DMMC Post-Graduate Course Techniques and Strategies in Molecular Medicine
10th December 2007

Differential Gene Expression: Overview of Relevant Methods

Prof. William M. Gallagher
Associate Professor of Cancer Biology

UCD School of Biomolecular and Biomedical Science
Conway Institute, University College Dublin

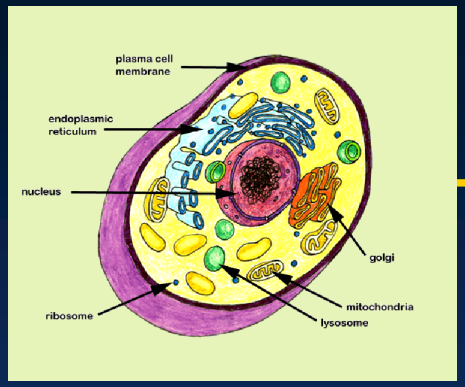
DIFFERENTIAL GENE EXPRESSION

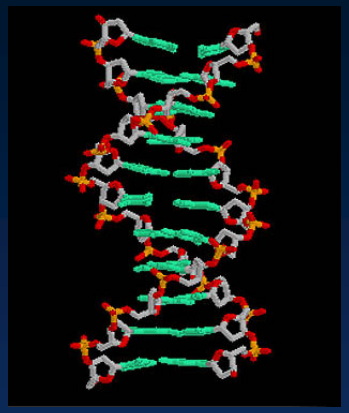
Examine differential patterns of expression between

- different tissues
- different development stages
- normal and disease states
- response of cells to drugs
- different physiological conditions

Why?

- Biological information
- Changes in gene expression → Changes in physiology
- Provide direct clues to function of genes
- Provide information for signaling pathways involved
- To identify biomarkers of disease
- Identification of gene targets for therapeutics





DNA → ~30-40,000

Transcription

RNA → >80,000

Translation

Protein → >1,000,000

EXPRESSED GENES

RNA CLASS	% OF TOTAL RNA	COPIES/CELL
ABUNDANT	20%	1,000-12,000
MEDIUM	25%	100-1,000
LOW	50%	<13
Total number of transcripts		Approx 70,000-80,000

- Typically mammalian cell may express 20,000-30,000 genes
- Different organs: 2,000-4,000 genes may vary
- Related cell types (i.e B & T lymphocytes): < 500 genes may vary

RNA: THE ISSUES

Concentration

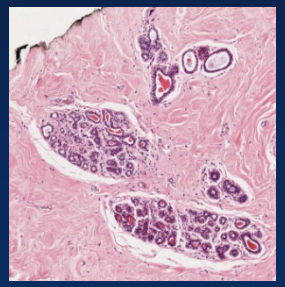
Stability

Integrity

Quantity

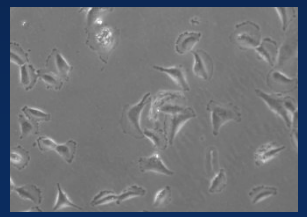
Translation to Protein

RNA Preparation: Sample Types



Clinical Tissue

Body Fluids



Cell lines

Animal Models

Issues

- Time interval to being frozen/extracted
- Use of formalin-fixed, paraffin-embedded tissues?
- Requirement for enriched cell populations? (laser capture microdissection)
- Requirement for amplification of target material (low yield)? (possible distortion of signal)

Gene expression analysis classically limited to gene-by-gene approach

Advent of new technologies facilitated high-throughput and global analysis of gene expression

HIGH-THROUGHPUT ANALYSIS OF GENE EXPRESSION

1. Sequencing of Expressed Sequence Tags (ESTs)
2. Differential display
3. Subtractive cloning
4. Serial analysis of gene expression (SAGE)
5. **DNA MICROARRAY HYBRIDISATION**

1. Expressed Sequence Tags (ESTs)

- EST is a short sequence of cDNA usually 200 to 500bp
- ESTs are produced by sequencing of a cDNA library

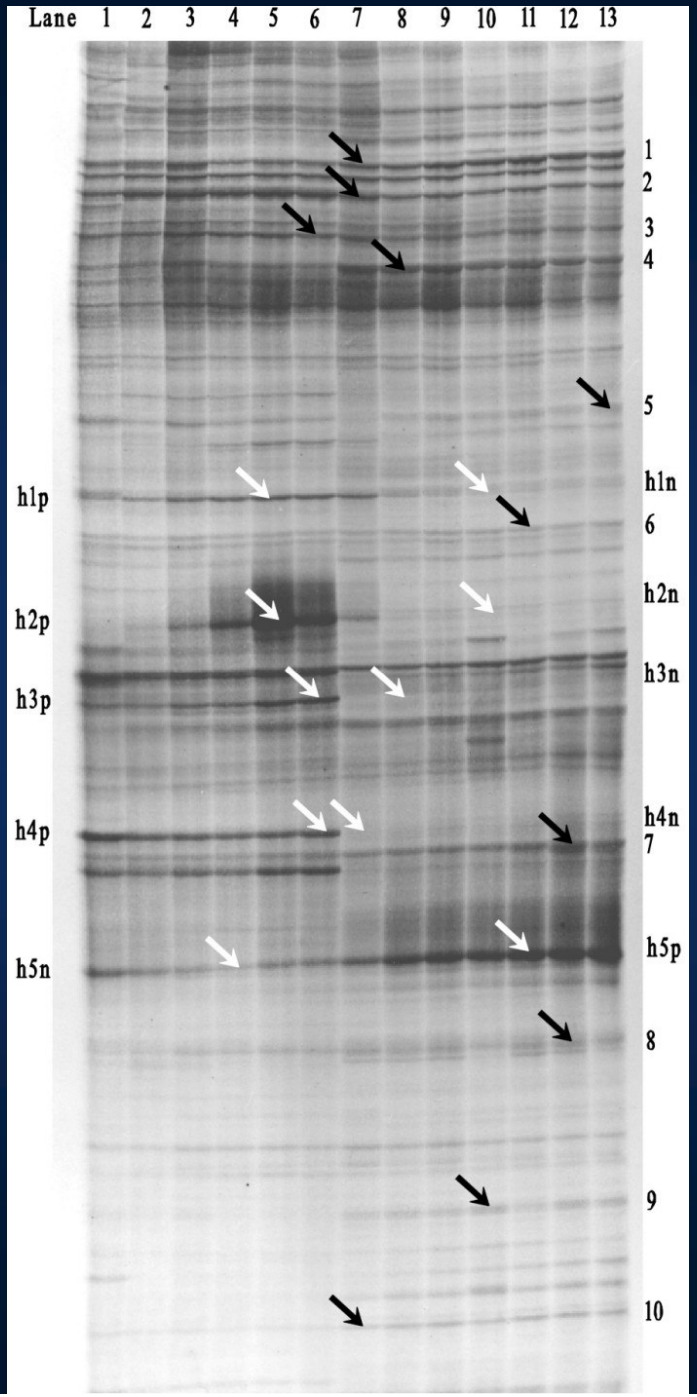
- They can be used to
  - discover new genes
  - obtain data on gene expression and regulation
  - for constructing genome maps
  - can be used to construct DNA microarrays
- Millions of ESTs available from various genomes in public databases (e.g. GenBank)

1. Expressed Sequence Tags (ESTs)

- Often single pass sequencing (therefore, error prone)
- Comparison of sequence databases – determine identity
- Occurrence of particular sequence – relative to concentration of original transcript
- Need efficient and cost-effective sequencing facility
- Low sampling rate limits utility to detection of highly up- or down-regulated transcripts

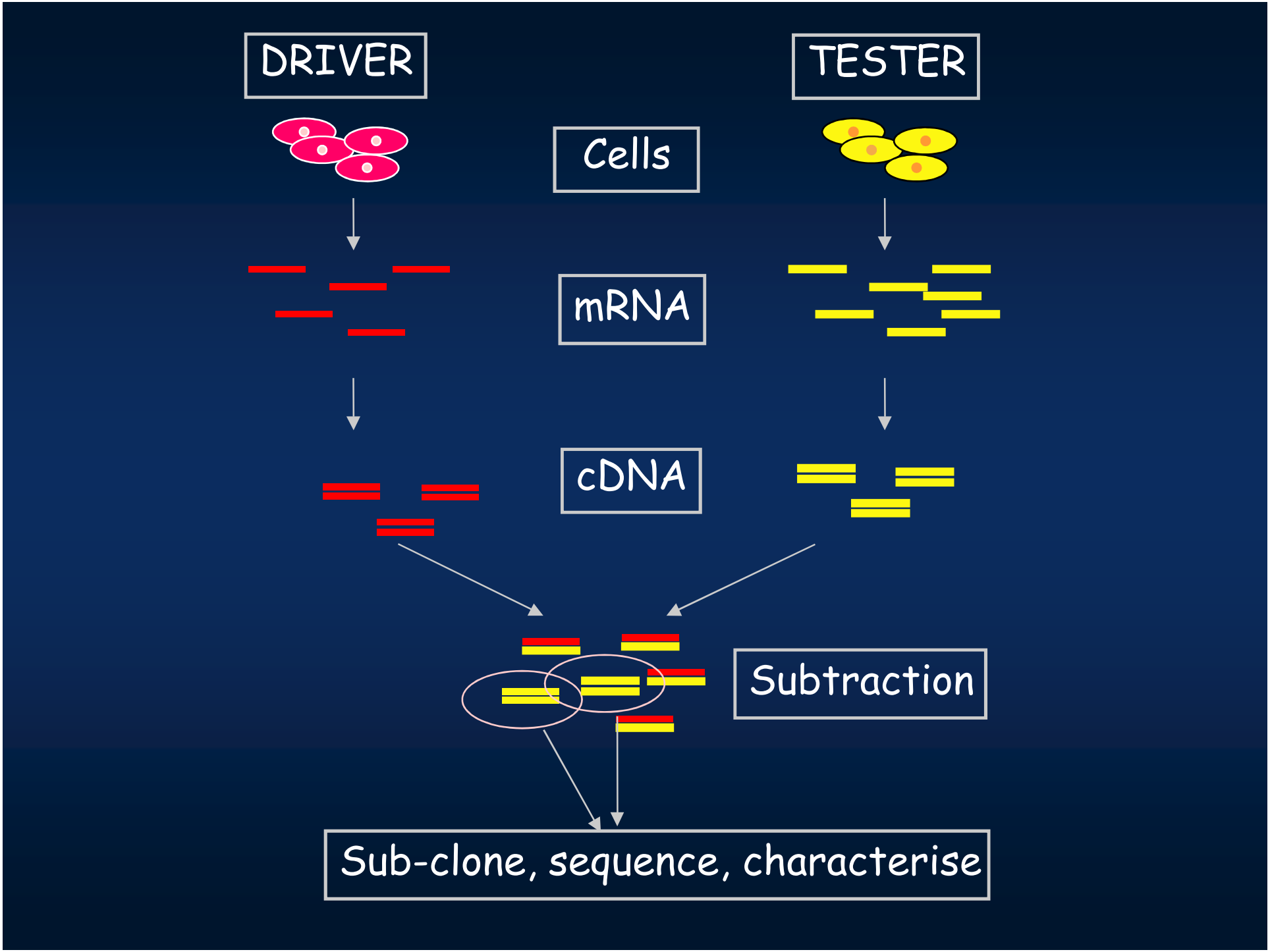
2. Differential Display

- Arbitrary or specific PCR amplification of cDNA
- Short PCR products displayed on PAGE gels
- Facilitates differential expression analysis across several samples
- Can identify novel transcripts
- Exact fold difference information not readily achievable


Lane 1 2 3 4 5 6 7 8 9 10 11 12 13
1
2
3
4
5
h1p
h1n
6
h2p
h2n
h3p
h3n
h4p
h4n
7
h5n
h5p
8
9
10

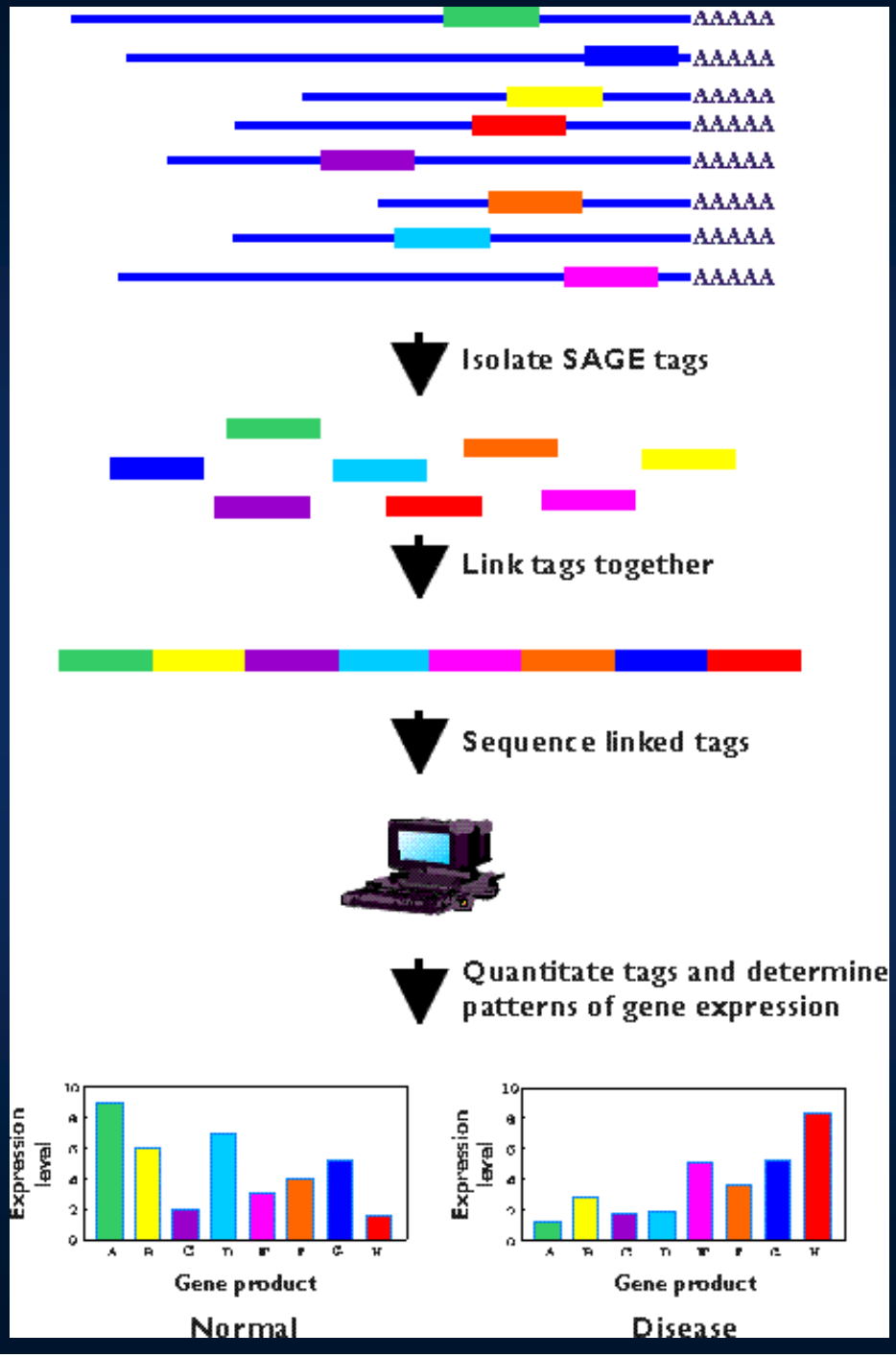
3. Subtractive Hybridisation

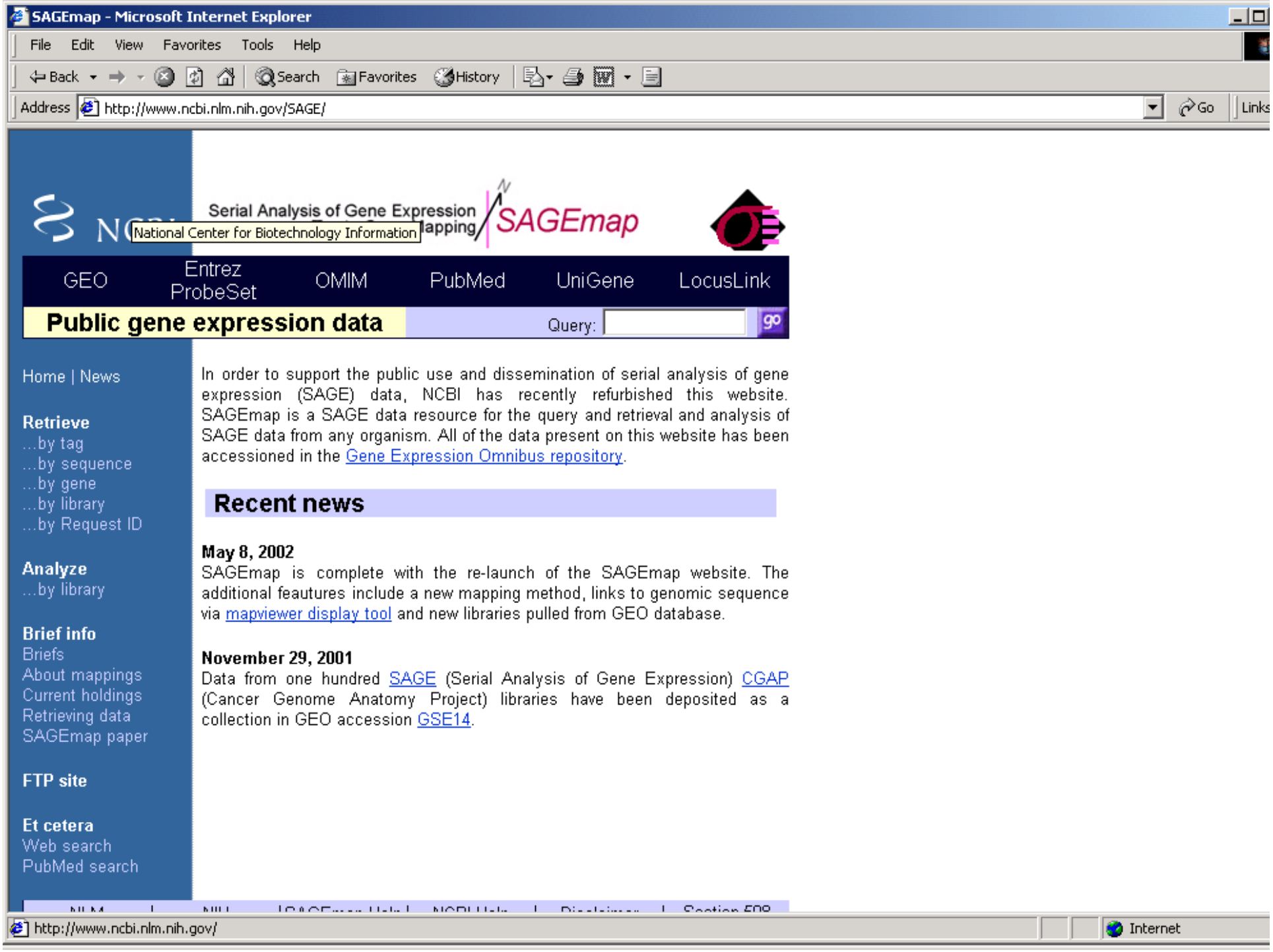
- Identifies differentially expressed genes between two samples
- Differential hybridisation of cDNA libraries
- Selective PCR amplification of discrete, enriched subsets (e.g. up- or down-regulated genes)
- Identification of transcripts via sequencing
- Normally, combined with other technique for validation purposes


DRIVER
TESTER
Cells
mRNA
cDNA
Subtraction
Sub-clone, sequence, characterise

4. Serial Analysis of Gene Expression (SAGE)

- Can be used to identify differential gene expression and novel genes
- Short sequence tags (10-14bp) of each gene is created from cDNA using restriction digestion
- Sequence tags are then ligated together to form serial molecules
- Tags sequenced in a high-throughput manner
- Relative ratio of unique tags between samples reflective of initial transcript concentration
- Need extensive sequencing facility and genome sequence info.


AAAAA
AAAAA
AAAAA
AAAAA
AAAAA
AAAAA
AAAAA
AAAAA
Isolate SAGE tags
Link tags together
Sequence linked tags
Quantitate tags and determine patterns of gene expression
Expression level
Gene product
Normal
Expression level
Gene product
Disease


SAGEmap - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites History
Address http://www.ncbi.nlm.nih.gov/SAGE/
Go Links
National Center for Biotechnology Information
Serial Analysis of Gene Expression
SAGEmap
GEO
Entrez ProbeSet
OMIM
PubMed
UniGene
LocusLink
Public gene expression data
Query:
go
Home | News
Retrieve
...by tag
...by sequence
...by gene
...by library
...by Request ID
Analyze
...by library
Brief info
Briefs
About mappings
Current holdings
Retrieving data
SAGEmap paper
FTP site
Et cetera
Web search
PubMed search
In order to support the public use and dissemination of serial analysis of gene expression (SAGE) data, NCBI has recently refurbished this website. SAGEmap is a SAGE data resource for the query and retrieval and analysis of SAGE data from any organism. All of the data present on this website has been accessioned in the Gene Expression Omnibus repository.
Recent news
May 8, 2002
SAGEmap is complete with the re-launch of the SAGEmap website. The additional feautures include a new mapping method, links to genomic sequence via mapviewer display tool and new libraries pulled from GEO database.
November 29, 2001
Data from one hundred SAGE (Serial Analysis of Gene Expression) CGAP (Cancer Genome Anatomy Project) libraries have been deposited as a collection in GEO accession GSE14.
http://www.ncbi.nlm.nih.gov/
Internet

5. DNA Microarray Hybridisation

- DNA microarrays are used to determine the differential expression of several thousand known genes and unknown genes (ESTs)
- There are several different types of DNA microarrays but the same principles apply

RNA

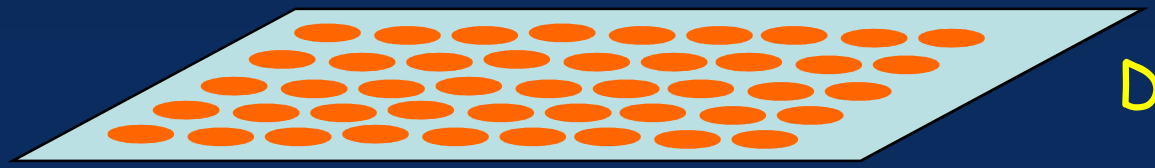
↓

cDNA

↓

Label cDNA (single stranded)

↓ Hybridise to array



DNA Microarray

Each spot contains single stranded cDNA/oligonucleotides from different genes

The amount of labeled cDNA that binds to each spot reflects the amount in the original sample

Samples can be analysed on different microarrays and compared to identify differentially expressed genes

DNA Microarray Hybridisation

- The two main are:

  A. Spotted DNA micro-arrays

  B. In situ synthesised arrays (e.g. GeneChips from Affymetrix)

- Spotted DNA micro-arrays: cDNA or Oligonucleotides representing each gene is spotted onto a solid support (glass slide) using Microarrayer

- GeneChips: Oligomers (20-25 bases) synthesised *in situ* on glass wafers using photolithography and combinatorial chemistry

What do DNA microarrays look like?

Spotted DNA micro-arrays

Glass slides

GeneChips:

Microfluidics system

Spotted DNA microarrays

- There are several advantages of spotted cDNA microarrays:

  - Can choose the genes you want to analyse, from tens to thousands

  - Generally cheaper than GeneChips

  - If you have the equipment and probes, several different arrays can be made in-house

  - Or alternatively specialised arrays can be purchased commercially

Spotted DNA microarrays

- There are several considerations for spotted arrays

  1. Type of solid support for DNA probes (nylon or glass)

  2. Type of probes (oligonucleotides, cDNA or PCR products)

  3. Optimisation of spotting probes to array (need Microarrayer)

  4. Optimization of RNA labelling and hybridisation procedure

  5. Require high resolution scanner to read microarray

  6. Require appropriate software to analyse data

Spotted DNA microarrays

- Non-porous support facilitates miniaturisation, i.e Glass slide or nylon
- 1,000-10,000 PCR products/oligonucleotides spotted per array
- PCR products (>100bps)
- Oligonucleotides (20-70mers)
- Fluorescence based detection (*e.g.* Cy3/Cy5) or radioactive based detection

Probe DNA Sources

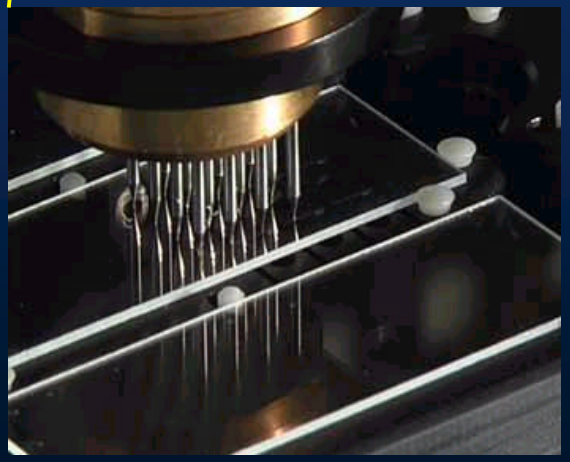
cDNA / PCR products

- 100 – 2,000 bp
- Cross-hybridisation
- Buy or Create library
- PCR optimisation

Oligonucleotides

- 20 - 70 bp probes
- Greater specificity
- Purchased Commercially
- Automated

Microarray Fabrication Process



Source DNA in buffer

↓

Spotting of Array

↓

Immobilisation by UV/baking

↓

Blocking

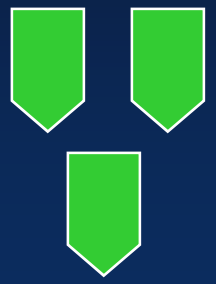
↓

Array Processing

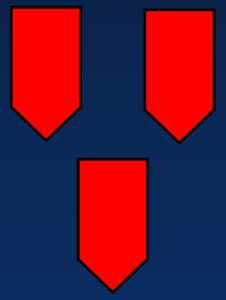
Spotted DNA microarrays

Label Control RNA with Fluorescent tag (eg Cy3)

Control RNA



Test RNA

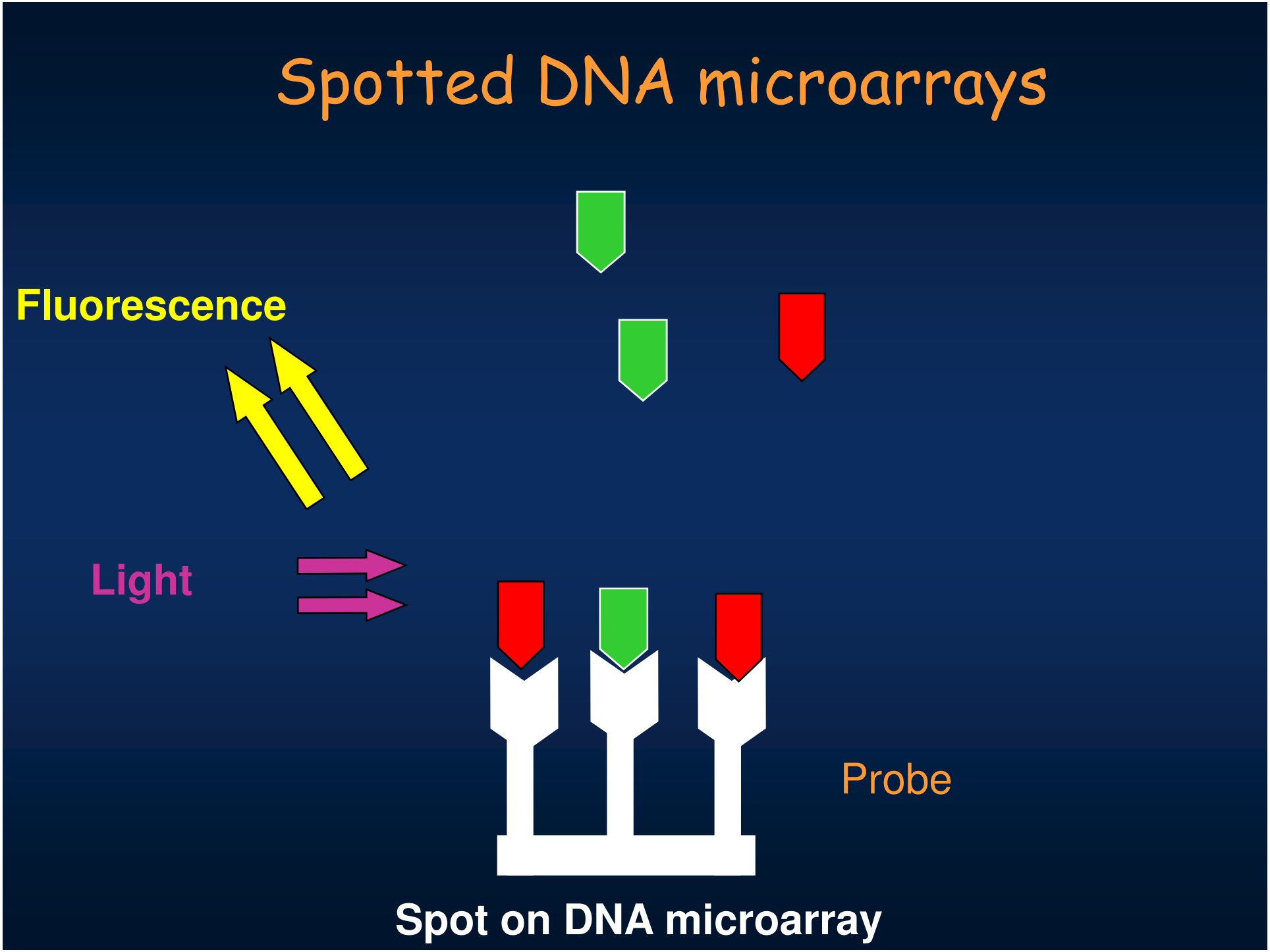


Label Test RNA with Fluorescent tag (eg Cy5)

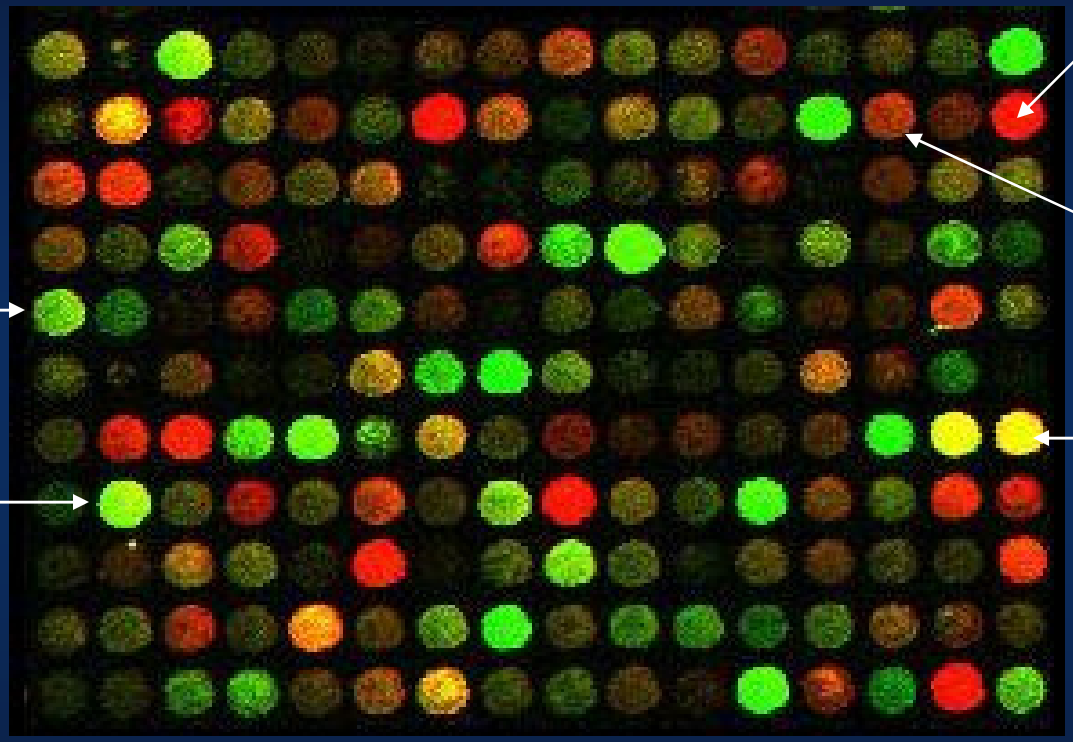
Probe

Spot on DNA microarray

Spotted DNA microarrays

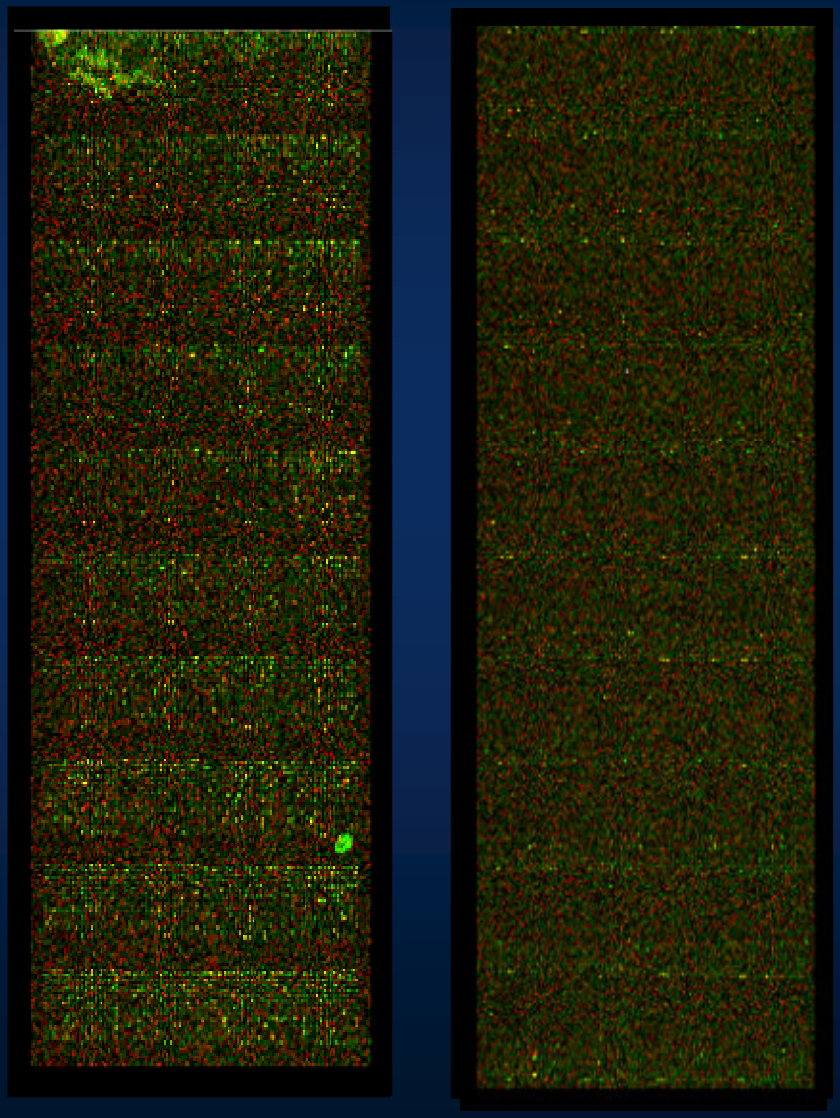


Hybridised Microarray Image



Example: 30K MWG oligo arrays

- A Human 30K (Hu30K) oligonucleotide set was obtained from MWG, and corresponding DNA microarrays fabricated onto glass slides
- The probe length of each oligonucleotide is 50bp in length, with one oligonucleotide per gene
- Hybridised with control RNA and melanoma cell line line (test)



Affymetrix GeneChip system

- There are several advantages of GeneChips:

  - Arrays are fabricated by Affymetrix to contain almost all known genes

  - All protocols and kits are optimised to give the best results- thus saving time and effort

  - Multiple controls on the array to improve quality of data

  - Substantial information about the genes and the probes used, including exact sequence and location on gene on their webpage

Affymetrix GeneChip system

- Oligomers (20-25 bases) synthesised *in situ* on glass wafers

- Several types of arrays available for different organisms and species (human, mouse, rat etc)

- Human RNA Expression Arrays:

  - HuGeneFL Array: 5,600 full-length human genes.

  - Human Cancer G110: 1,700 full-length human genes implicated in cancer

  - Human Genome U133: 47,000 transcripts

  - Human Exon 1.0 ST Array: 1.4 million exons on 300,000 transcripts

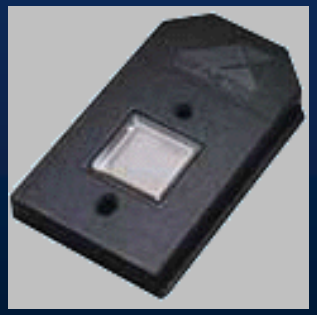
Affymetrix GeneChip system

Hybridisation oven

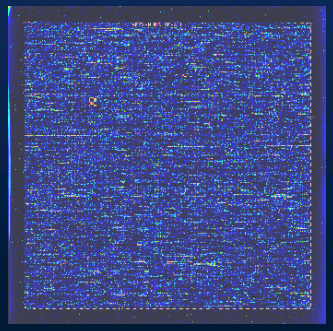
Scanner

GeneChip Fluidics Station

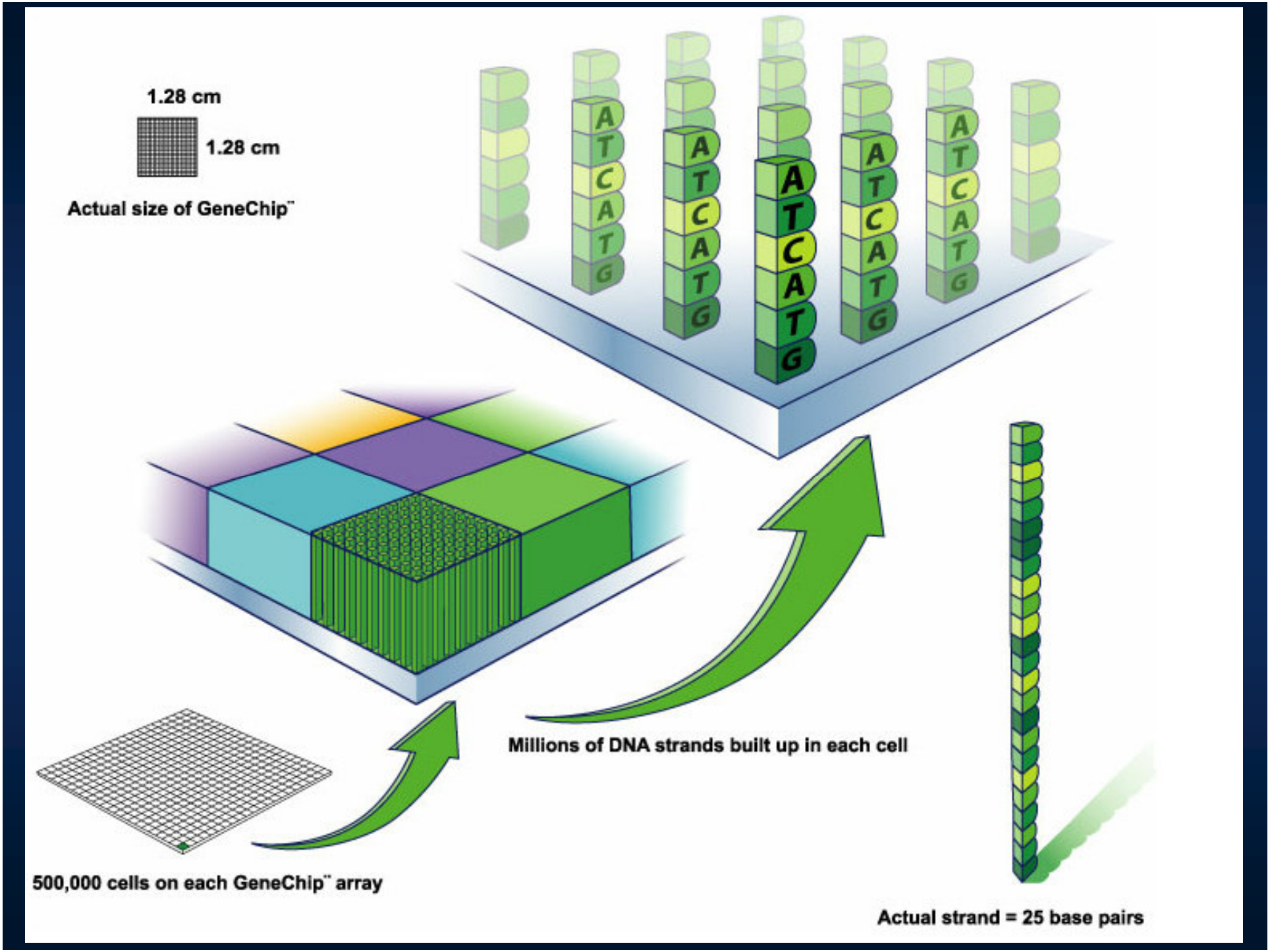
GeneChip

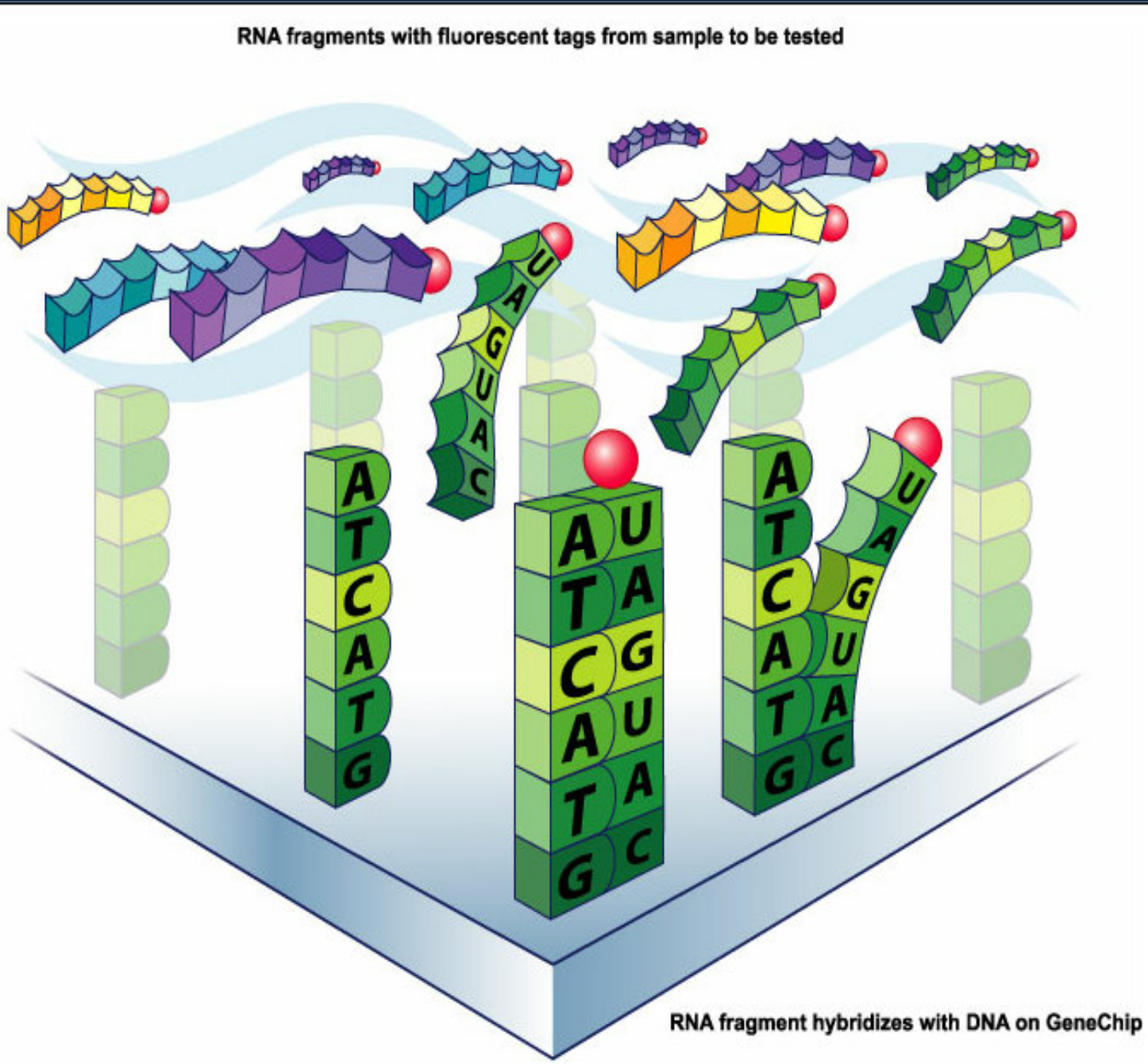


Representative Image



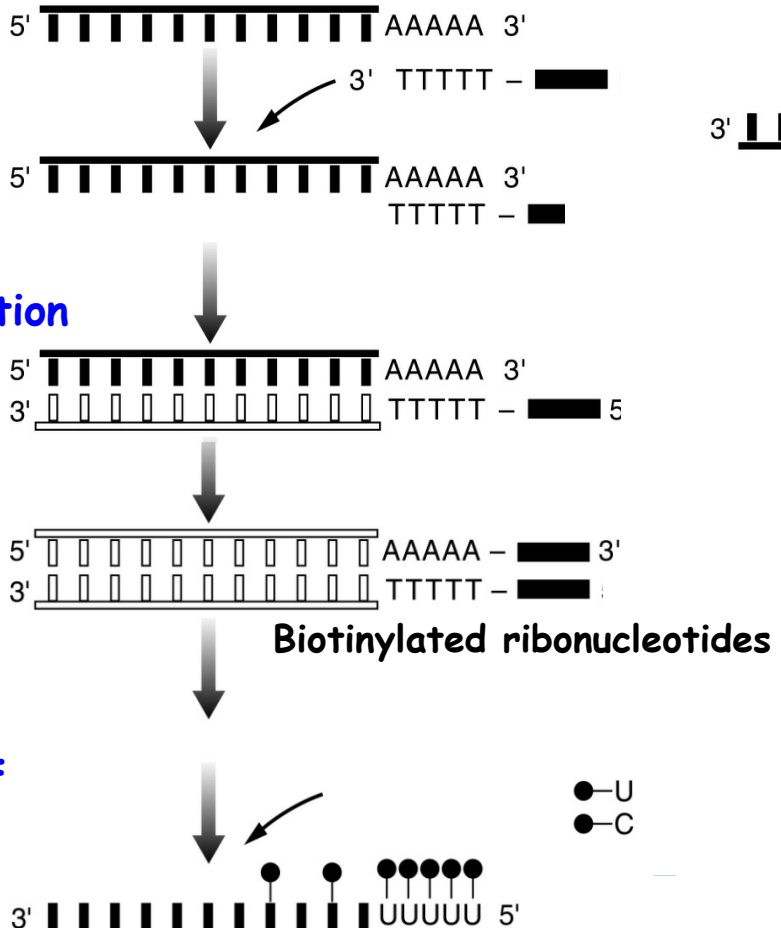
[Ms. Alison Murphy – Conway]

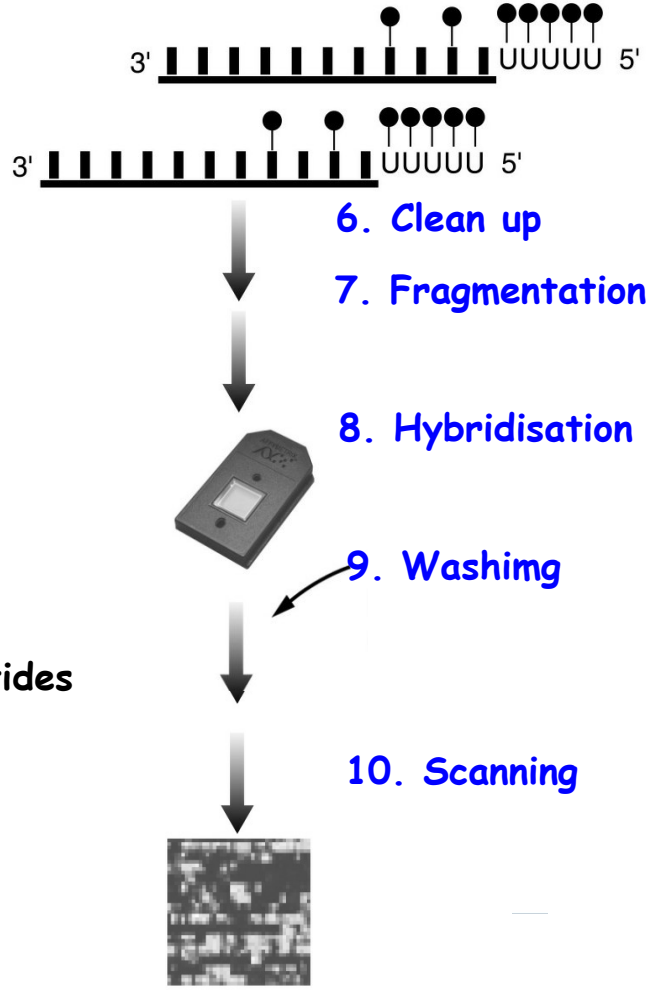

1.28 cm
1.28 cm
Actual size of GeneChip™
Millions of DNA strands built up in each cell
500,000 cells on each GeneChip™ array
Actual strand = 25 base pairs


RNA fragments with fluorescent tags from sample to be tested
RNA fragment hybridizes with DNA on GeneChip

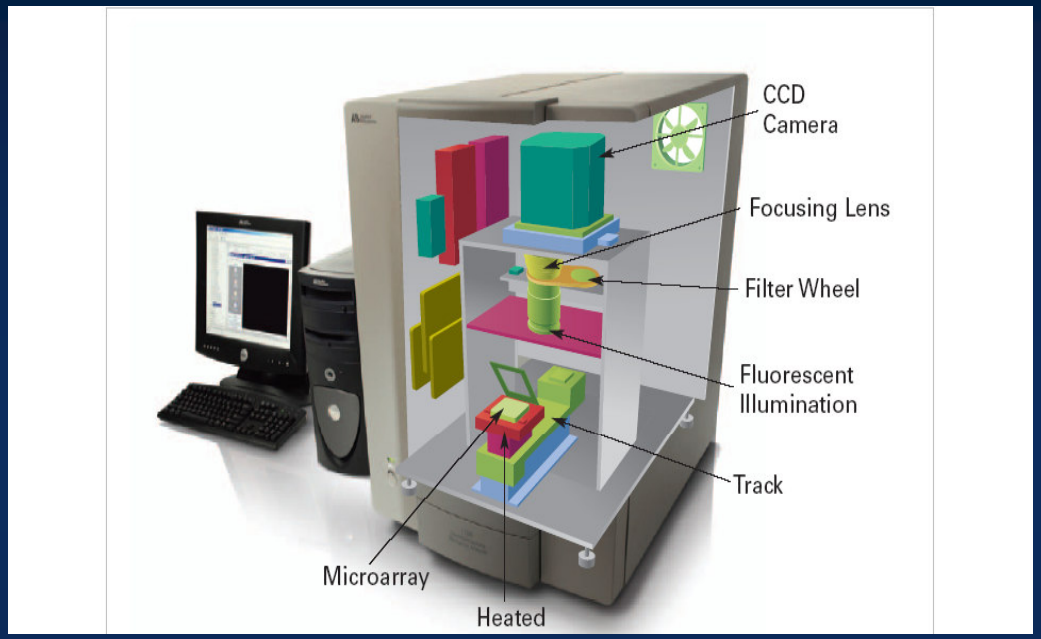
Target labelling for GeneChips

1. Total RNA
2. Primer binding
3. Reverse Transcription
4. cDNA Synthesis
5. Amplification & biotin labeling of antisense cRNA





AB 1700 Human Expression Array System



- 31,097 probes
- Curated oligonucleotides
- Chemiluminescence detection

Data mining, analysis & management

- Images obtained in digitised form; remove artifacts

- Further analysed using a variety of software programs

- Controls (replicates, colour reversal, control cDNA/oligos)

- Massive amounts of data generated

AMAD (www.microarray.org). Multivariate statistical analysis using Cluster and Treeview (FREE)

GeneSpring (Silicon Genetics). Analysis & visualisation tool. (Commercial Option)

Pattern Recognition

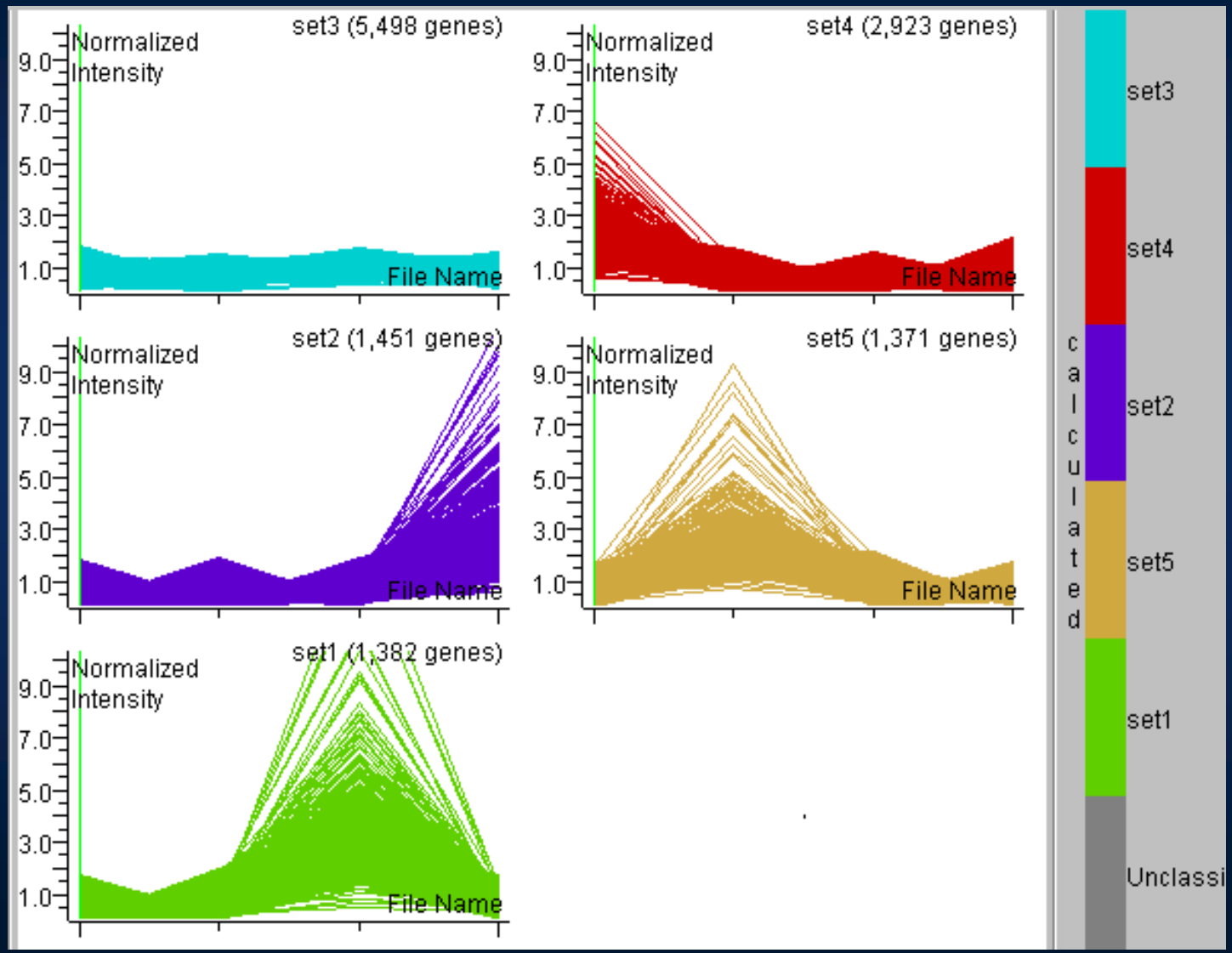
- Pattern recognition analysis to identify genes that are regulated in a similar manner across many experiments

- Performed with multivariate analytical algorithms such as hierarchial clustering, k-means clustering and self-organising maps

- Highlights a relatively small number of transcripts, that can be further verified by other means

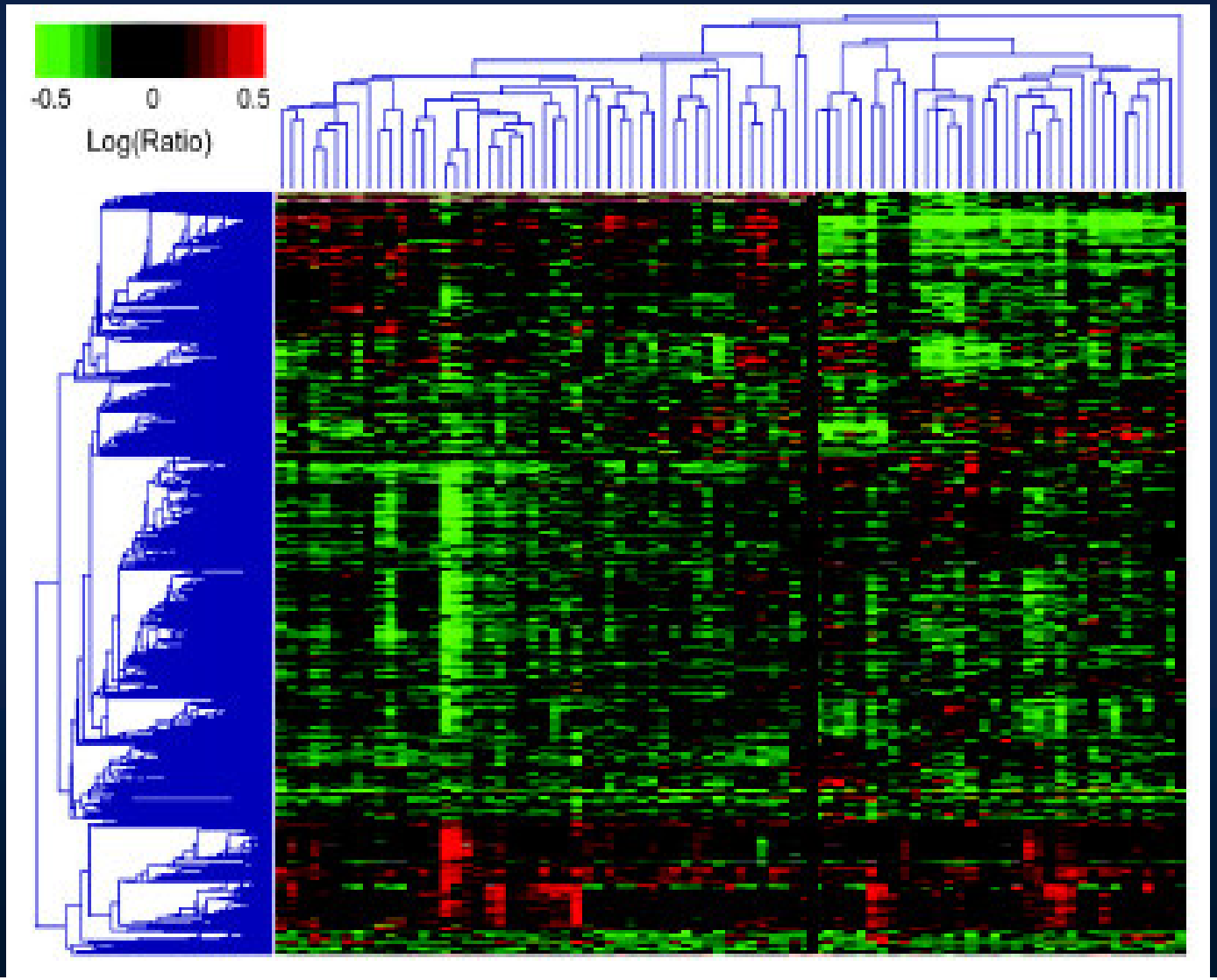
- Public access to results (www.ebi.ac.uk/microarray/MGED/)

Pattern Recognition- GeneSpring



Pattern Recognition- Clustering

Samples



Genes

Gene expression profiling

- Gene function identification
- Disease subclassification (*e.g.* cancer)
- Elucidate molecular determinants of biological processes (model systems/tissue samples)
- Drug discovery and target validation

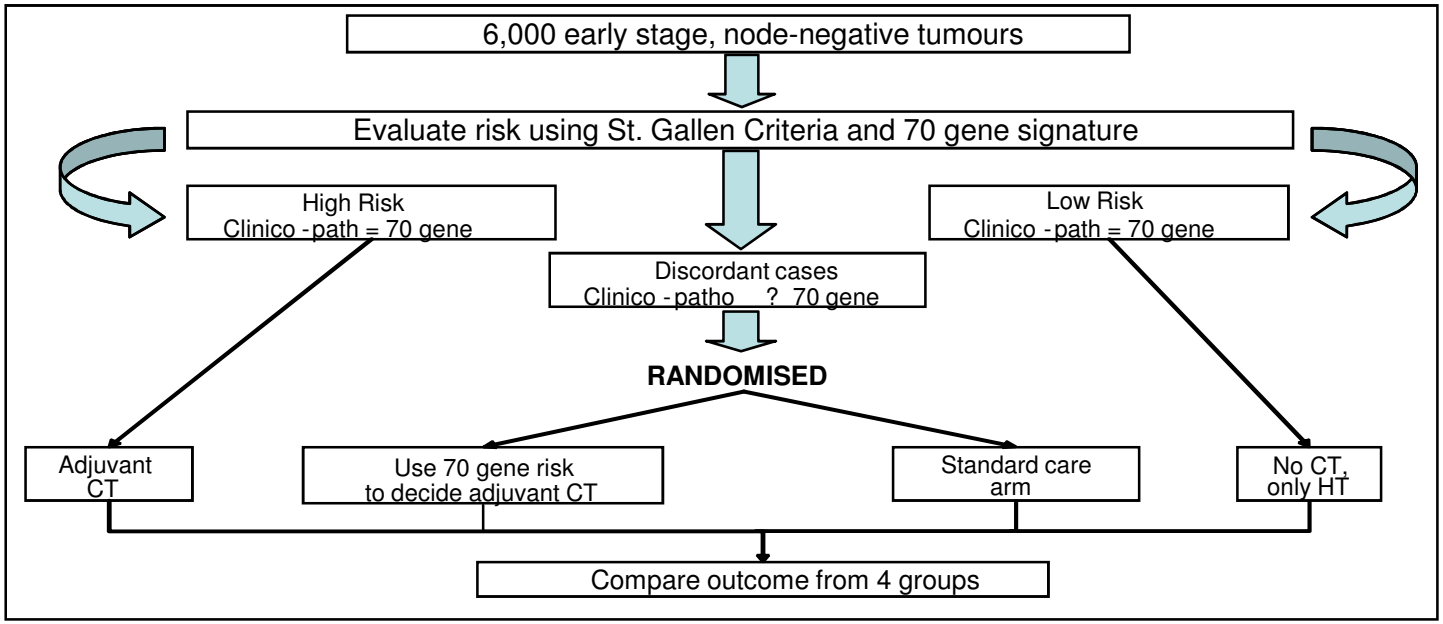
Prognosis prediction

- *Nature*, Vol 415, pp. 530-536 (2002)
- Node-negative breast cancer patients:

  70% are cured by surgery, remaining 30% have aggressive disease
- Van't Veer et al. examined expression of 25,000 transcripts; 78 node-negative patients
- Found 70 genes that could predict outcome in 65/78 patients
- Most useful genes for prediction involved in cell cycle, signal transduction, invasion and metastasis

MINDACT Prospective Study

Microarray In Node-Negative Disease may Avoid ChemoTherapy



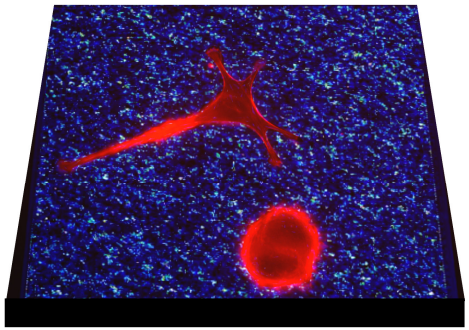
Brennan DJ et al. *Expert Opin. Biol. Ther.*, 5(8), 1069-1083 (2005)

Finding therapeutic targets

- *Nature Genetics*, Vol 29. 1st October, 2001
- Medulloblastoma (childhood tumour)
- 23 primary tumours (10 metastatic, 13 non-metastatic)
- Using gene expression profiling, found 85 genes that could accurately predict metastasis
- Found upregulation of PDGFRalpha and RAS/MAPK pathway in metastatic tumours
- Inhibitors of this pathway inhibited processes involved in metastasis

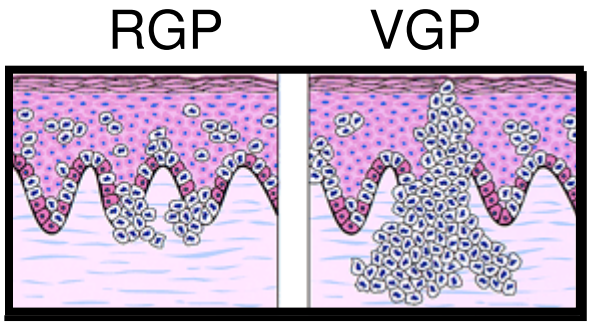
Cell-Biomaterial Interaction

PNAS, 2003



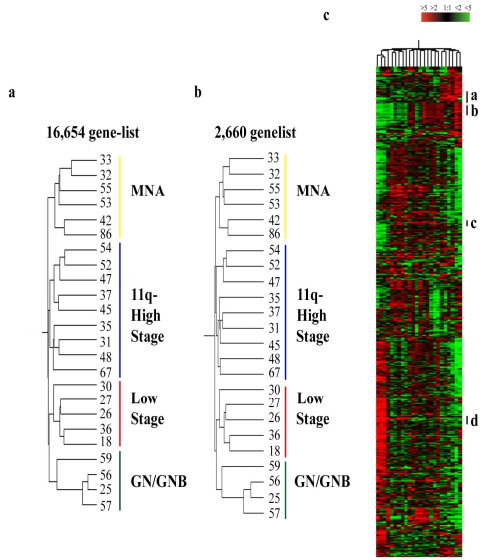
Malignant Melanoma

Carcinogenesis, 2005



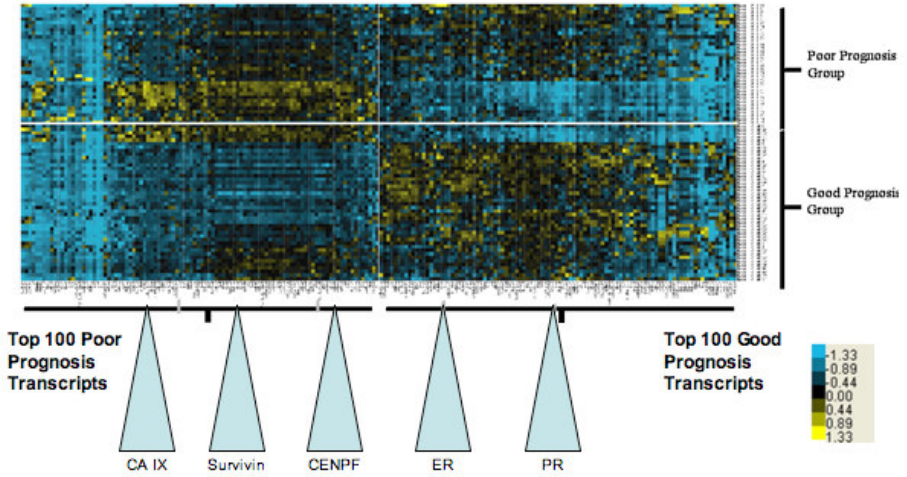
Neuroblastoma

Carcinogenesis, 2004



Breast Cancer

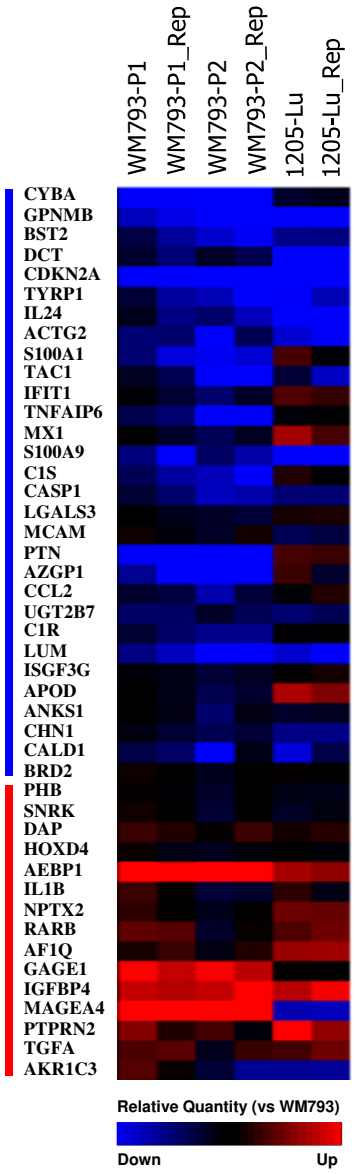
International Journal of Cancer, 2007



TaqMan Cards (Low Density Arrays)

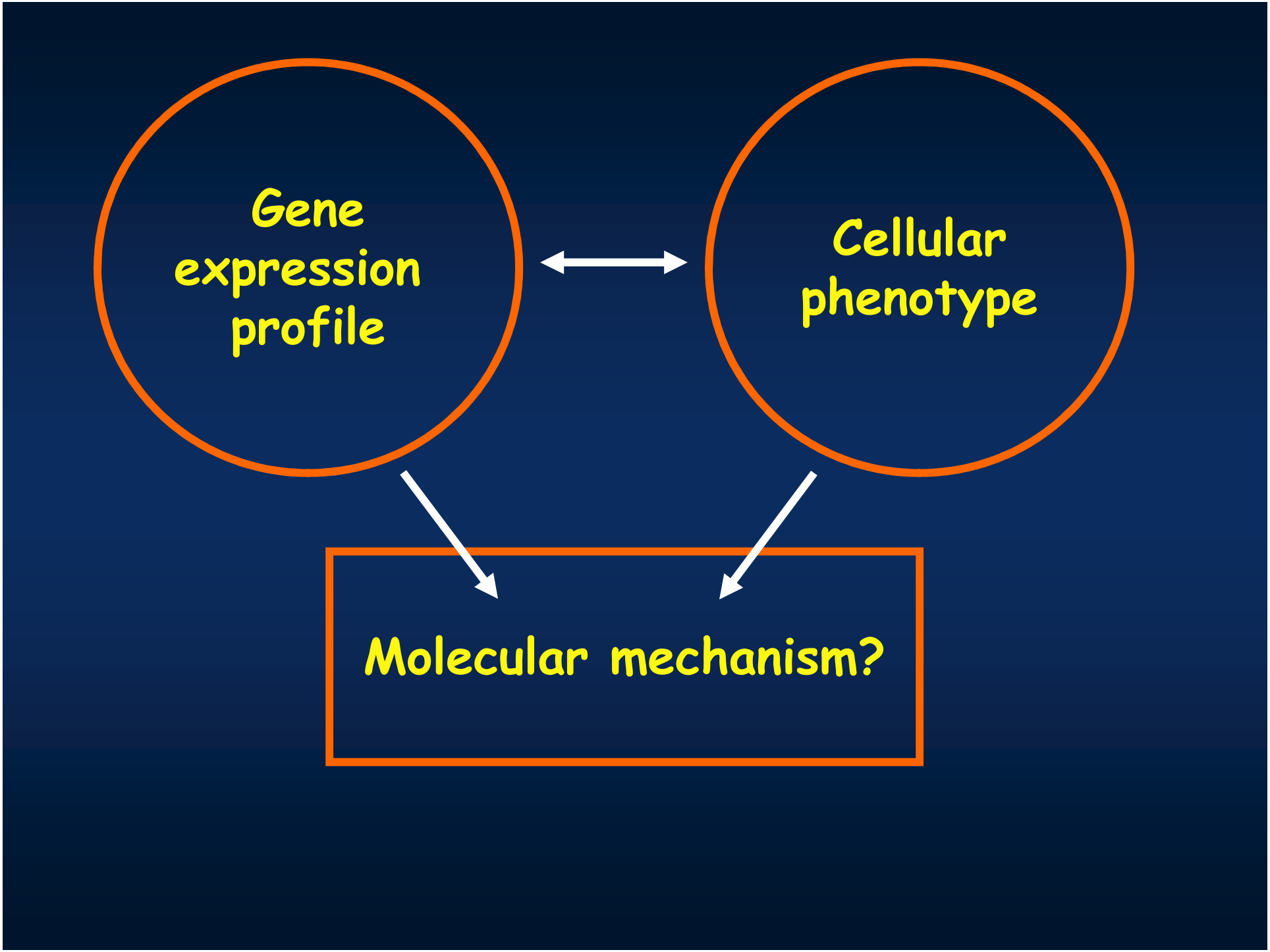
- More genes, less sample
- qRT-PCR
- Pre-designed probes and primers
- Convenient 384 assay format
  - 95 genes (duplicate) X 2 samples
  - 18 S rRNA control
- Minimises liquid handling, reagent consumption
- Cost effective for large-scale screening projects

45 out of 66 genes



Relative quantities

WM793 vs WM793-P1	WM793 vs WM793-P1	WM793 vs 1205-Lu
3.53E-06	7.84E-06	0.5
0.02	0.001	0.001
0.1	0.01	0.08
0.2	0.3	0.002
0.0005	0.0002	4.26E-06
0.2	0.02	0.02
0.3	0.08	0.004
0.1	0.1	0.01
0.06	0.01	2.4
0.3	5.79E-05	0.2
0.6	0.3	3.8
0.2	0.002	0.8
0.7	0.3	13.4
0.05	0.1	0.004
0.2	0.02	1.6
0.2	0.03	0.1
0.7	0.4	1.7
1.04	1.0	0.2
0.0009	0.001	3.6
0.03	0.001	1.8
0.3	0.2	1.5
0.1	0.3	0.16
0.24	0.07	1
0.05	0.003	0.01
0.7	0.4	1.4
0.8	0.3	18.5
0.8	0.4	0.5
0.5	0.3	0.08
0.2	0.3	0.1
0.1	0.8	1.2
1	0.7	0.6
1.3	0.6	0.6
1.6	2.3	1.9
1.0	0.7	0.9
810	309	20
2.32	0.37	1.60
1.9	0.7	7.7
5.9	0.8	6.7
1.5	1.4	19.5
134.7	69.4	N/A
35.9	83.5	66.6
775.2	987.8	0.033
7.2	0.4	134.4
5	1.8	6
3.2	0.19	0.06


Gene expression profile
Cellular phenotype
Molecular mechanism?

Summary

- Overall different techniques such as ESTs, SAGE, differential display, subtractive hybridisation and DNA microarrays can be used to identify genes with differential expression

- DNA microarrays, in particular, provide detailed information on thousands of genes, and differential expression between samples can be used to sub-classify diseases and identify potential targets for therapeutics

Literature Sources

Review articles

Lennon, G. G. (2000). High-throughput gene expression analysis for drug discovery. Drug Discovery Today, vol. 5, no. 2, pp. 59-66.

Schulze, A. and Downward, J. (2001). Navigating gene expression using microarrays – a technology review. Nature Cell Biology, vol. 3, pp. E190-E195.

Methods in Enzymology (2006). Volume 410. DNA microarrays, Part A: Array platforms and wet-bench protocols.