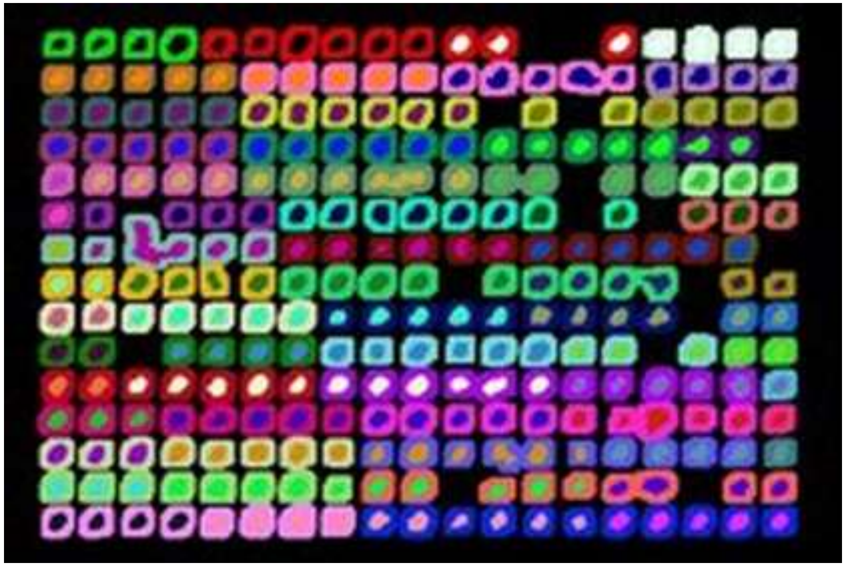
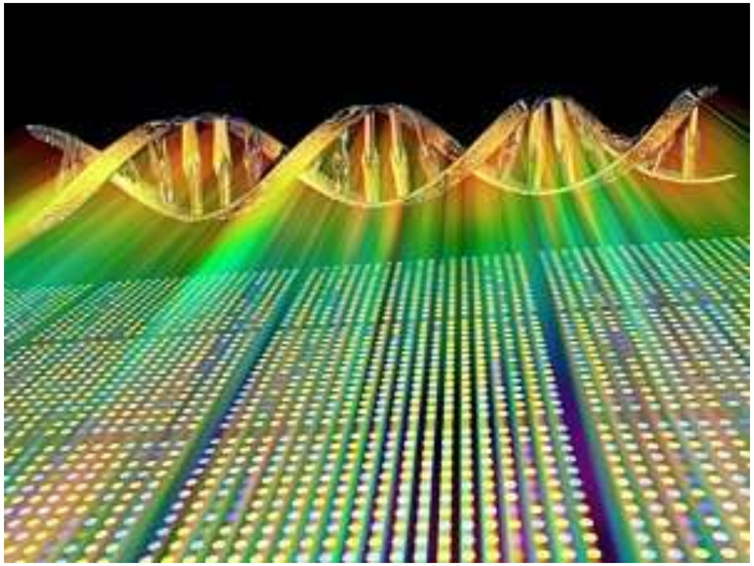
ADAM MICKIEWICZ UNIVERSITY IN POZNAŃ

Faculty of Biology

DNA MICROARRAY

“Gene Expression Profiling in Health and Disease”

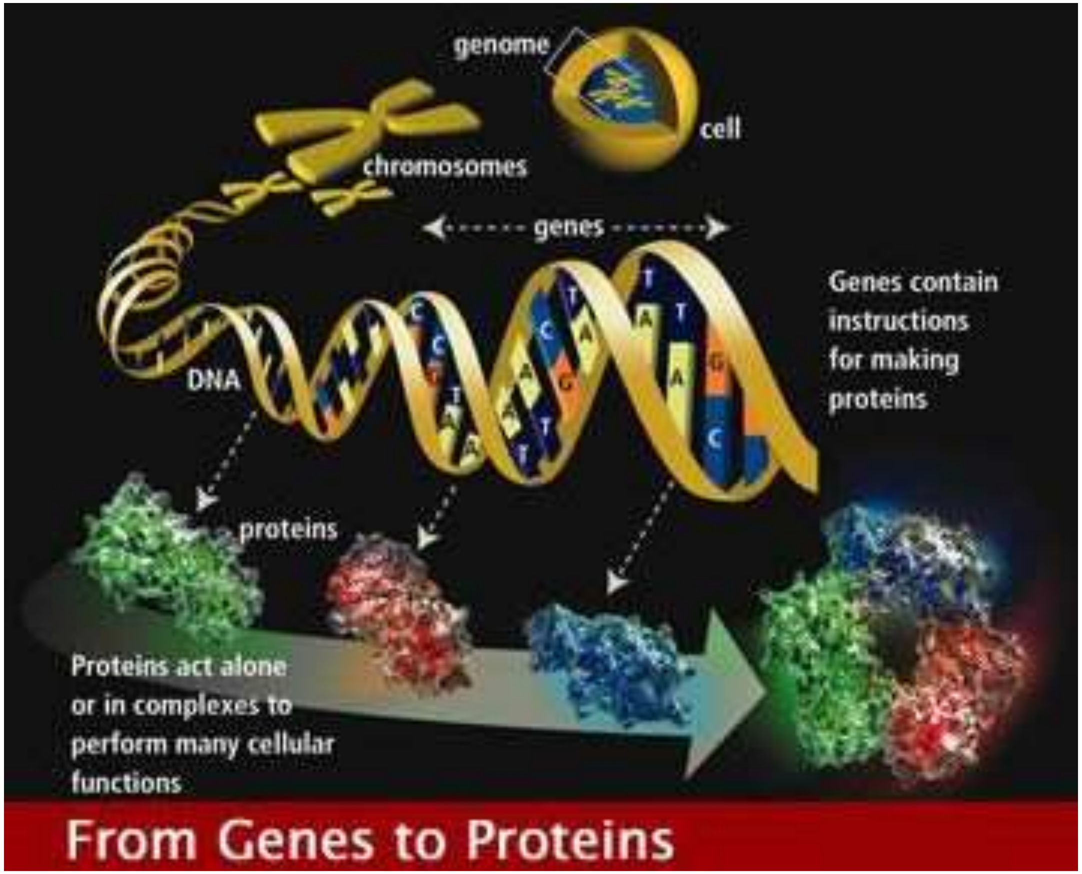




Hans Bluyssen
04-11-2020

www.biologia.amu.edu.pl

Genome & Genes



Gene

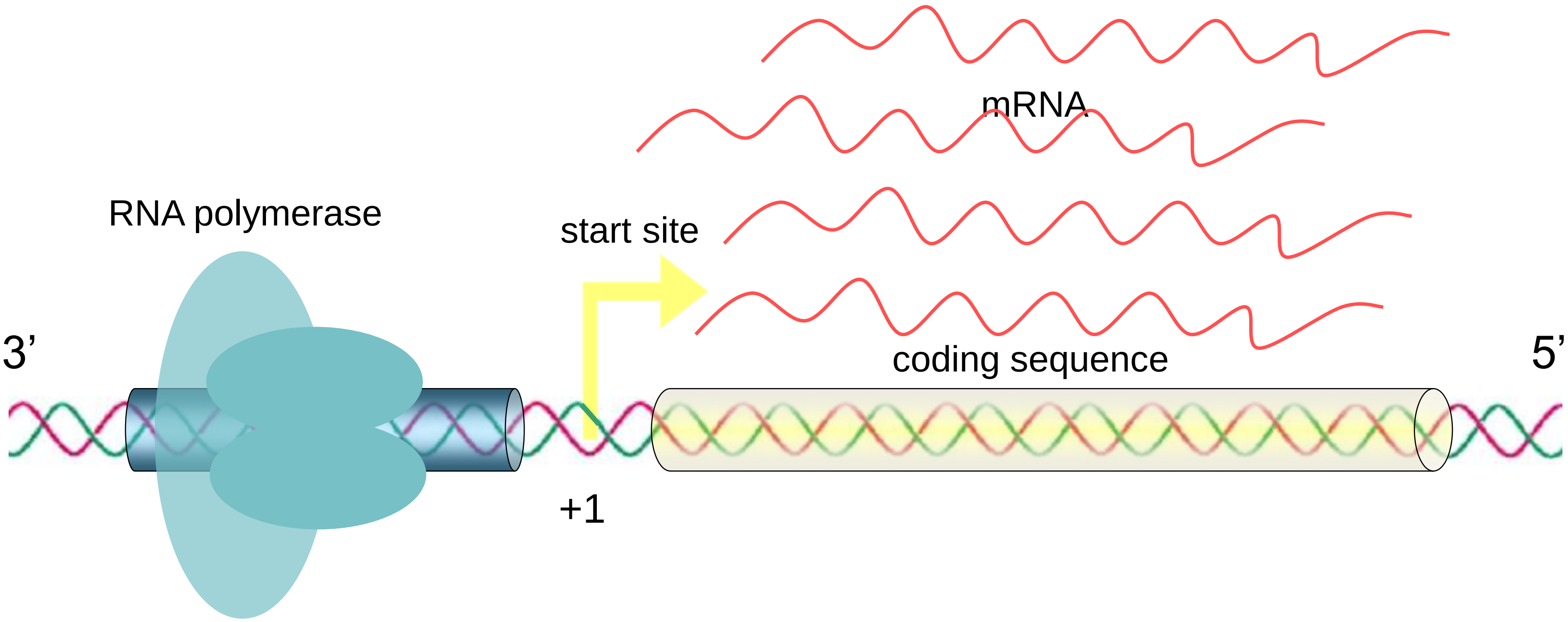
transcription

RNA

translation

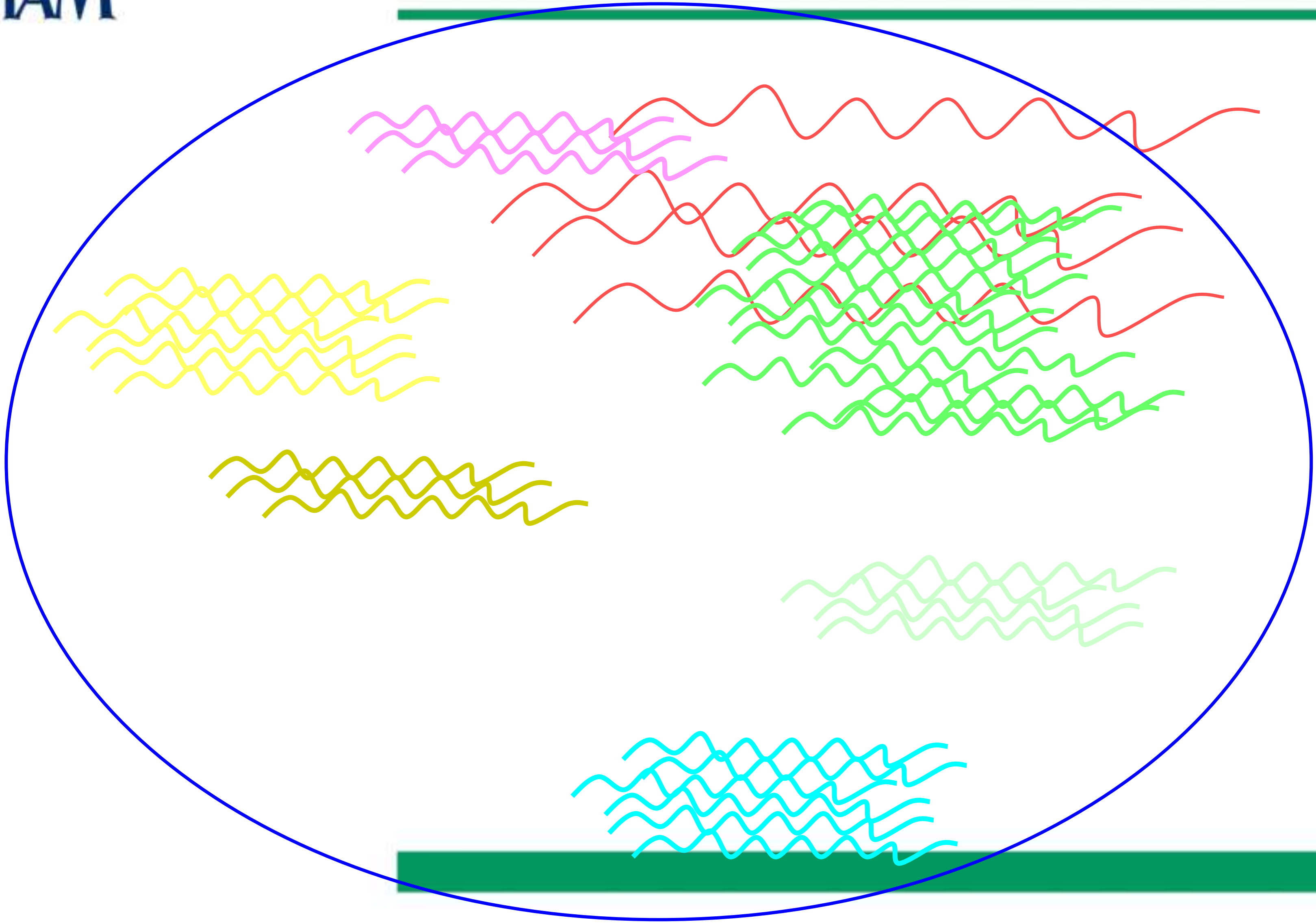
Protein

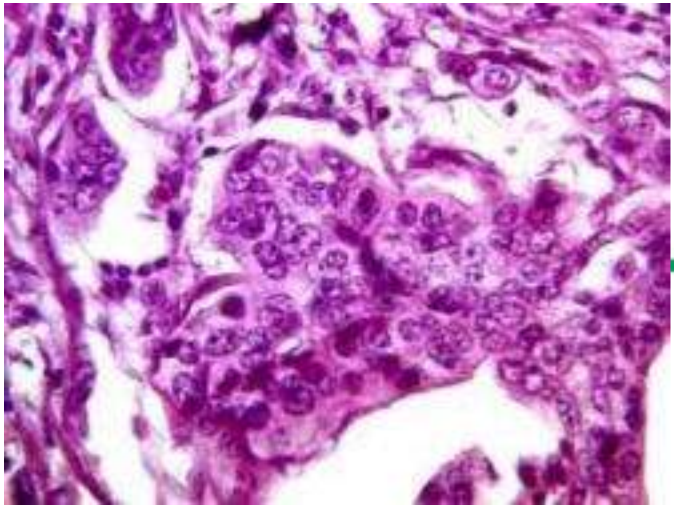
RNA Transcription



generation of mRNA from genomic DNA

The Transcriptome





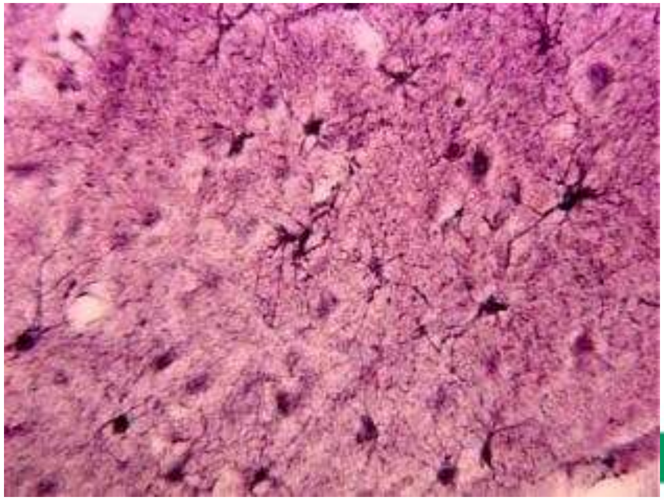
Breast Cancer cells

Red Blood cells

Sperm cells

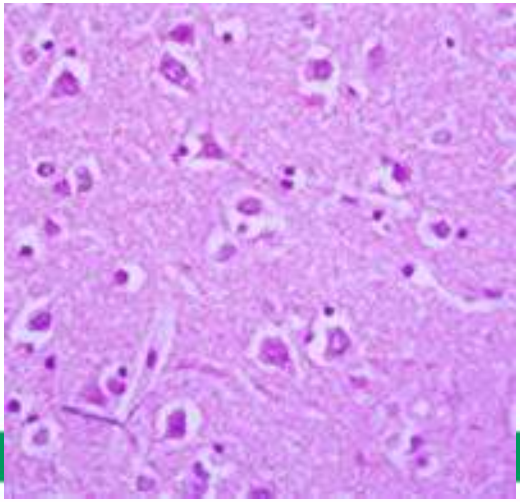
Transcriptional Program & Cellular Fate

Astrocytes



Smooth Muscle cells

Neuronal cells



Transcriptomics

Parallel monitoring of relative levels of thousands of mRNA species at one time point or condition:
expression profiling

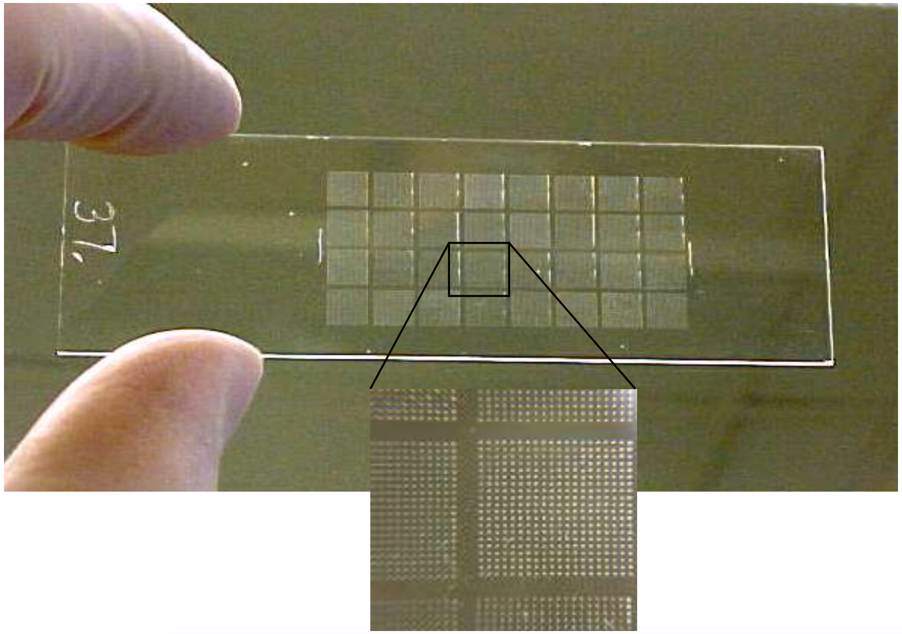
DNA Microarray
RNAseq

DNA Microarrays: Basics

- Put a large number (~30K) of cDNA sequences or synthetic DNA oligomers onto a glass slide (or other substrate) in known locations on a grid.
- Label an RNA sample and hybridize
- Measure amounts of RNA bound to each square in the grid
- Make comparisons
  - Cancerous vs. normal tissue
  - Treated vs. untreated
  - Time course
- Many applications in both basic and clinical research

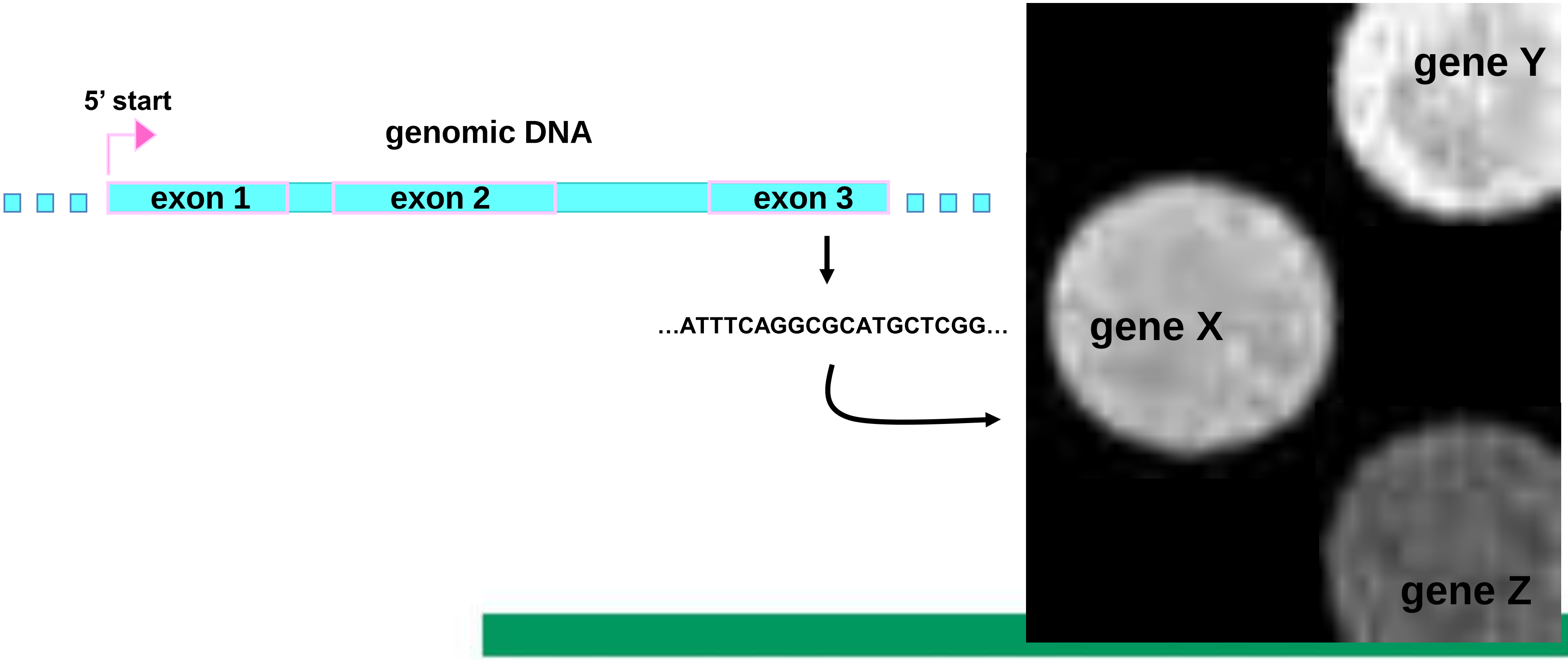
What is a DNA microarray?

DNA microarrays are ordered assemblies of DNA sequences immobilized on a solid support (such as chemically modified glass).



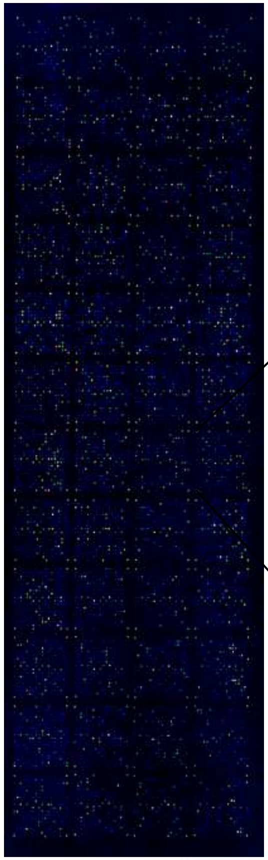
What is a DNA microarray?

The DNA sequences (e.g. PCR products or oligos) correspond to the transcribed regions of genes.

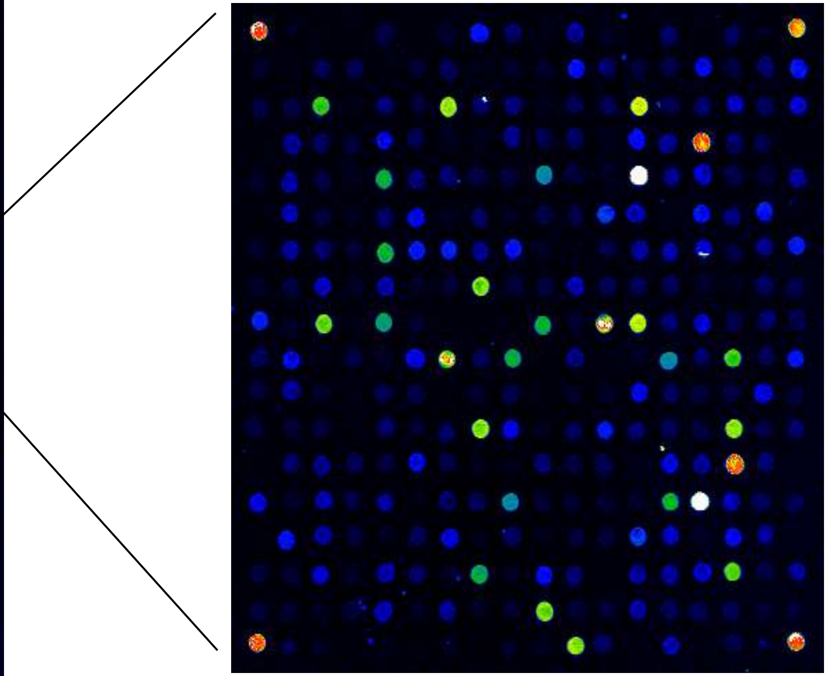


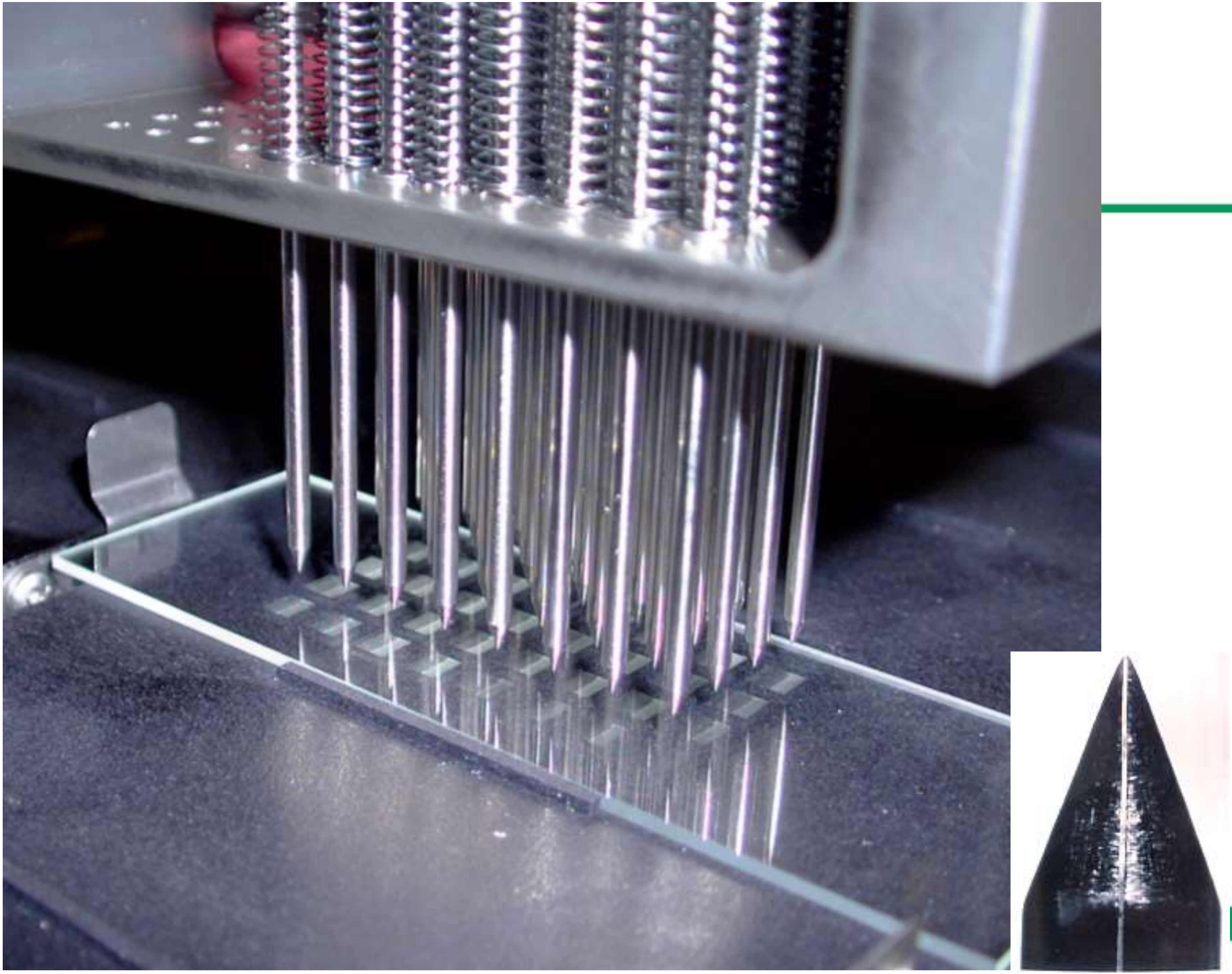
Long oligo arrays

Each gene represented by single long oligo (60 - 70-mer)



Oligo’s spotted by robot

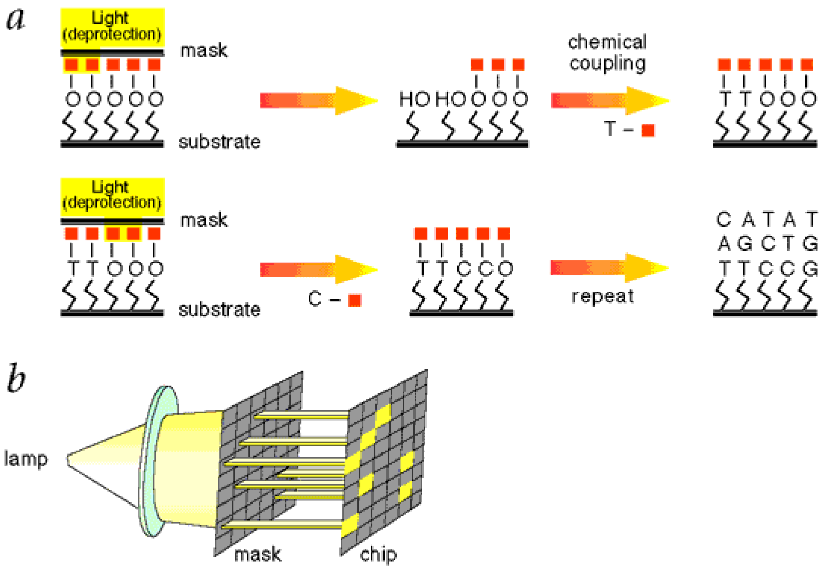


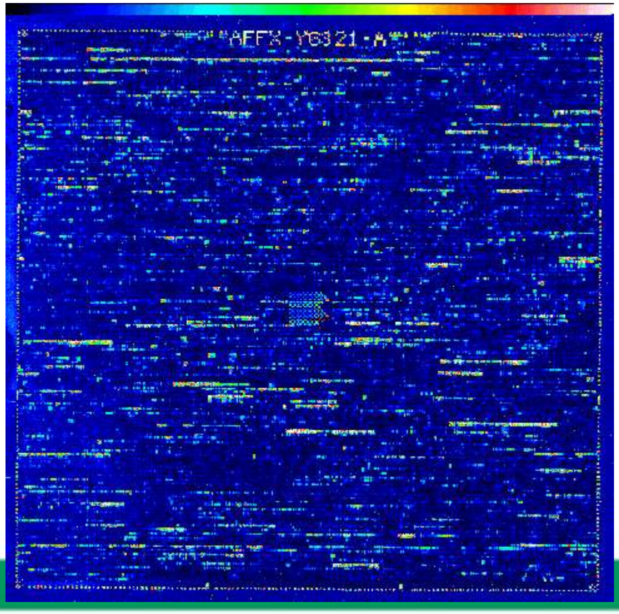


Affymetrix chips

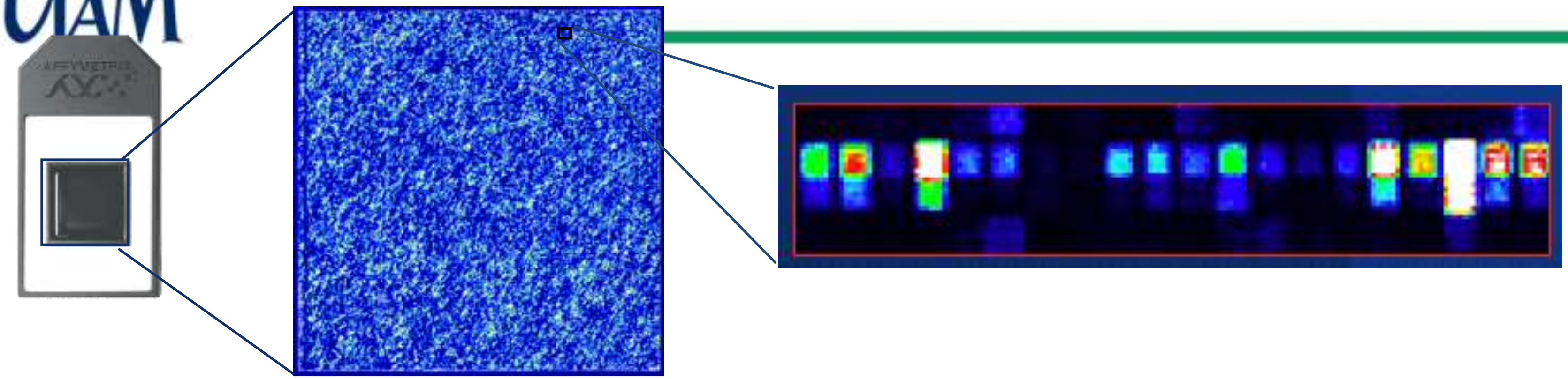
Each gene represented by 20 different short (25-mer) oligonucleotides and 20 mismatch controls

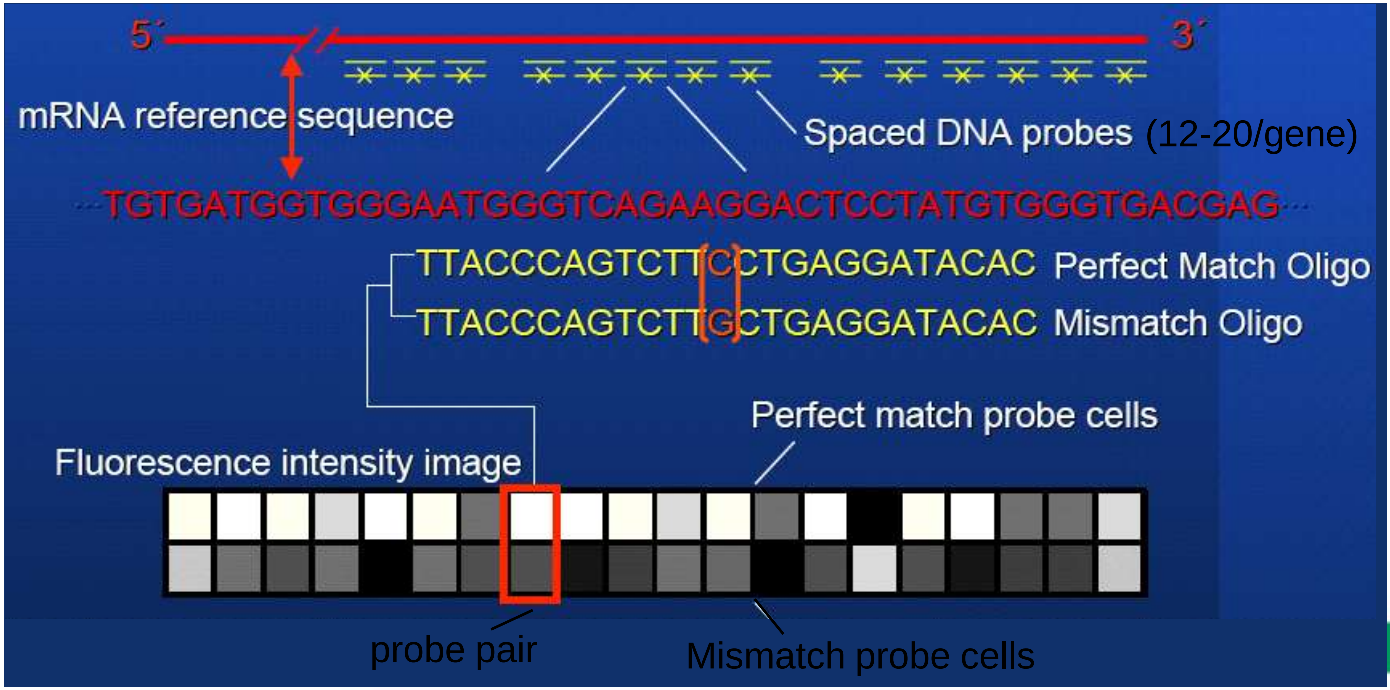
Oligonucleotides synthesized on chip by photolithographic masking



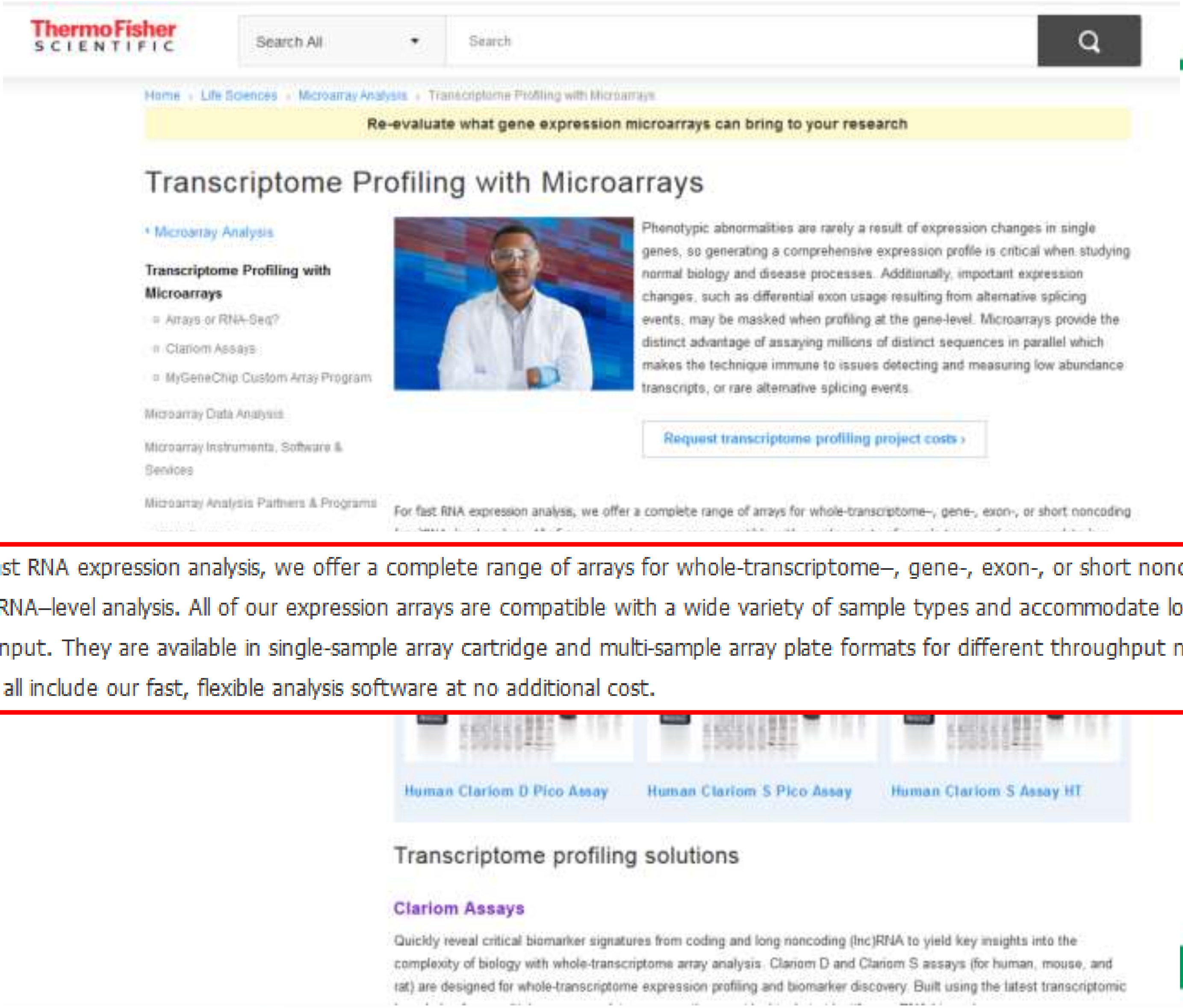


Affymetrix chips





Affymetrix chips



Re-evaluate what gene expression microarrays can bring to your research

Transcriptome Profiling with Microarrays

Phenotypic abnormalities are rarely a result of expression changes in single genes, so generating a comprehensive expression profile is critical when studying normal biology and disease processes. Additionally, important expression changes, such as differential exon usage resulting from alternative splicing events, may be masked when profiling at the gene-level. Microarrays provide the distinct advantage of assaying millions of distinct sequences in parallel which makes the technique immune to issues detecting and measuring low abundance transcripts, or rare alternative splicing events.

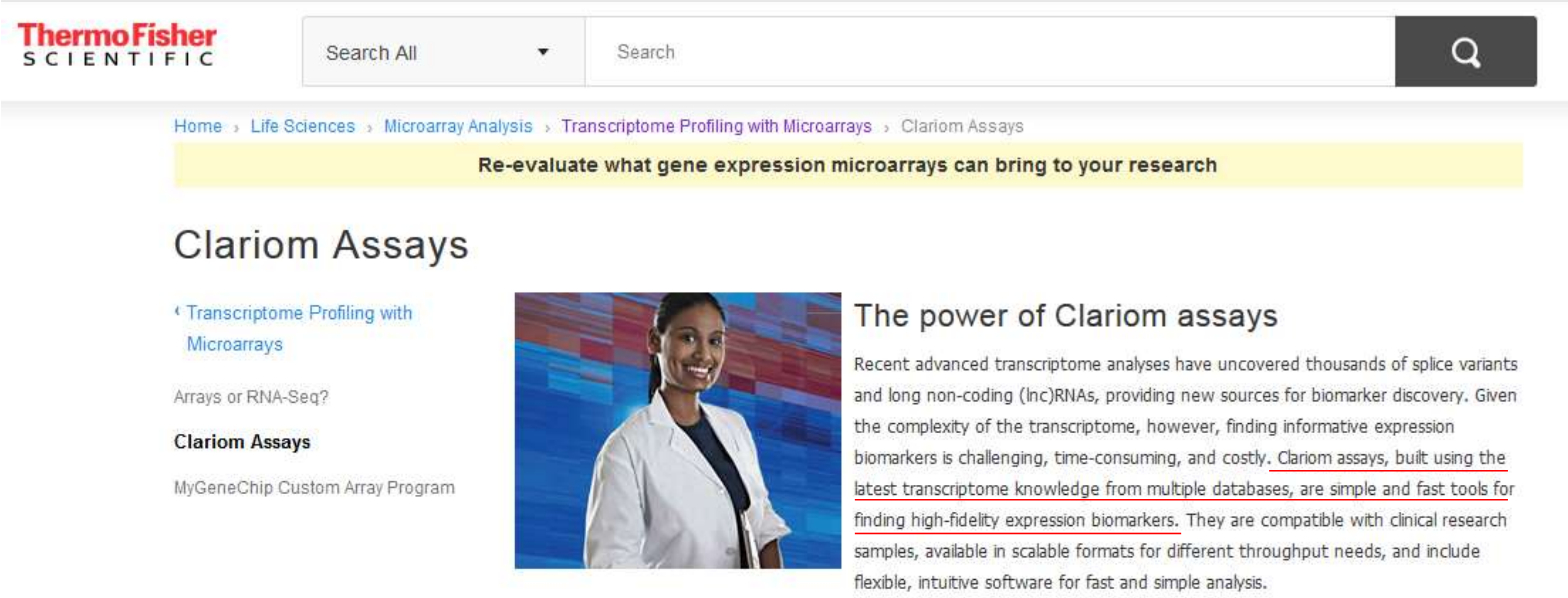
For fast RNA expression analysis, we offer a complete range of arrays for whole-transcriptome–, gene-, exon-, or short noncoding (snc)RNA–level analysis. All of our expression arrays are compatible with a wide variety of sample types and accommodate low RNA input. They are available in single-sample array cartridge and multi-sample array plate formats for different throughput needs. They all include our fast, flexible analysis software at no additional cost.

Transcriptome profiling solutions

Clariom Assays

Quickly reveal critical biomarker signatures from coding and long noncoding (lnc)RNA to yield key insights into the complexity of biology with whole-transcriptome array analysis. Clariom D and Clariom S assays (for human, mouse, and rat) are designed for whole-transcriptome expression profiling and biomarker discovery. Built using the latest transcriptomic

Affymetrix chips



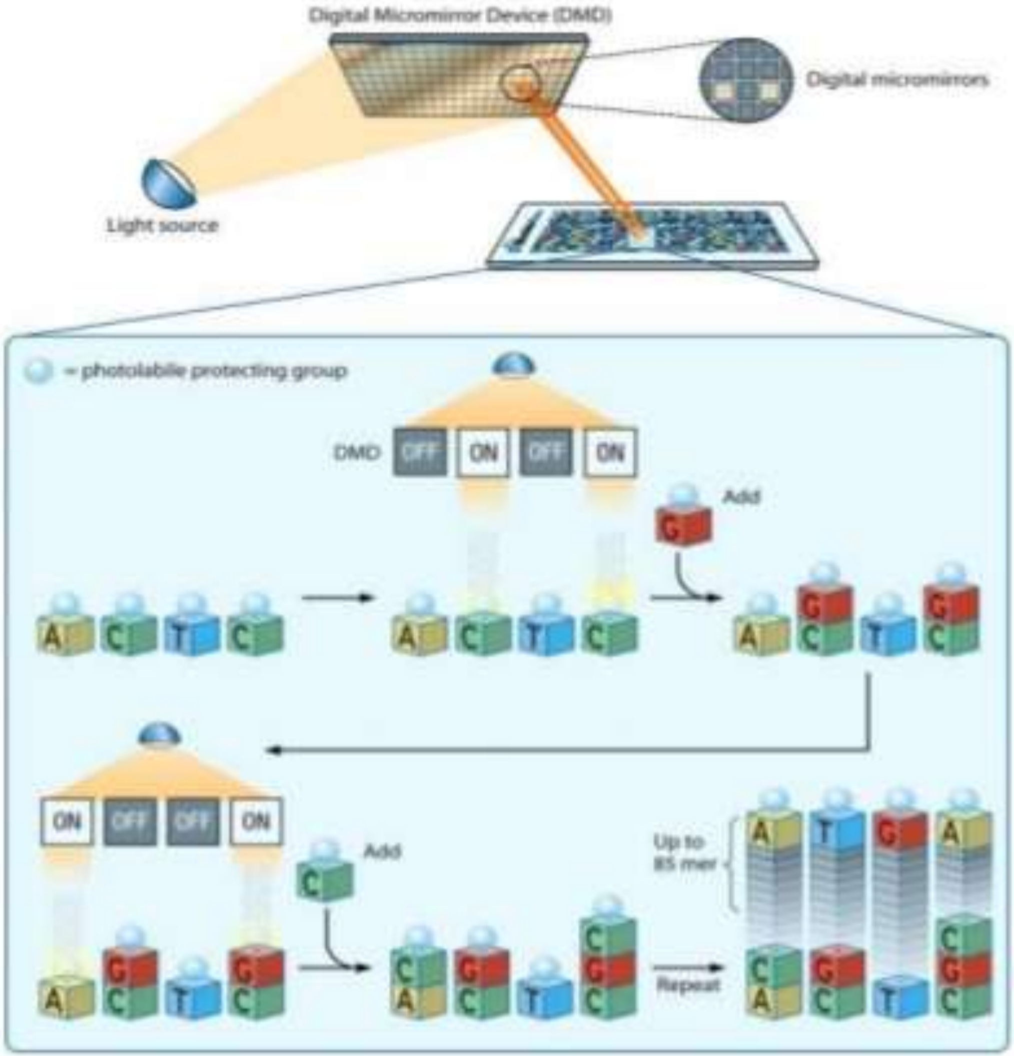
Re-evaluate what gene expression microarrays can bring to your research

Clariom Assays

The power of Clariom assays

Recent advanced transcriptome analyses have uncovered thousands of splice variants and long non-coding (lnc)RNAs, providing new sources for biomarker discovery. Given the complexity of the transcriptome, however, finding informative expression biomarkers is challenging, time-consuming, and costly. Clariom assays, built using the latest transcriptome knowledge from multiple databases, are simple and fast tools for finding high-fidelity expression biomarkers. They are compatible with clinical research samples, available in scalable formats for different throughput needs, and include flexible, intuitive software for fast and simple analysis.

Roche Nimblegen Oligonucleotide Microarray photolitography

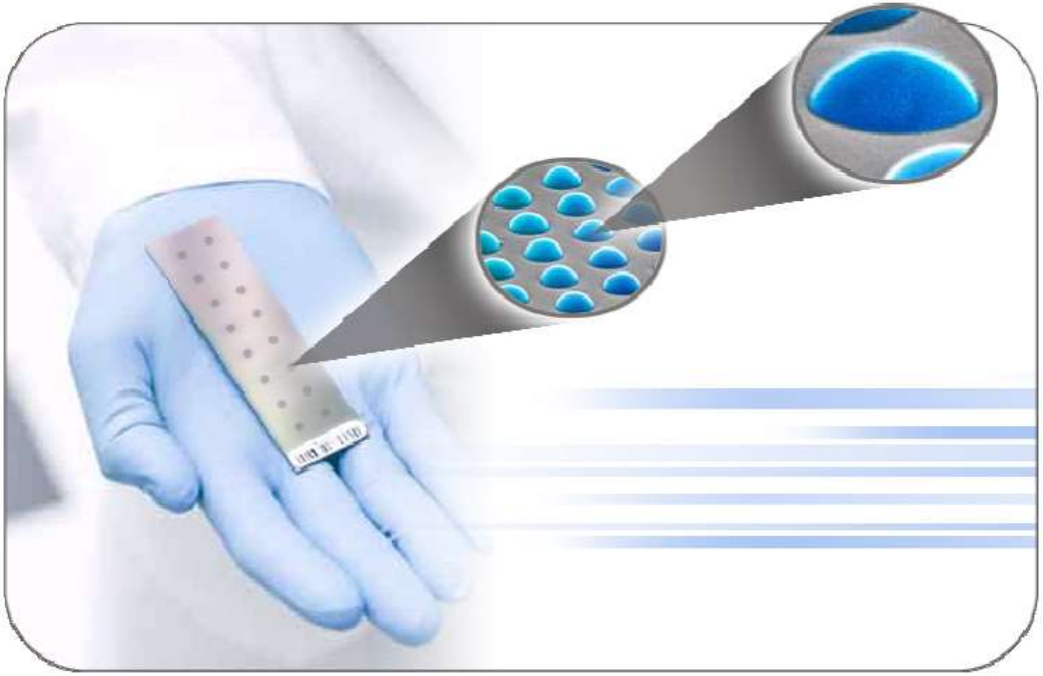


Utilizes a digital micromirror device (DMD) to create virtual masks. The DMD forms the pattern of UV light needed to direct the specific nucleic acid addition during photo-mediated synthesis. UV light removes the photolabile protecting group (circles) from the microarray surface, allowing the addition of a single protected nucleotide to the growing oligonucleotide chain. The cycling of DMD filtering, light deprotection, and nucleotide addition creates oligonucleotide features 60 to 100 bp in length on the NimbleGen microarray

NimbleGen Arrays

ILLUMINA

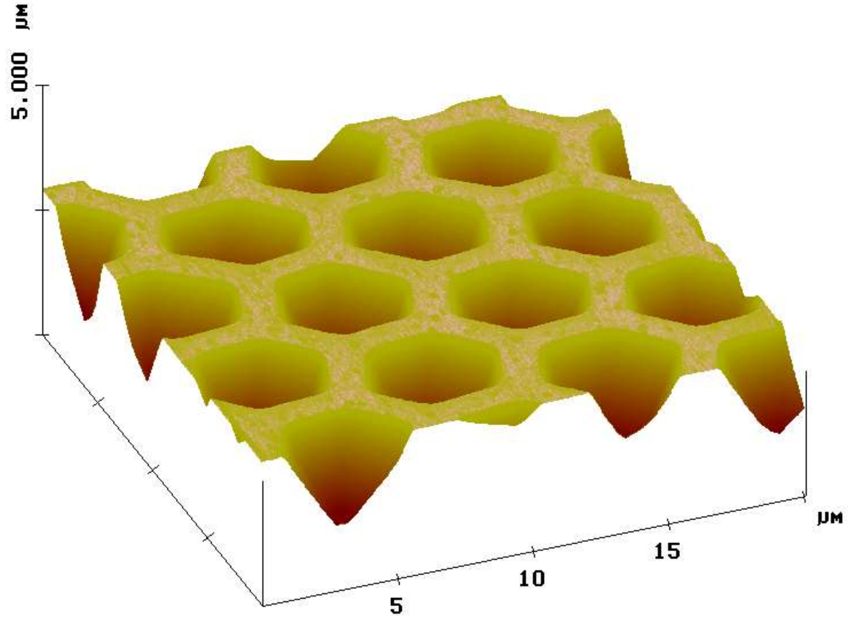
BeadArray^{TM} Technology

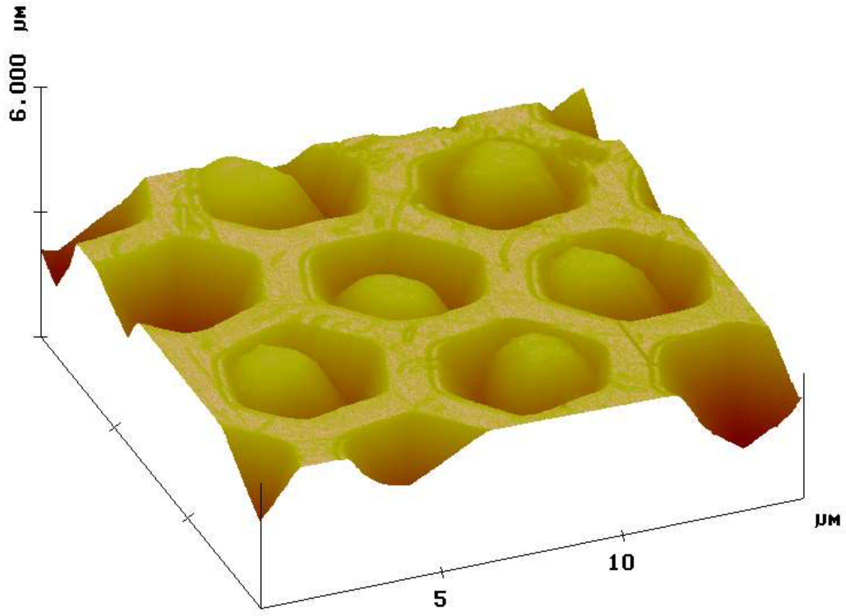


3-micron silica beads that self assemble in microwells on fiber optic bundles or planar silica slides.

When randomly assembled the beads have a uniform spacing of ~5.7 microns.

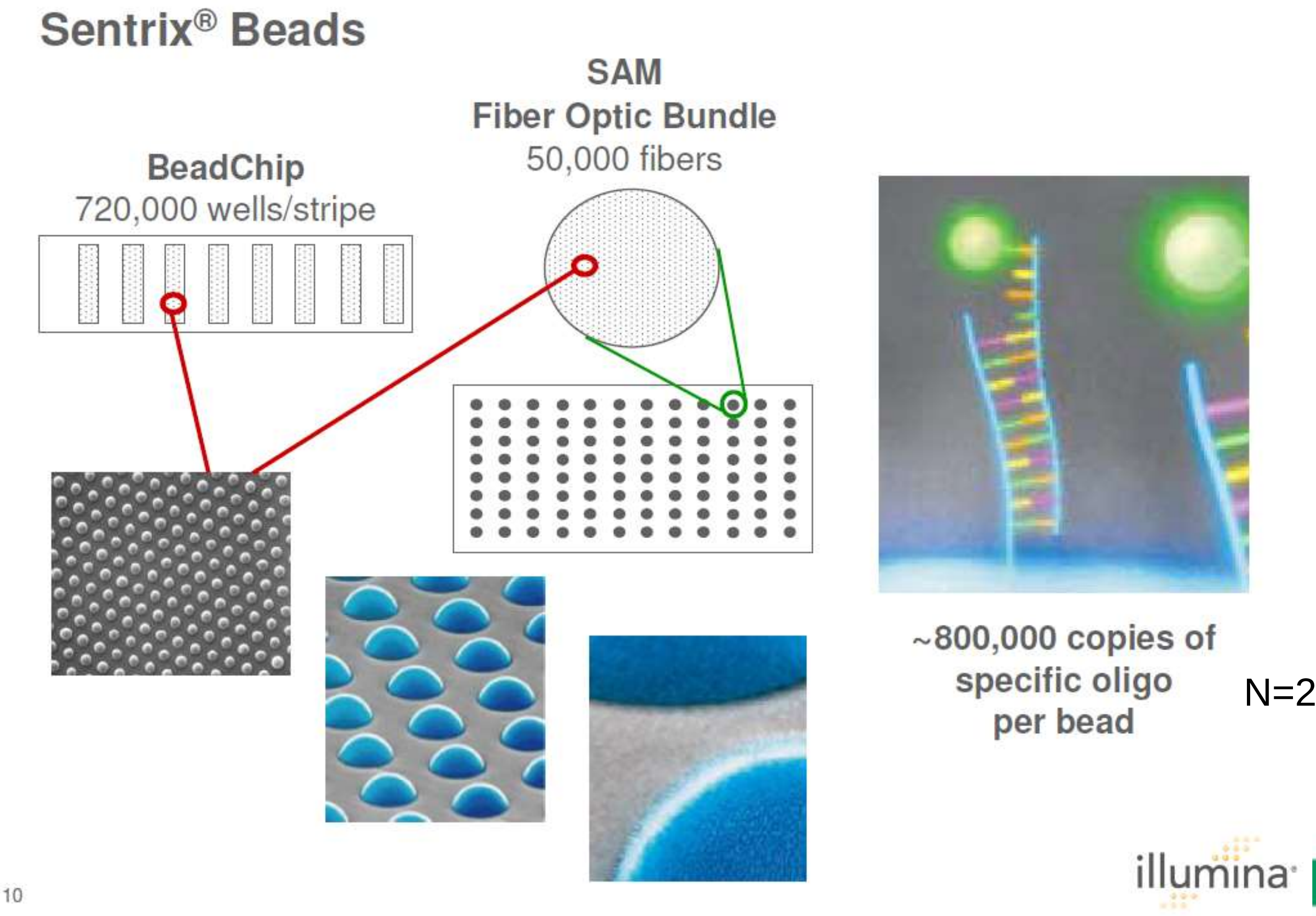
Optical fiber with wells…





… and BEADS!

ILLUMINA



Sentrix® Beads

SAM
Fiber Optic Bundle
50,000 fibers

BeadChip
720,000 wells/stripe

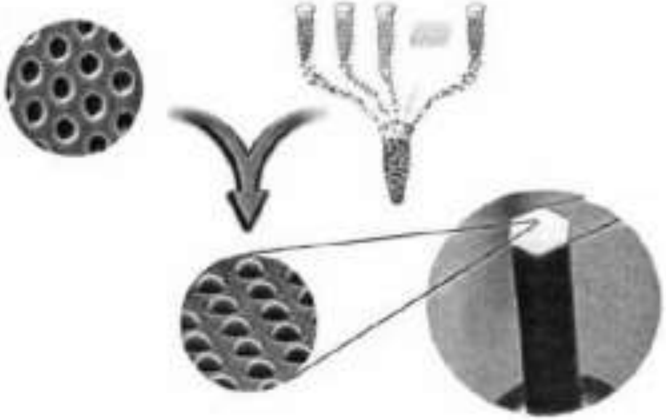
~800,000 copies of specific oligo per bead

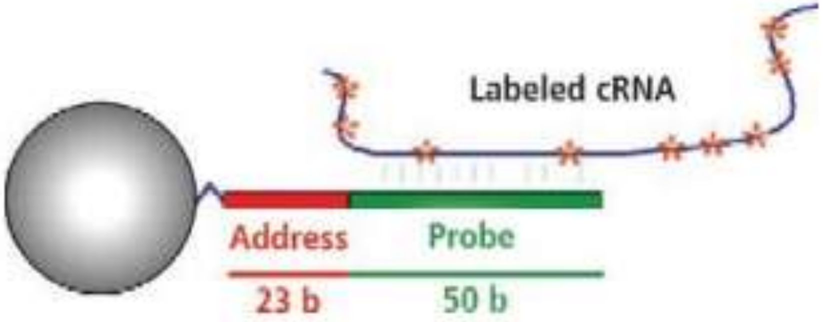
N=20

ILLUMINA

Manufacturing the arrays

- Long oligonucleotides attached to glass beads
- 50 nucleotide probe
- 23 nucleotide address (bead ID)
- Pooled together in beadpools
- Random assembly in etched wells





ILLUMINA

Direct Hyb: Whole Genome Expression Arrays

HumanRef-8
Mouse-Ref8
RatRef-12

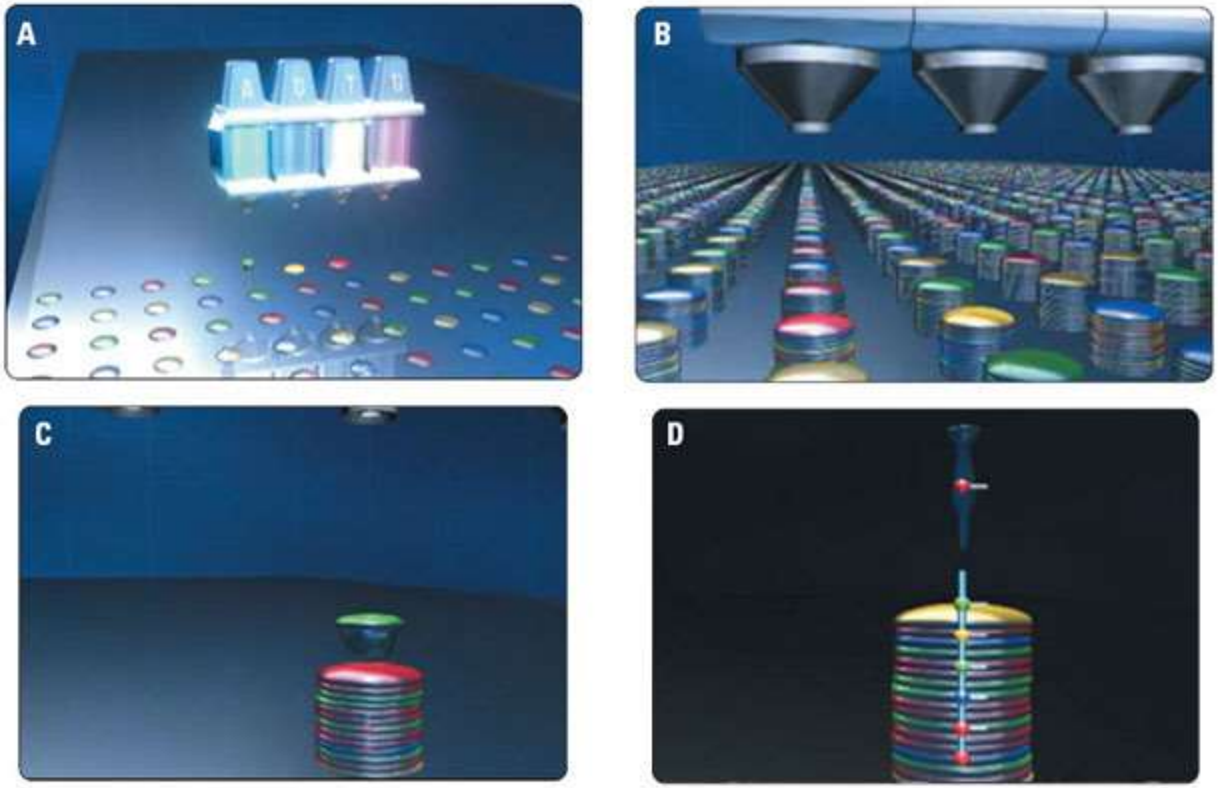
- RefSeq BeadChip
- 8 or 12 samples per slide
- >22,000 sequences

HumanWG-6
MouseWG-6

- Whole Genome BeadChip
- 6 samples per slide
- >46,000 sequences
- >10 million features

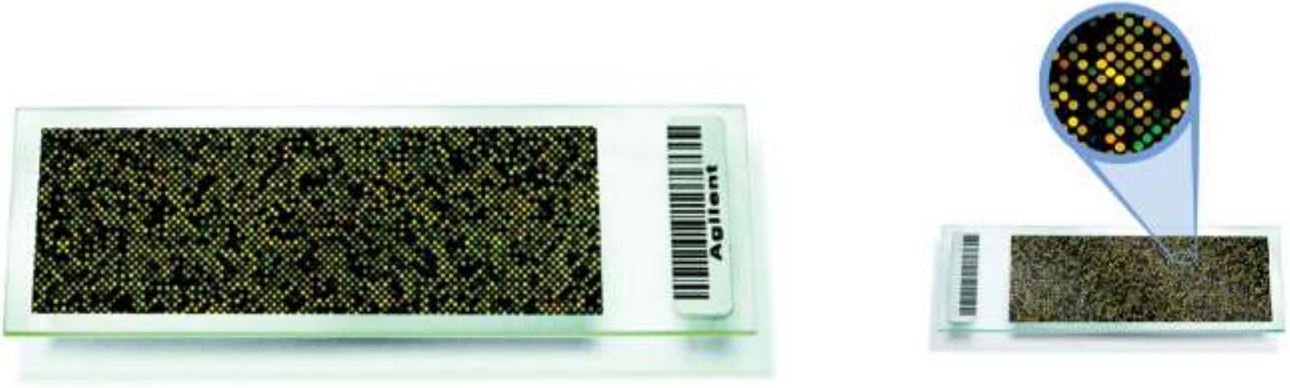
Agilent arrays

Oligo’s synthesized on chip by ink-jet printing

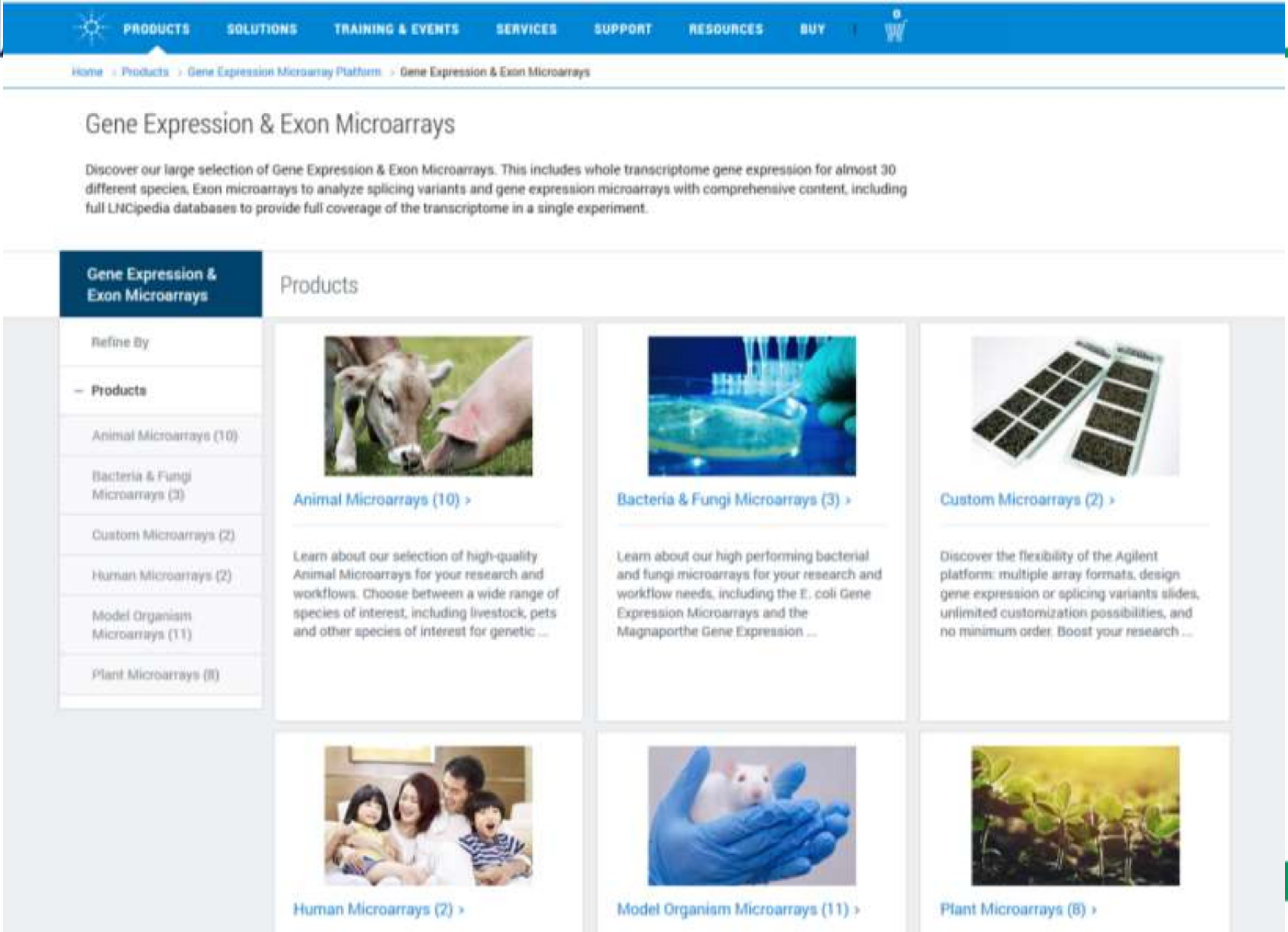


Agilent arrays

Each gene represented by single long oligo (60 - 70-mer)



Agilent arrays



Gene Expression & Exon Microarrays

Discover our large selection of Gene Expression & Exon Microarrays. This includes whole transcriptome gene expression for almost 30 different species, Exon microarrays to analyze splicing variants and gene expression microarrays with comprehensive content, including full LNCipedia databases to provide full coverage of the transcriptome in a single experiment.

Products

Animal Microarrays (10) ›

Learn about our selection of high-quality Animal Microarrays for your research and workflows. Choose between a wide range of species of interest, including livestock, pets and other species of interest for genetic ...

Bacteria & Fungi Microarrays (3) ›

Learn about our high performing bacterial and fungi microarrays for your research and workflow needs, including the E. coli Gene Expression Microarrays and the Magnaporthe Gene Expression ...

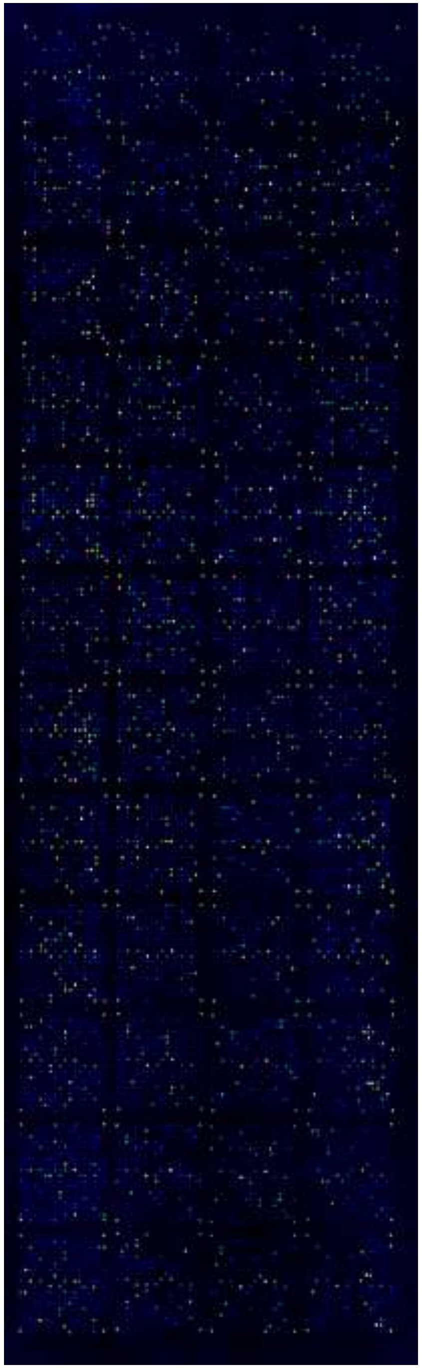
Custom Microarrays (2) ›

Discover the flexibility of the Agilent platform: multiple array formats, design gene expression or splicing variants slides, unlimited customization possibilities, and no minimum order. Boost your research ...

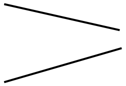
Human Microarrays (2) ›

Model Organism Microarrays (11) ›

Plant Microarrays (8) ›



Dual color
Microarray hybridizations

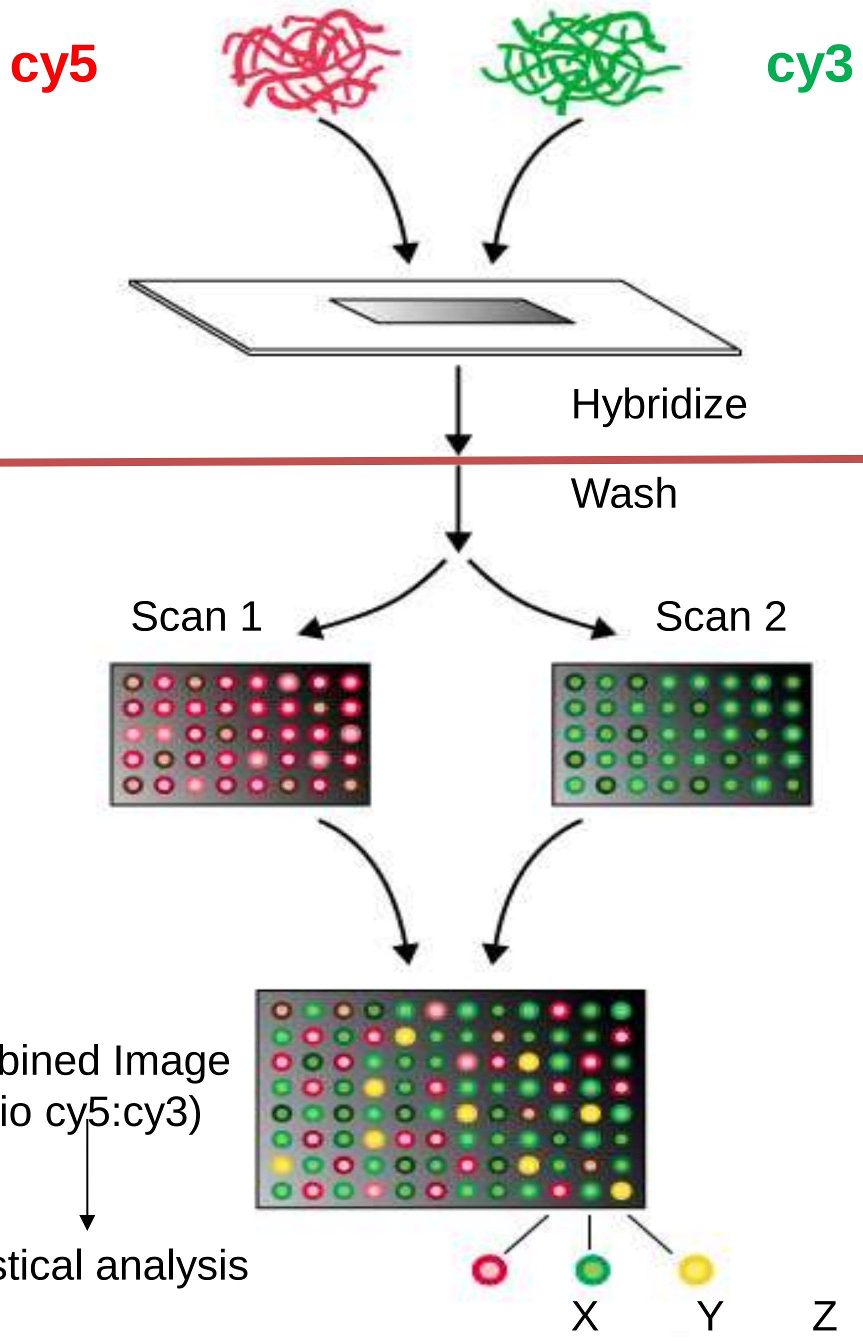


1.

Sample A	cDNA A Cy3	one microarray
Sample B	cDNA B Cy5	

2.

Measurements are <u>relative</u> i.e. a change is measured, not an absolute amount



cy5

cy3

Hybridize

Wash

Scan 1

Scan 2

Combined Image
(ratio cy5:cy3)

Statistical analysis

X Y Z

🟡 **equal expression**

🟢 **higher expression in Cy3**

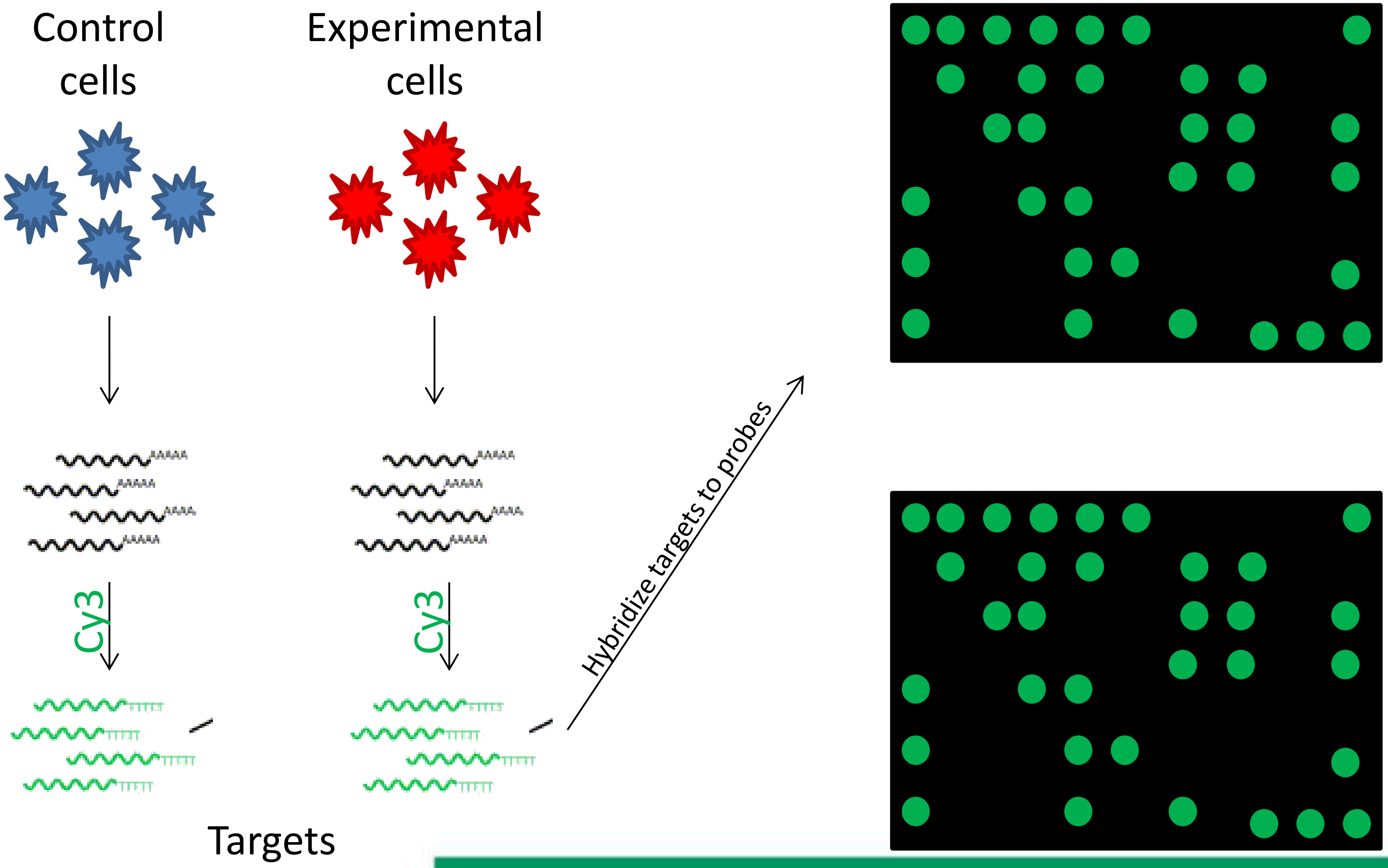
🔴 **higher expression in Cy5**

Microarray: Workflow

Overview of data capture

- two different mRNA populations, labeled with different fluors
- excited by a laser
- each fluor excites at a different wavelength, which is captured using a photo detector attached to a filter tuned to the particular fluor

ILLUMINA



Control cells

Experimental cells

Cy3

Cy3

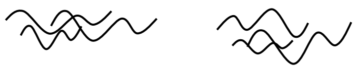
Hybridize targets to probes

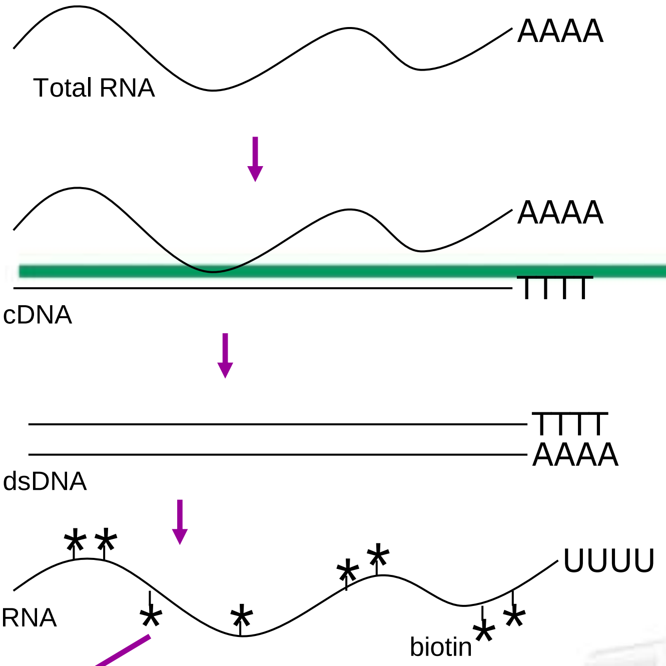
Targets

Microarray

Illumina DirectHyb

Sample 1 Sample 2

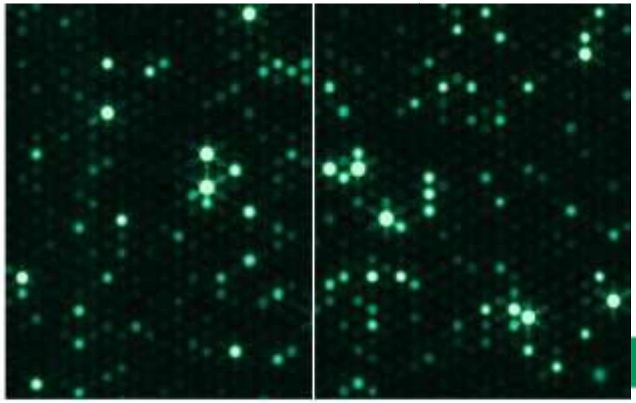




Cy-3-Streptavidin

Scanner

24 min/BeadChip

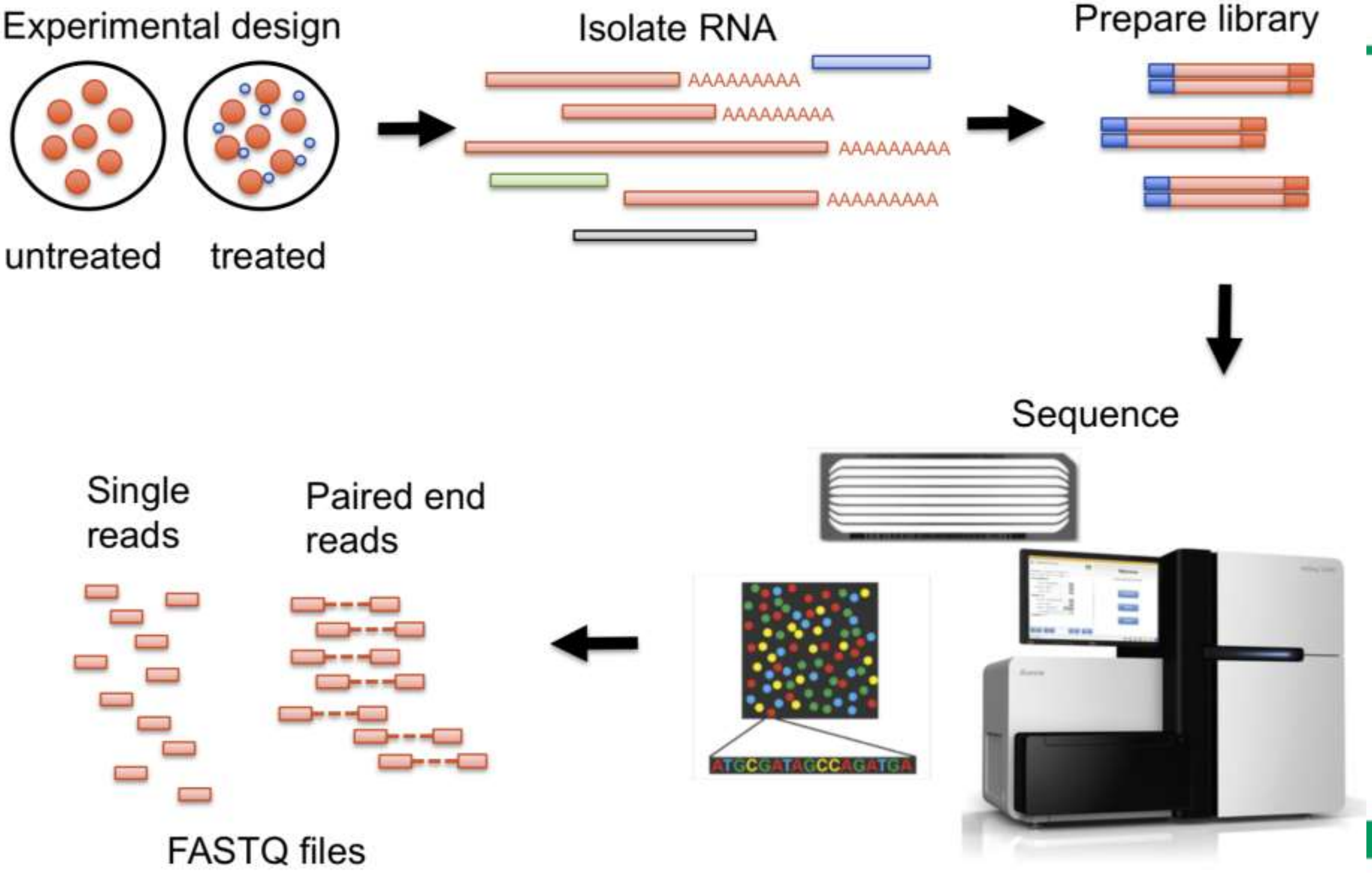


GenomeStudio

Differential gene expression


0
1
2

illumina®



Experimental design

untreated

treated

Isolate RNA

Prepare library

Sequence

Single reads

Paired end reads

FASTQ files

RNA-seq: overview

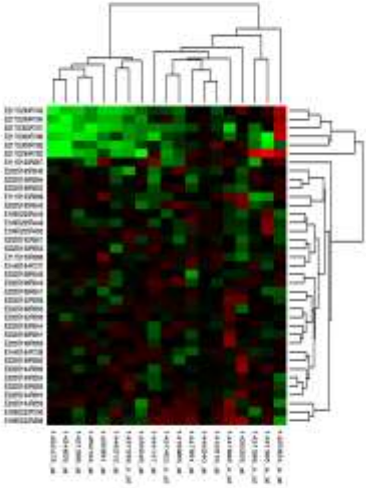
- Next generation sequencing approach that offers a snapshot of the entire transcriptome or messenger RNA (mRNA) profile at a given moment in time.
- The term RNA-Seq is frequently inaccurately used, as RNA is not directly sequenced.
- Single RNA strands are converted to complementary DNAs (cDNA) and then turned into double stranded DNA before being sequenced. So while the initial starting input material is RNA, material loaded on the sequencing instrument is DNA.

https://genohub.com/rna-seq-library-preparation/

RNA-Seq: general workflow

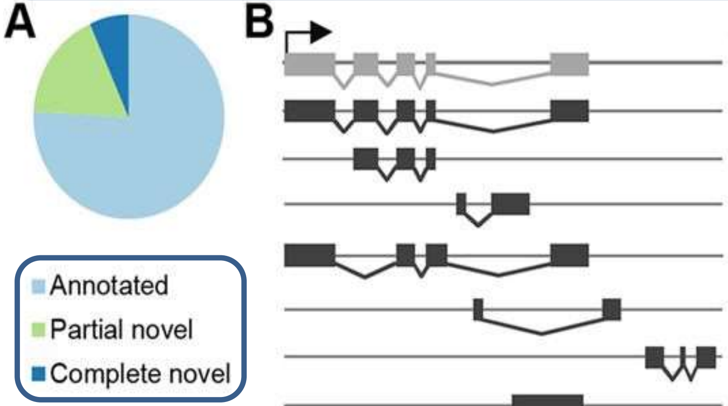
- RNA isolation
- RNA quality and quantity assesment
- Library preparation
- Library QC, quantification and pooling
- Sequencing run
- QC of the run
- Post-run QC of libraries
- Data analysis

Goals of sequencing the transcriptome



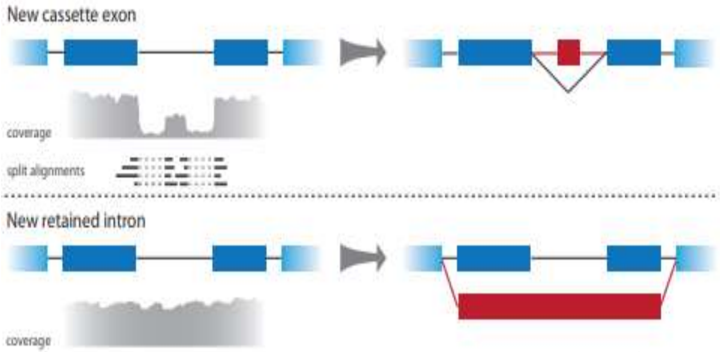
Quantification
Abundance of transcripts
between different conditions/tissues

Annotation – identify genes, exons, detect novel transcripts, transcription start and end sites etc.



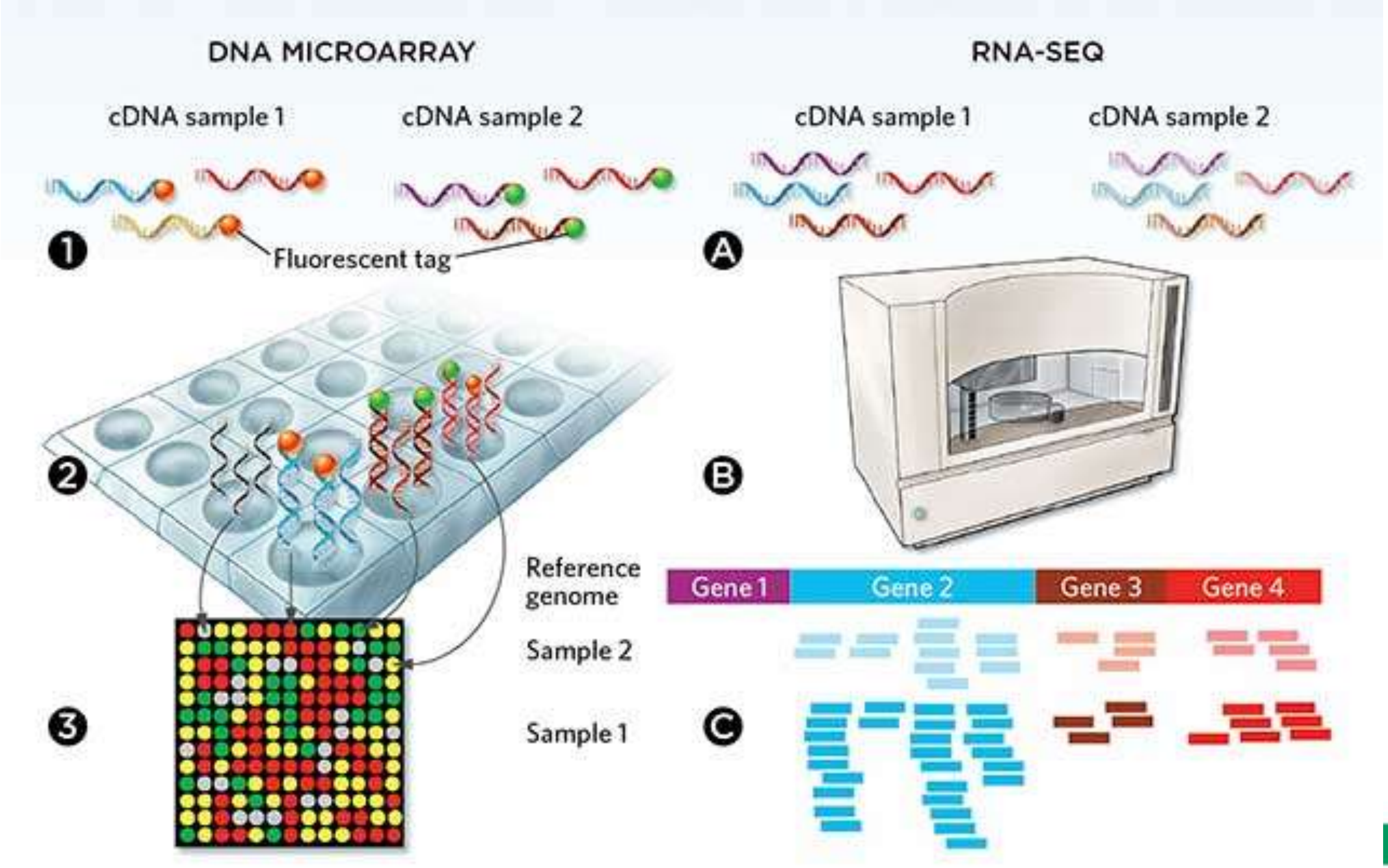
Werber M 2014

Annotation - identify splicing events

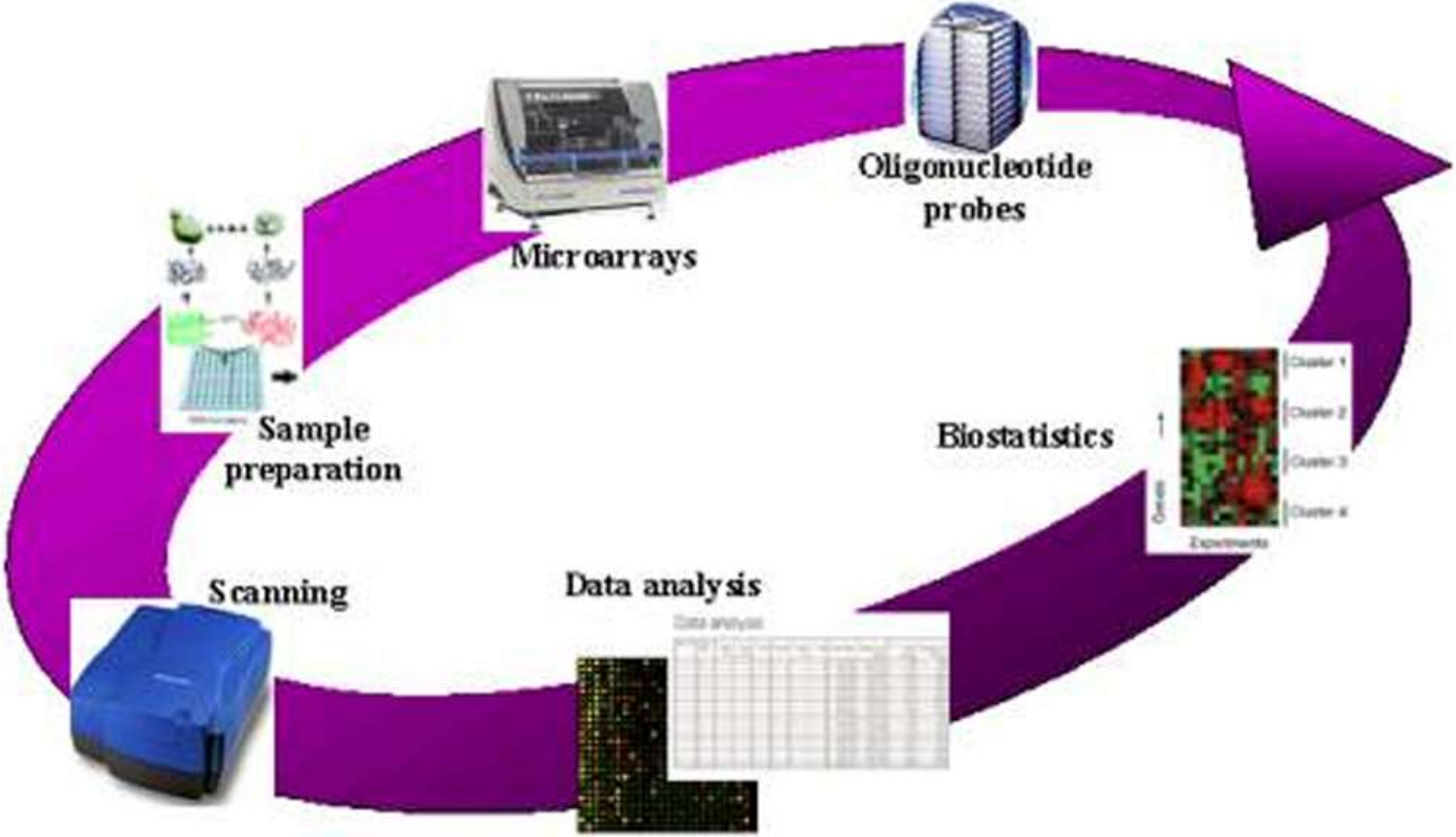


Kahles A 2016

Microarray vs NG-Sequencing



Microarray: Workflow



Goals of a Microarray Experiment

1. **Find the genes** that change expression between experimental and control samples
2. **Classify samples** based on a gene expression profile
3. **Find patterns:** Groups of biologically related genes that change expression together across samples/treatments
4. **Correlate expression profile** to disease state, diagnosis/prognosis or treatment

Microarray Experiment Design

- **Type I**: (n = 2)
  - How is this gene expressed in target 1 as compared to target 2?
  - Which genes show up/down regulation between the two targets?
- **Type II**: (n > 2)
  - How does the expression of gene A vary over time, tissues, or treatments?
  - Do any of the expression profiles exhibit similar patterns of expression?

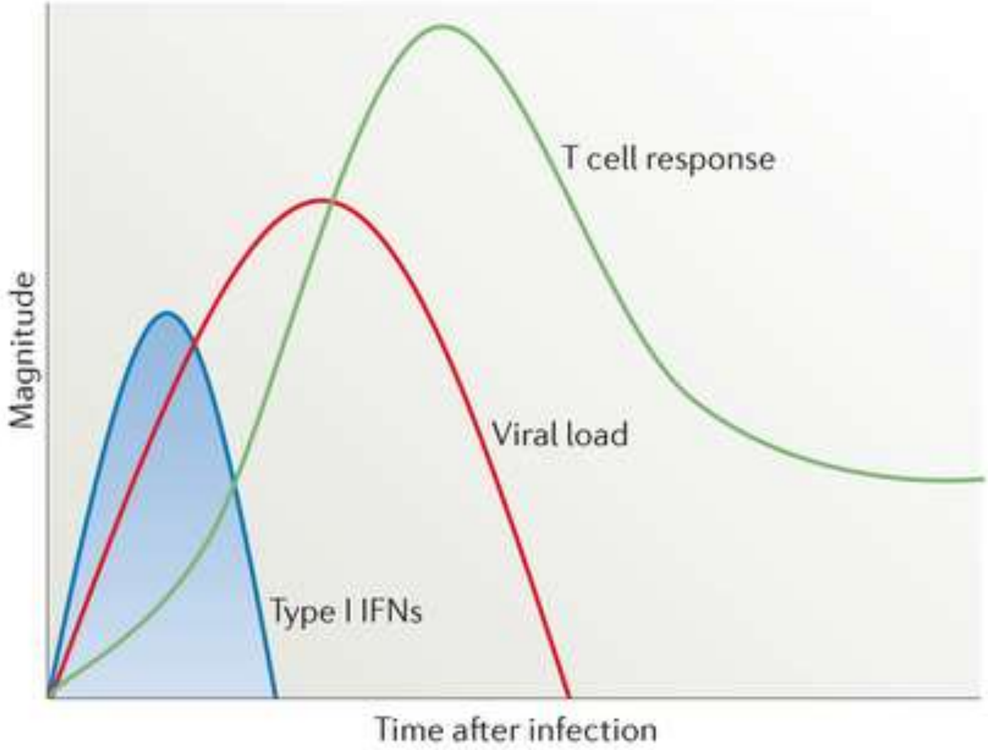
Viral Infection

Small pox

Herpes

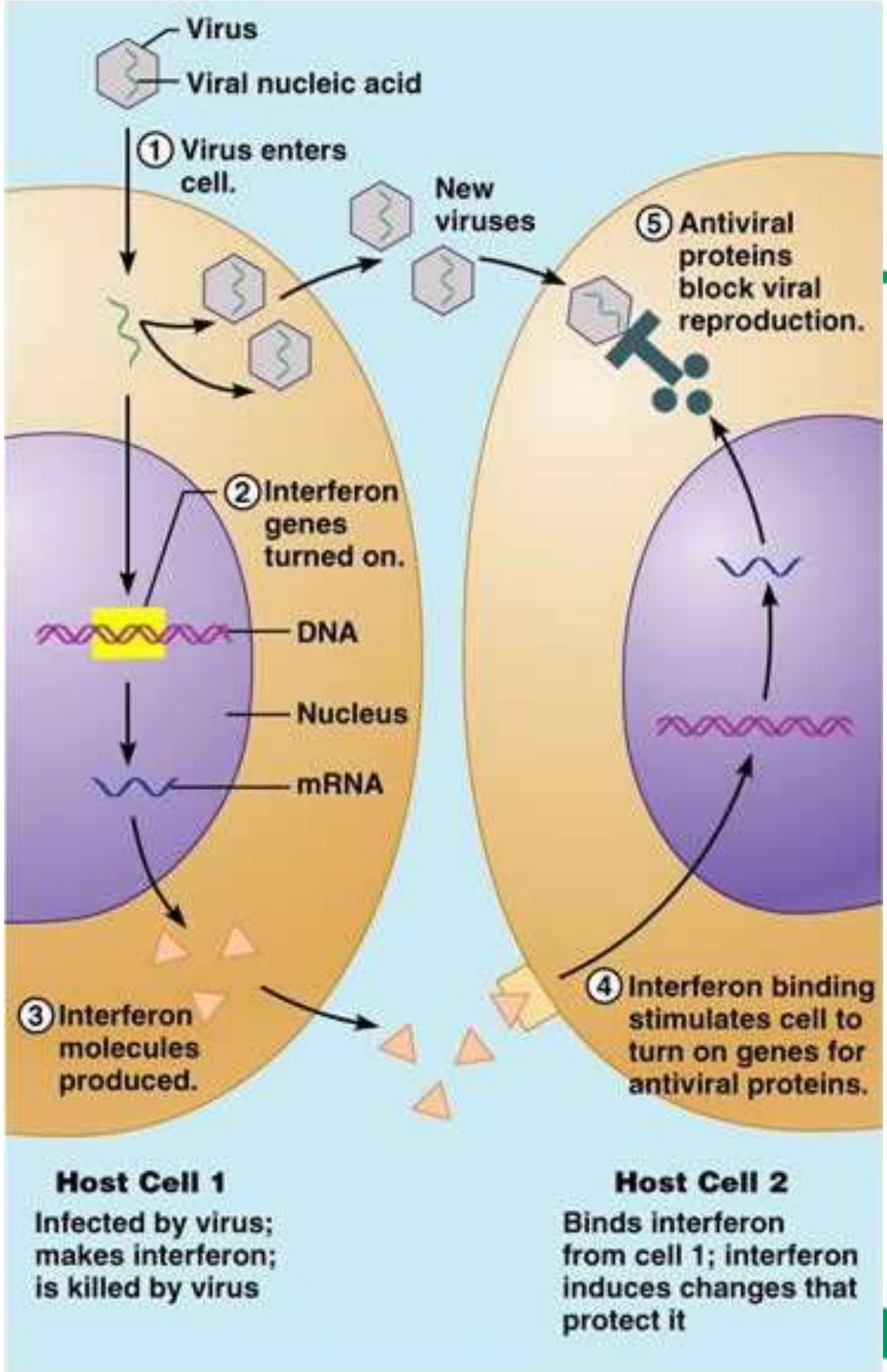
Influenza

Anti-viral Response



Nature Reviews | Immunology

Type I IFN Production & action



Inhibition viral replication

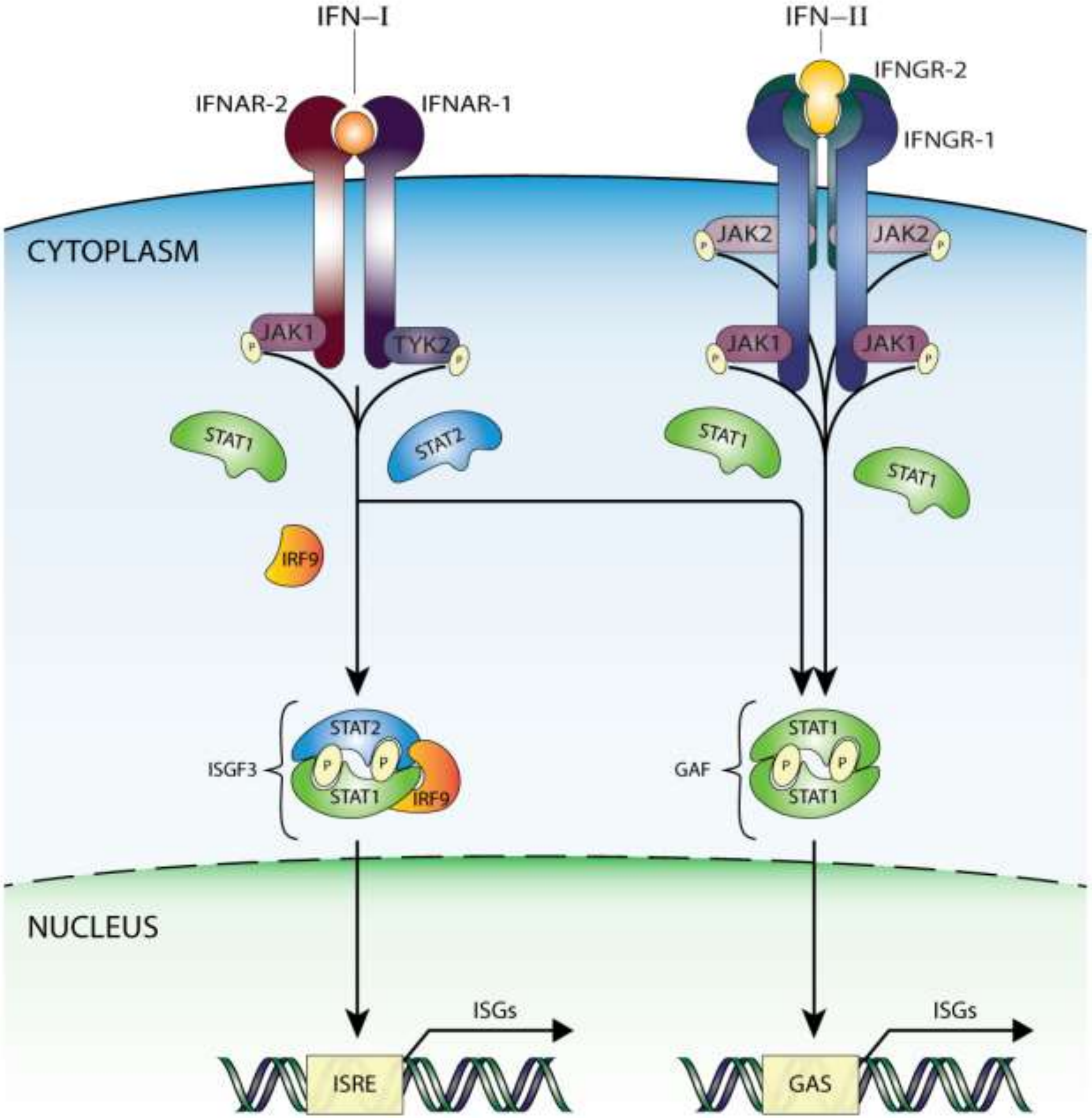
Inhibition cell growth

Activation immune system

↓

Anti-viral State
Adaptive immune response

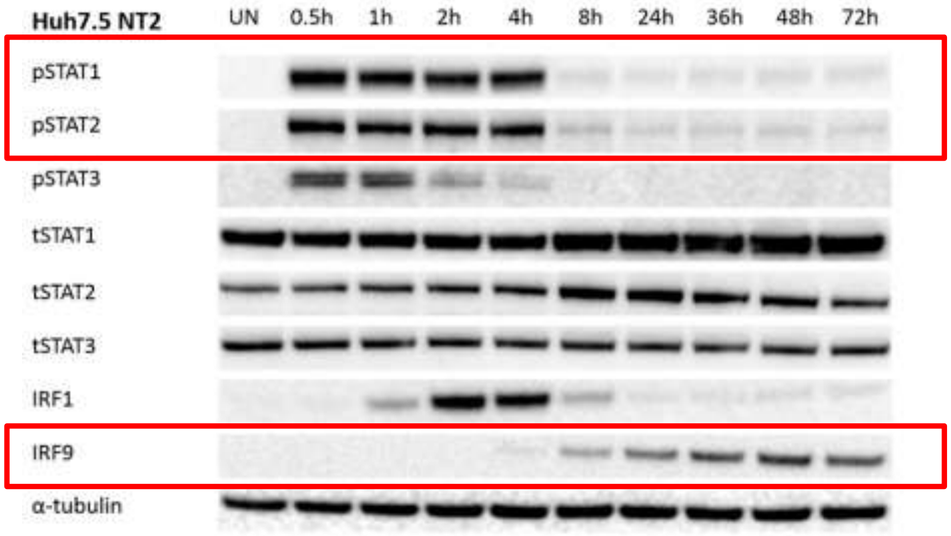
Canonical IFN-signaling (1990)

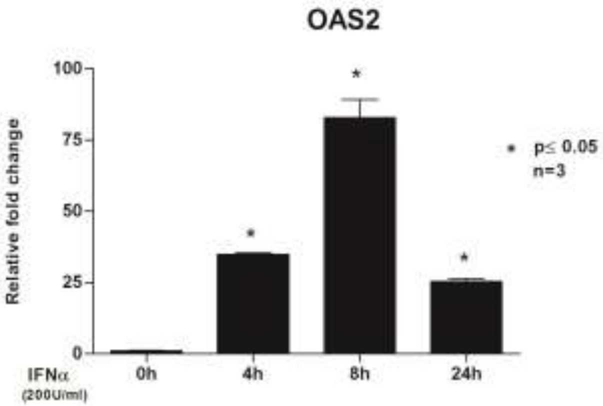


Michalska et al. Front in Immunol. 2018

IFN-signaling:
pSTAT1, pSTAT2 & IRF9

IFN-I



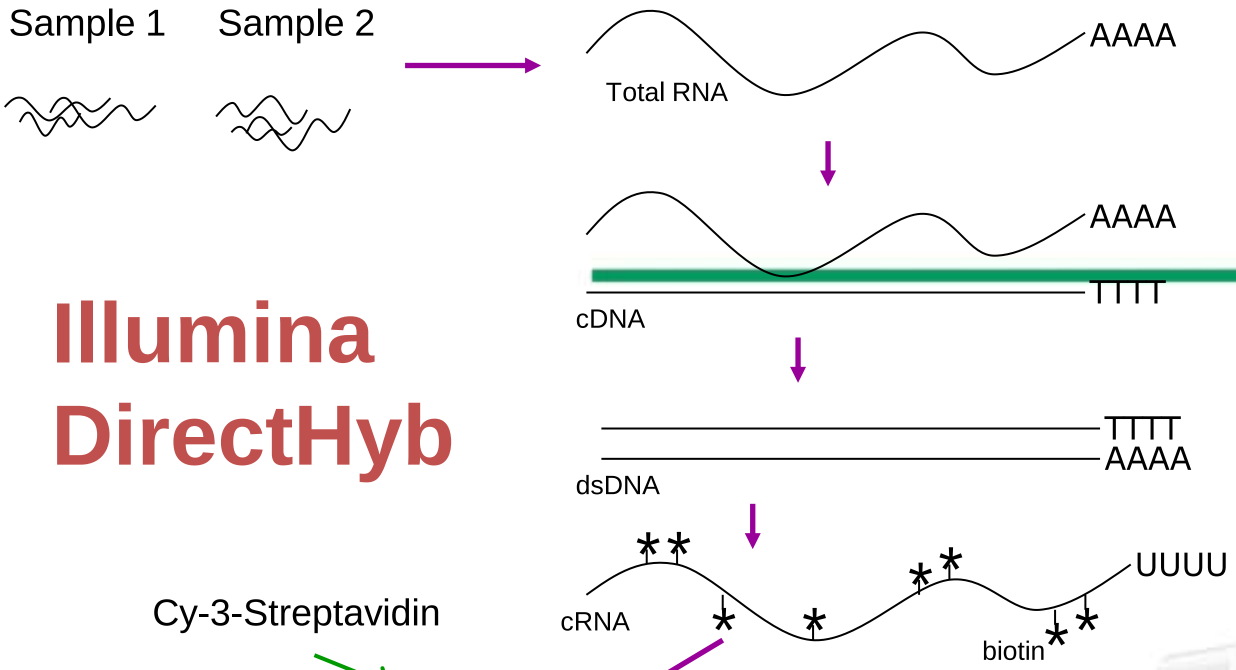


Yamauchi et al., Scientific Rep 2016

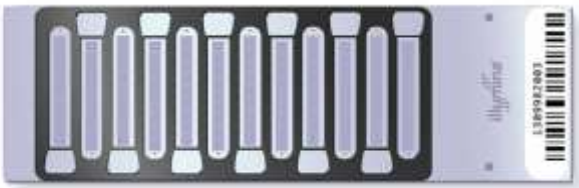
Michalska et al. 2018

Microarray

Illumina DirectHyb

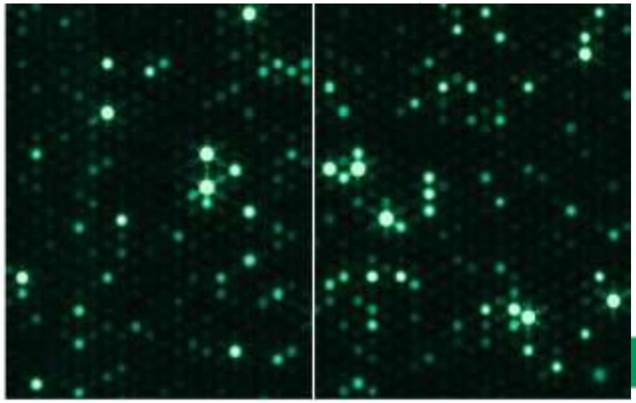


Cy-3-Streptavidin



Scanner

24 min/BeadChip



GenomeStudio

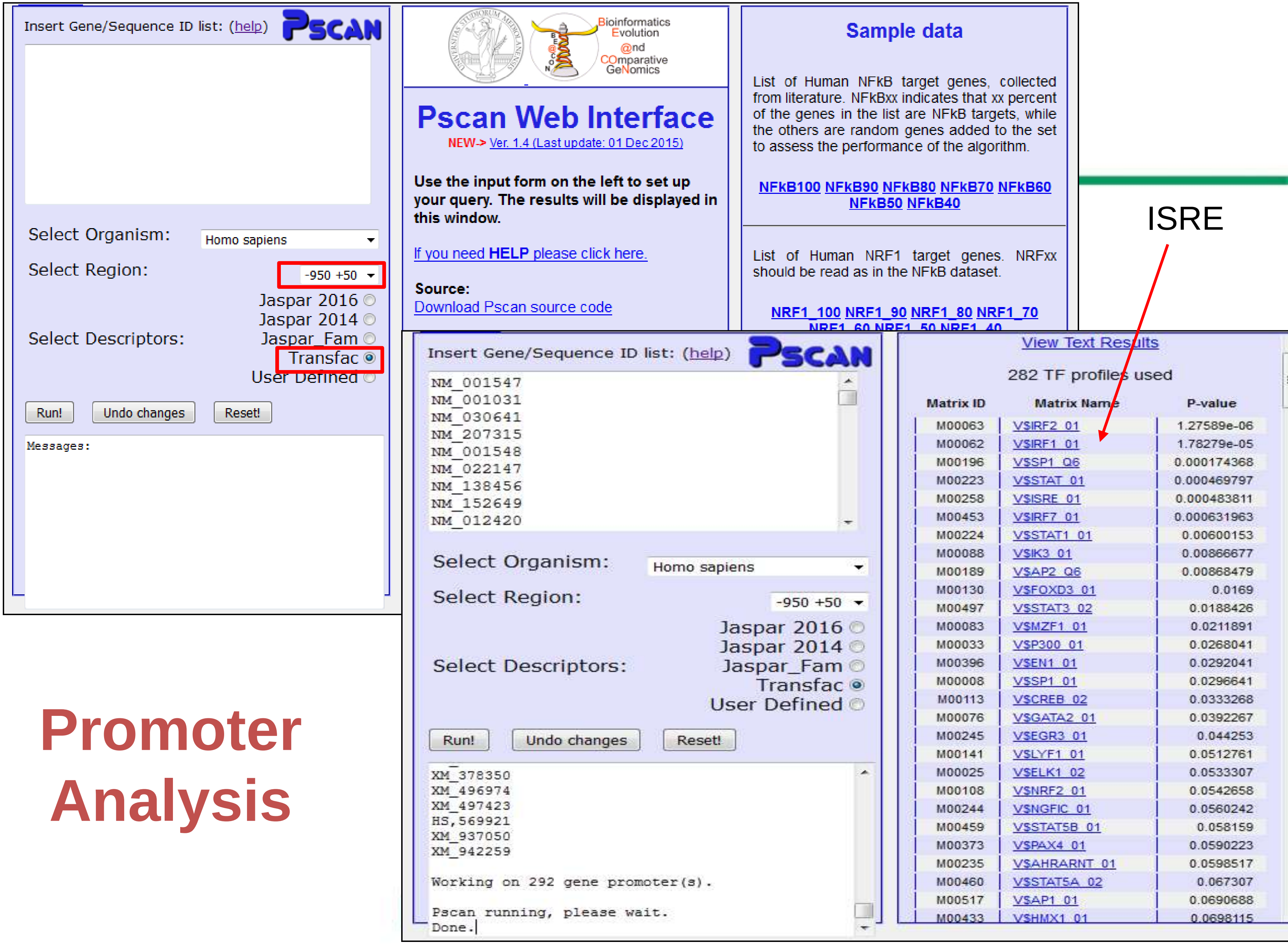
Differential gene expression

Microarray on MEF WT treated with IFNA

SYMBOL	Ratio 0 vs 4	Ratio 0 vs 8	Ratio 0 vs 24	t-test 0 vs 4h	t-test 0 vs 8h	t-test 0 vs 24h	AVG detection p_value 4h	AVG detection p_value 8h	AVG detection p_value 24h
Cxcl9	262.44	136.17	30.09	0.00	0.00	0.01	0.00	0.00	0.03
Batf2	150.80	71.53	30.59	0.00	0.00	0.01	0.00	0.00	0.03
Oas1a	131.77	92.41	135.83	0.00	0.00	0.00	0.00	0.00	0.00
Oasl1	109.70	58.69	44.61	0.01	0.01	0.01	0.00	0.00	0.00
Rsad2	109.01	71.63	53.57	0.00	0.00	0.00	0.00	0.00	0.00
Serpina3f	84.20	10.51	1.89	0.04	0.11	0.85	0.00	0.01	0.31
Ifi47	79.59	32.72	21.20	0.00	0.01	0.01	0.00	0.00	0.00
Oas2	68.25	112.57	147.72	0.01	0.01	0.00	0.00	0.00	0.00
Gbp10	65.69	45.31	36.51	0.03	0.03	0.04	0.00	0.00	0.00
Tgtp	64.77	21.92	10.95	0.05	0.08	0.11	0.00	0.00	0.00
Gbp5	58.87	17.03	2.46	0.00	0.00	0.42	0.00	0.08	0.40
Cxcl10	57.24	22.72	13.14	0.00	0.00	0.01	0.00	0.00	0.00
Gbp6	51.22	27.93	26.31	0.02	0.04	0.03	0.00	0.00	0.00
LOC435565	48.21	43.05	44.27	0.01	0.01	0.01	0.00	0.00	0.00
Tyki	42.93	25.75	17.71	0.01	0.01	0.01	0.00	0.00	0.00
D14Ertd668e	42.56	35.27	29.67	0.00	0.00	0.00	0.00	0.00	0.00
Mx2	40.47	31.18	30.51	0.00	0.01	0.01	0.00	0.00	0.00
Ifi203	40.37	26.68	15.41	0.00	0.00	0.00	0.01	0.03	0.09
Oas1b	30.35	27.09	22.41	0.03	0.03	0.03	0.00	0.00	0.00

🟥 Anti-viral ISRE containing genes

Promoter Analysis



Insert Gene/Sequence ID list: (help)

Select Organism: Homo sapiens

Select Region: -950 +50

Select Descriptors:
Jaspar 2016 ○
Jaspar 2014 ○
Jaspar_Fam ○
Transfac ●
User Defined ○

Messages:

Bioinformatics Evolution @nd COmparative GeNomics

Pscan Web Interface

NEW-> Ver. 1.4 (Last update: 01 Dec 2015)

Use the input form on the left to set up your query. The results will be displayed in this window.

If you need **HELP** please click here.

Source:
Download Pscan source code

Sample data

List of Human NFkB target genes, collected from literature. NFkBxx indicates that xx percent of the genes in the list are NFkB targets, while the others are random genes added to the set to assess the performance of the algorithm.

NFkB100 NFkB90 NFkB80 NFkB70 NFkB60 NFkB50 NFkB40

List of Human NRF1 target genes. NRFxx should be read as in the NFkB dataset.

NRF1_100 NRF1_90 NRF1_80 NRF1_70 NRF1_60 NRF1_50 NRF1_40

Insert Gene/Sequence ID list: (help)

```
NM_001547
NM_001031
NM_030641
NM_207315
NM_001548
NM_022147
NM_138456
NM_152649
NM_012420
```

Select Organism: Homo sapiens

Select Region: -950 +50

Select Descriptors:
Jaspar 2016 ○
Jaspar 2014 ○
Jaspar_Fam ○
Transfac ●
User Defined ○

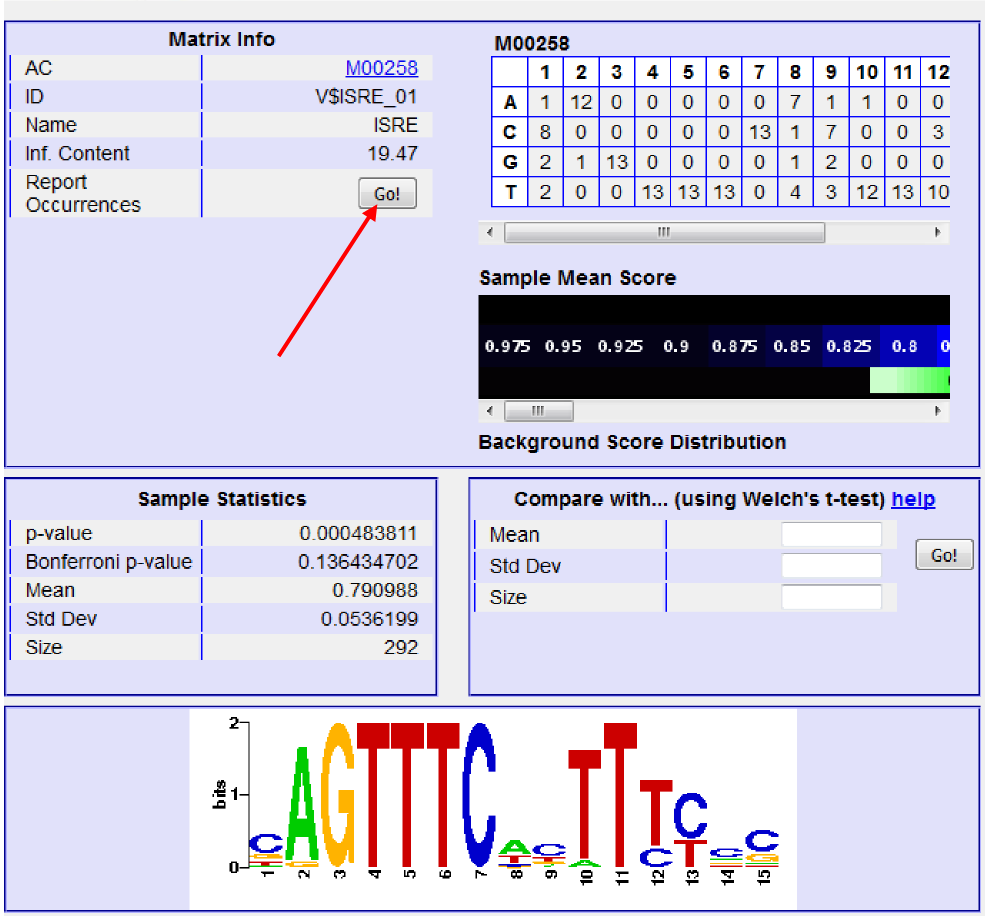
```
XM_378350
XM_496974
XM_497423
HS,569921
XM_937050
XM_942259

Working on 292 gene promoter(s).

Pscan running, please wait.
Done.
```

View Text Results

282 TF profiles used

Matrix ID	Matrix Name	P-value
M00063	V$IRF2_01	1.27589e-06
M00062	V$IRF1_01	1.78279e-05
M00196	V$SP1_Q6	0.000174368
M00223	V$STAT_01	0.000469797
M00258	V$ISRE_01	0.000483811
M00453	V$IRF7_01	0.000631963
M00224	V$STAT1_01	0.00600153
M00088	V$IK3_01	0.00866677
M00189	V$AP2_Q6	0.00868479
M00130	V$FOXD3_01	0.0169
M00497	V$STAT3_02	0.0188426
M00083	V$MZF1_01	0.0211891
M00033	V$P300_01	0.0268041
M00396	V$EN1_01	0.0292041
M00008	V$SP1_01	0.0296641
M00113	V$CREB_02	0.0333268
M00076	V$GATA2_01	0.0392267
M00245	V$EGR3_01	0.044253
M00141	V$LYF1_01	0.0512761
M00025	V$ELK1_02	0.0533307
M00108	V$NRF2_01	0.0542658
M00244	V$NGFIC_01	0.0560242
M00459	V$STAT5B_01	0.058159
M00373	V$PAX4_01	0.0590223
M00235	V$AHRARNT_01	0.0598517
M00460	V$STAT5A_02	0.067307
M00517	V$AP1_01	0.0690688
M00433	V$HMX1_01	0.0698115

ISRE containing genes



Matrix Info

AC	M00258
ID	V$ISRE_01
Name	ISRE
Inf. Content	19.47
Report Occurrences	Go!

M00258

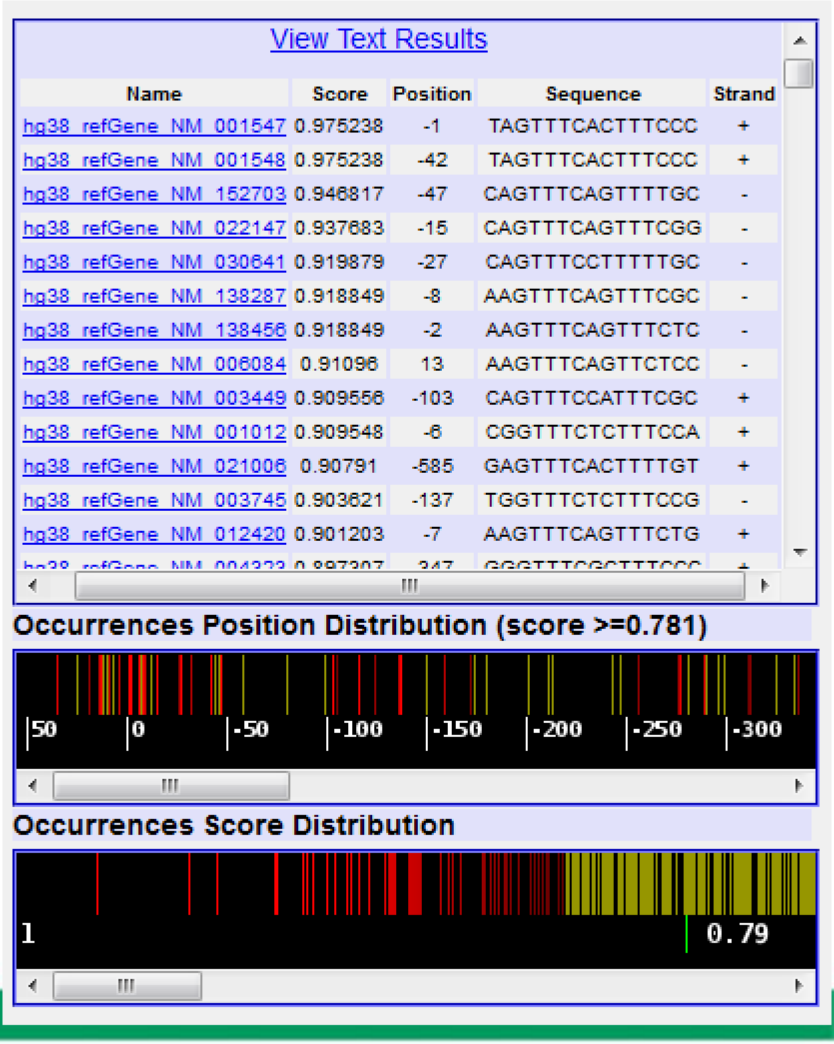
	1	2	3	4	5	6	7	8	9	10	11	12
A	1	12	0	0	0	0	0	7	1	1	0	0
C	8	0	0	0	0	0	13	1	7	0	0	3
G	2	1	13	0	0	0	0	1	2	0	0	0
T	2	0	0	13	13	13	0	4	3	12	13	10

Sample Statistics

p-value	0.000483811
Bonferroni p-value	0.136434702
Mean	0.790988
Std Dev	0.0536199
Size	292

Compare with... (using Welch's t-test) help

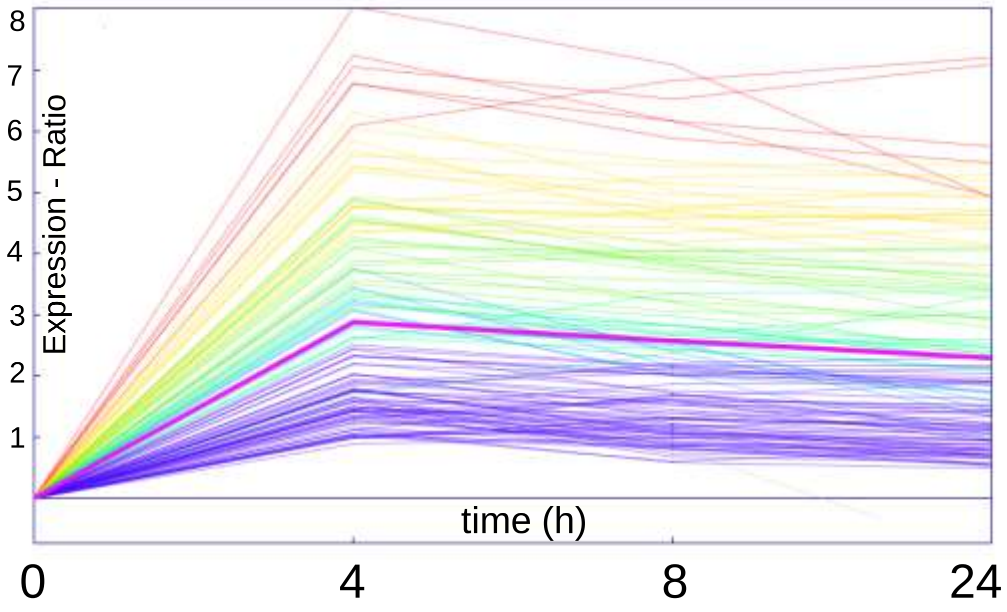
Mean	
Std Dev	
Size	



View Text Results

Name	Score	Position	Sequence	Strand
hg38_refGene_NM_001547	0.975238	-1	TAGTTTCACTTTCCC	+
hg38_refGene_NM_001548	0.975238	-42	TAGTTTCACTTTCCC	+
hg38_refGene_NM_152703	0.946817	-47	CAGTTTCAGTTTTGC	-
hg38_refGene_NM_022147	0.937683	-15	CAGTTTCAGTTTCGG	-
hg38_refGene_NM_030641	0.919879	-27	CAGTTTCCTTTTTGC	-
hg38_refGene_NM_138287	0.918849	-8	AAGTTTCAGTTTCGC	-
hg38_refGene_NM_138456	0.918849	-2	AAGTTTCAGTTTCTC	-
hg38_refGene_NM_006084	0.91096	13	AAGTTTCAGTTCTCC	-
hg38_refGene_NM_003449	0.909556	-103	CAGTTTCCATTTCGC	+
hg38_refGene_NM_001012	0.909548	-6	CGGTTTCTCTTTCCA	+
hg38_refGene_NM_021006	0.90791	-585	GAGTTTCACTTTTGT	+
hg38_refGene_NM_003745	0.903621	-137	TGGTTTCTCTTTCCG	-
hg38_refGene_NM_012420	0.901203	-7	AAGTTTCAGTTTCTG	+

ISG Expression in IFN-I treated cells



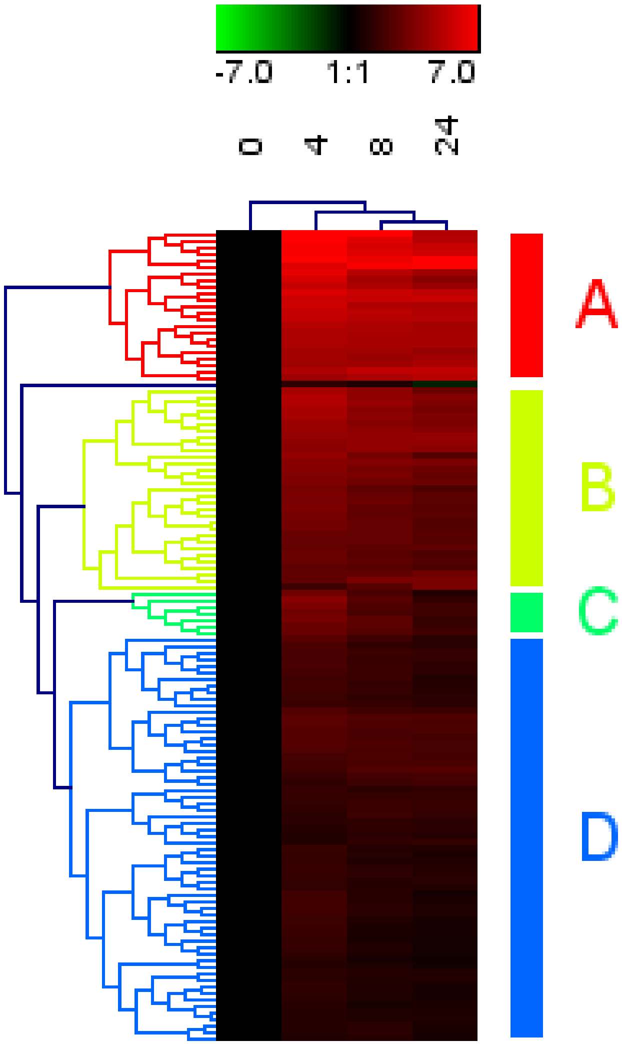
Expression view – 124 up regulated genes

Gene Ontology

ter_ID	Description	$\log_{10}$ p-value
GO:0002376	immune system process	-29.12
GO:0051607	defense response to virus	-28.38
GO:0002252	immune effector process	-25.61
GO:0006955	immune response	-21.24
GO:0045087	innate immune response	-19.97
GO:0019882	antigen processing and presentation	-10.24
GO:0002682	regulation of immune system process	-10.01
GO:0042089	cytokine biosynthetic process	-5.04
GO:0042107	cytokine metabolic process	-4.91
GO:0009617	response to bacterium	-4.44
GO:0032608	interferon-beta production	-4
GO:0032606	type I interferon production	-4
GO:0045343	regulation of MHC class I biosynthetic process	-3.23

Anti viral response

ISG Expression in MEF WT after IFNA treatment



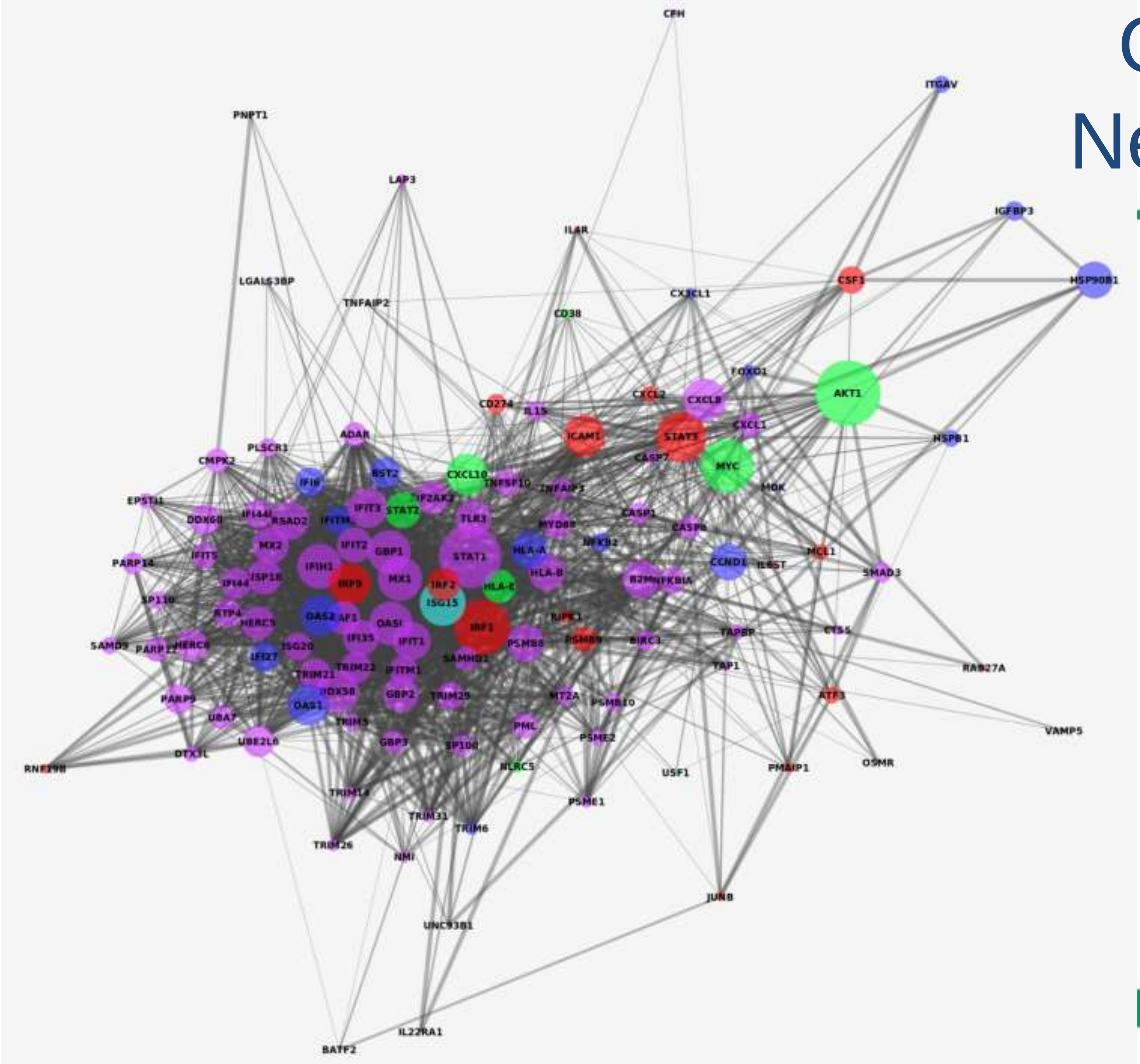
Cluster	Genes
A	Bst2
	Oasl1
	Rsad2
	Oas1a
	Oas2
	Ifi47
	Cxcl10
	Tykl
	Gbp10
	LOC435565
	Gbp6
	D14Ertd668e
	Mx2
	Oas1b
	Irf7
	LOC100048346
	Usp18
	LOC667370
	Trim30
	Oasl2
	Sp100
	Oas1g

Er O Terms within specific cluster

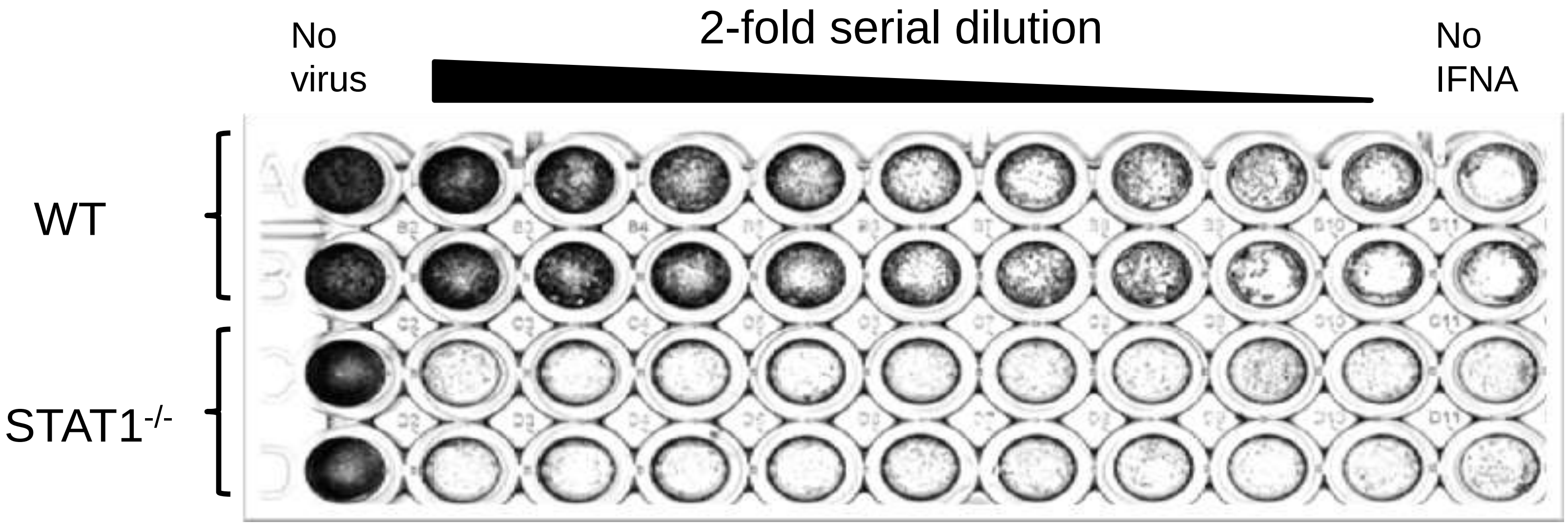
Cluster	GO Term		p-value
A	defense response		4.72E-17
	defense response to virus		1.63E-15
	response to virus		1.39E-14
B	cellular response to interferon-beta		5.73E-13
	response to interferon-beta		1.86E-12
	Response to type I interferon		4.85E-07
C	immune system process		3.76E-06
	immune response		3.36E-04
	negative regulation of type I nterferon-mediated signaling pathway		8.17E-04
D	ntigen processing and presentation of endogenous peptide antigen		2.44E-09
	ntigen processing and presentation of peptide antigen via MHC class I		2.44E-09
	regulation of T cell mediated cytotoxicity		7.26E-07

Protection against viral infection

Gene Network



Anti-viral assay



Cell lines treated with 200U/ml mIFNA 2-fold dilution (24h)
EMCV MOI=0.1 (18h)

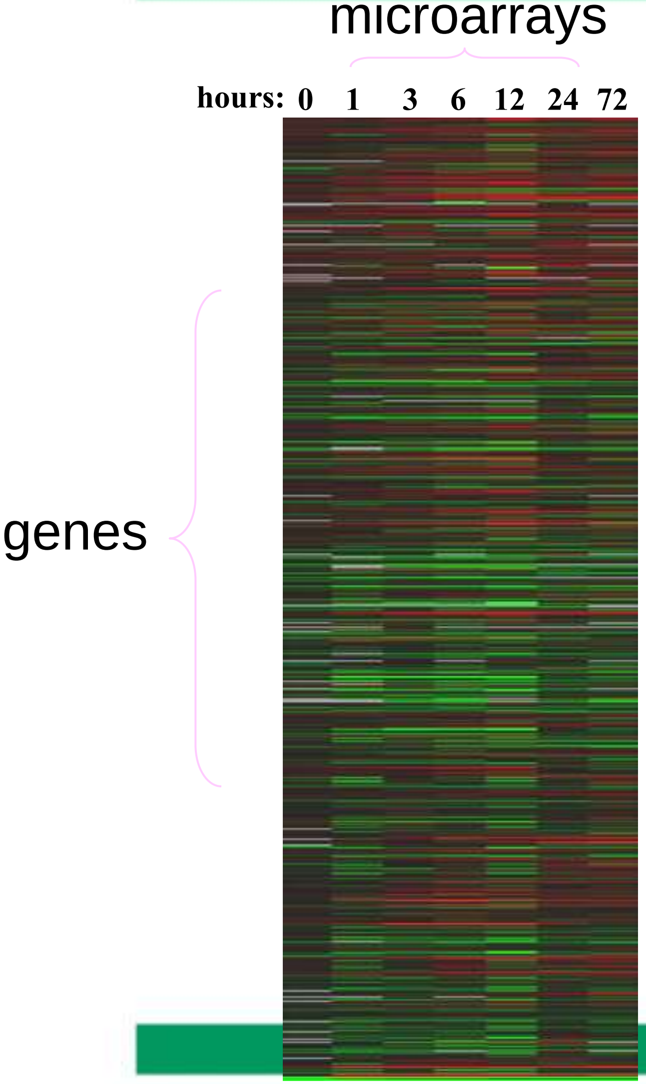
ENCEPHALOMYOCARDITIS VIRUS - EMCV

The transcriptional program in the response of human fibroblasts to serum

Iyer et al. (1999) *Science*, 283: 83

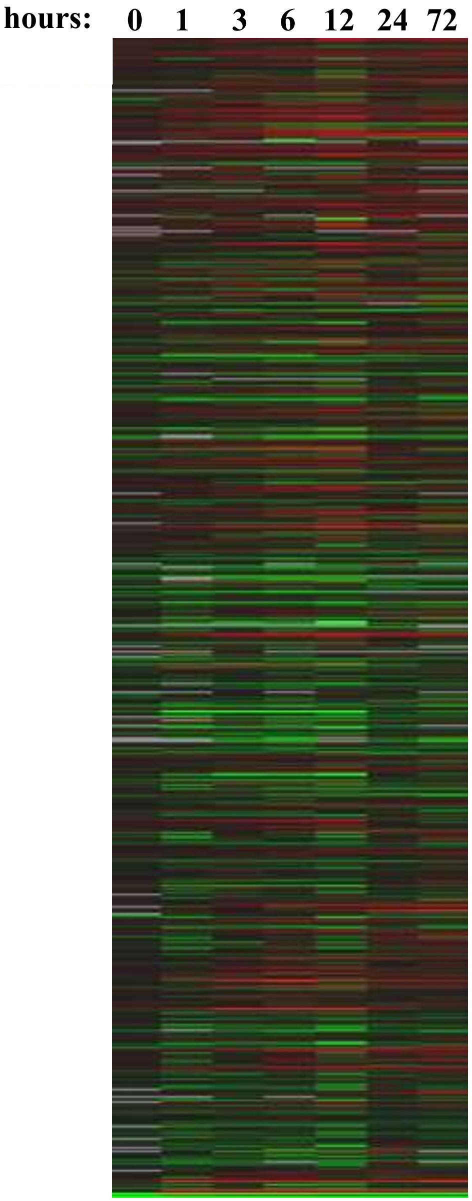
- Identify genes with similar expression
- Grouping unknown genes with known genes may provide insight into function of unknown genes
- Cluster genes by similar changes - only really meaningful across multiple treatments or time points
- Cluster samples by similar gene expression profiles

Basics of Data Filtering and Visualization

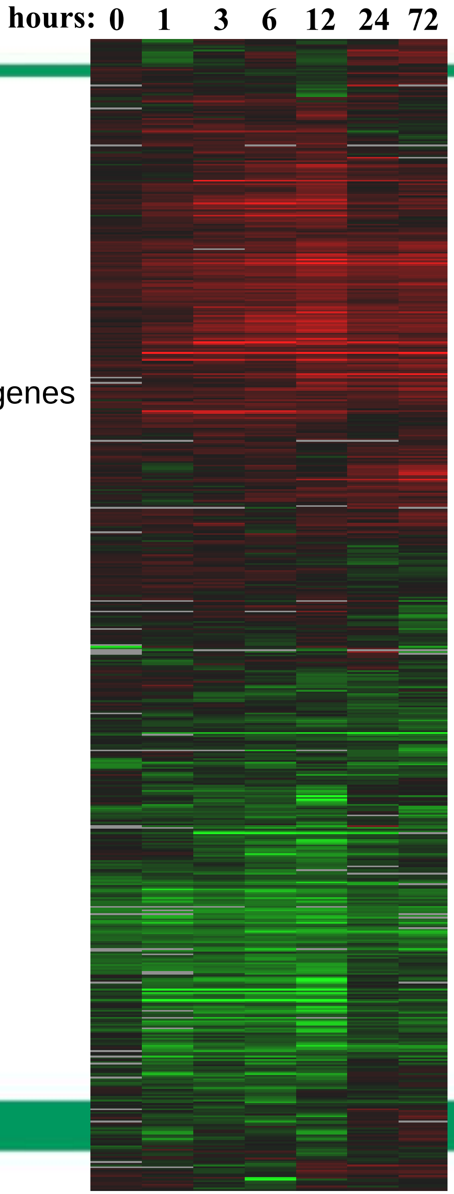


0 vs 1
0 vs 3
0 vs 6
etc..

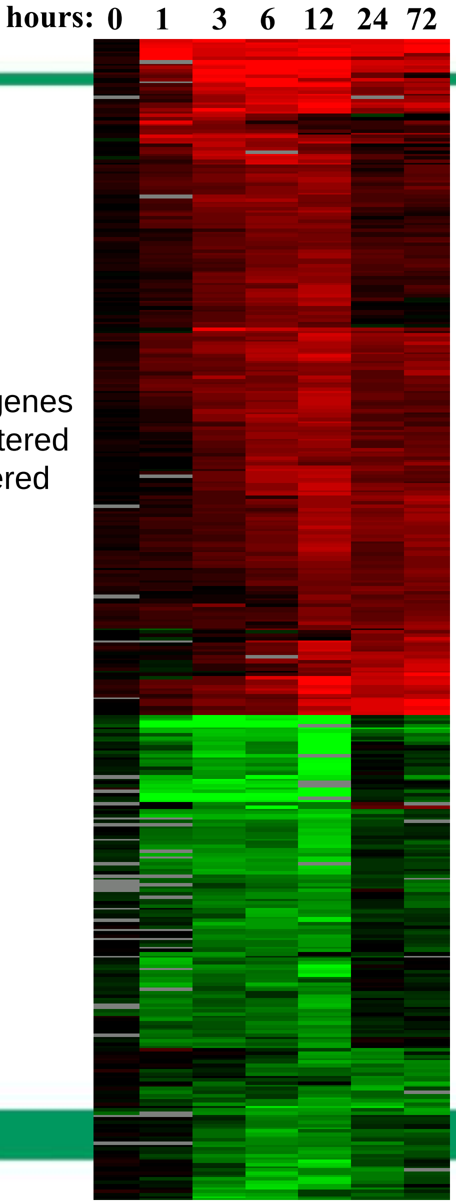
Grouping genes: clustering



4,416 genes

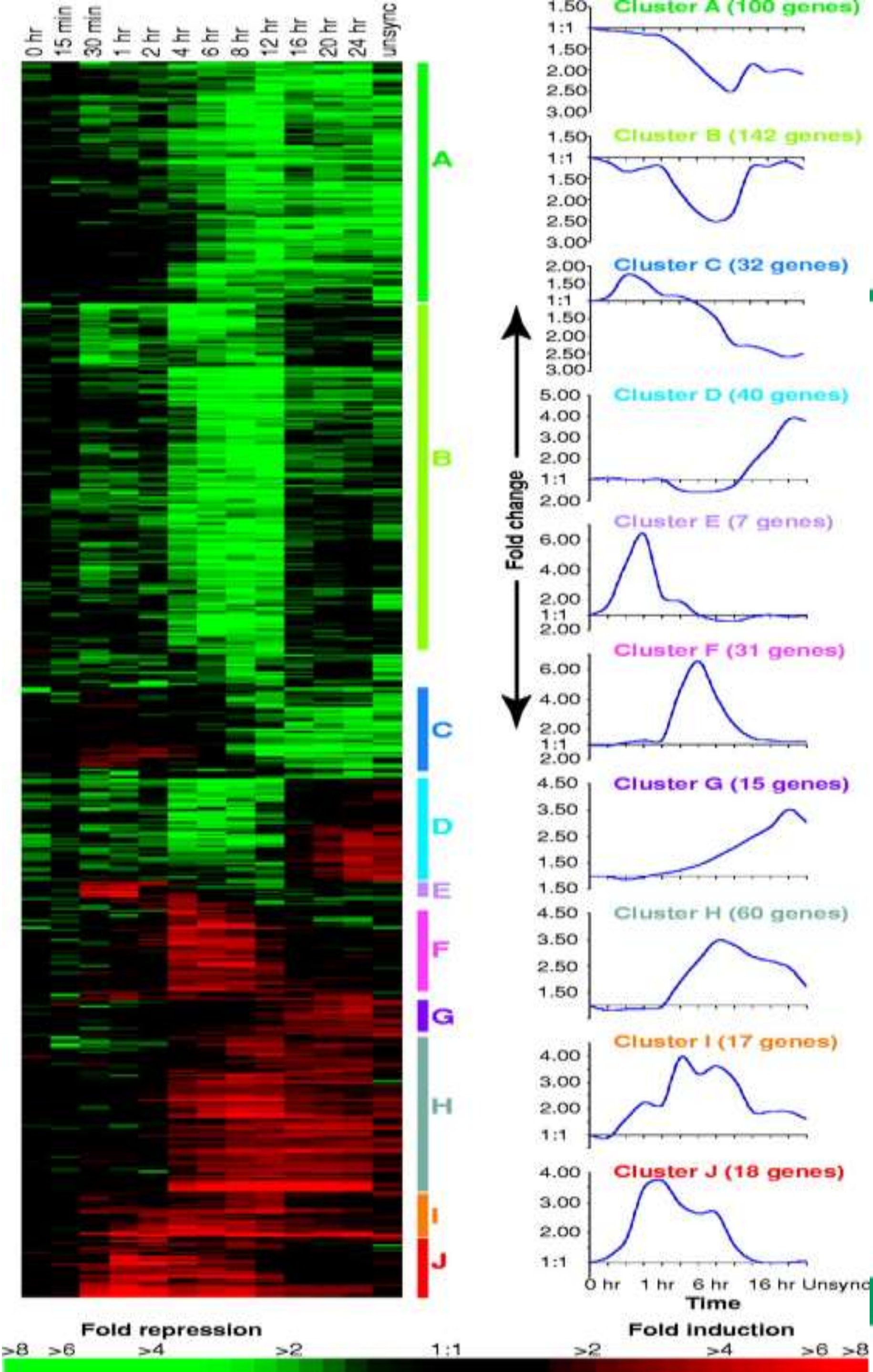


4,124 genes
spot-filtered
clustered



464 genes
spot-filtered
ratio-filtered
clustered

Microarray (Clustering)



8600 cDNA clones

- Coordinated gene expression
- Differential gene expression

Biological information

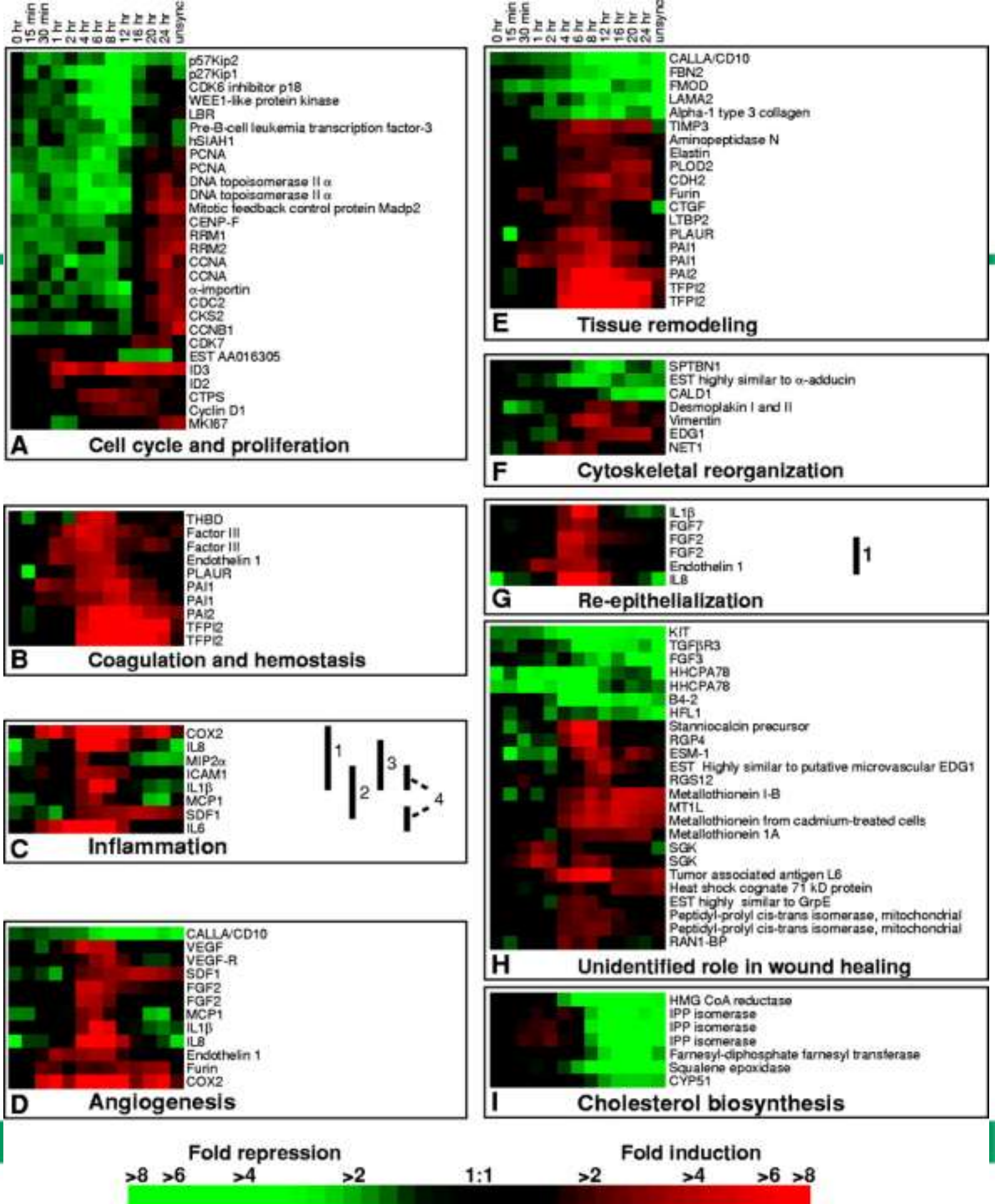
Iyer et al. (1999) *Science*, 283: 83

Expression signatures

Couple expression to GO

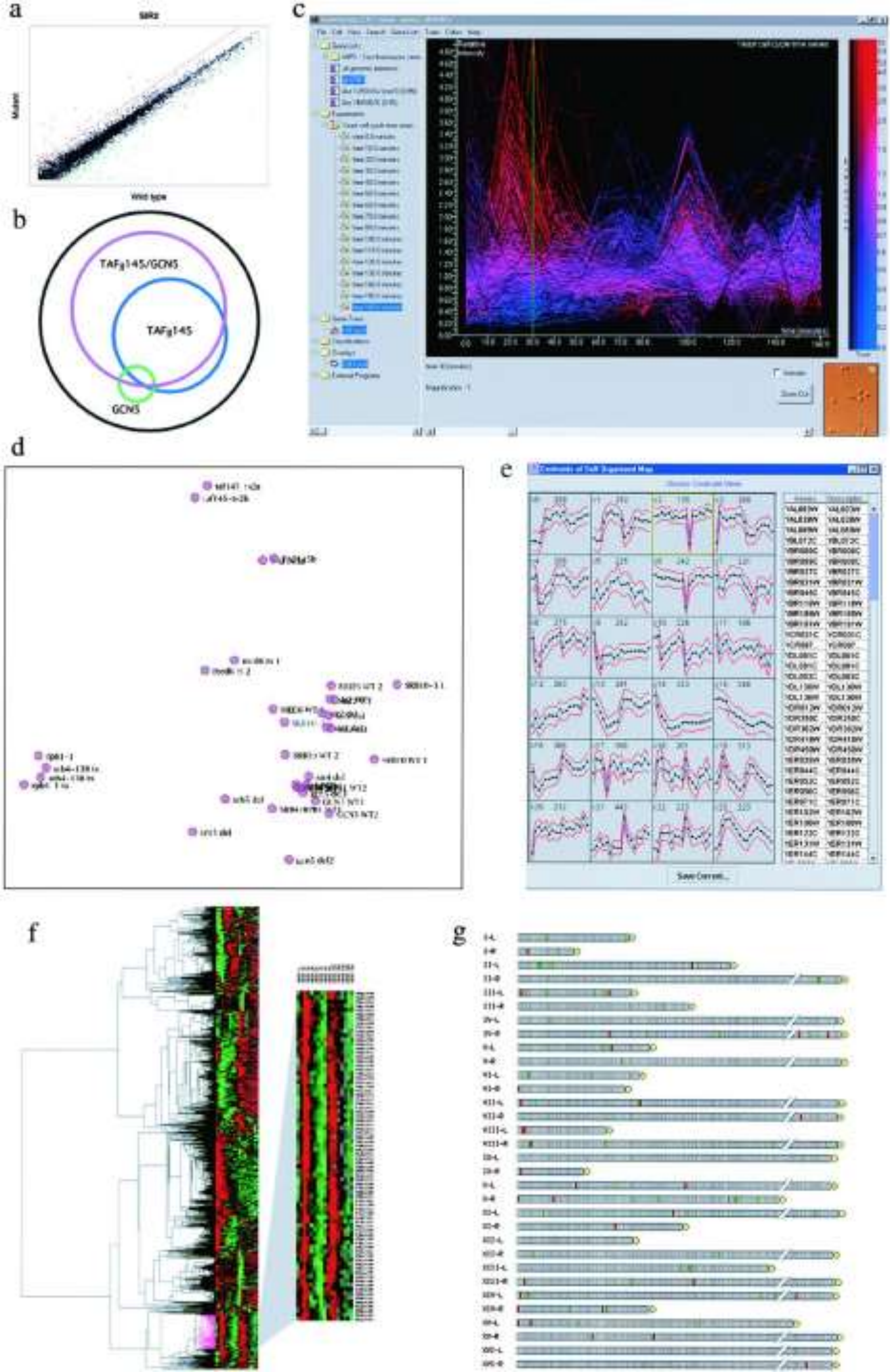
Serum treatment ‘*in vitro*’

Wound healing ‘*in vivo*’

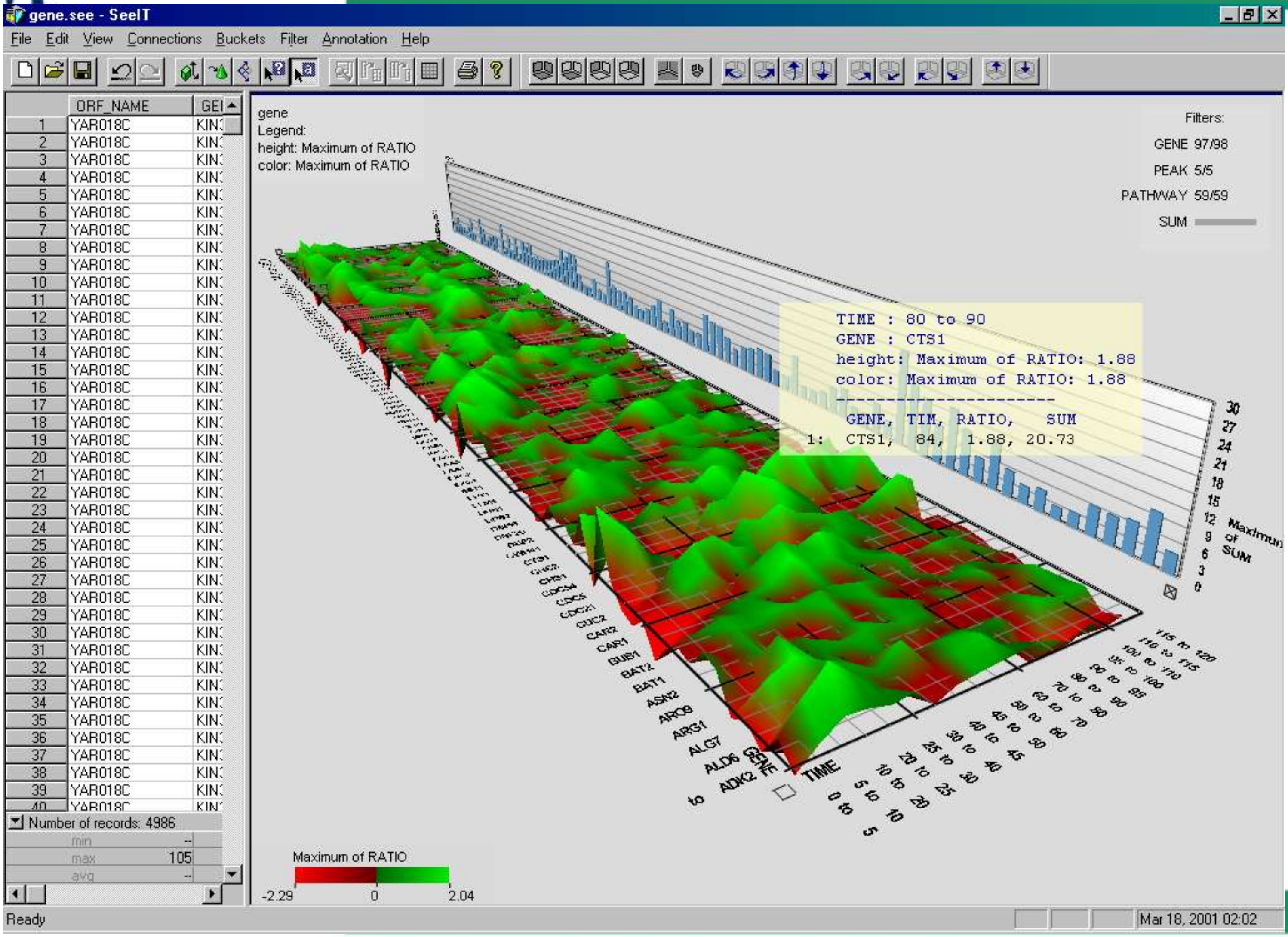


Visualisation

GeneSpring



Expression Landscape of cell-cycle regulated genes in yeast



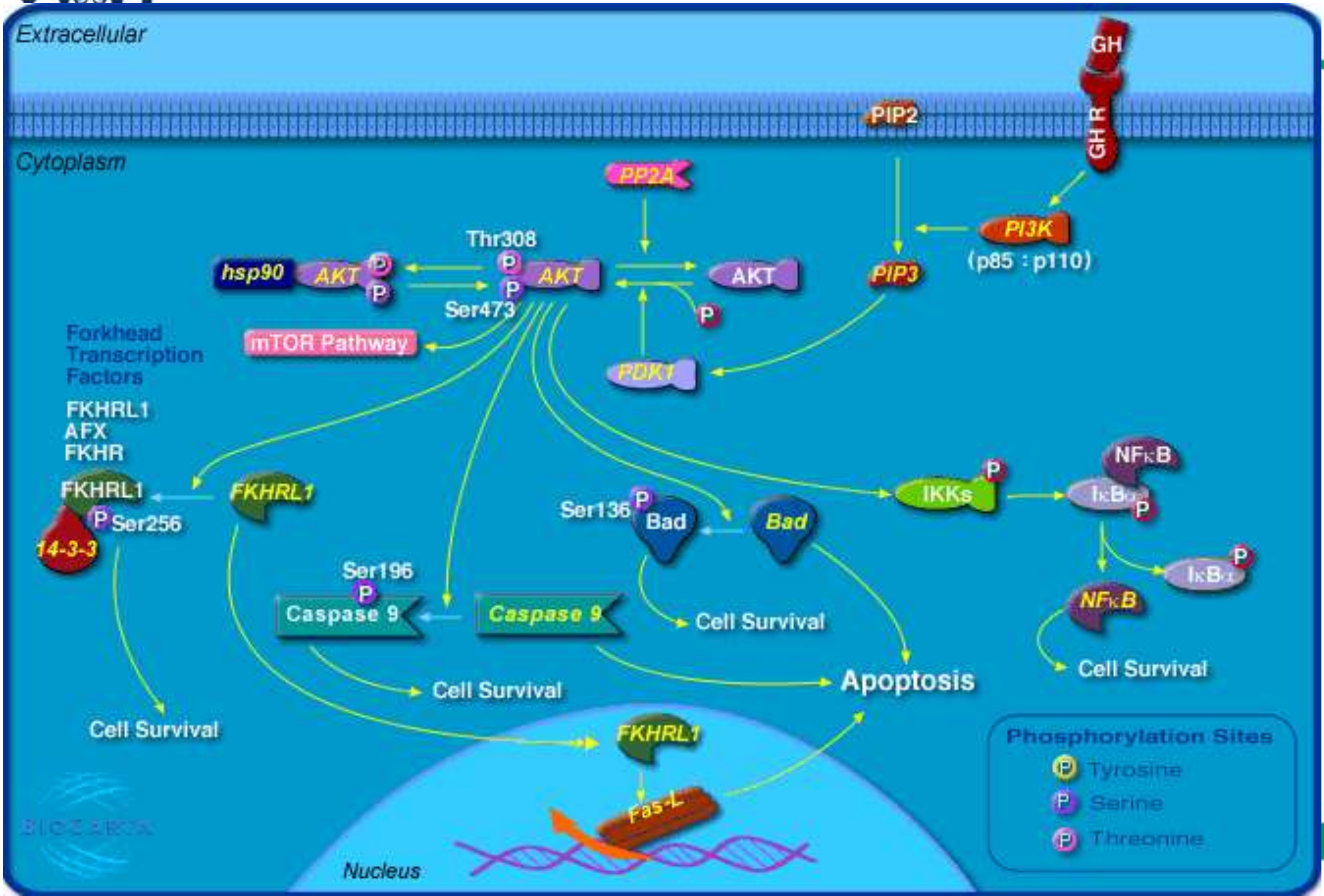
	ORF_NAME	GEI
1	YAR018C	KIN
2	YAR018C	KIN
3	YAR018C	KIN
4	YAR018C	KIN
5	YAR018C	KIN
6	YAR018C	KIN
7	YAR018C	KIN
8	YAR018C	KIN
9	YAR018C	KIN
10	YAR018C	KIN
11	YAR018C	KIN
12	YAR018C	KIN
13	YAR018C	KIN
14	YAR018C	KIN
15	YAR018C	KIN
16	YAR018C	KIN
17	YAR018C	KIN
18	YAR018C	KIN
19	YAR018C	KIN
20	YAR018C	KIN
21	YAR018C	KIN
22	YAR018C	KIN
23	YAR018C	KIN
24	YAR018C	KIN
25	YAR018C	KIN
26	YAR018C	KIN
27	YAR018C	KIN
28	YAR018C	KIN
29	YAR018C	KIN
30	YAR018C	KIN
31	YAR018C	KIN
32	YAR018C	KIN
33	YAR018C	KIN
34	YAR018C	KIN
35	YAR018C	KIN
36	YAR018C	KIN
37	YAR018C	KIN
38	YAR018C	KIN
39	YAR018C	KIN
40	YAR018C	KIN

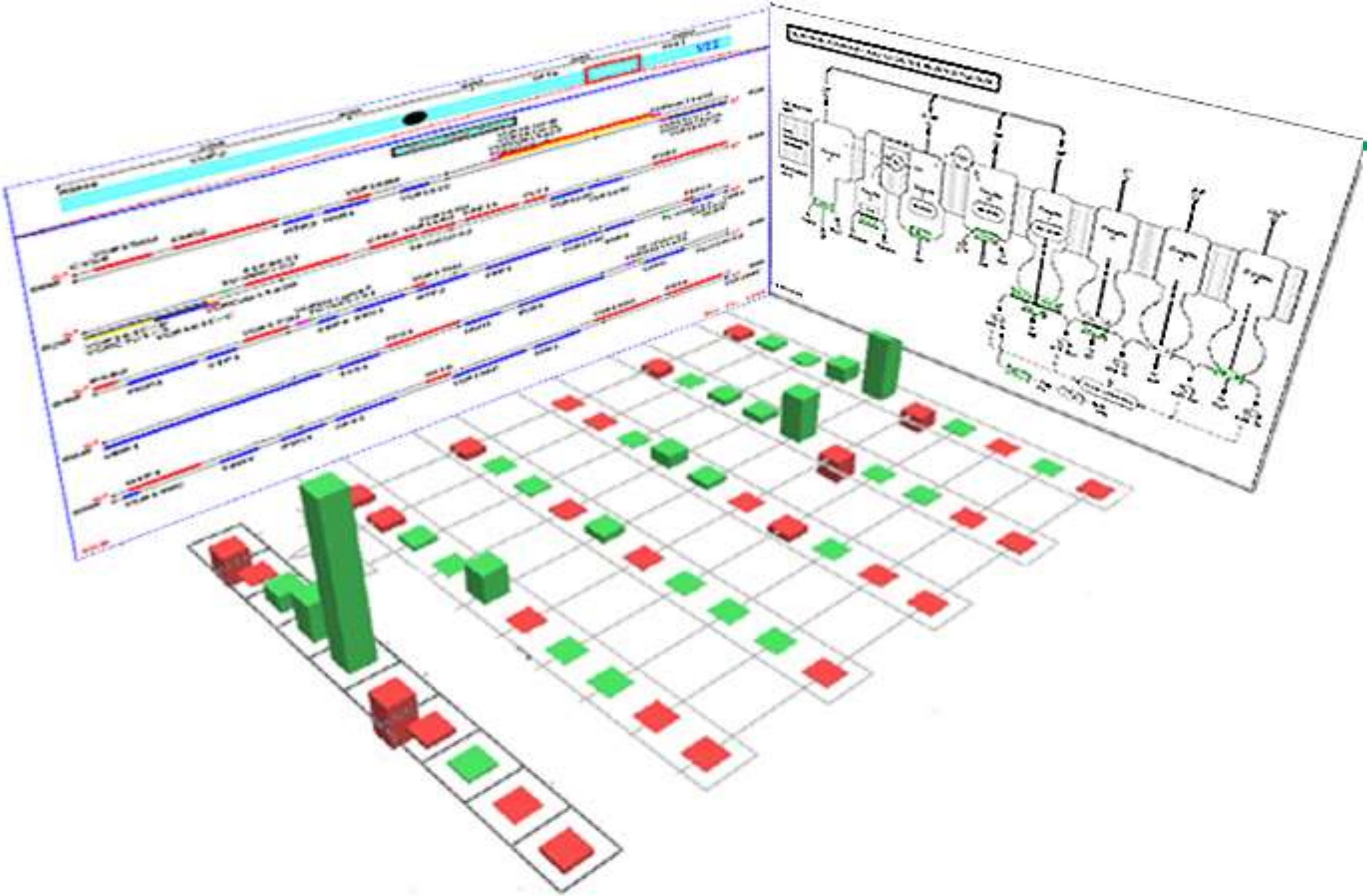
Number of records: 4986

min --
max 105
avg --

```
TIME : 80 to 90
GENE : CTS1
height: Maximum of RATIO: 1.88
color: Maximum of RATIO: 1.88
----------------------
    GENE, TIM, RATIO,   SUM
1:  CTS1,  84,  1.88, 20.73
```

GenMapp: Biological Pathways





Microarray and cancer

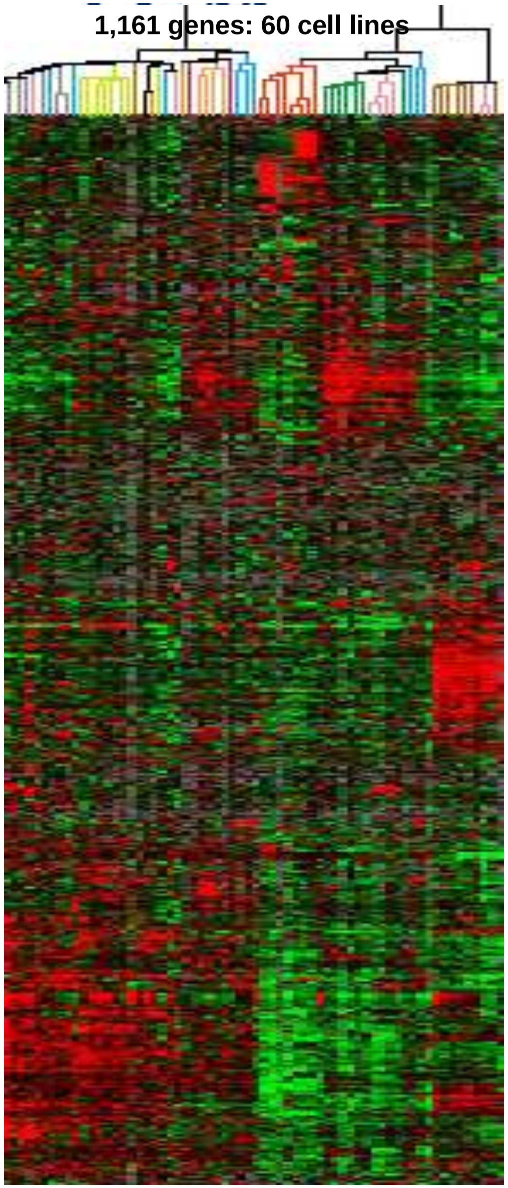
- Identification of prognostic biomarkers specific to onset and progression
- Disease classification
- Development of drug resistance
- Risk of relapse assessment
- Metastasis
- Response to treatment
- Survival

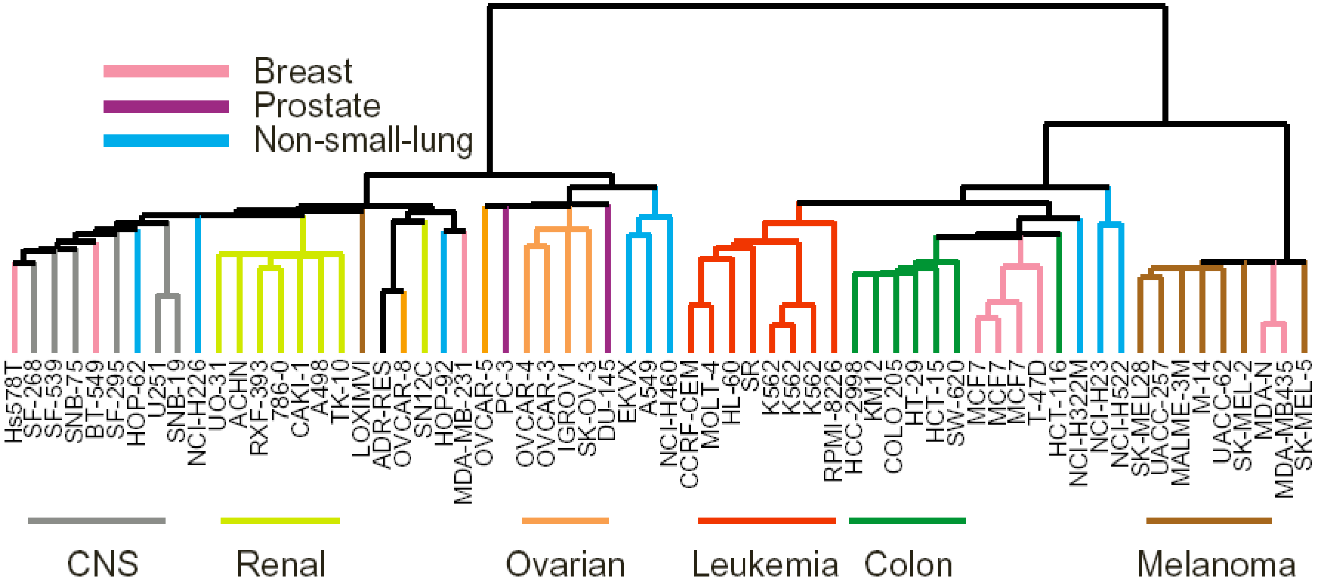
Variation in Gene Expression Patterns in Human Cancer Cell Lines

Ross DT, et. al., *Nature Genetics*, (24): 2000, 227-235.

Aim:
to explore the variation in gene expression of ~ 8,000 genes among 60 human cancer cell lines (spanning 9 distinct tissues)

Hierarchical Clustering of Gene Expression Patterns Groups Cell Lines According to Tissue of Origin





- •Relationship between expression profile and tissue of origin
- •Recognize previously incorrect classified outliers
- •Recognize relationships to tumors *in vivo*

Distinct Types of Diffuse Large B-cell Lymphoma (DLBCL)

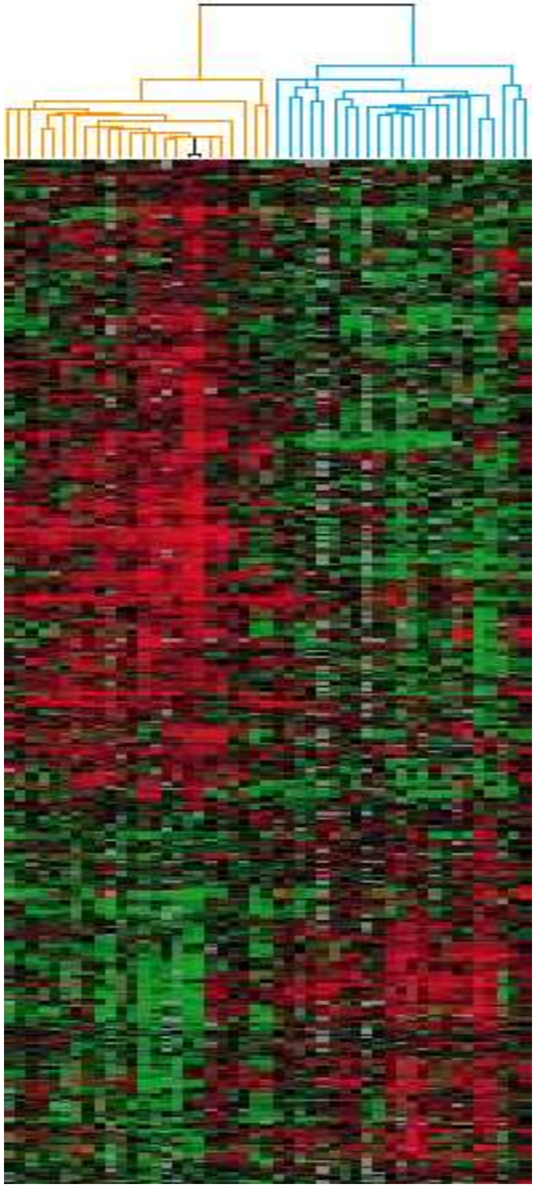
Alizadeh AA, et. al., *Nature*, (403): 2000,503-511.

Aim:
to determine whether gene expression profiling could subdivide DLBCL – a clinically heterogeneous diagnostic category – into molecularly distinct diseases with more homogeneous clinical behaviors

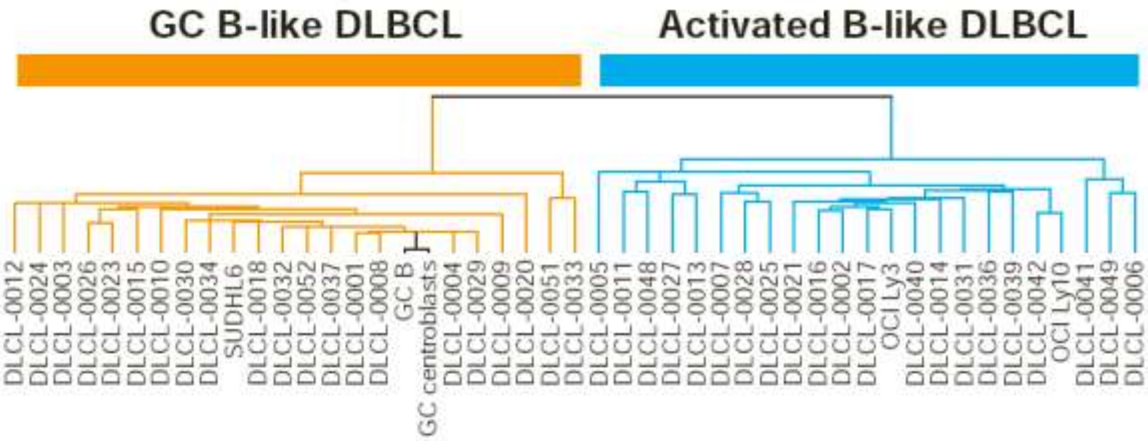
Only 40% of patients respond well to therapy

“Lymphochip”: -17.856 cDNA clones
-lymphoid cell origin
-cancer + immunology

Clustering Identifies 2 Major Subgroups of DLBCL



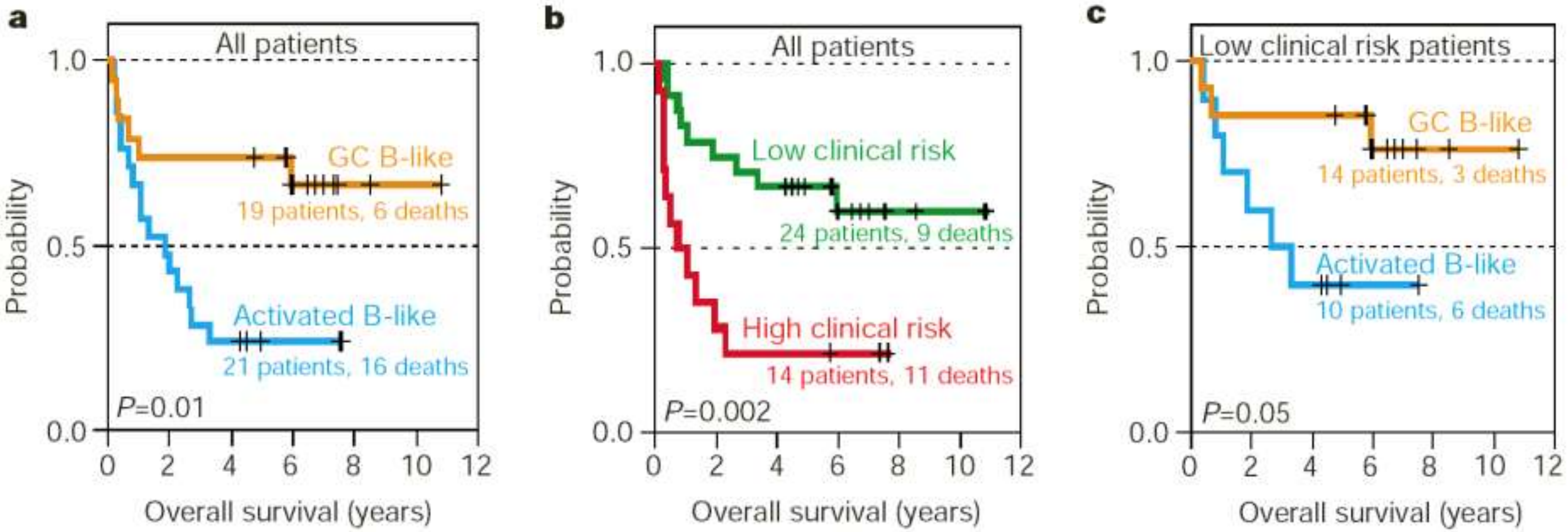
45 DLBCL biopsies



Different B-cell differentiation stage

Set of ~3000 genes

DLBCL Subgroups Define Prognostic Categories



a. Kaplan-Meier plot of overall survival of DLBCL patients grouped on the basis of gene expression profiling.

b. Kaplan-Meier plot of overall survival of DLBCL patients grouped according to the International Prognostic Index.

c. Kaplan-Meier plot of overall survival of low clinical risk DLBCL patients grouped on the basis of gene expression profiles.

Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications

Therese Sørlie[a,b,c], Charles M. Perou[a,d], Robert Tibshirani[e], Turid Aas[f], Stephanie Geisler[g], Hilde Johnsen[b], Trevor Hastie[e], Michael B. Eisen[h], Matt van de Rijn[i], Stefanie S. Jeffrey[j], Thor Thorsen[k], Hanne Quist[l], John C. Matese[c], Patrick O. Brown[m], David Botstein[c], Per Eystein Lønning[g], and Anne-Lise Børresen-Dale[b,n]

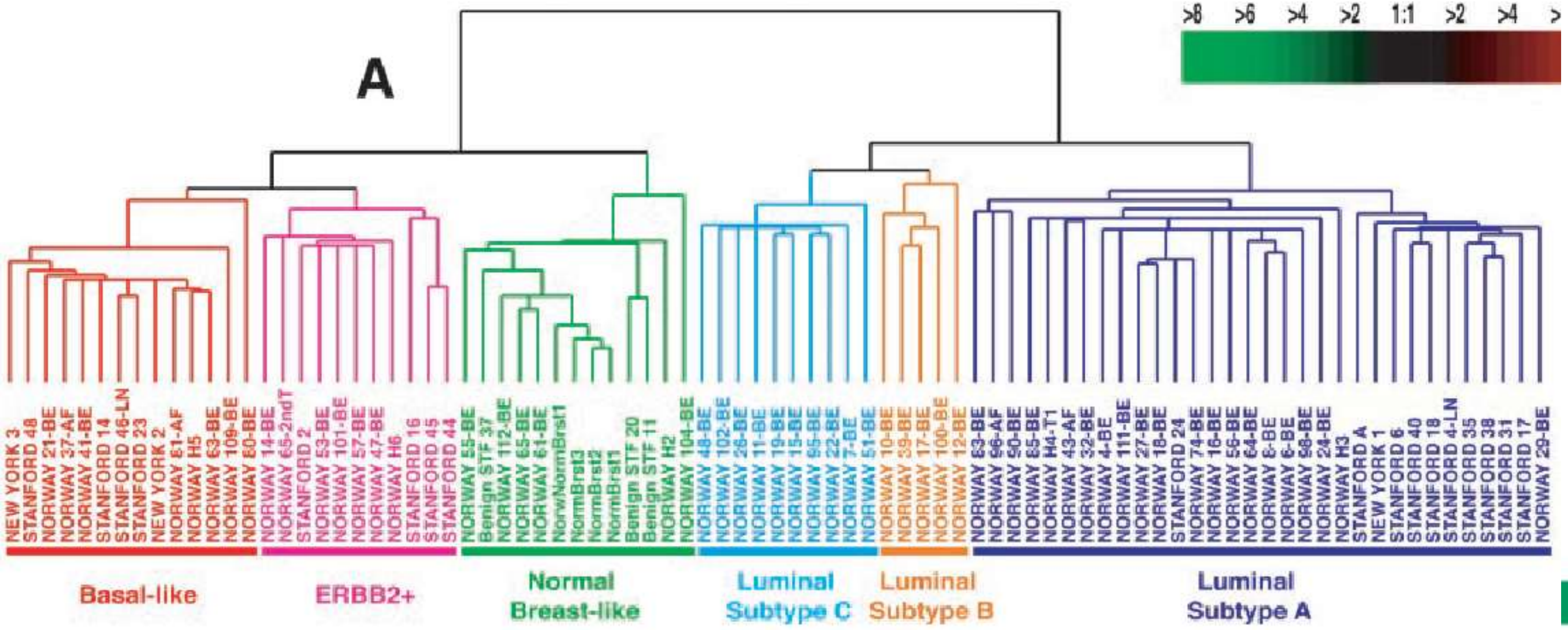
2001, *PNAS*

Aim:
To classify breast carcinoma’s based on expression profiling and to correlate these to clinical outcome

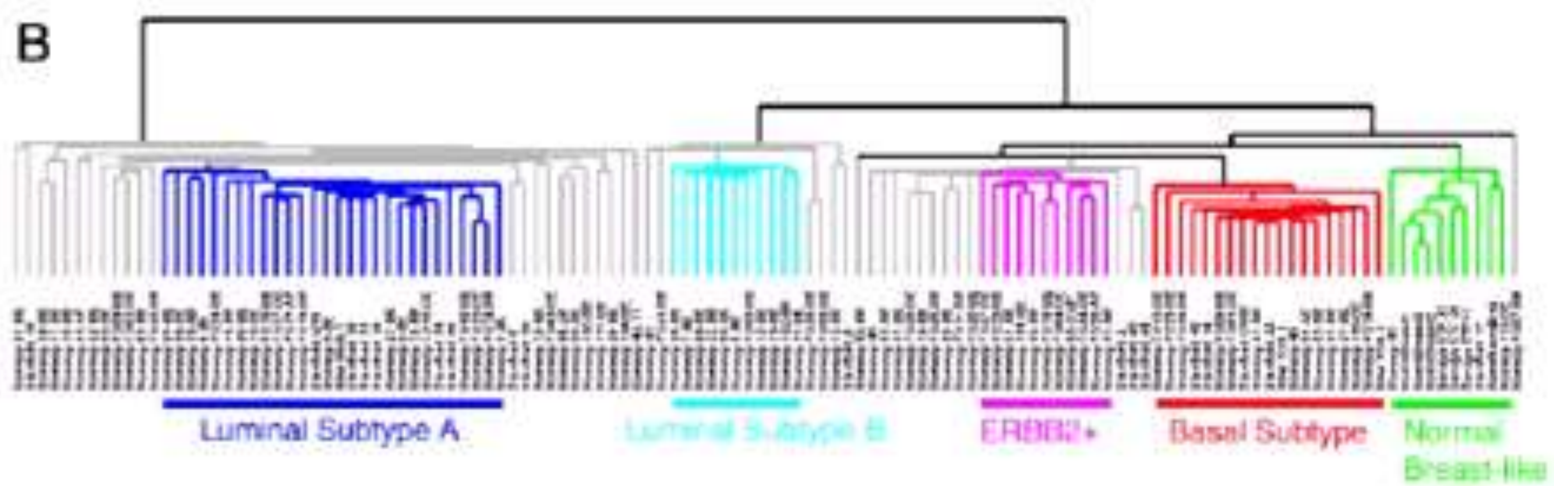
Clustering Identifies novel and existing Subgroups of Breast cancer

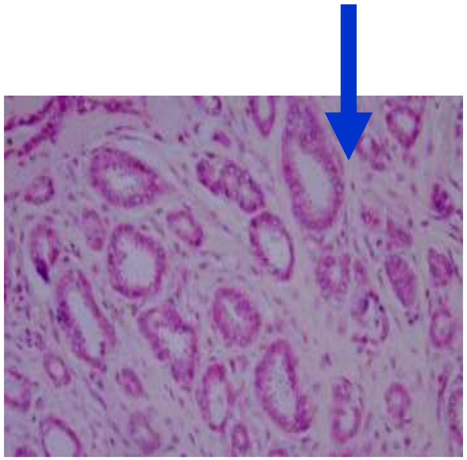
Differential expressed genes: 476
- tumor properties
- patient outcome

85 biopsy samples

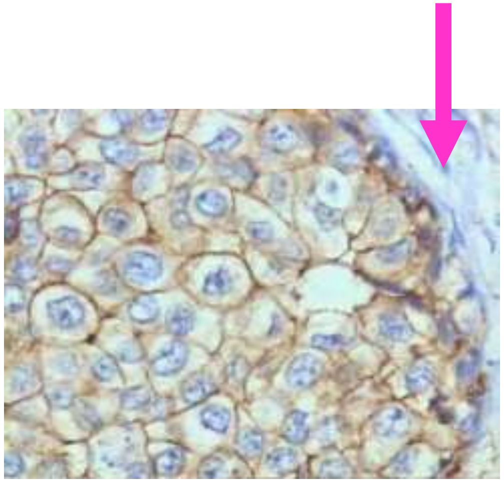


Sorlie *et al.* PNAS 2003

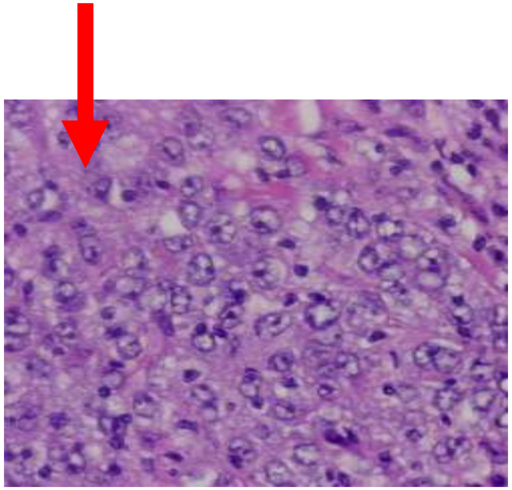




ER++, PR++, G1,2



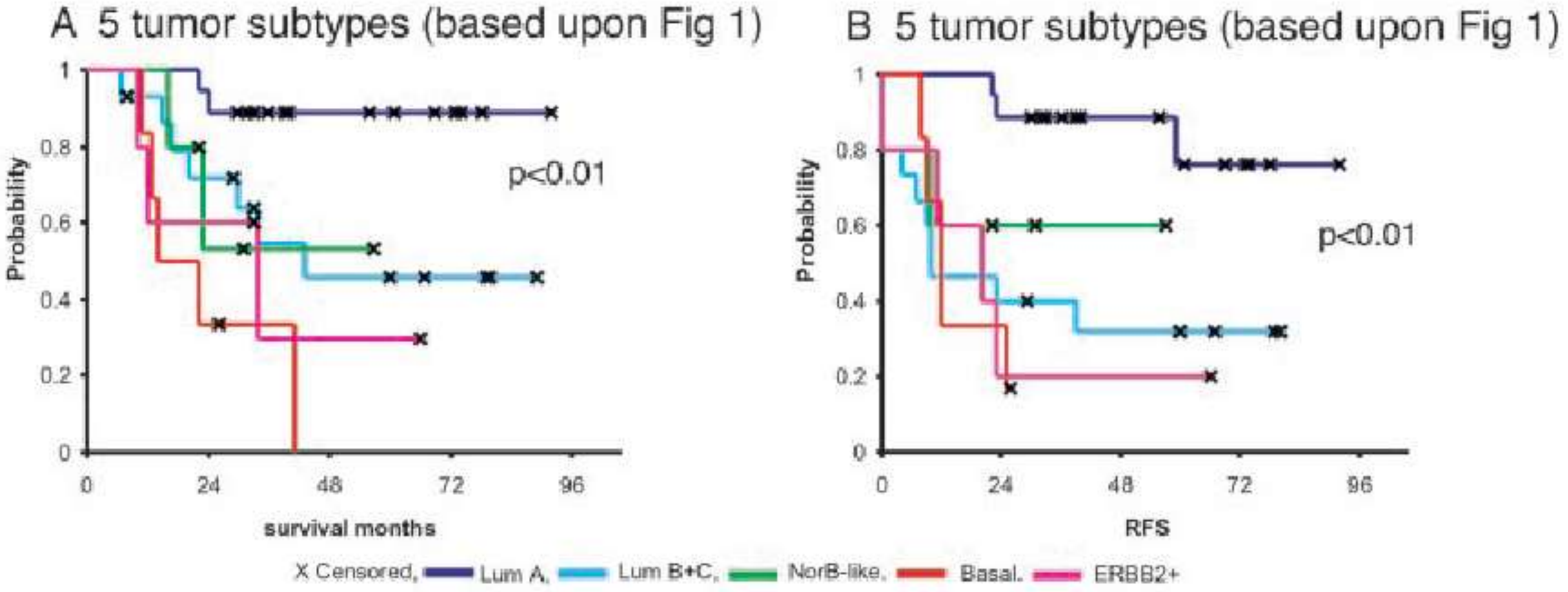
HER2 ISH pos



“triple neg,” CK5/6+

Molecular classes are predictive of outcome

overall survival: relapse-free survival:



BioArray News (2, no. **35**, 2002)
Arrays Hold Promise for Cancer Diagnostics

Oncologists would like to use arrays to predict whether or not a cancer is going to spread in the body, how likely it will respond to a certain type of treatment, and how long the patient will probably survive.

It would be useful if the gene expression signatures could distinguish between subtypes of tumours that standard methods, such as histological pathology from a biopsy, fail to discriminate, and that require different treatments.

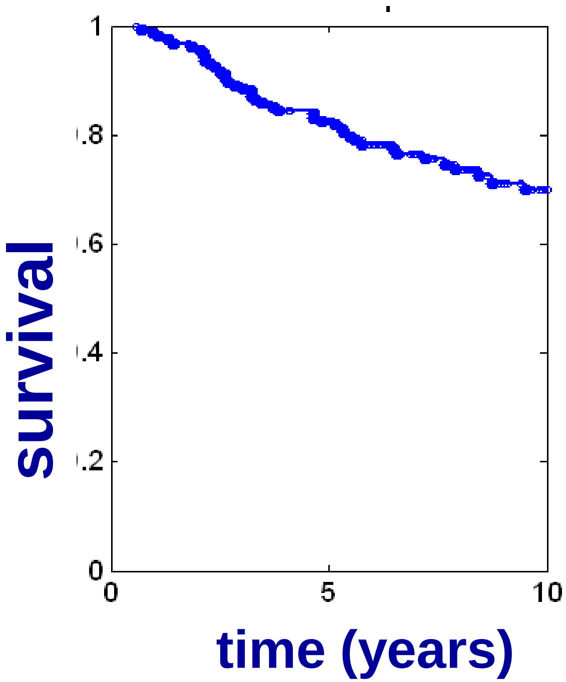
Gene expression profiling predicts clinical outcome of breast cancer

Van ‘t Veer, et. al., *Nature*, (415): 2002,530-536.

Aim:
to determine whether gene expression profiling could predict disease outcome and provide a strategy to select patients who would benefit from adjuvant therapy (metastasis)

Breast Cancer – Survival Pre-menopausal patients, lymph node negative

traditional diagnostics



~30% die <10 year

~70% survive >10 year

Everyone receives chemotherapy...!

Breast Cancer – Survival Pre-menopausal patients, lymph node negative

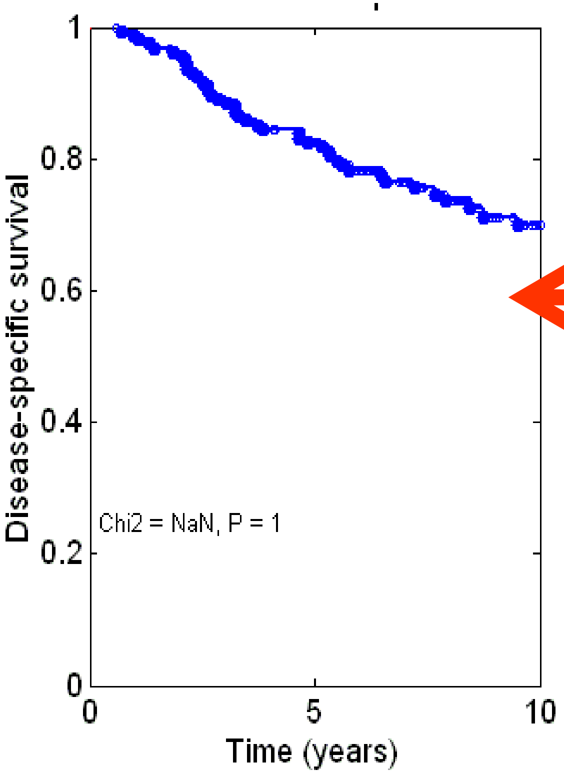
Current adjuvant treatment selection criteria:

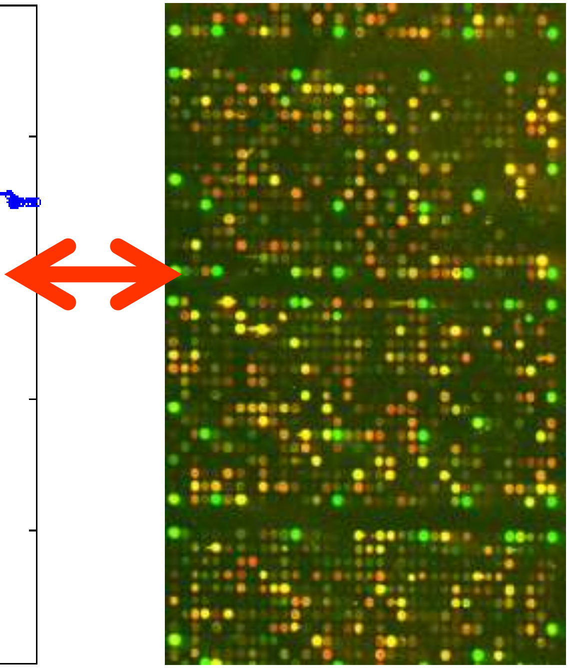
- NIH (US) consensus criteria: > 95%
- St Gallen (EU) consensus criteria: > 80%

  receive adjuvant chemo- and hormonal therapy

As only 30% of these patients develop distant metastases, some 50-65% of patients are over-treated with adjuvant (chemo)therapy

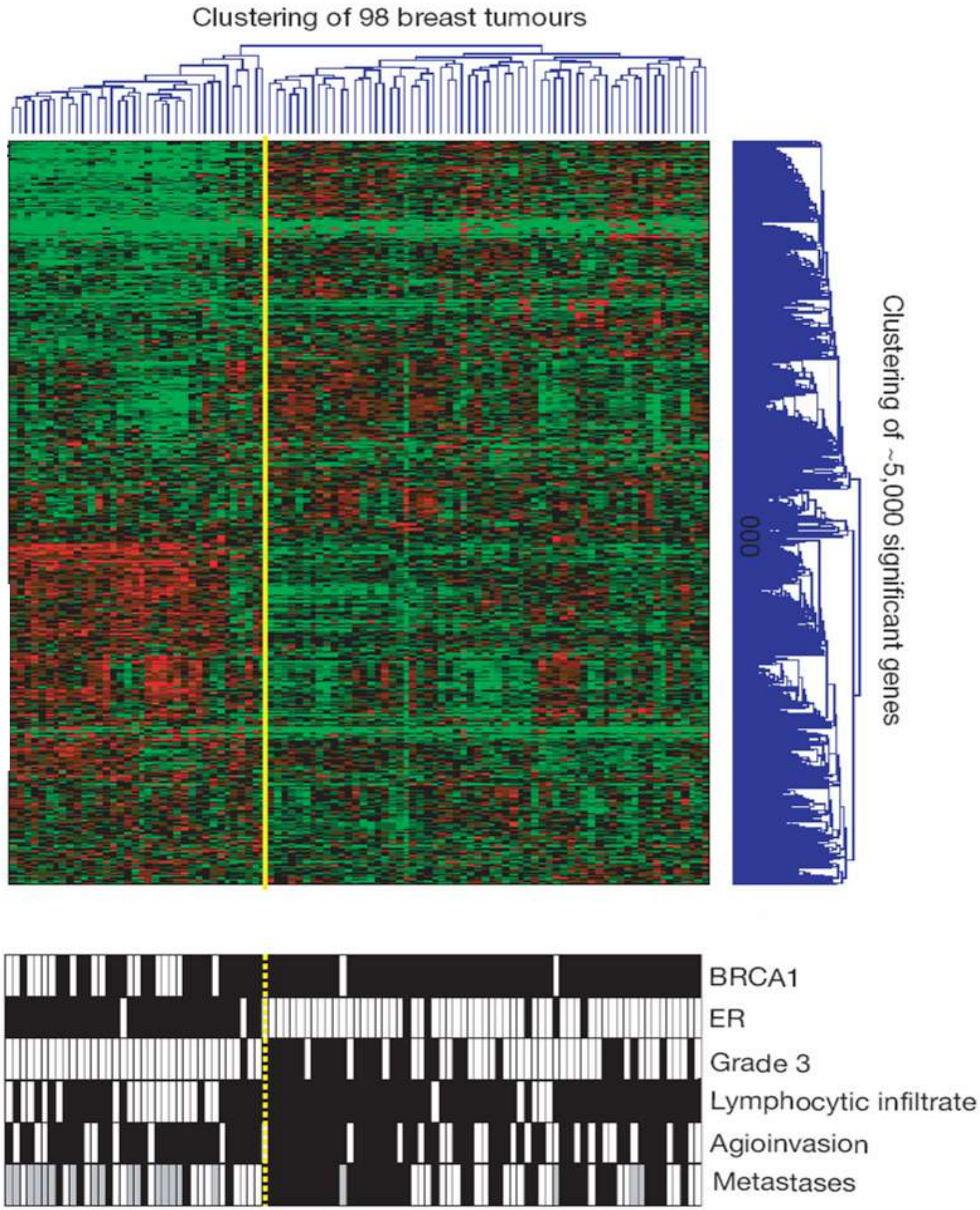
Identification of gene expression changes in breast cancer





- analyse 98 breast tumors
- 34 metastases-positive <5 year
  - bad prognosis
- 44 metastases-negative >5 year
  - good prognosis
- 18 BRCA1 +
- 2 BRCA2 +

'sporadic'

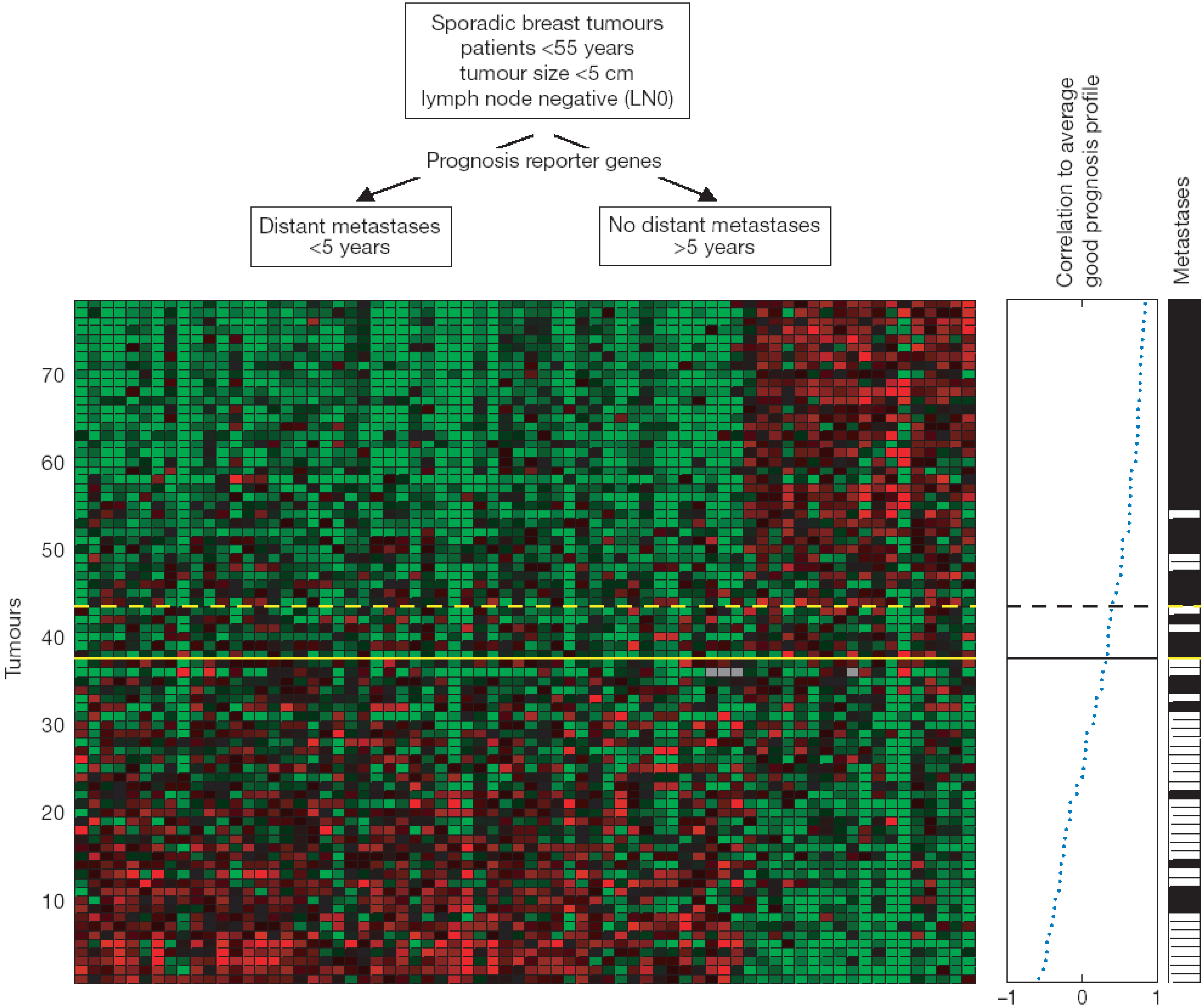


- 98 breast tumors analysed
  - 34 ‘bad’ vs. 44 ‘good’
  - 18 BRCA1 +
  - 2 BRCA2 +
- microarray with 24.000 genes
- 5.000 genes showed expressional changes in tumors

Different classes of breast tumors...!

70-gene prognosis classifier for predicting risk of distant metastasis within 5 years

Supervised clustering



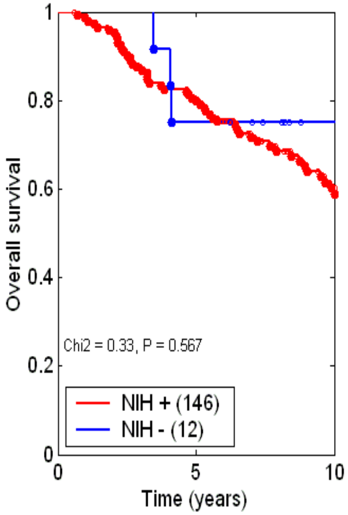
Good prognosis

Poor prognosis

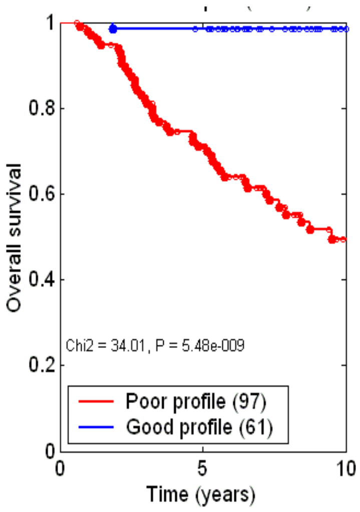
Microarray classification vs. NIH classification

5 % low risk
95 % high risk

39 % low risk
61 % high risk



Classical
NIH classification

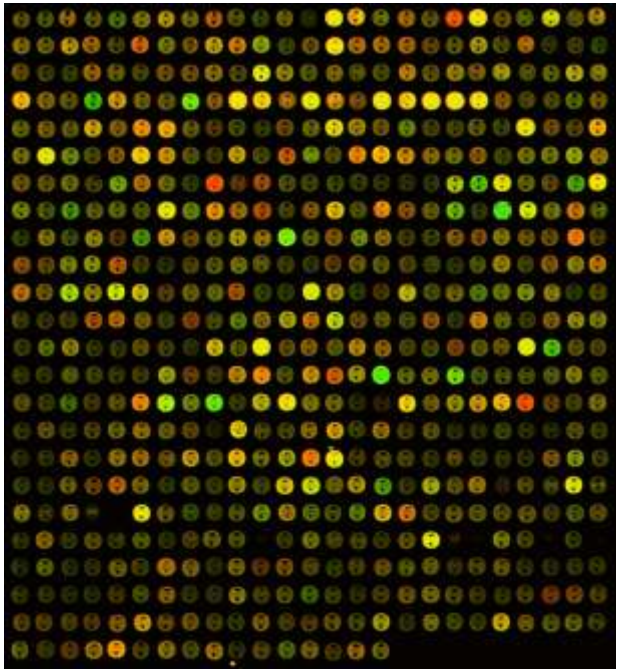


Classification based
on microarray

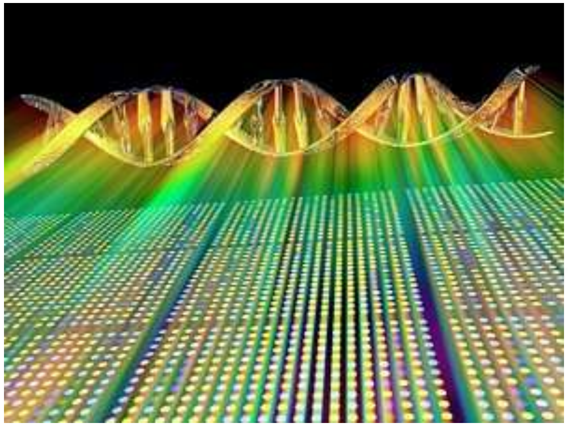
- Classification of 158 breast cancer tumors
- Less unnecessary chemo-therapy
- Identification of genes playing a role in breast cancer

Microarray to be used as routine clinical screen

by C. M. Schubert
Nature Medicine
9, ***9,*** *2003.*



The Netherlands Cancer Institute in Amsterdam is the first institution in the world to use microarray techniques for the routine prognostic screening of cancer patients. Aiming for a June 2003 start date, the center will use a panoply of 70 genes to assess the tumor profile of breast cancer patients and to determine which women will receive adjuvant treatment after surgery.

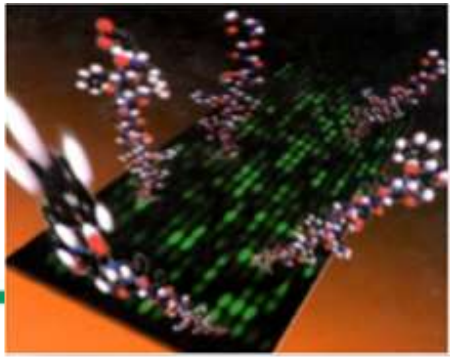


Expression profiling & clinical application

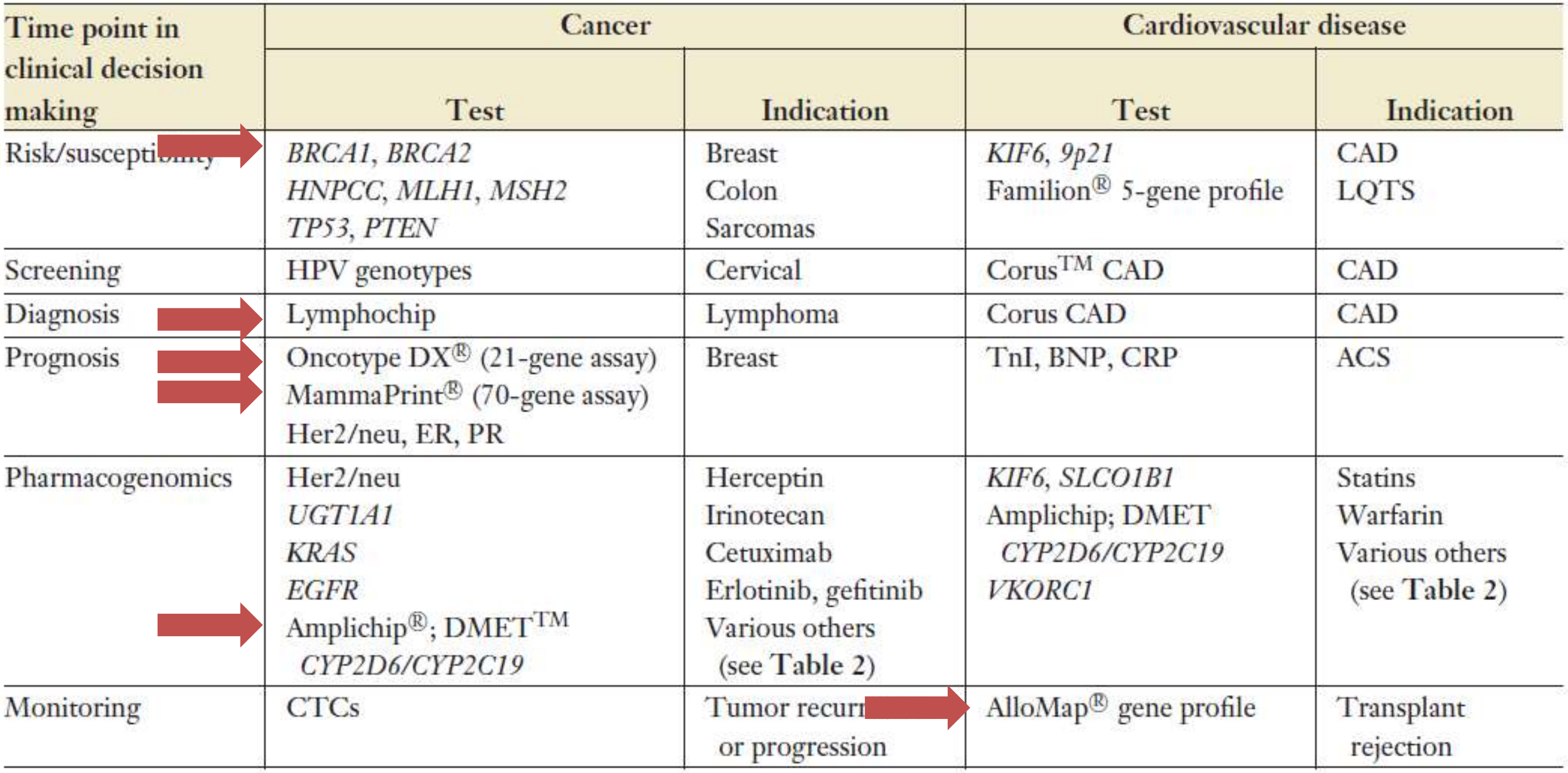
“Though each tumor is molecularly unique, there exist common transcriptional cassettes that underlie biological and clinical properties of tumors that may be of diagnostic, prognostic and therapeutic significance”.

→ Also true for other complex diseases

Clinically Available Molecular Diagnostics



Diagnostic Kits
Laboratory-developed-tests (LDTs)



Time point in clinical decision making	Cancer		Cardiovascular disease	
	Test	Indication	Test	Indication
Risk/susceptibility	*BRCA1, BRCA2* *HNPCC, MLH1, MSH2* *TP53, PTEN*	Breast Colon Sarcomas	*KIF6, 9p21* Familion® 5-gene profile	CAD LQTS
Screening	HPV genotypes	Cervical	Corus™ CAD	CAD
Diagnosis	Lymphochip	Lymphoma	Corus CAD	CAD
Prognosis	Oncotype DX® (21-gene assay) MammaPrint® (70-gene assay) Her2/neu, ER, PR	Breast	TnI, BNP, CRP	ACS
Pharmacogenomics	Her2/neu *UGT1A1* *KRAS* *EGFR* Amplichip®; DMET™ *CYP2D6/CYP2C19*	Herceptin Irinotecan Cetuximab Erlotinib, gefitinib Various others (see Table 2)	*KIF6, SLCO1B1* Amplichip; DMET *CYP2D6/CYP2C19* *VKORC1*	Statins Warfarin Various others (see Table 2)
Monitoring	CTCs	Tumor recurr or progression	AlloMap® gene profile	Transplant rejection

(Chan & Ginsburg, 2011)