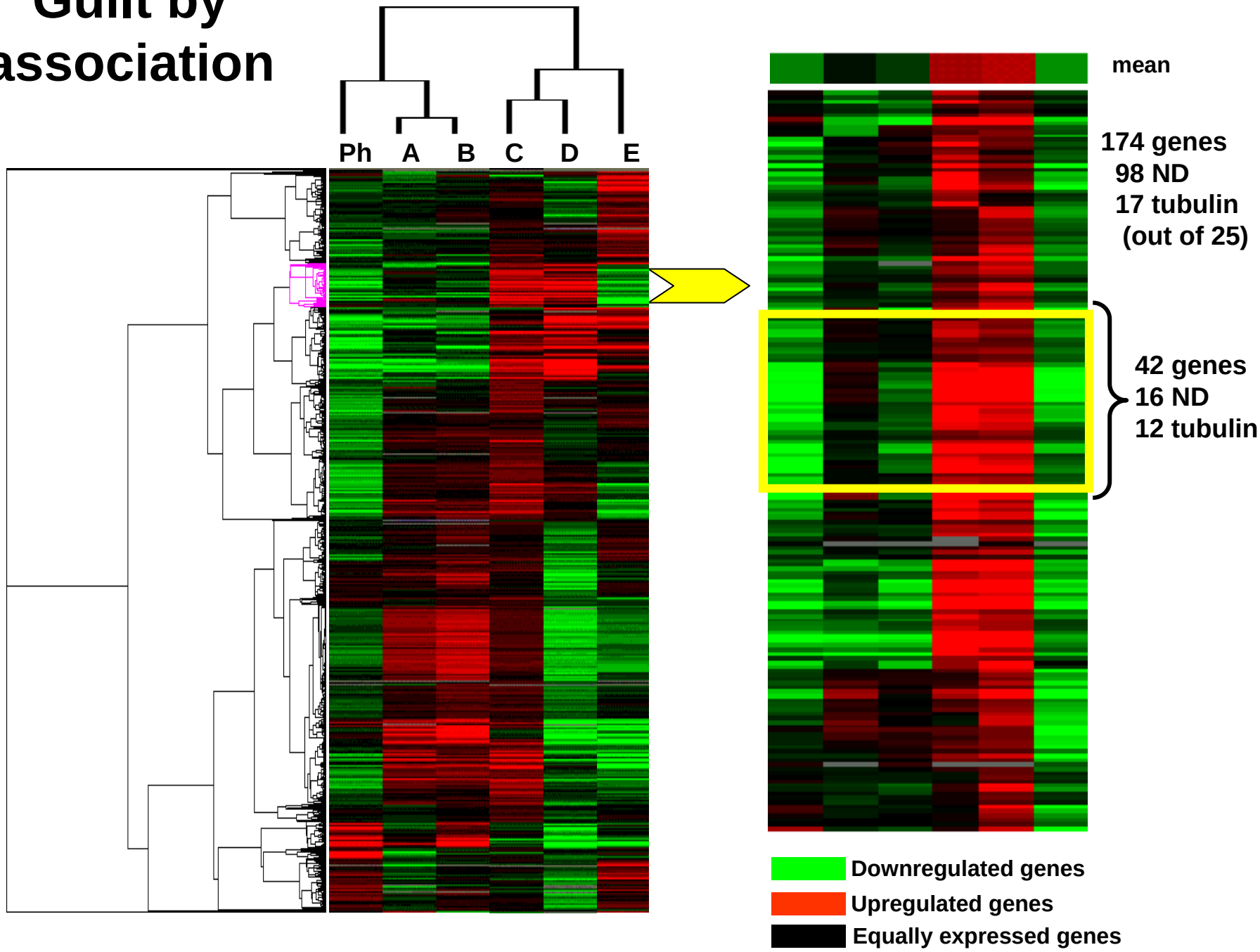
Clustering; grouping of similar expresion patterns



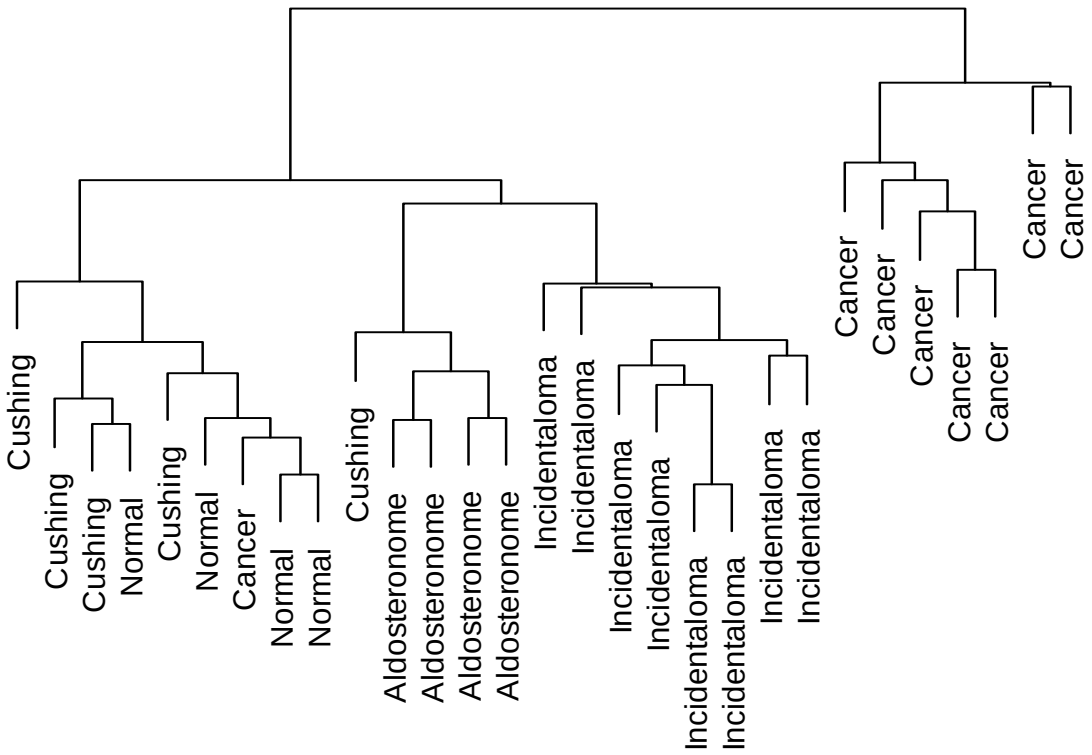
Profile plots

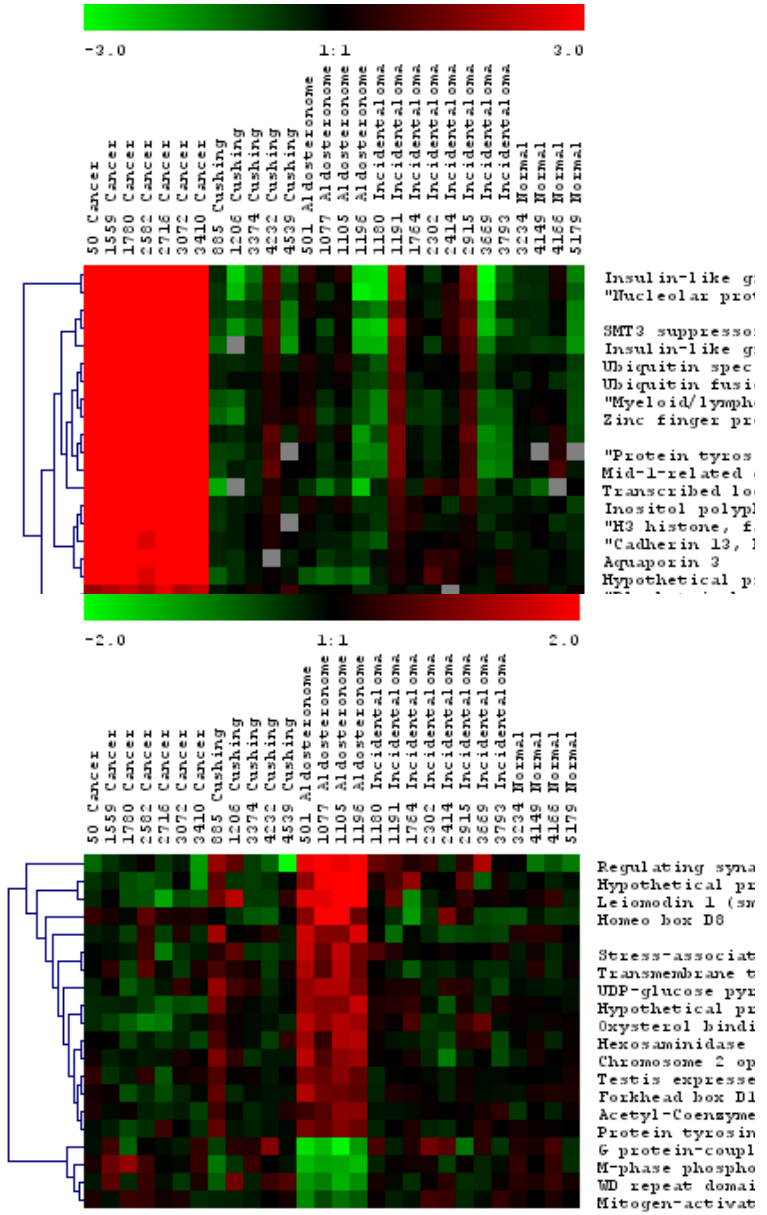


Guilt by association



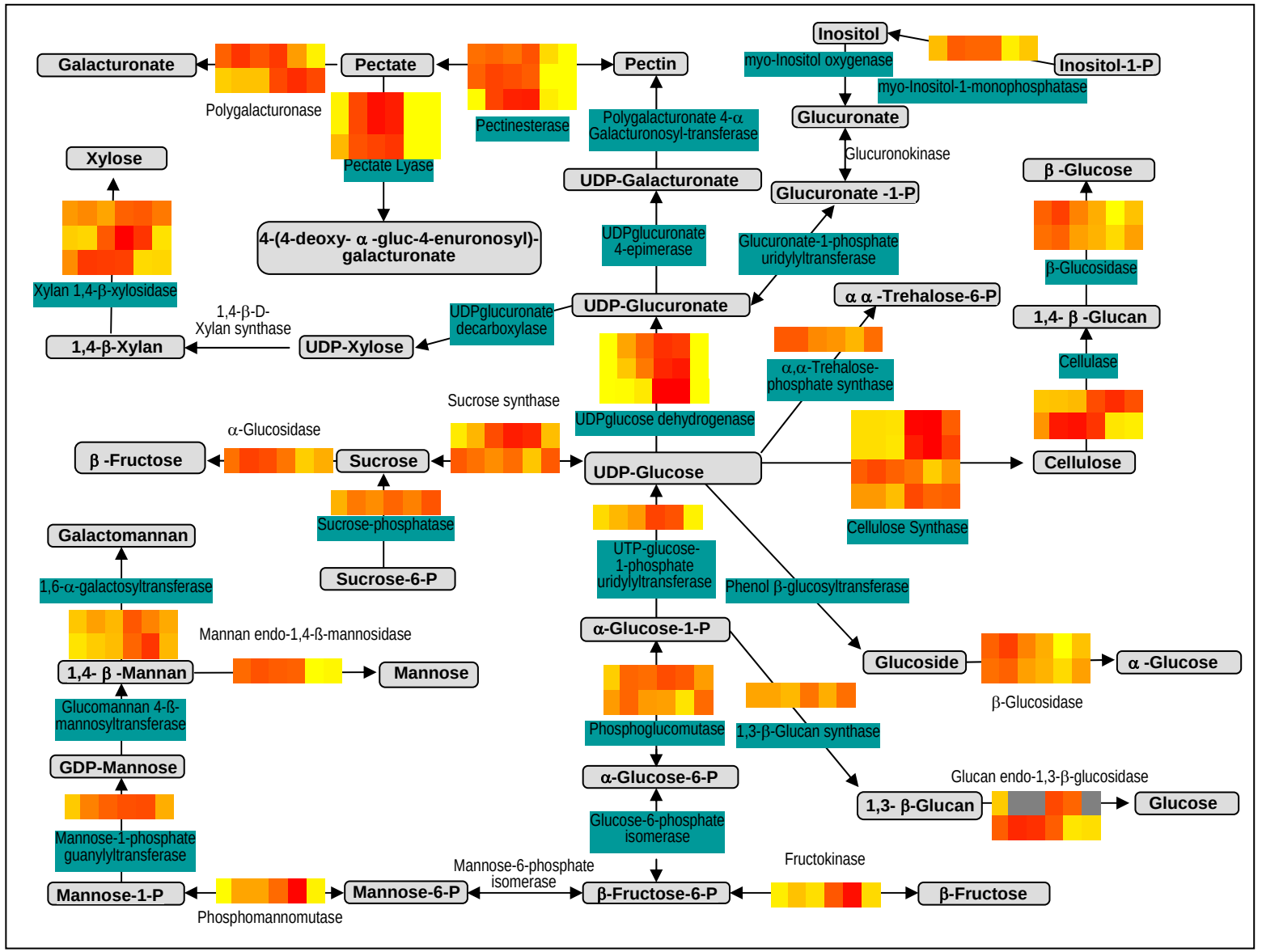
Expression Profiling of Adrenocortical Tumors suggests a Molecular Signature of Malignancy





Differential Gene Expression in Wood-formation

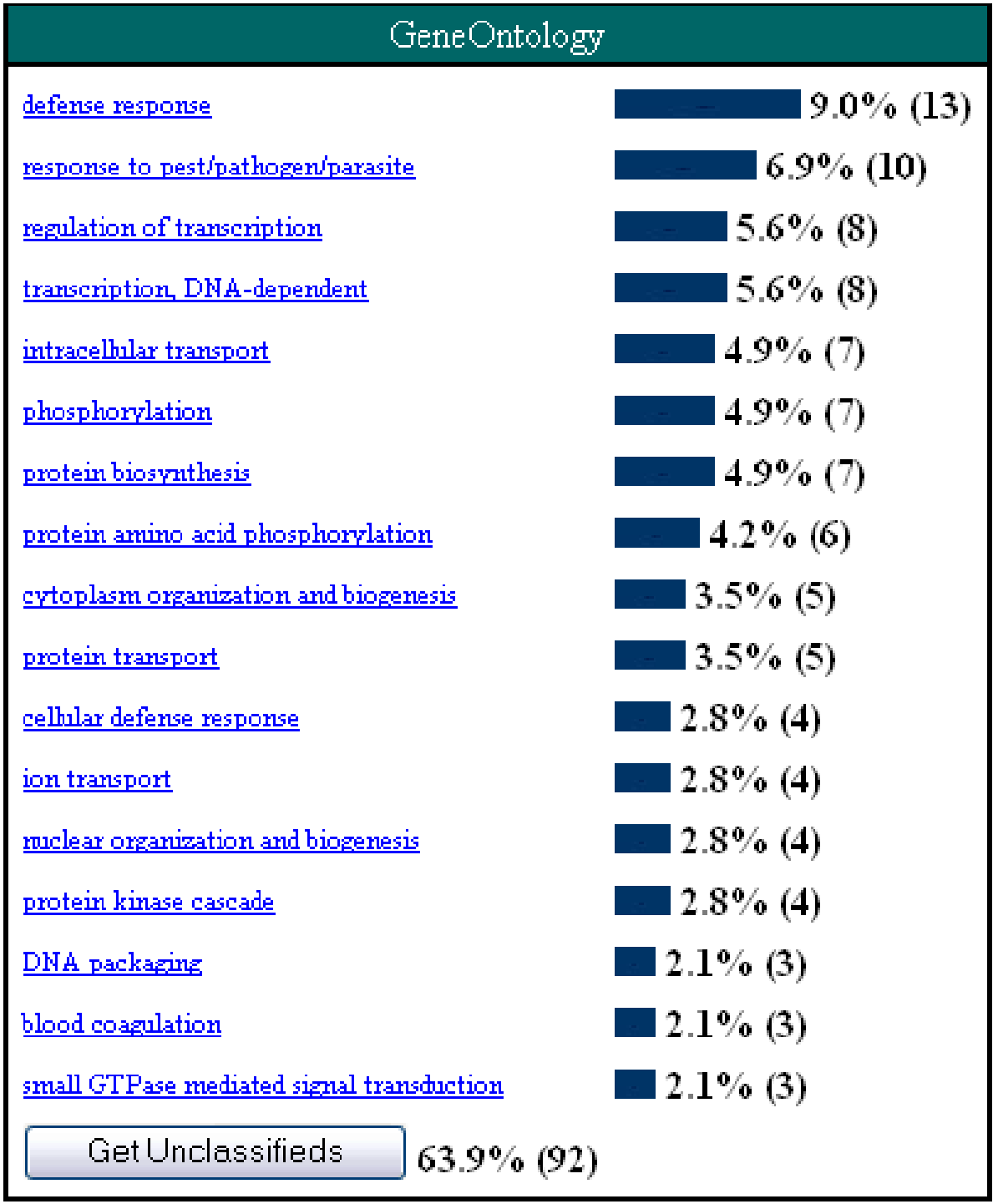
Expression profiling in methabolic pathways over the tissue sections



UDP-Glucose Pathway

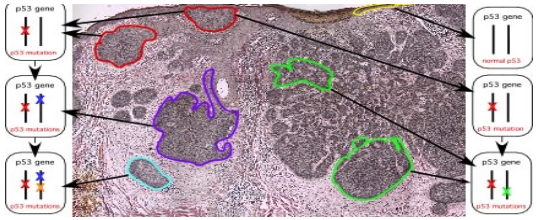
Hertzberg *et al* (2001) Proc Natl Acad Sci USA

DE genes based
on foldchange and B top

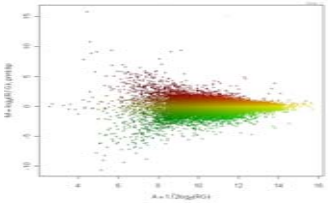


Biology

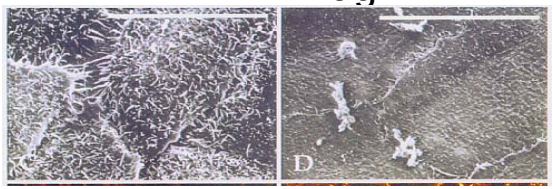
Cancer



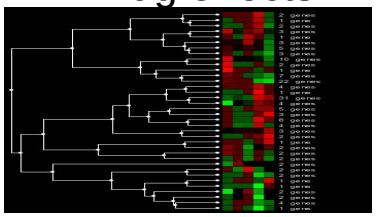
Molecular signature of cancer



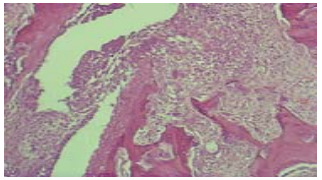
Drug

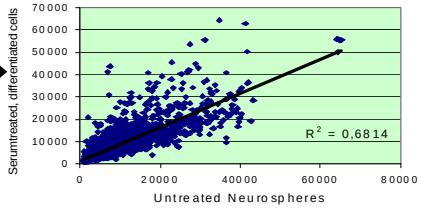


Drug effects

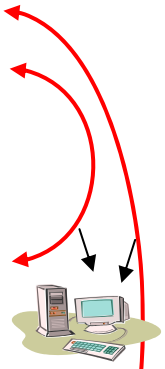


Inflammation

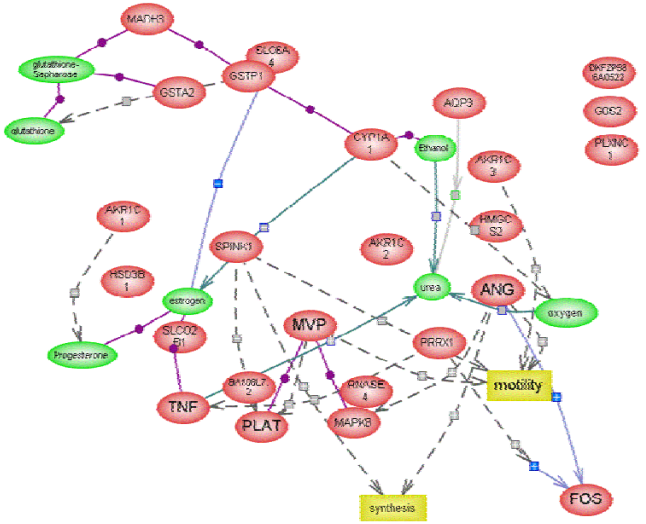




Molecular signature of inflammatory sites

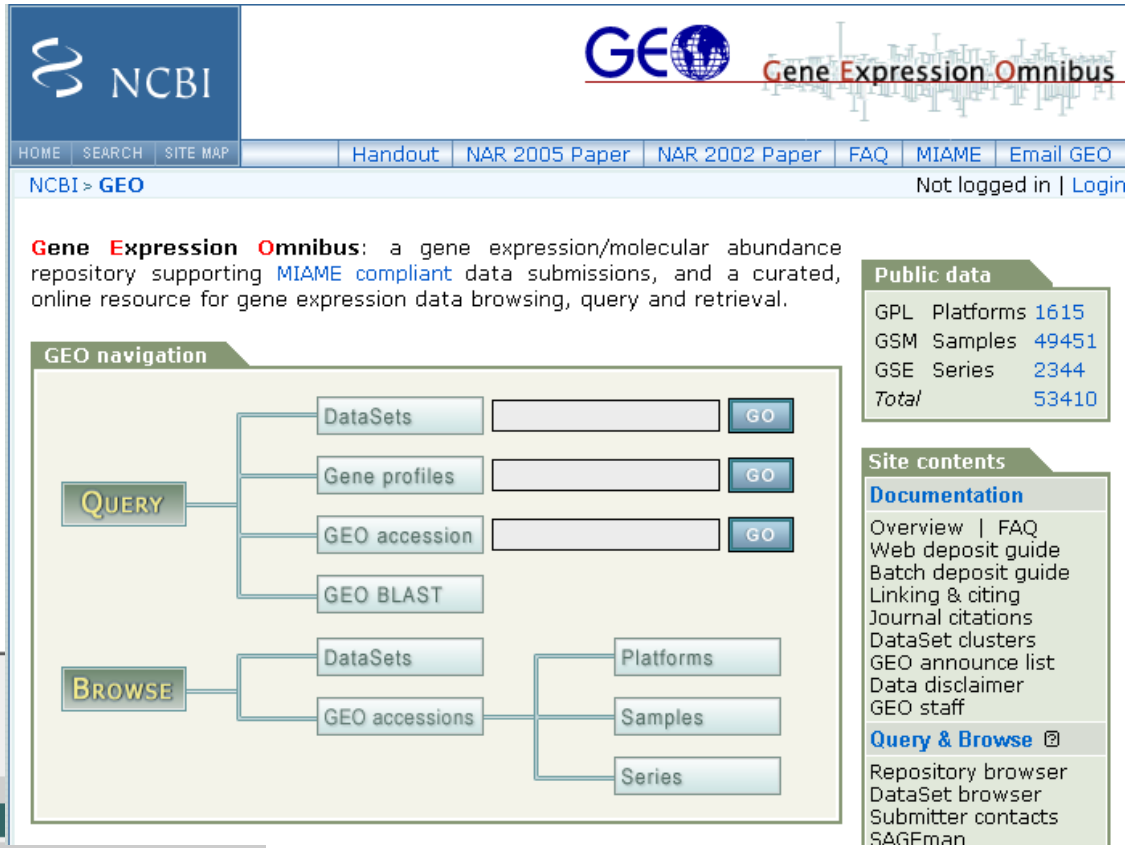


Other sources of information .... (metabolomic, proteomics data)

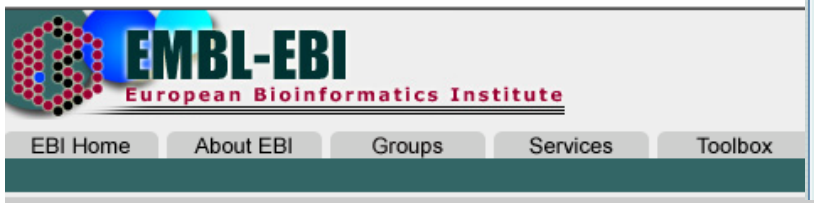


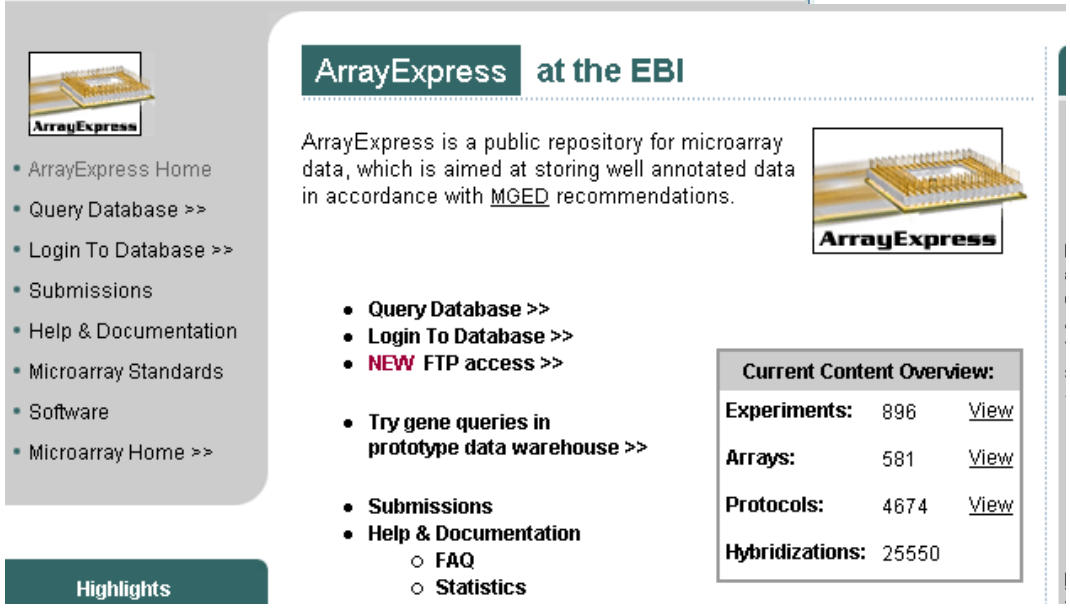
Public databases for microarray data

USA

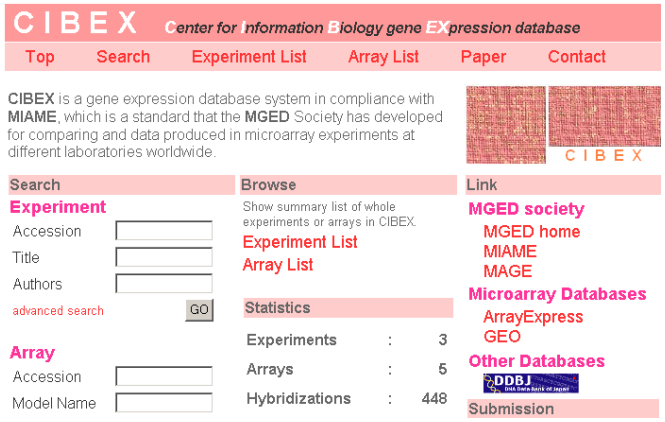


Europe





Japan



Open source data analysis software

The R Project for Statistical Computing
www.r-project.org

BIOCONDUCTOR
open source software for bioinformatics

BioConductor is an open source and open development software project for the analysis and comprehension of genomic data.

www.bioconductor.org

KTH-package for microarray data analysis
www.ktharray.se

www.tigr.org/software

TM4: A package of Open Source software programs for Microarray analysis

TIGR Microarray Data Analysis System (**MIDAS**) is a microarray data quality filtering and normalization tool that allows raw experimental data to be processed through various data normalizations, filters, and transformations via a user-designed analysis pipeline. Currently implemented normalization and data analysis algorithms include total-intensity normalization, Lowess (Locfit) normalization, flip-dye consistency checking, replicates analysis, intensity-dependent z-score filtering (slice analysis), etc. MIDAS is implemented by Java language and thus a platform-independent application. It requires JDK v1.3 or higher. Refer to the included manual for details.

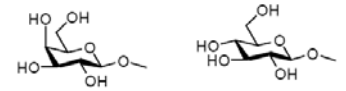
MADAM

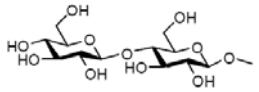
Microarray experiments produce large amounts of data for even the simplest of experiments. In order to analyze data from many experiments that data must be stored in an accessible form, such as in a database. **MADAM** (MicroArray DAta Manager) is a java-based application designed to load and retrieve microarray data to and from a database (also supplied with the software). MADAM provides data entry forms, data report forms and additional applications necessary to maintain microarray data for further analysis. Madam requires JRE 1.3.1.

TIGR MultiExperiment Viewer (MEV) is a Java application designed to allow the analysis of microarray data to identify patterns of gene expression and differentially expressed genes. Numerous normalization, clustering and distance algorithms have been implemented, along with a variety of graphical displays to best present the results. MEV was written to be flexible and expandable, and supports a variety of input and output formats. MEV requires version 1.2 or higher of Sun's JRE and J3D package.

TIGR **Spotfinder** is a software tool designed for Microarray image processing using the TIFF image files generated by most microarray scanners. TIGR Spotfinder was written in C/C++ for PCs running Windows NT/2000/ME/XP.

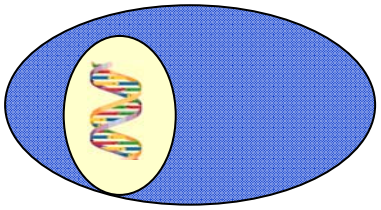
Microarray-based tool box





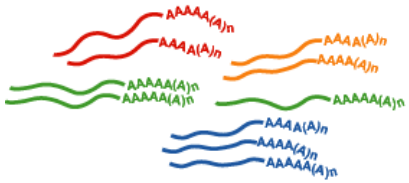
Carbohydrates

DNA



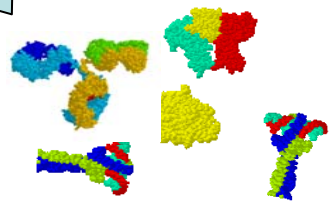
Genome

RNA

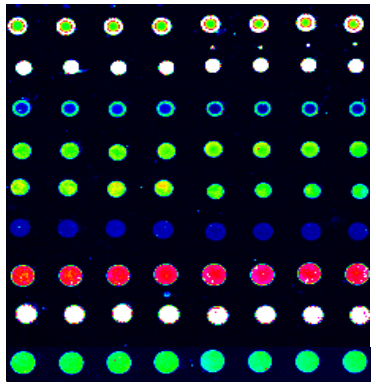


Transcriptome

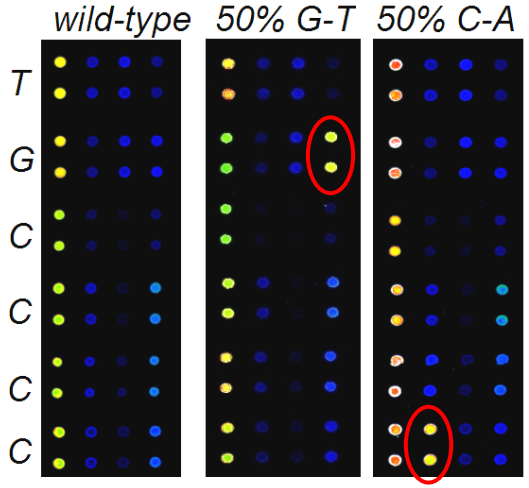
Protein



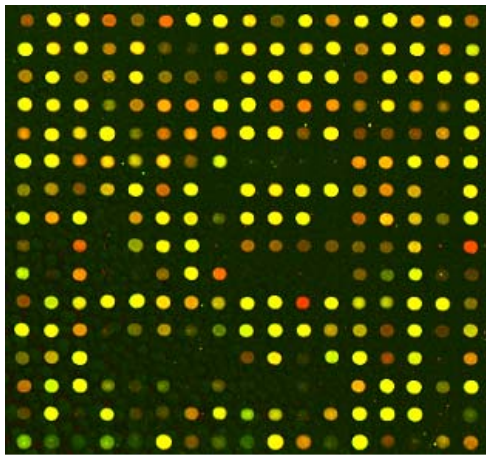
Proteome



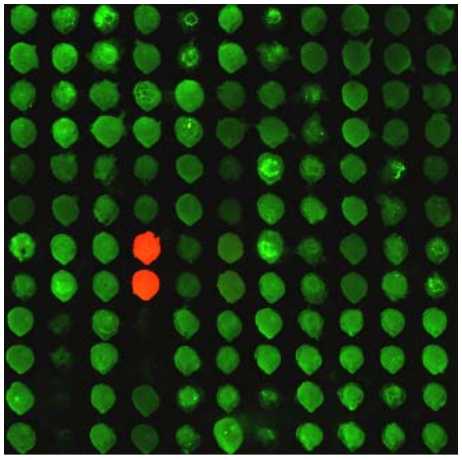
Antibody array



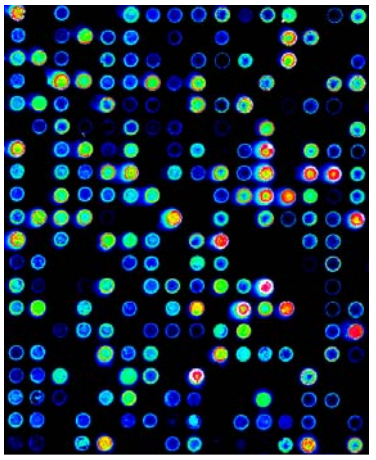
Genotyping array



cDNA/oligo array

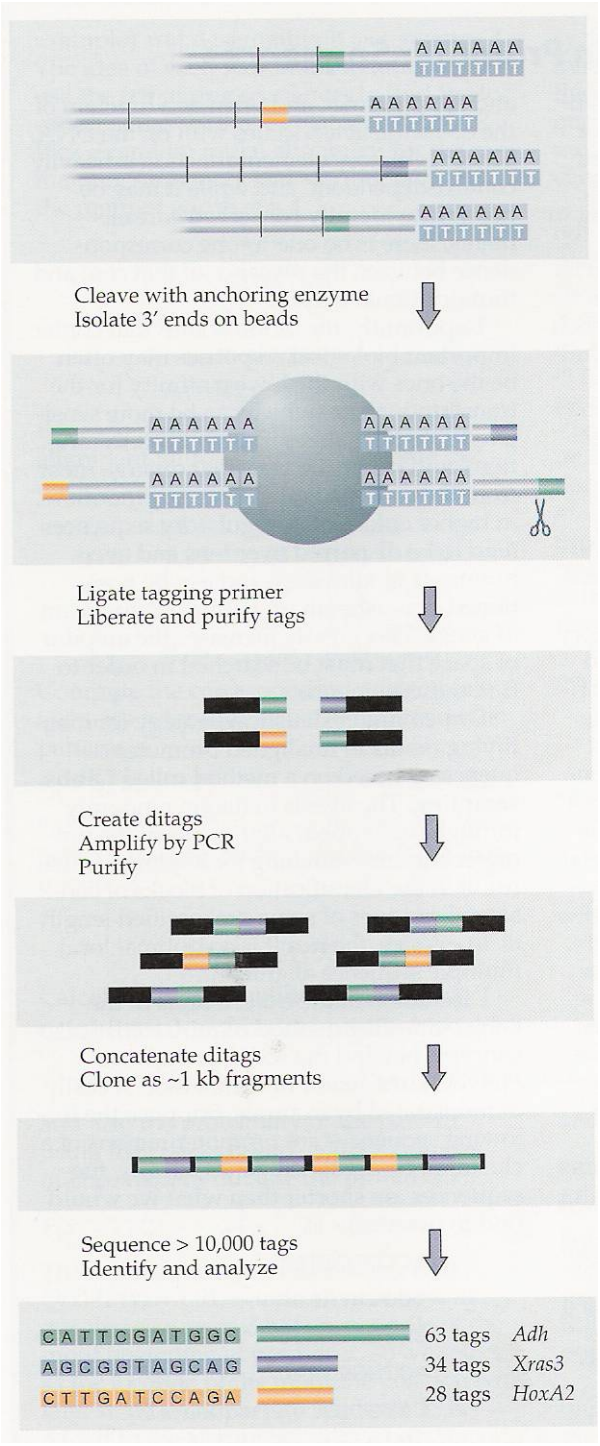


Protein array



Serum array

Principle of SAGE
Serial Analysis of Gene Expression



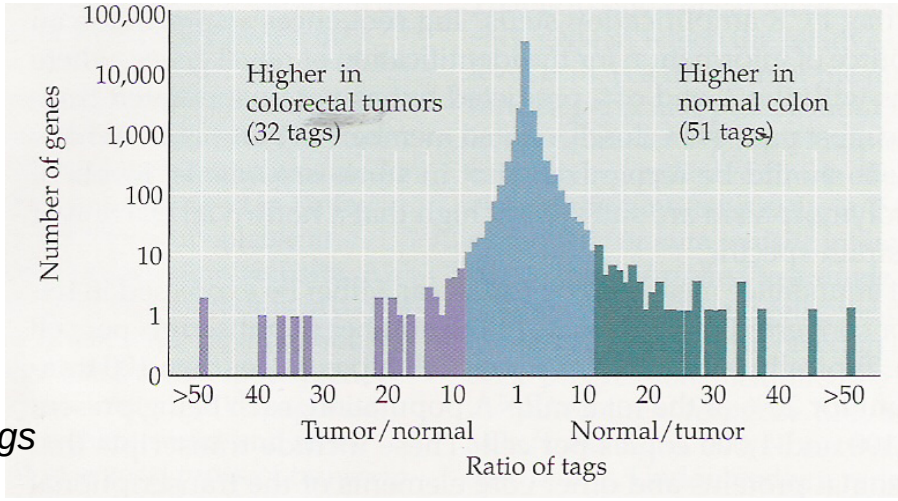
- *convert mRNA to cDNA*
- *Fragmentize*

- *capture mRNA*
- *add linker*
- *remove from bead*

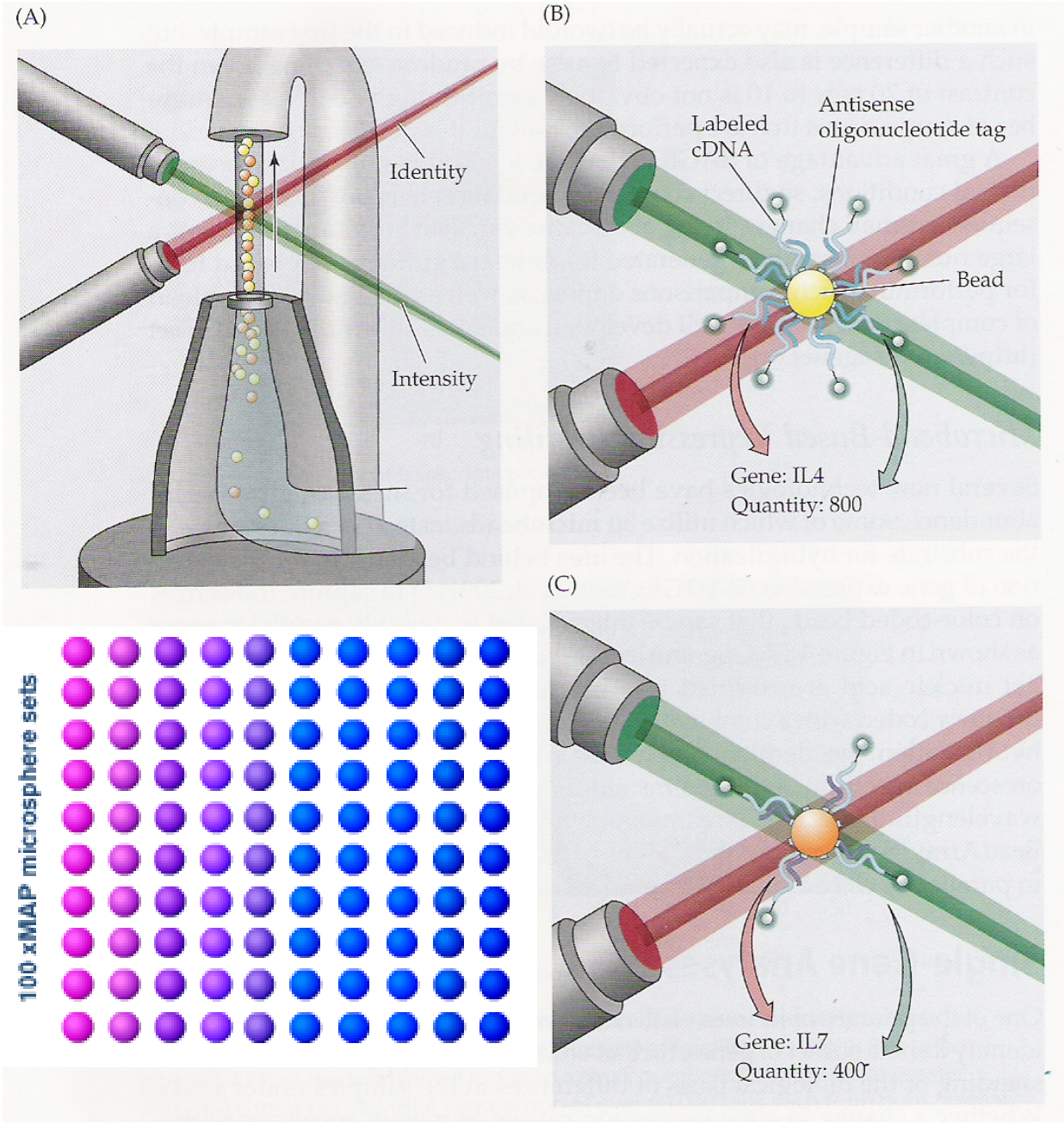
- *ligate two tags (15+15 bp)*
- *amplify with PC*

- *ligate many di-tags (1000 bp)*
- *clone into a plasmid*

- *Sequence >10.000 tags*
- *identify and count the tags*



Microbead-based expression profiling



THE POWER INSIDE Luminex | xMAP TECHNOLOGY Because you want to know more

BeadArray™ Technology Overview

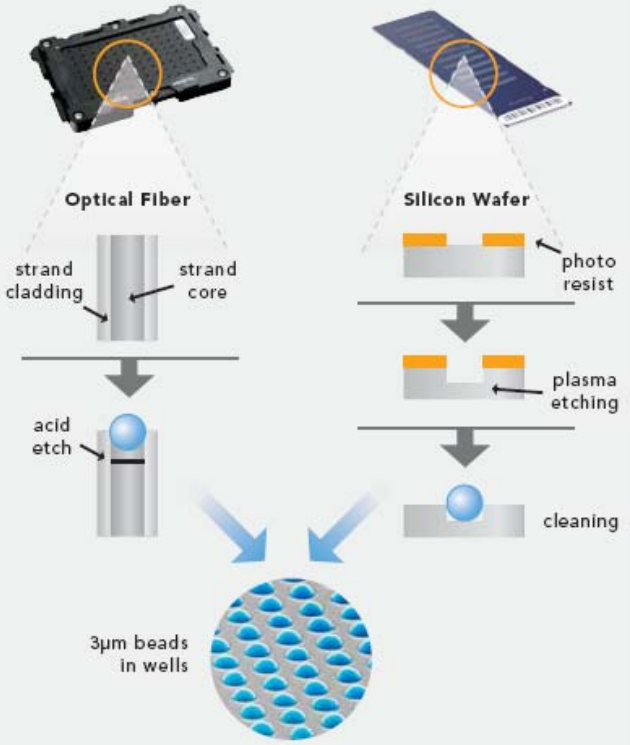
Overview

Illumina's BeadArray Technology is based on 3-micron silica beads that self assemble in microwells on either of two substrates: fiber optic bundles or planar silica slides. When randomly assembled on one of these two substrates, the beads have a uniform spacing of ~5.7 microns. Each bead is covered with hundreds of thousands of copies of a specific oligonucleotide that act as the capture sequences in one of Illumina's assays (*Figure 1*).

Figure 1

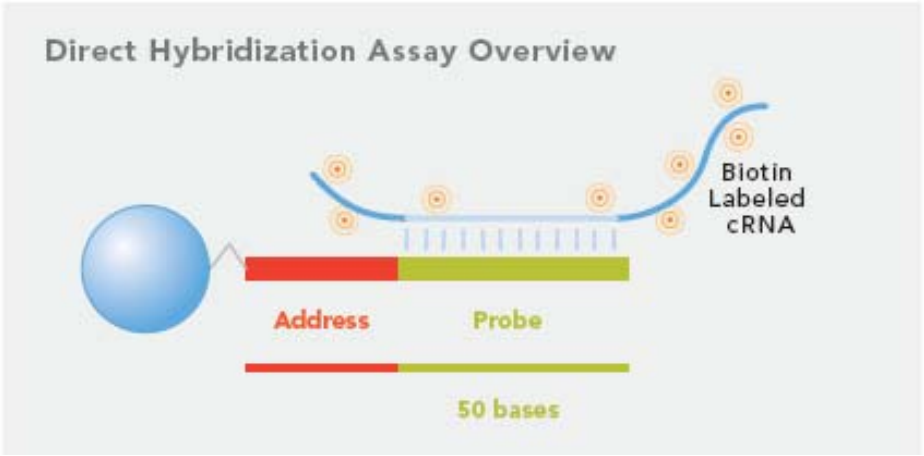
Illumina BeadArray Technology

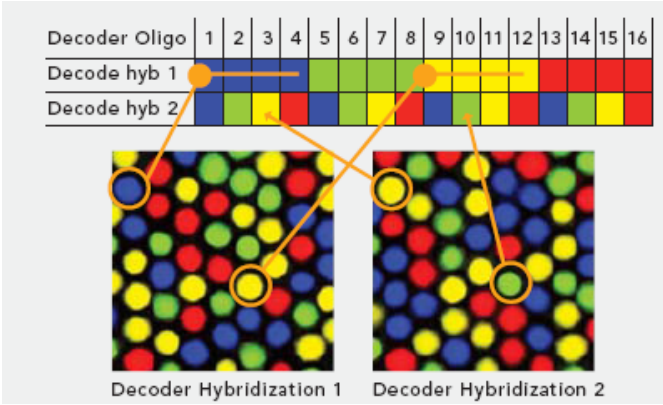
Figure 2



Illumina multi-sample array formats

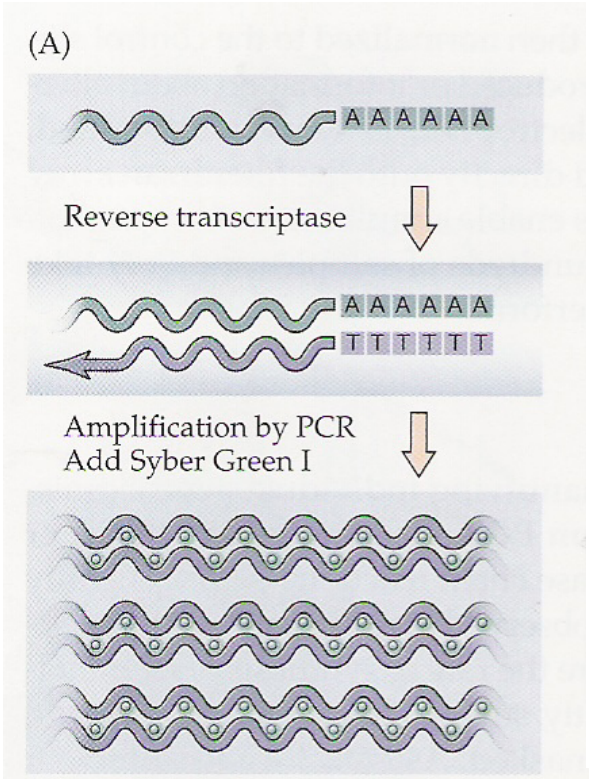
Direct Hybridization Assay

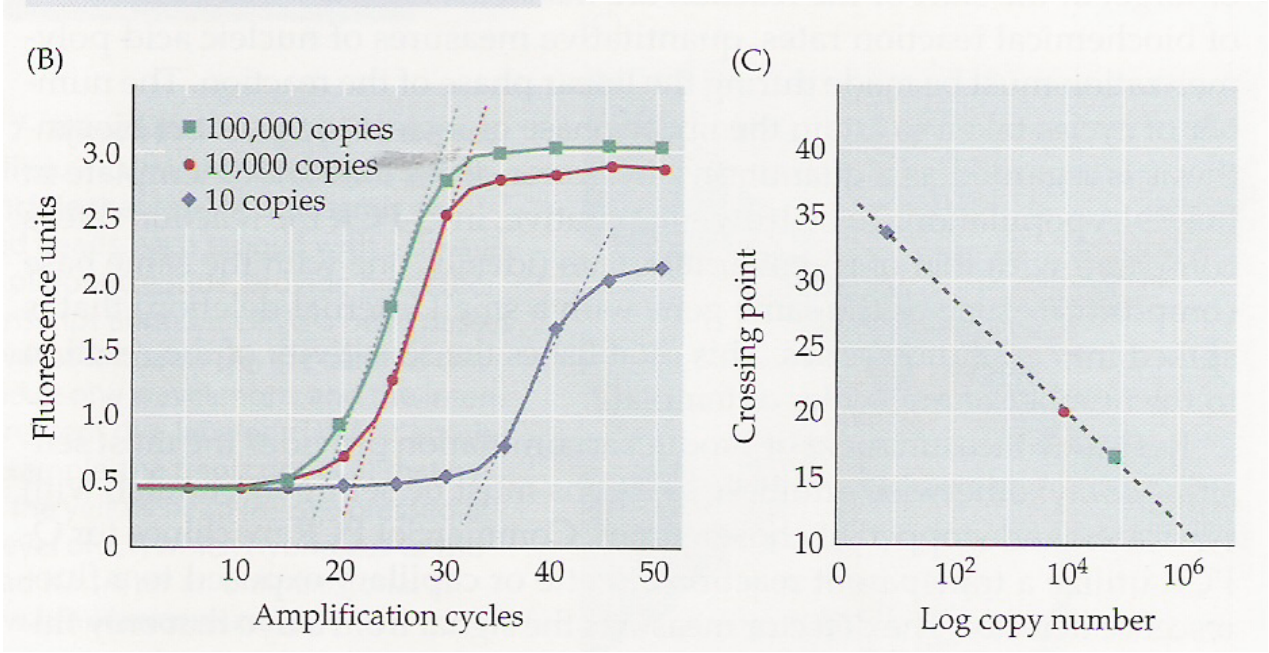




qPCR
quantitative reverse-transcription PCR (Q-RT-PCR)
real time PCR (RT-PCR)

Often used to verify microarray data, gene-by gene





Summary

- Transcriptome
- Microarrays
  - analysis of whole coding genome
  - fabrication and use
  - spotted vs in-situ synhtesized
  - transcript profiling
- Microarray data analysis
  - experimental design
  - image analysis
  - normalization
  - significance test of differentially expressed genes
  - clustring and visualization
  - data mining
  - open source software
- Other methods
  - Bead-based arrays
  - SAGE
  - qPCR


KTH
VETENSKAP
OCH KONST