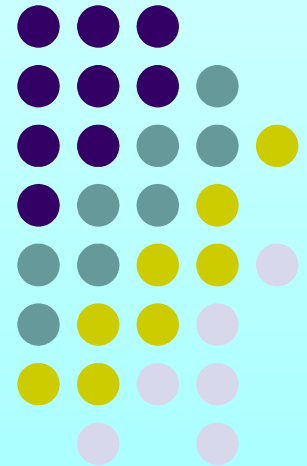


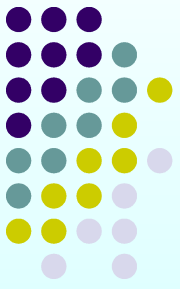
Profili di espressione genica nel tumore della mammella

Michele De Bortoli

Dipartimento di Scienze oncologiche
Università degli Studi di Torino,
Istituto per la Ricerca e la Cura del Cancro, Candiolo (TO)



Breast cancer is clinically heterogeneous



1st issue:

Prognosis ranges from complete cure after surgery to early relapse and death.

Main risk factor: invasion to lymph nodes

Still, one-third of N- patients will relapse

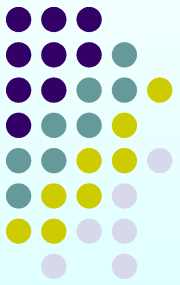
As a result, there is increasing tendency to adjuvant treatments in N- and consequently, 2/3 of them will be overtreated.

2nd issue:

A significant group of breast cancers can be successfully treated with endocrine therapy.

Main predictor: steroid receptors in the tumor.

Still, >one-third ER+/PgR+ patients will not respond.



Few prognostic factors, individually not strong enough

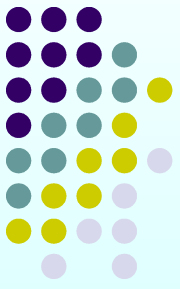
Goal:

to identify a set of “prognostic genes”, whose expression profile could be used as a marker for prognosis

Rationale:

*To seek for **combinatorial** (associative) value of many biomolecular markers (tens, hundreds), selected on the basis of **limited association** with clinical features*

ER
PgR
ERBB2/neu
MYC
P53
Cathepsin D
P21
EGFR
ERBB3
Nm23
MMP-2
E-cadherin
uPA
PSA
VEGF
Survivin
Telomerase
Serpine 1
Cyclin D1
Ki-67 antigen
Keratins



The second kind of information we expect from gene expression profiling is

*to discover **molecular lesions***

i.e. alterations to metabolic or signalling pathways that can be responsible of tumor development and that can be

target for therapeutics

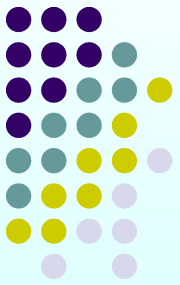
*This is accomplished by identifying clusters of genes segregating with **subgroups** of tumors and then looking for ontologies among genes, i.e. finding significant groups of genes, showing co-regulation in tumors and belonging to the same or correlated pathways.*

ANN image analysis

Gene alteration profiling (DNA) by CGH

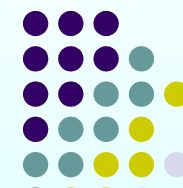
Gene expression profiling (RNA) on microarrays of RT-PCR cards

Proteomic analysis



... the question is just having a complex image,
enough to recognize the sample !

More on microarrays ([goto](#))



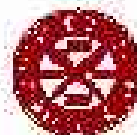
The New England Journal of Medicine

Copyright © 2002 by the Massachusetts Medical Society

VOLUME 347

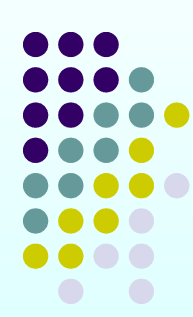
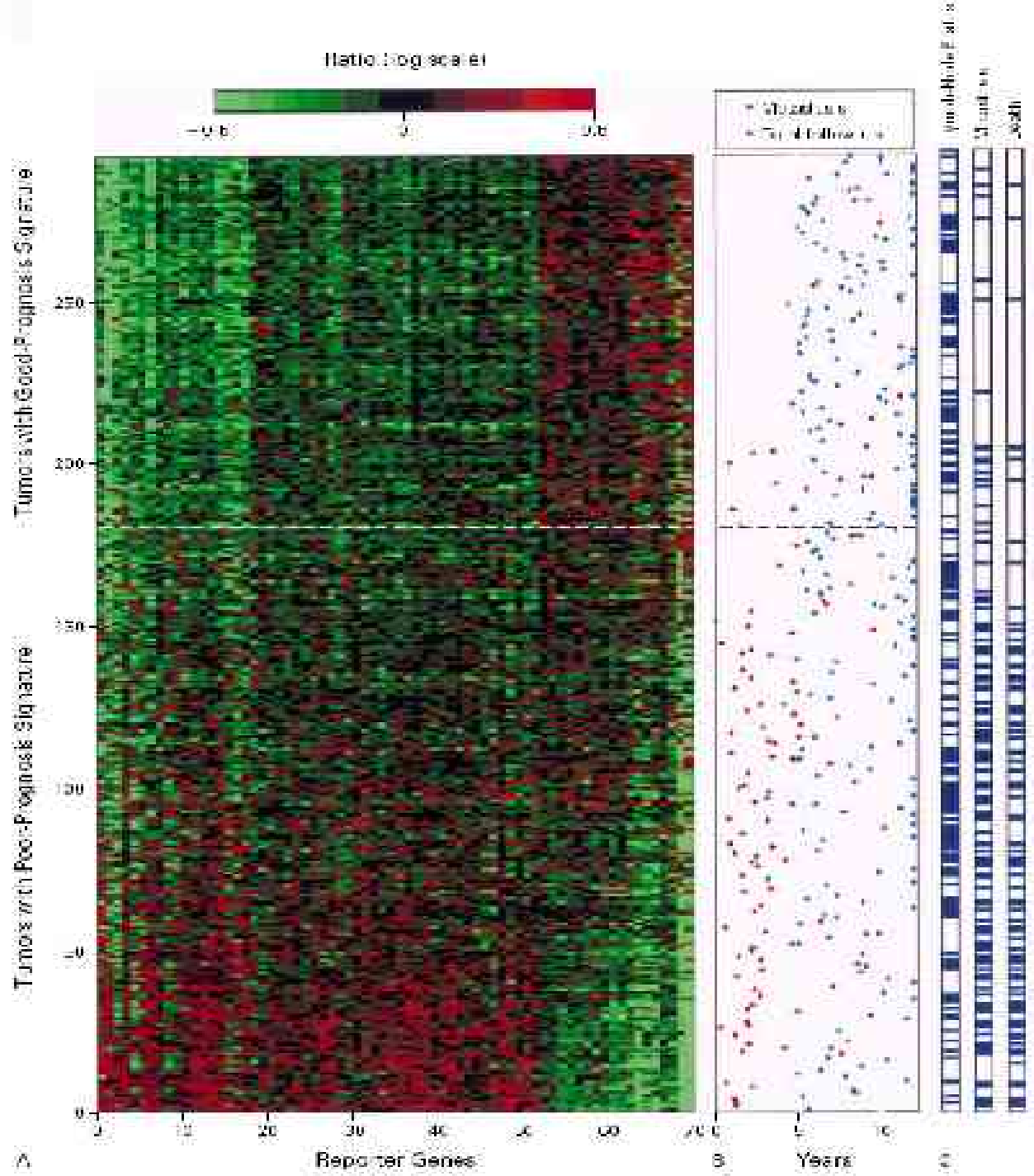
DECEMBER 19, 2002

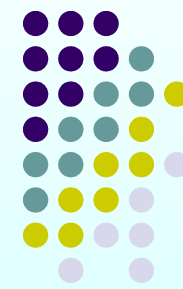
NUMBER 25



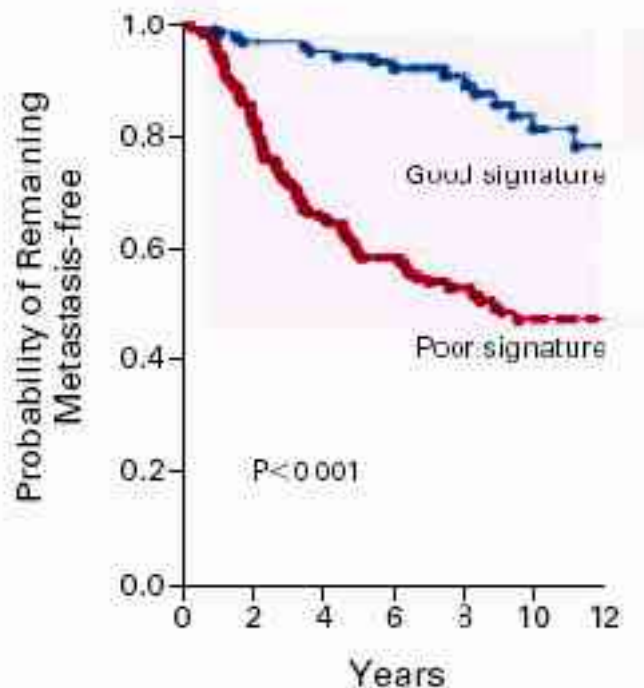
A GENE-EXPRESSION SIGNATURE AS A PREDICTOR OF SURVIVAL IN BREAST CANCER

MARC J. VAN DE VIJVER, M.D., PH.D., YUDONG D. HE, PH.D., LAURA J. VAN 'T VEER, PH.D., HONGYUE DAI, PH.D.,
AUGUSTINUS A.M. HART, M.Sc., DORIEN W. VOSKUIL, PH.D., GEORGE J. SCHREIBER, M.Sc., JOHANNES L. PETERSE, M.D.,
CHRIS ROBERTS, PH.D., MATTHEW J. MARTON, PH.D., MARK PARRISH, DOUWE ATSMA, ANKE WITTEVEEN,
ANNUSKA GLAS, PH.D., LEONIE DELAHAYE, TONY VAN DER VELDE, HARRY BARTELINK, M.D., PH.D.,
SJOERD RODENHUIS, M.D., PH.D., EMIEL T. RUTGERS, M.D., PH.D., STEPHEN H. FRIEND, M.D., PH.D.,
AND RENÉ BERNARDS, PH.D.





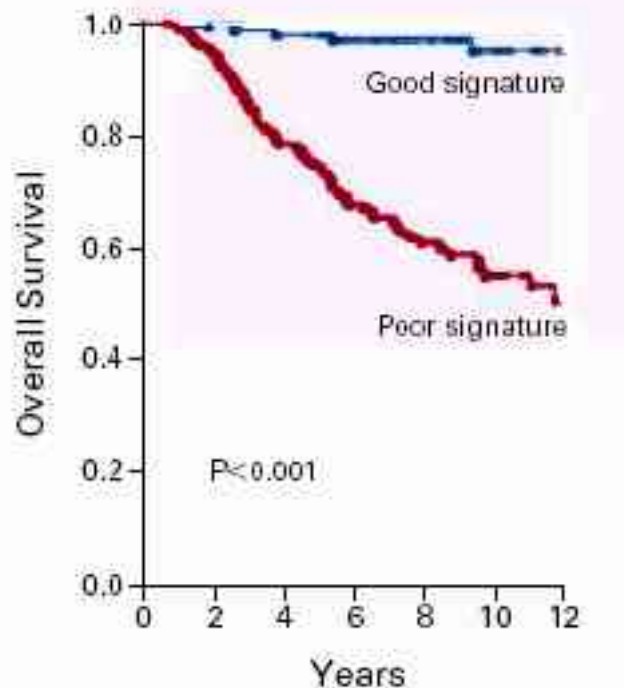
A All Patients



No. at Risk

Good signature	* 15	111	107	87	59	36	13
Poor signature	* 80	146	111	84	52	33	17

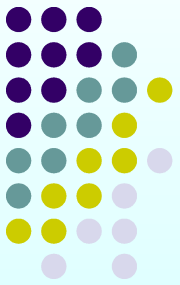
B All Patients



No. at Risk

Low risk	115	114	112	91	65	43	23
High risk	180	167	134	100	62	40	19

Is there any direct relationship between gene **copy** alterations and expression profiles?



Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors

Jonathan R. Pollack^{*,†}, Therese Sørlie[§], Charles M. Perou[‡], Christian A. Rees^{*,*}, Stefanie S. Jeffrey^{††}, Per E. Lonning[†], Robert Tibshirani^{§§}, David Botstein^{||}, Anne-Lise Borresen-Dale[§], and Patrick O. Brown^{††††}

Genomic DNA copy number alterations are key genetic events in the development and progression of human cancers. Here we report a genome-wide microarray comparative genomic hybridization (array CGH) analysis of DNA copy number variation in a series of primary human breast tumors. We have profiled DNA copy number alteration across 6,591 mapped human genes. In 44 predominantly advanced, primary breast tumors and 10 breast cancer cell lines, while the overall patterns of DNA amplification and deletion corroborate previous cytogenetic studies, the high-resolution (gene-by-gene) mapping of amplicon boundaries and the quantitative analysis of amplicon shape provide significant improvement in the localization of candidate oncogenes. Parallel microarray measurements of mRNA levels reveal the remarkable degree to which variation in gene copy number contributes to variation in gene expression in tumor cells. Specifically, we find that 62% of highly amplified genes show moderately or highly elevated expression, that DNA copy number influences gene expression across a wide range of DNA copy number alterations (deletion, low-, mid- and high-level amplification), that on average, a 2-fold change in DNA copy number is associated with a corresponding 1.5-fold change in mRNA levels, and that overall, at least 12% of all the variation in gene expression among the breast tumors is directly attributable to underlying variation in gene copy number. These findings provide evidence that widespread DNA copy number alteration can lead directly to global deregulation of gene expression, which may contribute to the development or progression of cancer.

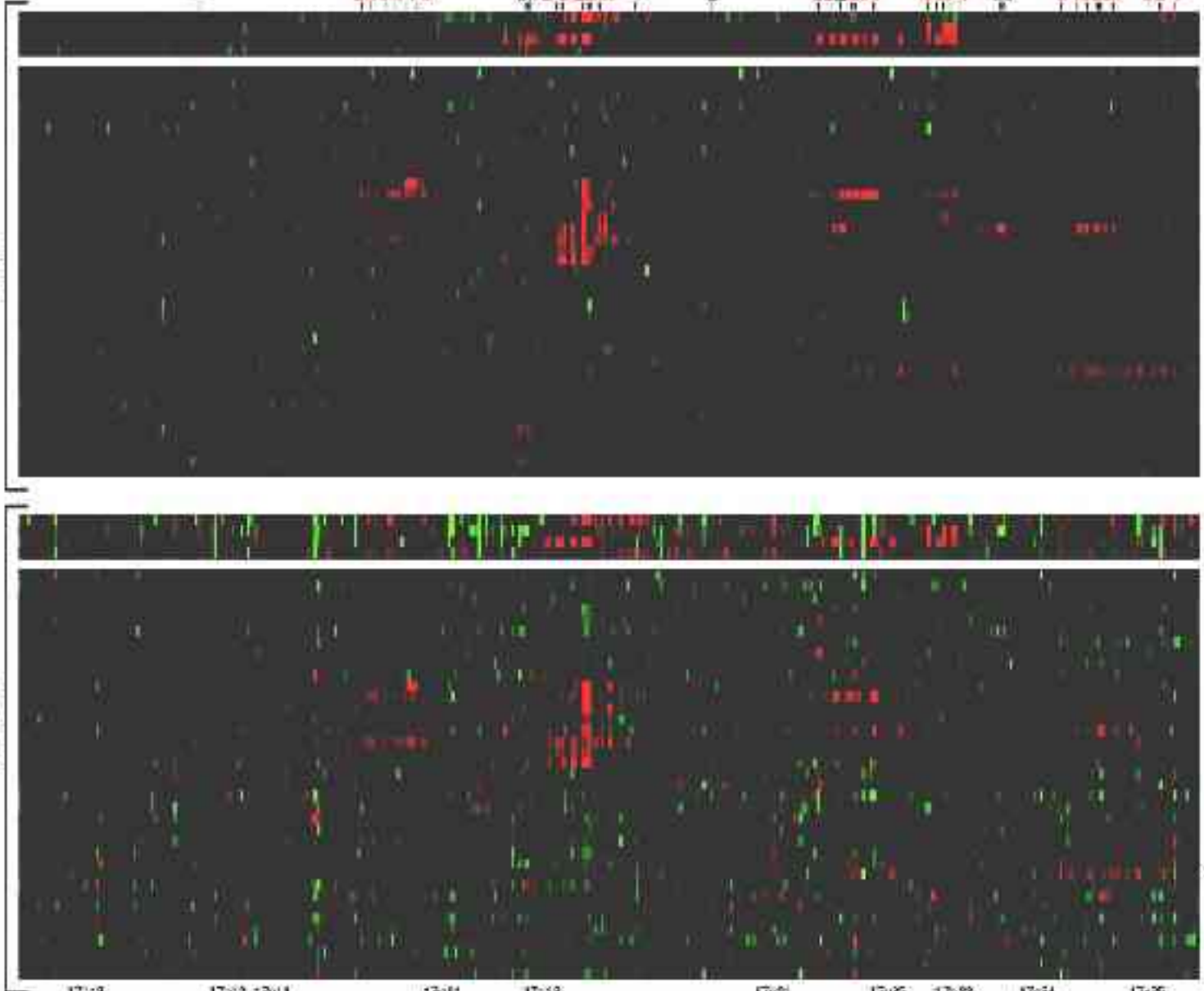
PNAS | October 1, 2002 | vol. 99 | no. 20 | 12963-12968

SHD78
MCE7
SF474
T470

NORWAY 27
NORWAY 7
NORWAY 100
NORWAY 39
NORWAY 105
NORWAY 171
NORWAY 172
STANFORD 36
STANFORD 35
NORWAY 172
STANFORD 35
STANFORD A
NORWAY 53
NORWAY 61
NORWAY 57
NORWAY 47
NORWAY 29
NORWAY 101
NORWAY 45
NORWAY 14
NORWAY 13
NORWAY 11
NORWAY 65
NORWAY 15
STANFORD 128
NORWAY 13
NORWAY 13
NORWAY 55
STANFORD 16
STANFORD 14
NORWAY 12
STANFORD 24
STANFORD 17
NORWAY 104
NORWAY 10
NORWAY 11
NORWAY 118

DNA

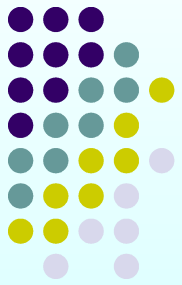
mRNA



test/ref ratio:
>4
2
1
0.5
0.25

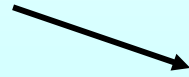
test/ref ratio:
>4
4
2
1
0.5
0.25
0.125

Mb 0 17q8 17q12 17q11 80 40 00 17q22 17q29 17q34 17q25 17pter can 17qter



Gene expression profiles associated with aggressiveness

*Archival tissues from patients with **short** vs **long** survival*



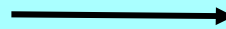
Constitute small groups of tumors (n=5) for each class



Microarray analysis of RNA pools



Search for subsets of genes segregating with survival



Analysis with real-time RT-PCR on individual tumor RNAs

Total RNA
extraction from 30
ductal breast
carcinoma biopsies



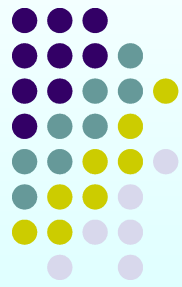
Prep of total RNA pools from two groups:
disease-free survival <72 months (PP, poor
prognosis) or >72 months (GP, good prognosis)
5 patients/pool
7 g RNA/patient



cRNA
prep



Duplicate
oligonucleotide array
analysis (AffyChip
HU133A)



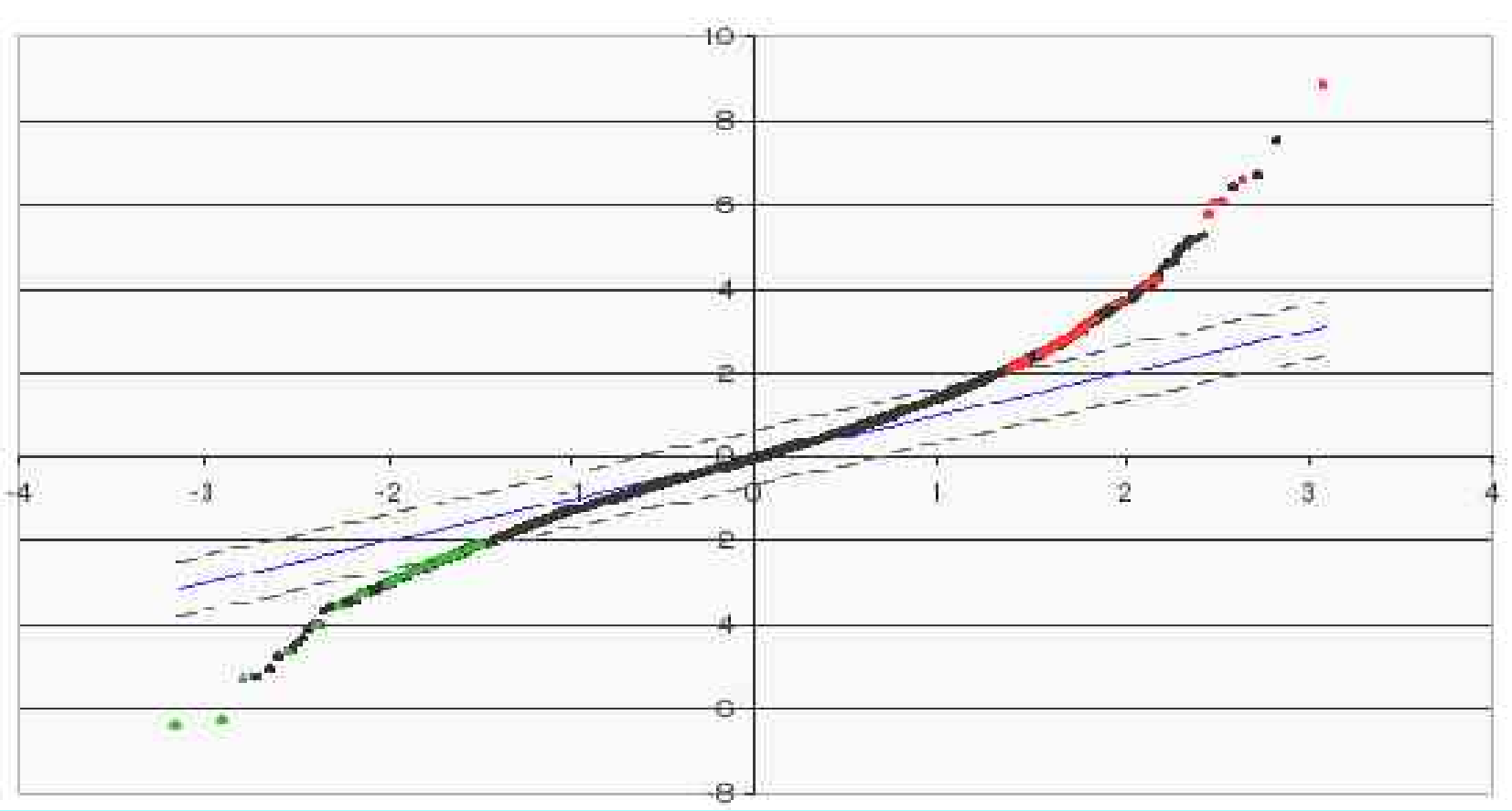
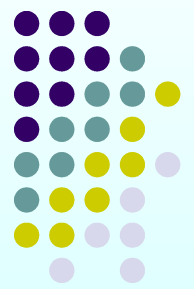
Statistical and bioinformatics analysis:

Data normalization and probeset intensities determined using Affymetrix Microarray Suite 5 (MAS) and Bioconductor.

Differential expression validation carried out using SAM and CyberT.

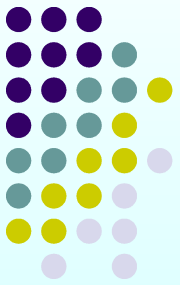
63 statistically significant genes found, using a fold change threshold of $|1.5|$ and a false discovery rate (FDR) of 1.2%.

SAM analysis



GREEN: good prognosis

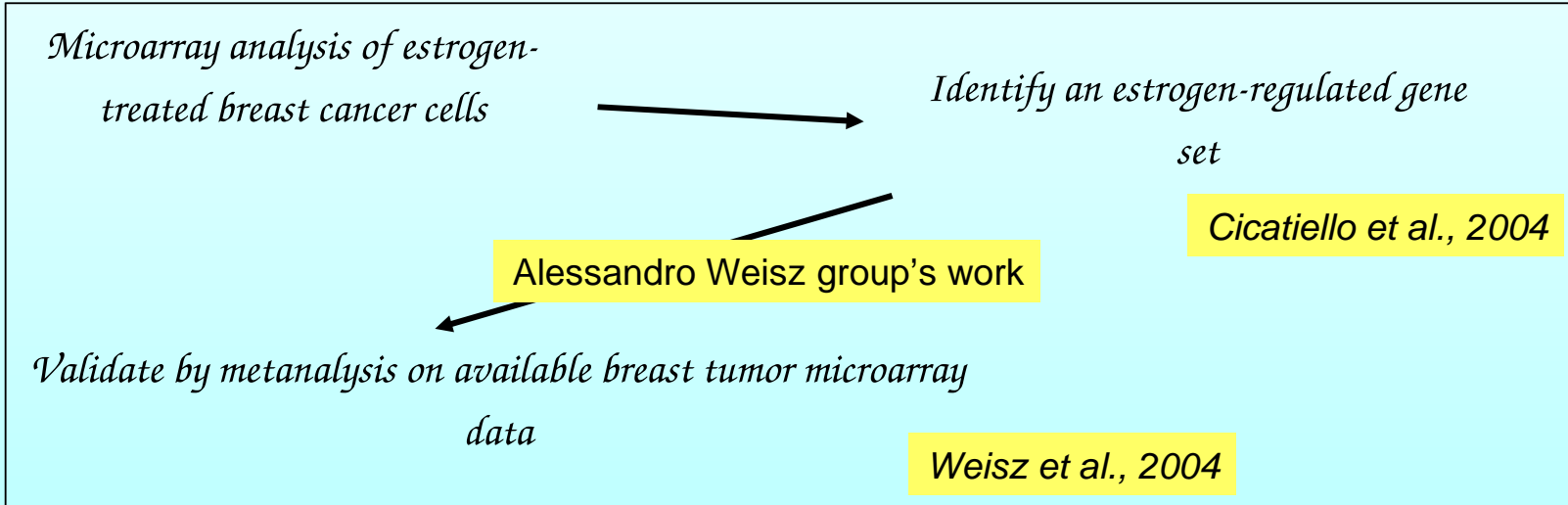
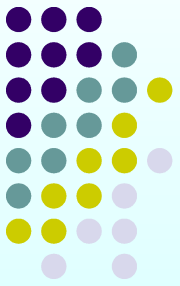
RED: poor prognosis



Genes identified by microarray analysis on pools are now validated by real-time RT-PCR using RNAs from individual tumors (same and independent series)

Generally, they individually display borderline association with prognosis

Gene expression profiles associated with endocrine responsiveness



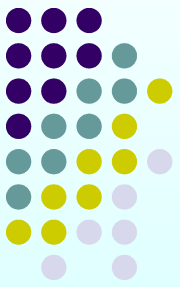
Correlation analysis with response to anti-estrogenic therapy

Analysis by real-time RT-PCR of subsets of ERGs in individual tumor

Adjuvant tamoxifen

Neo-adjuvant setting

Metastatic, first-line



Question: is the set of estrogen-regulated genes good for discriminating ER+ and ER- breast tumor biopsies?

Dataset analyzed: “Molecular Portrait of Human breast Tumors” (Perou et al. 2000) (expression analysis on 62 tumor specimens from 40 patients)

Method: unsupervised two-dimensional hierarchical clustering.

Result: the 218 genes in common with our 344 were NOT able to cluster tumors.

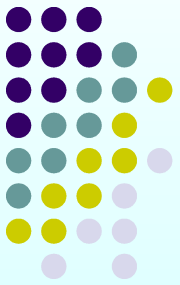
However:

Hierarchical clustering was again performed using a **new set** of supervised genes:

- ➡ the 19 genes identified by cell line analysis
- ➡ 8 known estrogen-responsive genes found within the tumor data set
- ➡ 4 additional genes consistently expressed in ER+ cells/tumors

This estrogen-responsive gene set DID cluster tumors

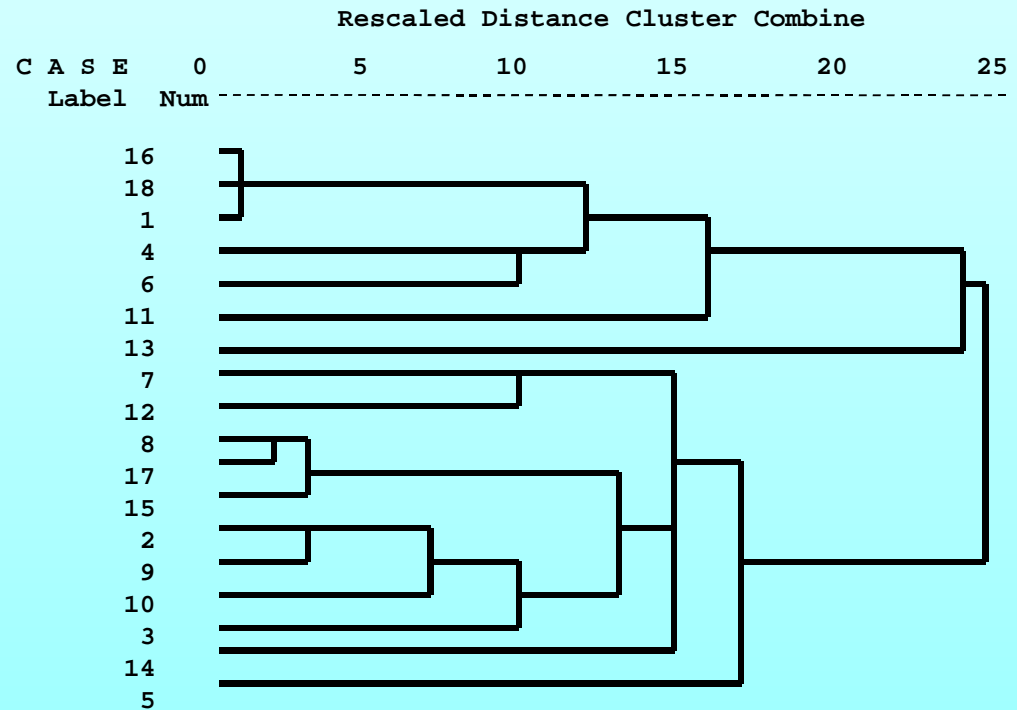
Question: could a small number of ERGs, measured in tumors and not including ER, together with common variables, discriminate the ER+ phenotype?



- 1) Select genes from the “in vitro” estrogen-regulated set.
- 2) Examine their expression level in randomly selected cases of primary invasive ductal breast cancer by quantitative RT-PCR.
- 3) Examine their combinatorial ability to discriminate
 - a) steroid receptor status
 - b) clinical outcome

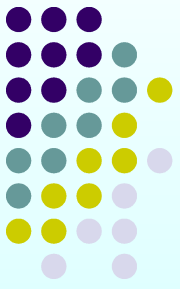
Example: 18 cases, 8 genes:

Unsupervised hierarchical clustering, two cluster solution. ER distribution in the two cluster is significantly different ($P < 0.02$)



Sorbello et al. “Quantitative real-time RT-PCR analysis of eight novel oestrogen-regulated genes in breast cancer”. *Int J Biol Markers* 18: 123-9 (2003).

Correlation analysis with response to anti-estrogenic therapy



Adjuvant setting:

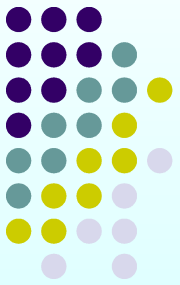
retrospective analysis in post-menopausal patients with ER+ tumors, given adjuvant tamoxifen or aromatase inhibitors. Compare gene expression in R+ and R-. Requires hundreds of cases.

Neo-adjuvant setting:

older patients with unoperable ER+ tumors, are given pre-operative tamoxifen for 3-4 months. Objective response is evaluated. Gene expression profile is measured on fine-needle biopsy.

Metastatic:

patients with metastasis after adjuvant therapy, given first-line tamoxifen. Objective response is evaluated. Gene expression profile is measured on the primary tumor tissues, the rationale being that overall expression profiles are conserved.



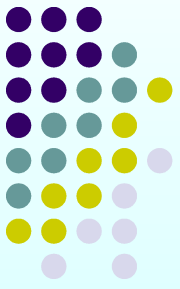
CONCLUSIONS

Estrogen-regulated genes are good “detectors” of the presence of ER in both cell lines and tumors

A combination of few ERGs can be used to support ER/PgR characterization in breast cancer

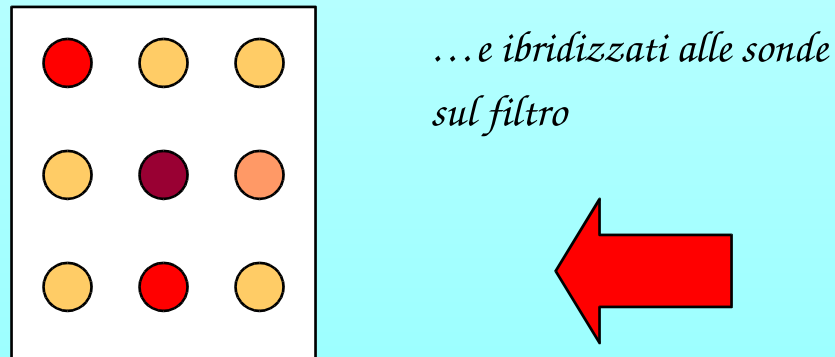
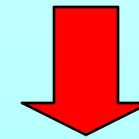
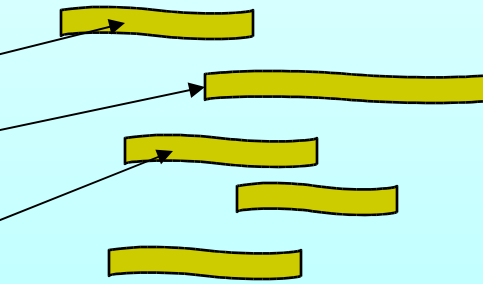
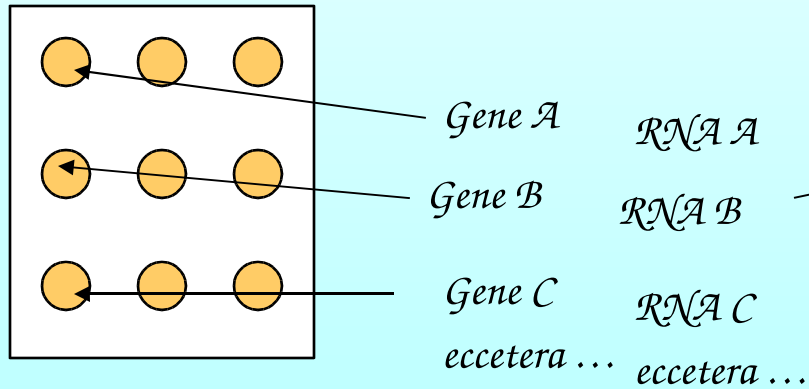
ERGs are now being tested as “detectors” of responsiveness to endocrine treatments in breast cancer

La tecnica di ibridazione su fase solida permette di verificare la presenza di molti geni contemporaneamente

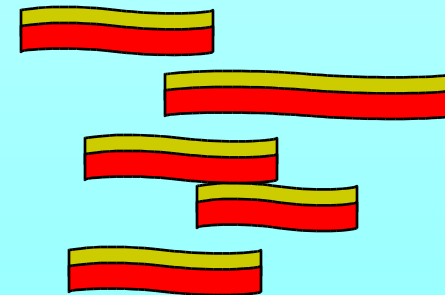


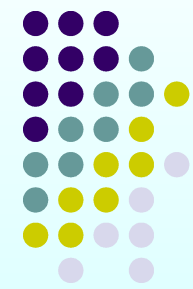
Sonde complementari a diversi geni vengono depositate su un filtro

L'RNA o il DNA vengono estratti dalle cellule



appositamente marcati



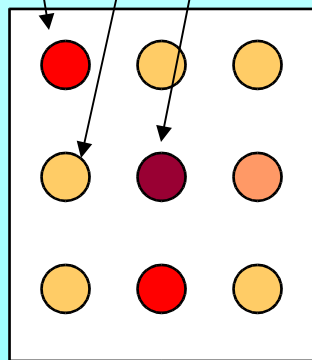


L'RNA o il DNA vengono estratti dalle cellule

Gene A: SI

Gene B: NO

Gene X: molto!

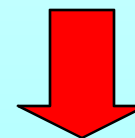
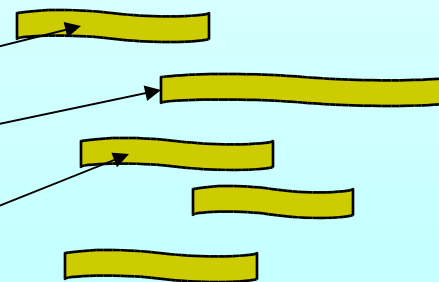


RNA A

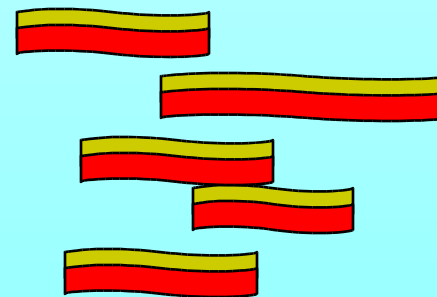
RNA B

RNA C

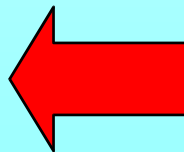
eccetera ...



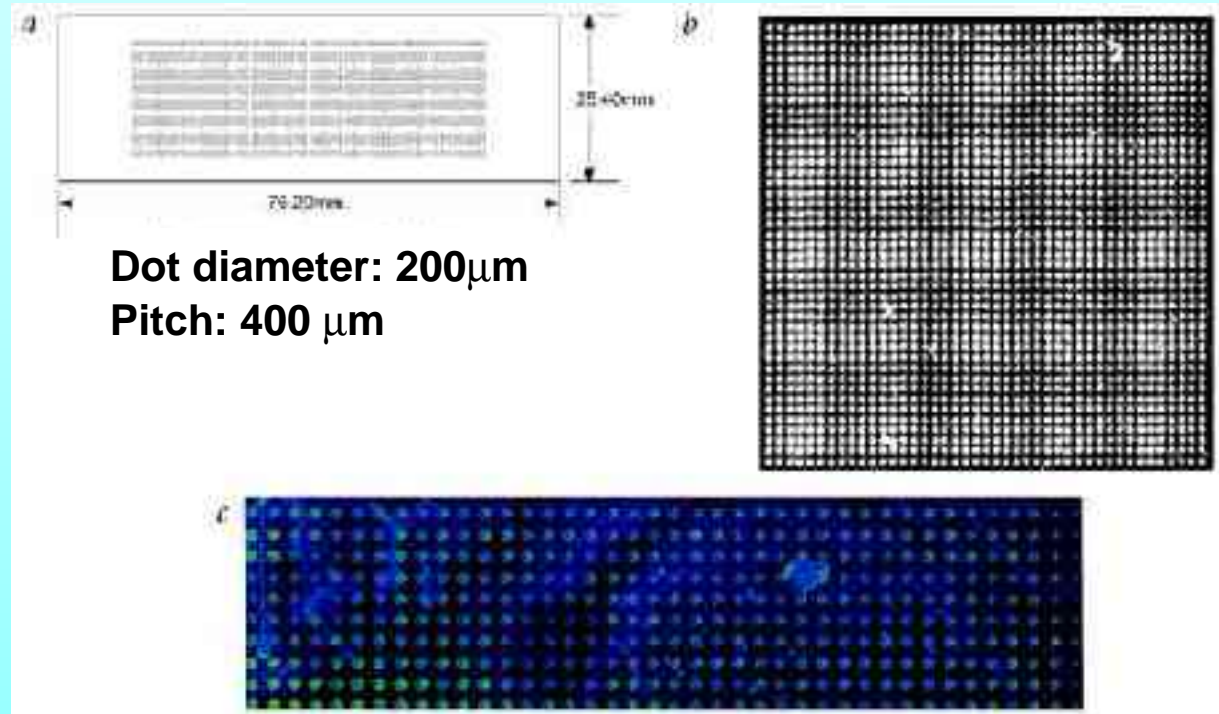
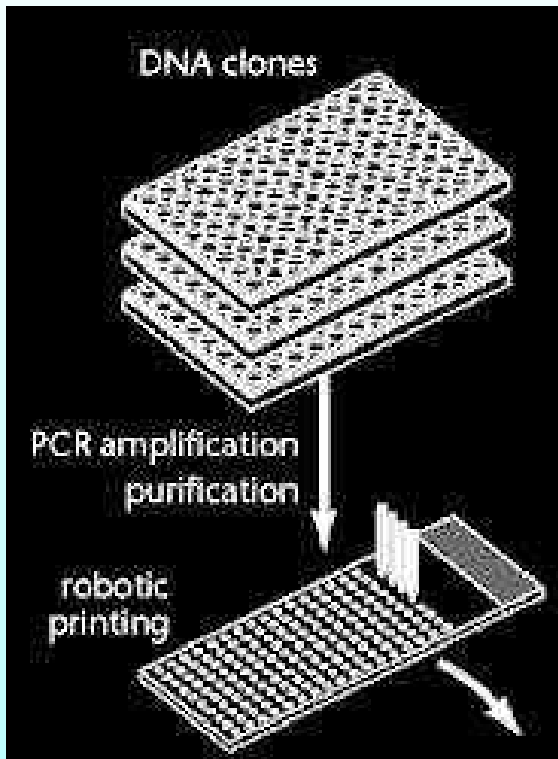
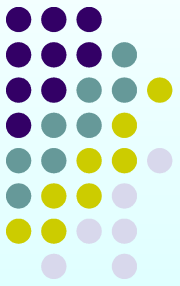
appositamente marcati



... e ibridizzati alle sonde sul filtro



Highly parallel probe arrays: “DNA chips”



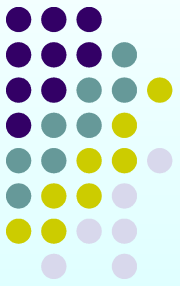
DNA chips: there are currently different types available:

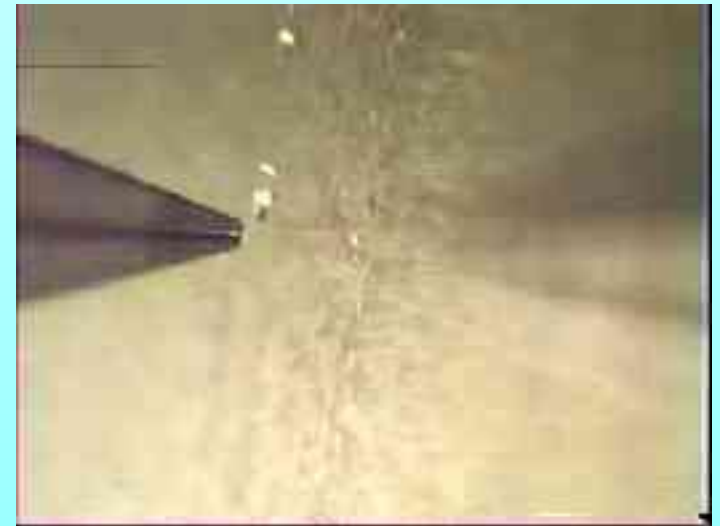
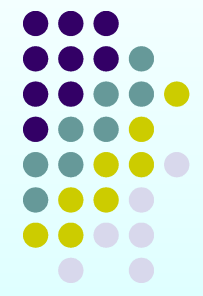
Spotted microarrays with cDNA probes

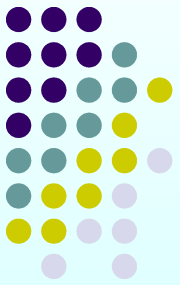
Spotted microarrays with long oligonucleotide probes

Short oligonucleotide microarrays (Affymetrix)

Ink-jet synthesized long oligonucleotide microarrays



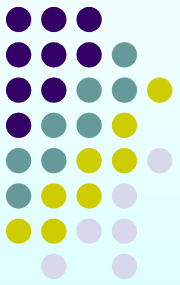




Available microarrays differ widely over a number of respects:

3. *The kind of probes*
4. *Specificity of the probes*
5. *Use as absolute versus relative measurement*
6. *Flexibility and cost*

8. *Sample preparation*



Probes

cDNA fragments, PCR products are low-cost, very flexible probes,

But

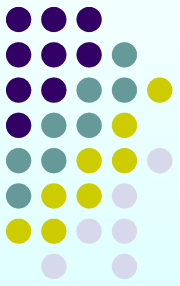
when dealing with probes longer than 60-70 nt, sub-section hybridization can give rise to false detection (i.e. nonspecific)

Probes may also present very different CG content in different sections.

Flexibility is somehow an advantage

But

*individual probe design make comparison between results obtained in different laboratories **very difficult***



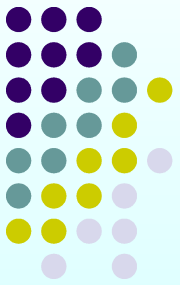
Specificity

For cDNA and PCR products, there are wide variations between producers

For long oligonucleotide probes, the specificity is usually checked a priori, i.e. oligo probes are bioinformatically designed to hybridize to unique sequences. In addition, the T_m of the probes is controlled by uniformity in CG content.

In the case of Affymetrix, the relative shortness of the oligos does not allow the use of one single probe / gene.

The specificity is, in this case, evaluated by interrogating each gene with multiple oligos: a "probeset".

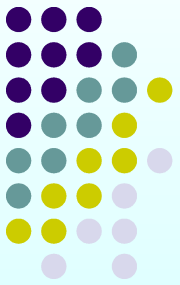


Absolute versus relative measurement, sample preparation

Using hybridization reaction as an absolute measurement of RNA requires that the amount of probe on each spot be uniform and reproducible.

This is a requirement fulfilled by Affymetrix arrays and by the latest generation of spotted long-oligo arrays (e.g. Amersham).

*In all the other cases, the amount of probe / spot is variable and unassessable, so that **relative** measurements are necessary.*



Sample preparation

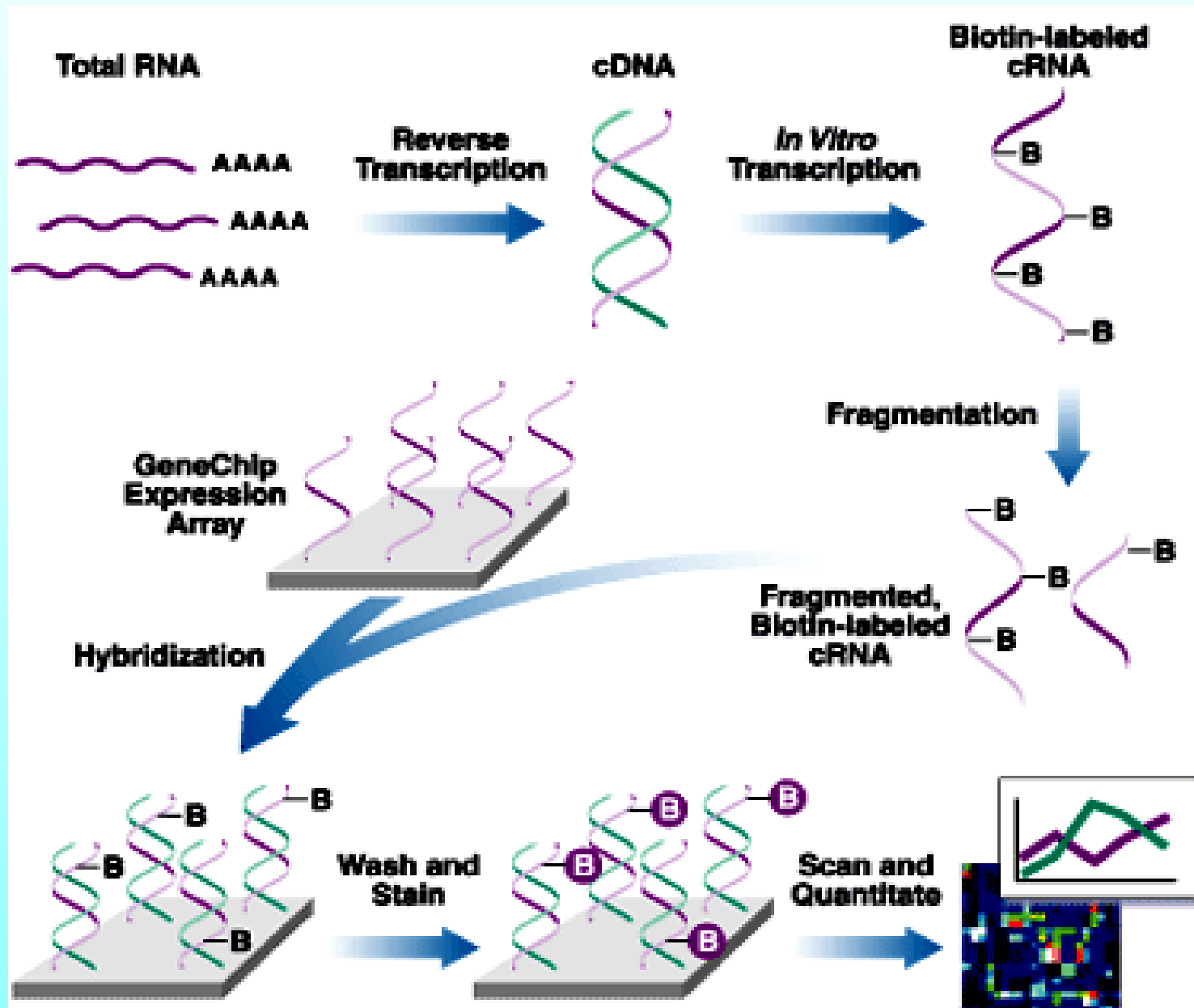
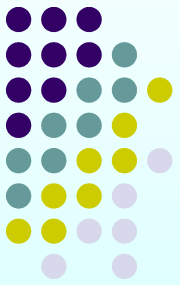
There are widely different methods to prepare and label the \mathcal{RNA} (the “complex probe”).

First, the use of total \mathcal{RNA} versus $\text{poly}(\mathcal{A}^+)\text{-}\mathcal{RNA}$

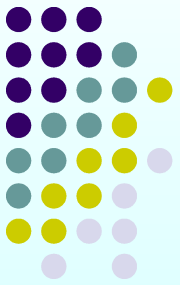
Second, nonamplified versus amplified

Third, direct versus indirect labelling

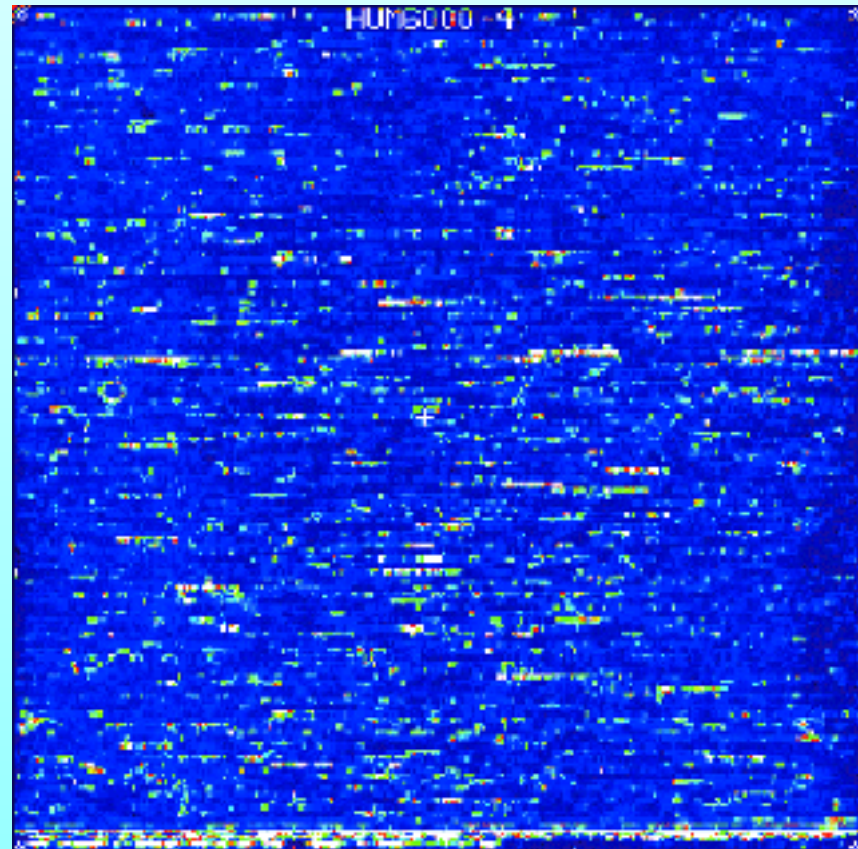
How an Affychip is used

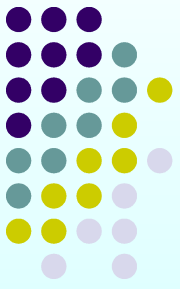


How an Affychip result looks like



An oligonucleotide array (Affychip®) hybridized to biotin-labelled cRNA and revealed with fluorochrome-conjugated avidin





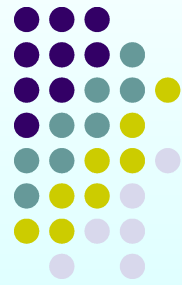
Spotted arrays, on the contrary, are commonly used in relative measurements, i.e. to compare gene expression between two biological samples.

RNA from **sample** and from **reference** are labeled by introducing two different fluorochromes.

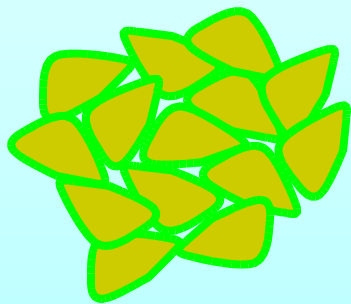
This allows co-hybridization of the two samples to the same chip, providing direct comparison by two-color analysis

The most common fluorochromes are the cyanines **Cy3** (red) and **Cy5** (green)

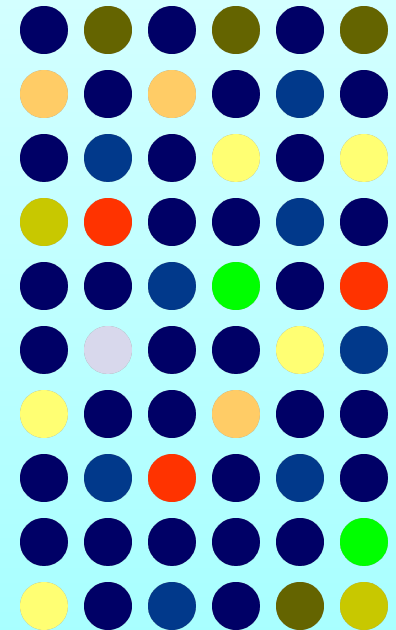
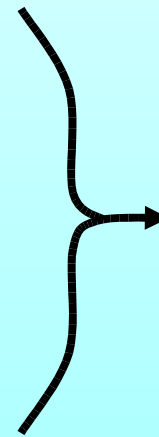
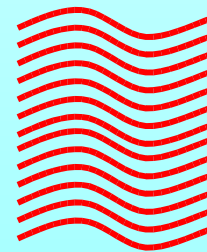
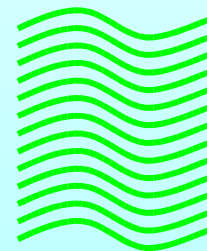
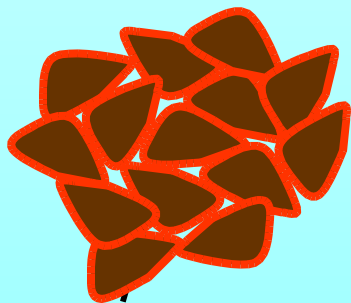
Co-hybridization of double-color labelled test and reference samples



“Test” sample (tumor tissue, stimulated cells)



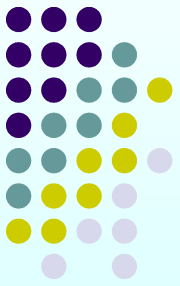
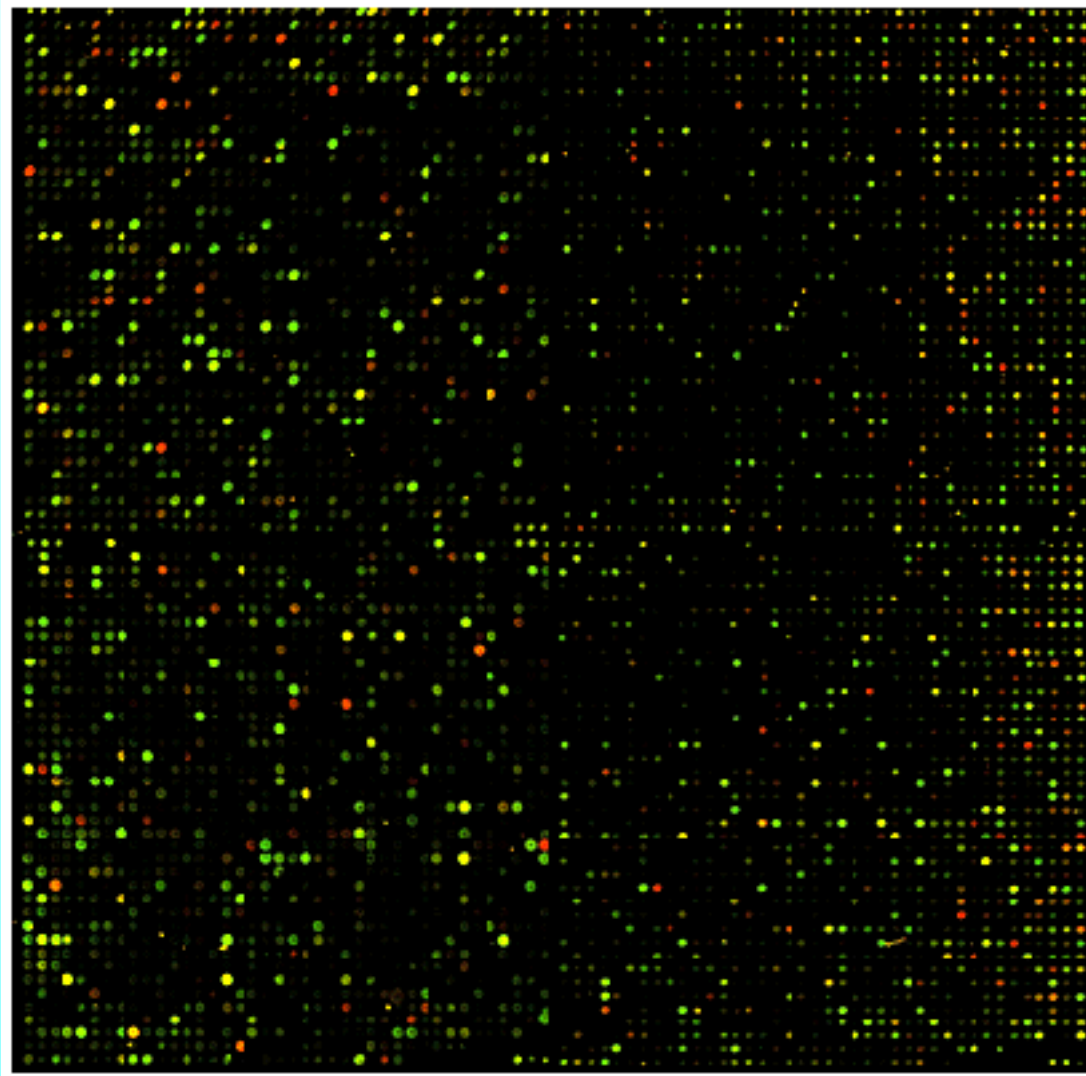
RNA Extraction,
cDNA Synthesis
and labelling

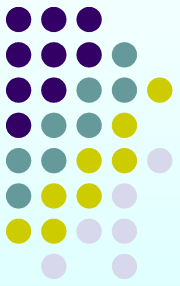


Hybridization

“reference” sample (mix of different tissues, unstimulated cells)

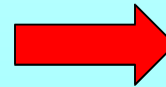
*How a spotted microarrays hybridized with two-colors probes
looks like*





Laser scanning ↙ intensities ↙ normalization ↙ data analysis

Expression ratios are then transformation to false-color codes for representation

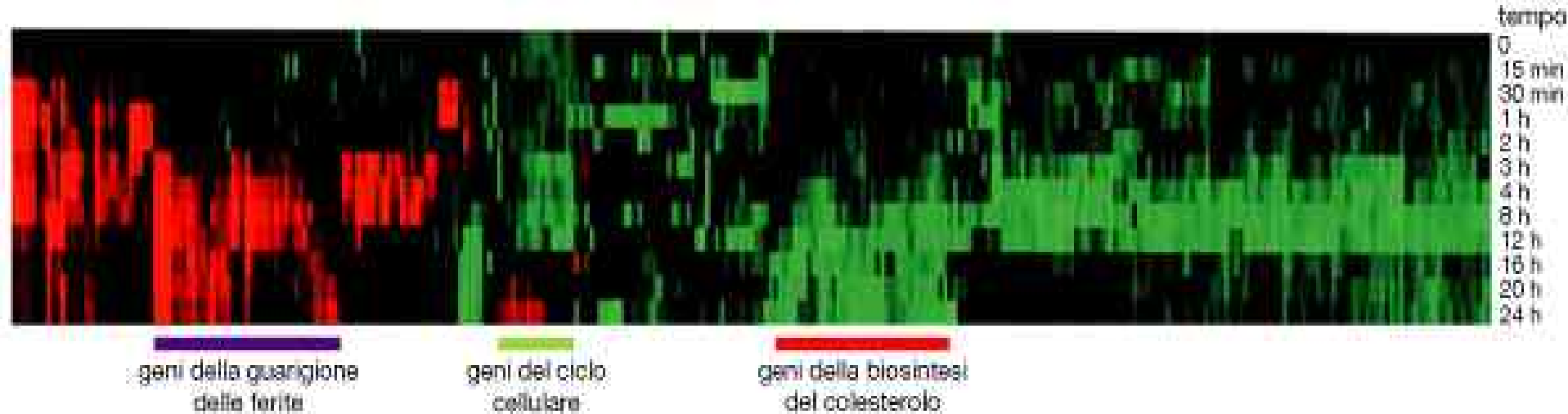
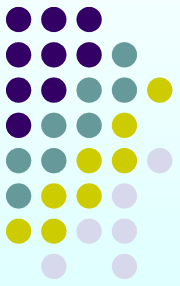


and subjected to statistical and clustering analysis

human fibroblasts: 48 hrs serum starvation, then serum back for the indicated times

RNA extracted at time points labelled with red dye

RNA extracted at time = 0 labelled with green dye



Back to the main course ([goto](#))