

Microarrays and genomics

CH 370 - Notes by:

Vishy Iyer
Molecular Genetics and Microbiology

inside, linked together by hydrogen bonds. This structure as described is rather ill-defined, and for this reason we shall not comment on it.



This figure is purely diagrammatic. The two ribbons symbolize the two phosphate-sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis.

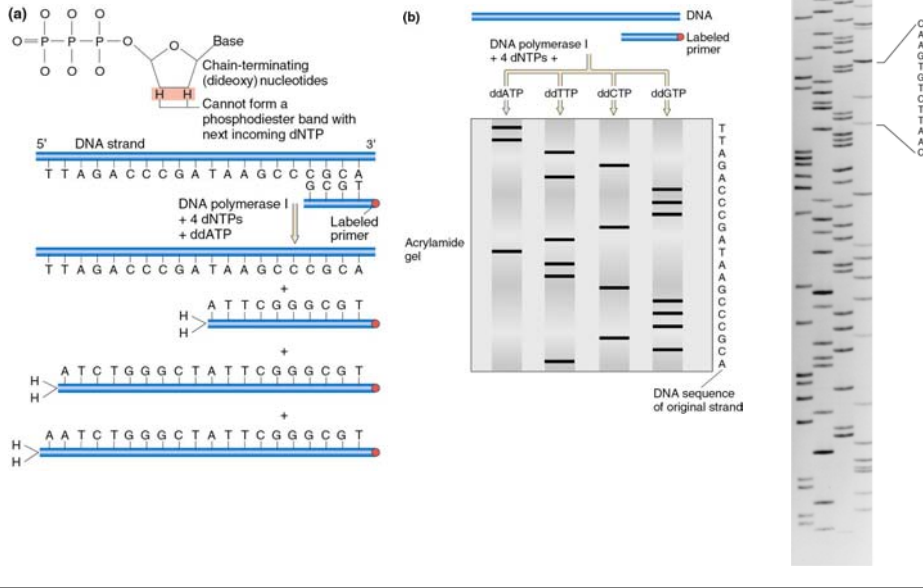
We wish to put forward a radically different structure for the salt of deoxyribose nucleic acid. This structure has two helical chains each coiled round the same axis (see diagram). We have made the usual chemical assumptions, namely, that each chain consists of phosphate diester groups joining β -D-deoxy-ribofuranose residues with 3',5' linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains follow right-handed helices, but owing to the dyad the sequences of the atoms in the two chains run in opposite directions. Each chain loosely resembles Furbert's model No. 1; that is, the bases are on the inside of the helix and the phosphates on the outside. The configuration of the sugar and the atoms near it is close to Furbert's 'standard configuration', the sugar being roughly perpendicular to the attached base. There

Nature – 1953



Nature – 2001

Dideoxy sequencing



Automated dye-terminator sequencing

4-fluorescently labelled dideoxy dye terminators

ddATP

ddGTP

ddCTP

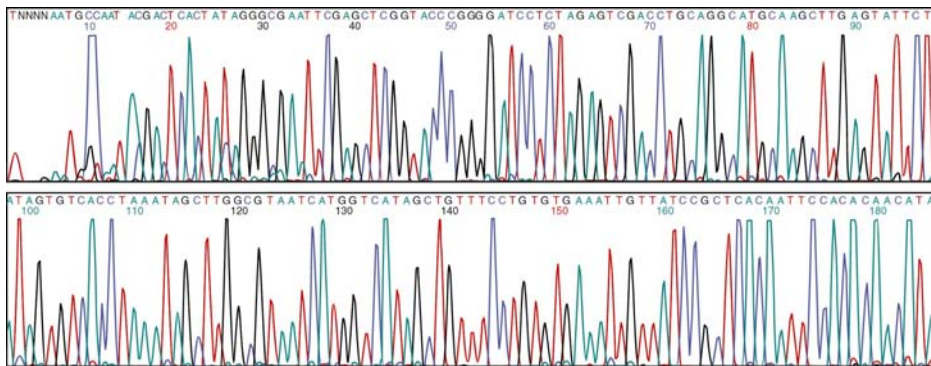
ddTTP

pool and load in a single well or capillary

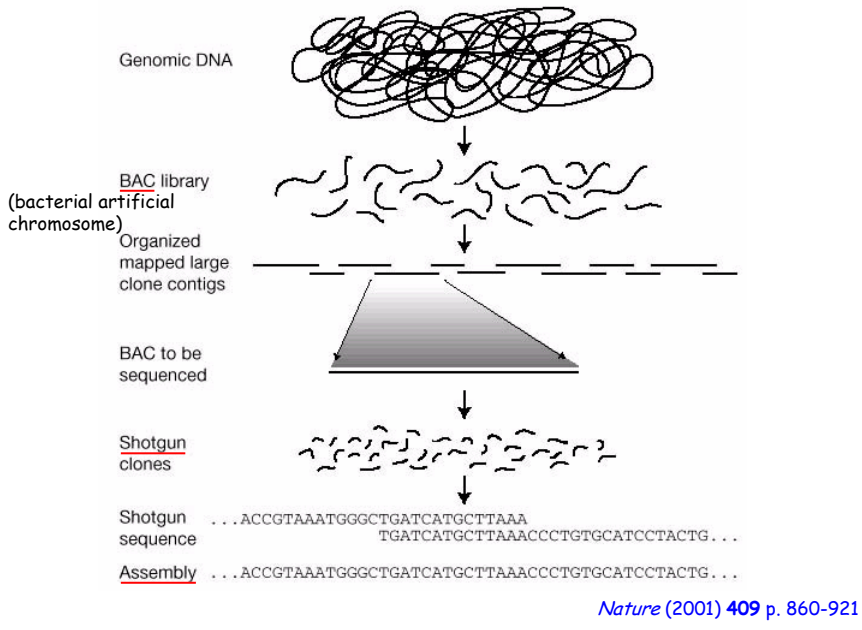
• scan with laser + detector specific for each dye

• automated base calling

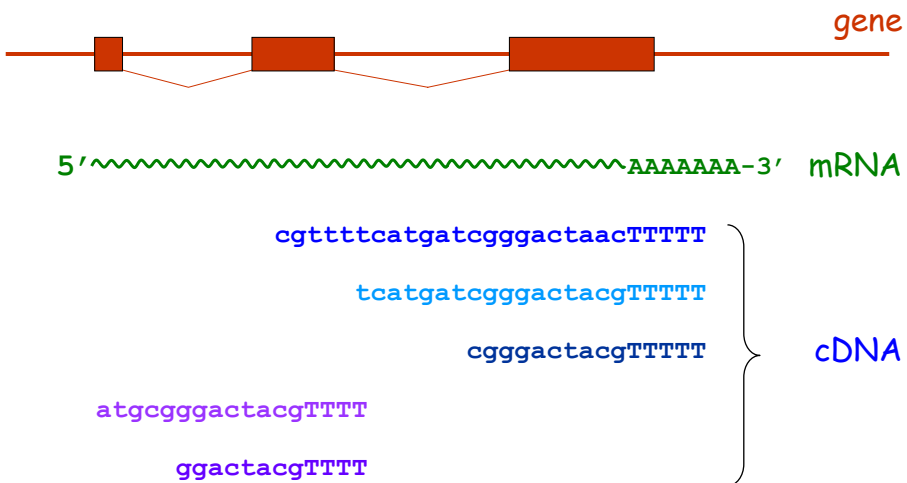
• very long reads (~ 1000 bases)/run



Physical mapping and sequencing of the human genome



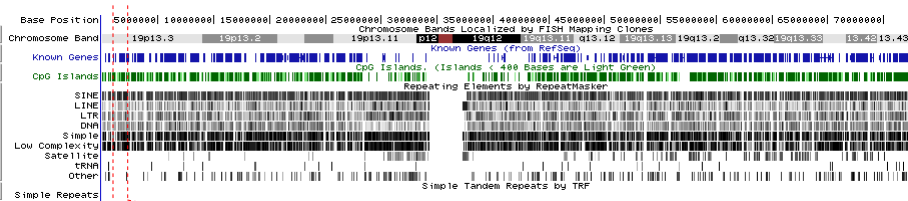
Genes can also be identified by sequencing cDNAs at random. The sequenced cDNAs are called **ESTs** (expressed sequence tags)



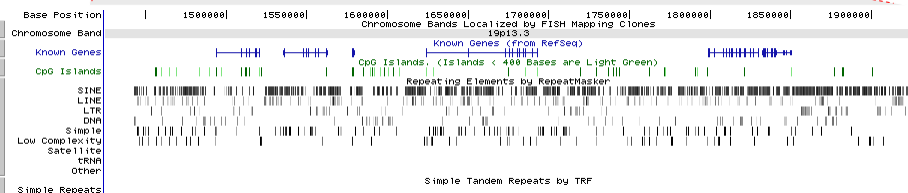
Finding genes in genomes

- look for open reading frames (works for prokaryotes and simple eukaryotes)
- compare to EST or cDNA sequence
- similarity to other genes and proteins
- Gene prediction algorithms (identifying splice sites, coding sequence bias, etc.)

Human chromosome 19: 1-72365612 (72 Mb)

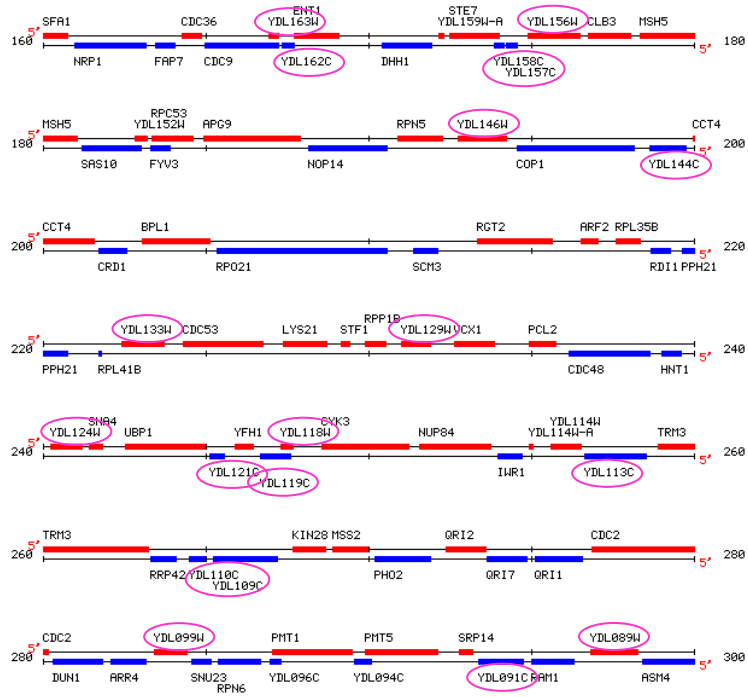


chr19: 1425001-1925000 (500 Kb)



<http://genome.ucsc.edu/>

Genes of unknown function in the yeast genome

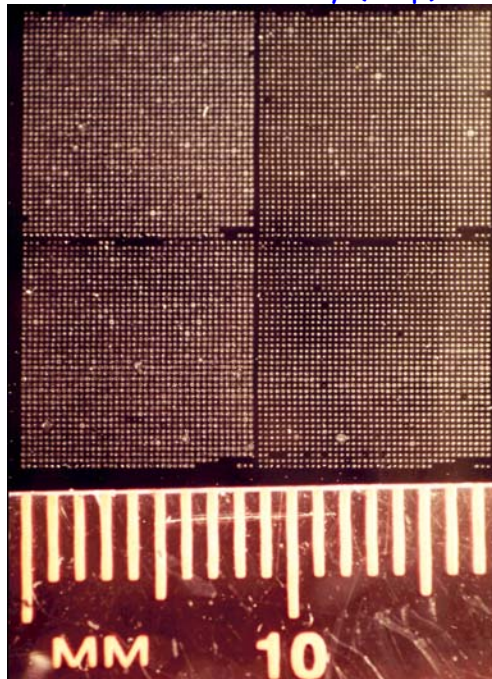


Drew Sheneman - The Newark Star Ledger

Functional genomics and proteomics

- Identify genes and proteins encoded by the genome (*Gene finding*)
- Measure gene expression on a genome-wide scale (microarrays)
- Identify protein function
30-50% of the genes in a genome are of unknown function
- Identify protein interactions, biochemical pathways, gene interaction networks inside cells

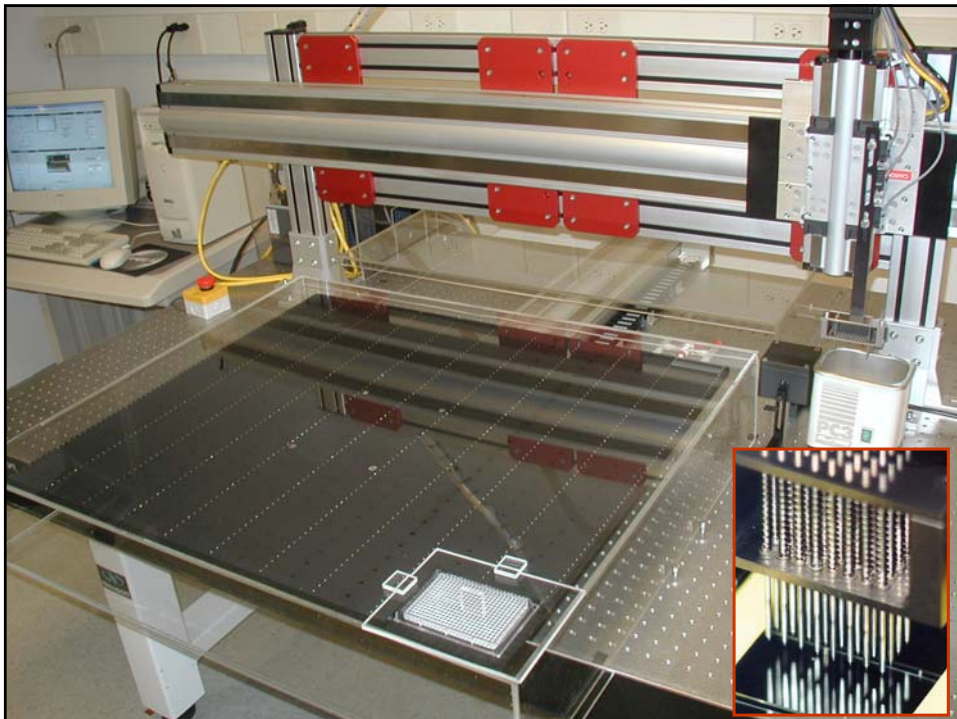
DNA microarray (chip)



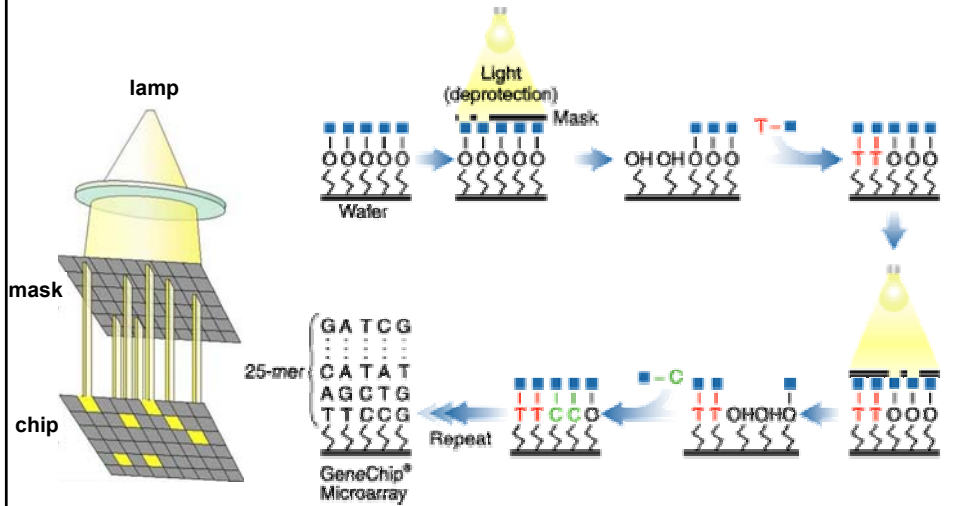
Methods of making microarrays

- Robotic spotting
 - using a printing tip
 - using inkjets
- Synthesis of oligonucleotides
 - photolithography (Affymetrix)
 - using inkjets
 - Digital Light Processor (DLP) or Digital Micromirror Device (DMD)

Microarrays can be used to study gene expression, DNA-protein interactions, mutations, protein-protein interactions, etc., all on a genome-wide scale

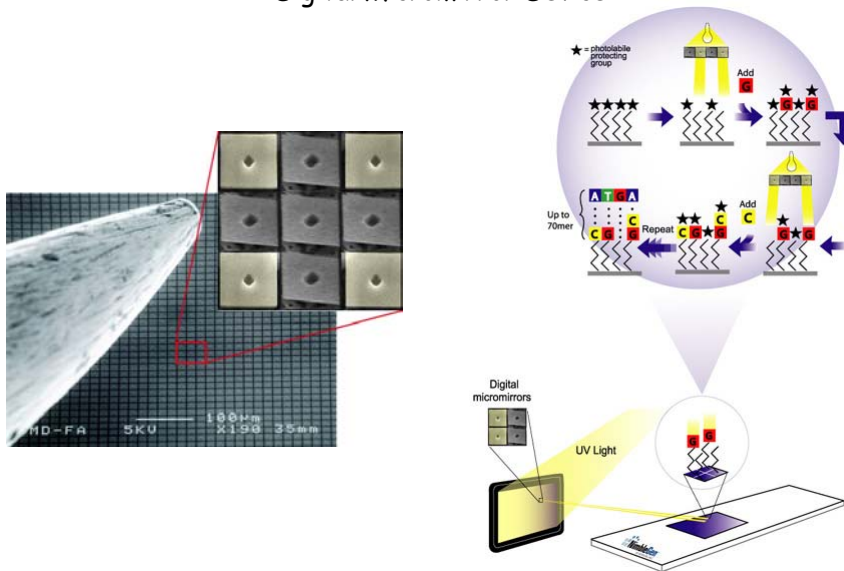


Affymetrix GeneChip

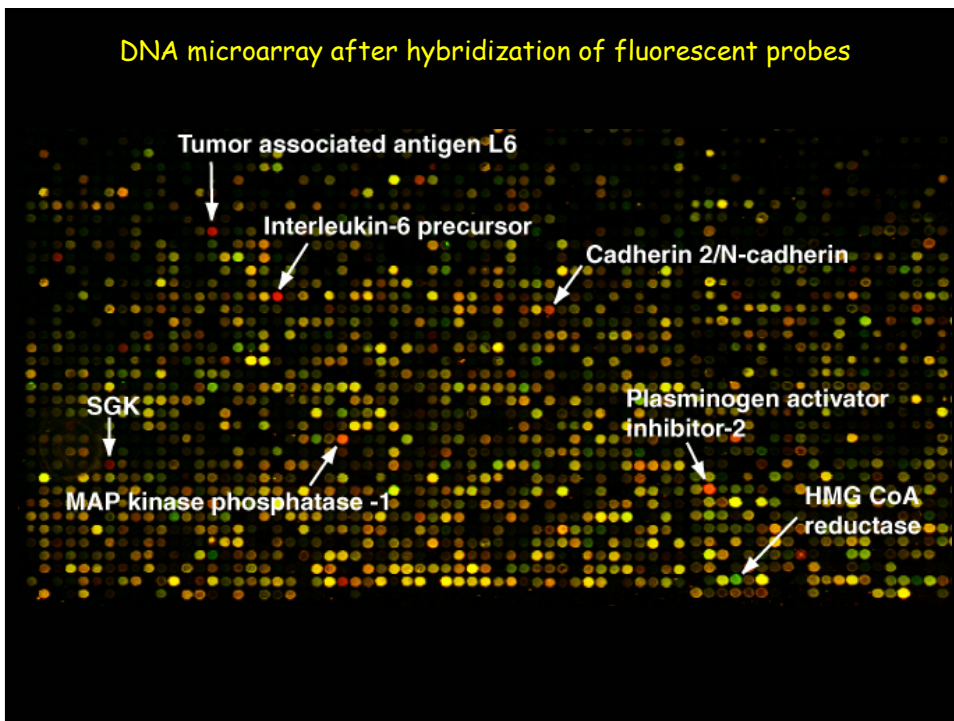
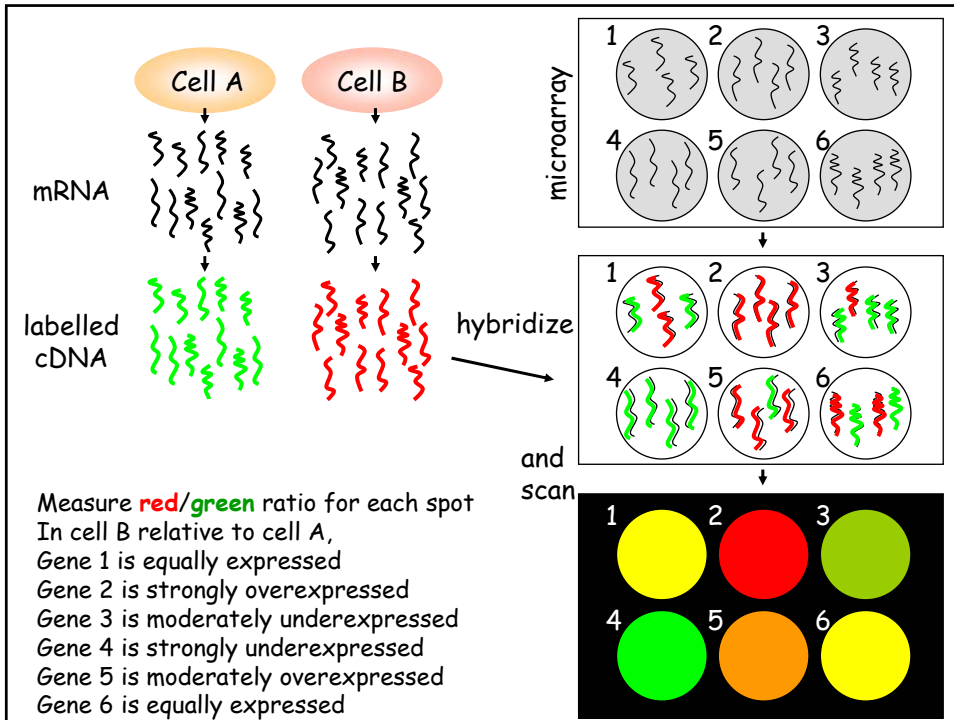


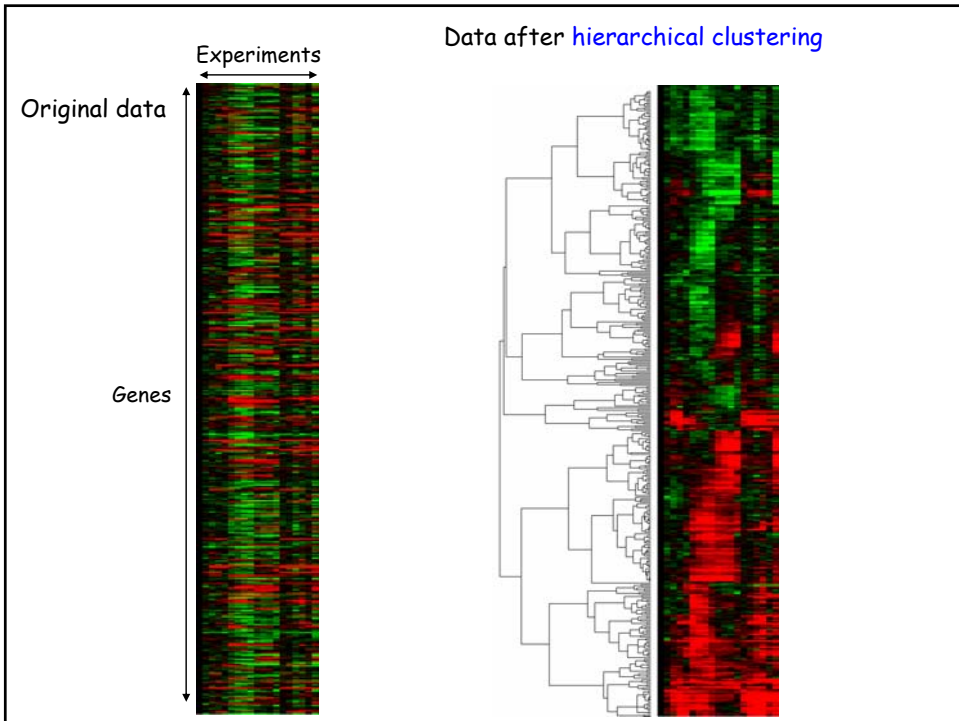
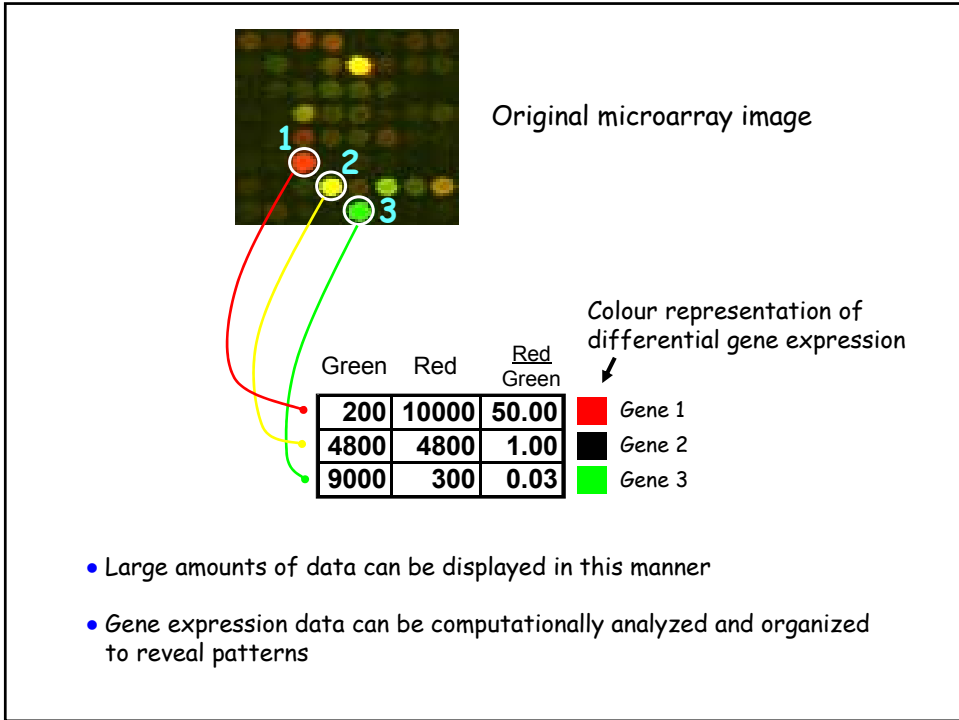
courtesy: www.affymetrix.com

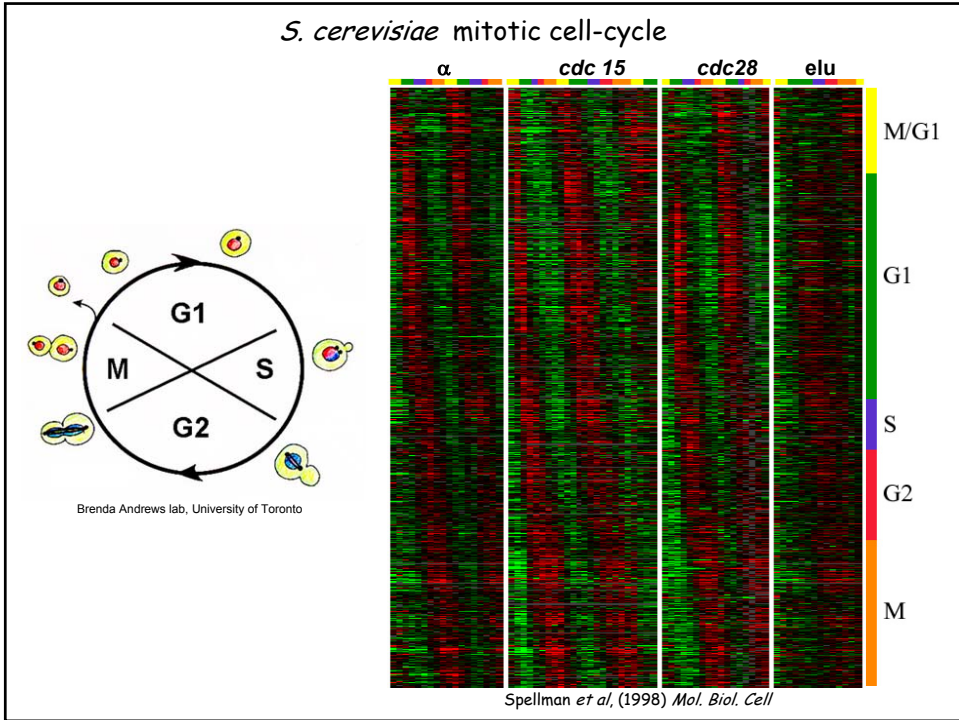
Digital Micromirror Device

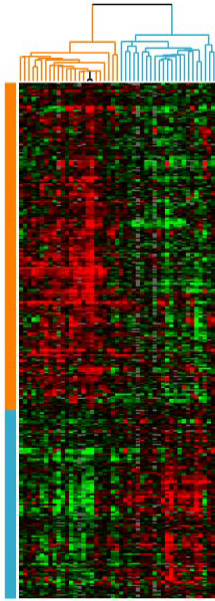


courtesy: www.nimblegen.com



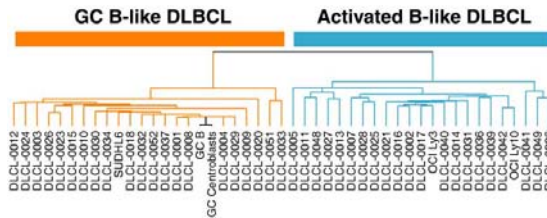




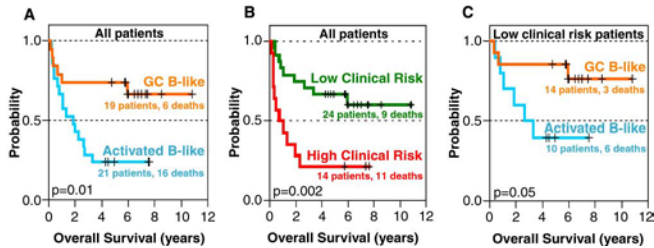


Clustering of tumour samples from cancer patients can be used for molecular classification of cancers. This may be useful for diagnosis and treatment

Subtypes of Diffuse Large B-Cell Lymphoma (DLBCL)



Nature (2000) 403: p 503

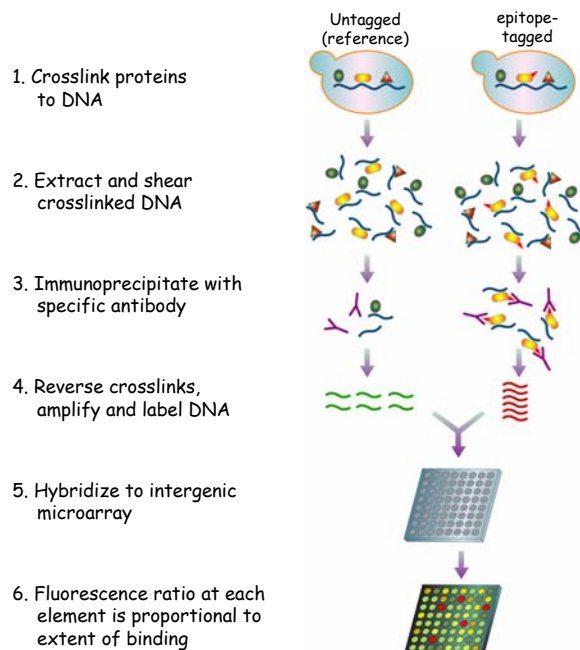


Nature (2000) 403: p 503

Other applications of microarrays

- DNA-protein interactions
mapping genome-wide distribution of proteins that interact with DNA
- Genomic amplifications and deletions
Comparative Genomic Hybridization
- RNA and protein localization
analysis of RNAs associated with membrane-associated ribosomes, polysomes, different sub-cellular fractions
- Polymorphisms
oligonucleotide (Affymetrix) arrays used for analyzing single nucleotide polymorphisms (SNPs) for linkage mapping and association studies
- Protein microarrays
detecting proteins in complex mixtures
- Tissue microarrays
high-throughput pathology

Mapping the binding distribution of proteins on the genome



Genome-wide map of DNA protein interactions



Discovering binding specificity

```
Simplified A 22239aa9
motif letter- C 7153:::
probability G 16341::
matrix T :::::1
```

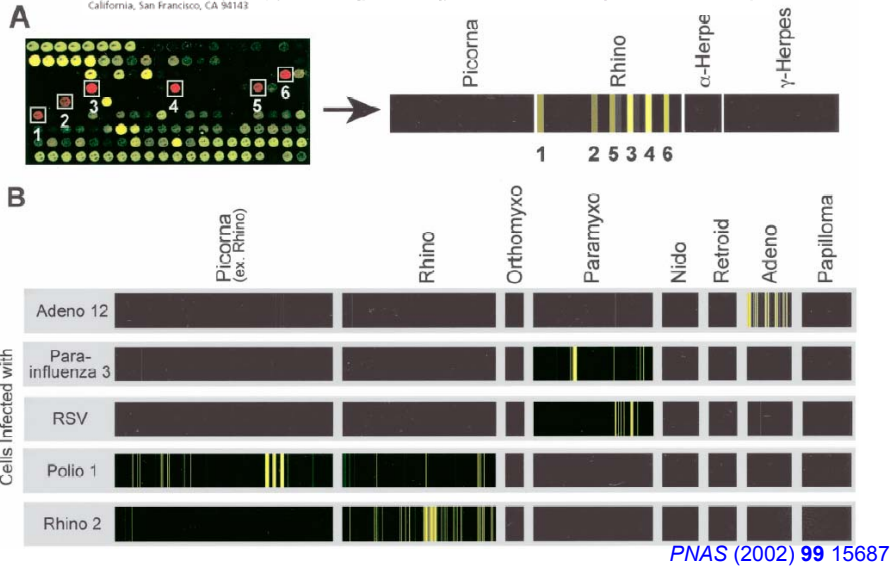
```
bits 2.2
      2.0
      1.7
      1.5 **
Information 1.3 ***
content 1.1 ****
(7.9 bits) 0.9 ****
           0.7 ** ****
           0.4 *** ****
           0.2 ****
           0.0 -----
```

```
Multilevel CGCGAAAA
consensus AGC
sequence AA
```

Microarray-based detection and genotyping of viral pathogens

David Wang¹, Laurent Coscoy¹, Maxine Zylberberg², Pedro C. Avila¹, Homer A. Boushey¹, Don Ganem^{1,5}, and Joseph L. DeRisi^{1*}

Departments of ¹Biochemistry and Biophysics, ²Microbiology and Immunology, and ³Medicine, and ⁴Howard Hughes Medical Institute, University of California, San Francisco, CA 94143



The role of Bioinformatics

- Databases
 - Genbank, SwissProt (DNA and protein sequence)
 - functional genomics and proteomics data (gene expression, protein profiles, drug data)
 - protein structure data (crystallography, NMR)
 - biomedical literature (PubMed)
- Analysis methods
 - finding patterns in expression data (clustering)
 - gene interaction networks
 - data mining
 - regulatory elements, novel genes etc.
 - visualization