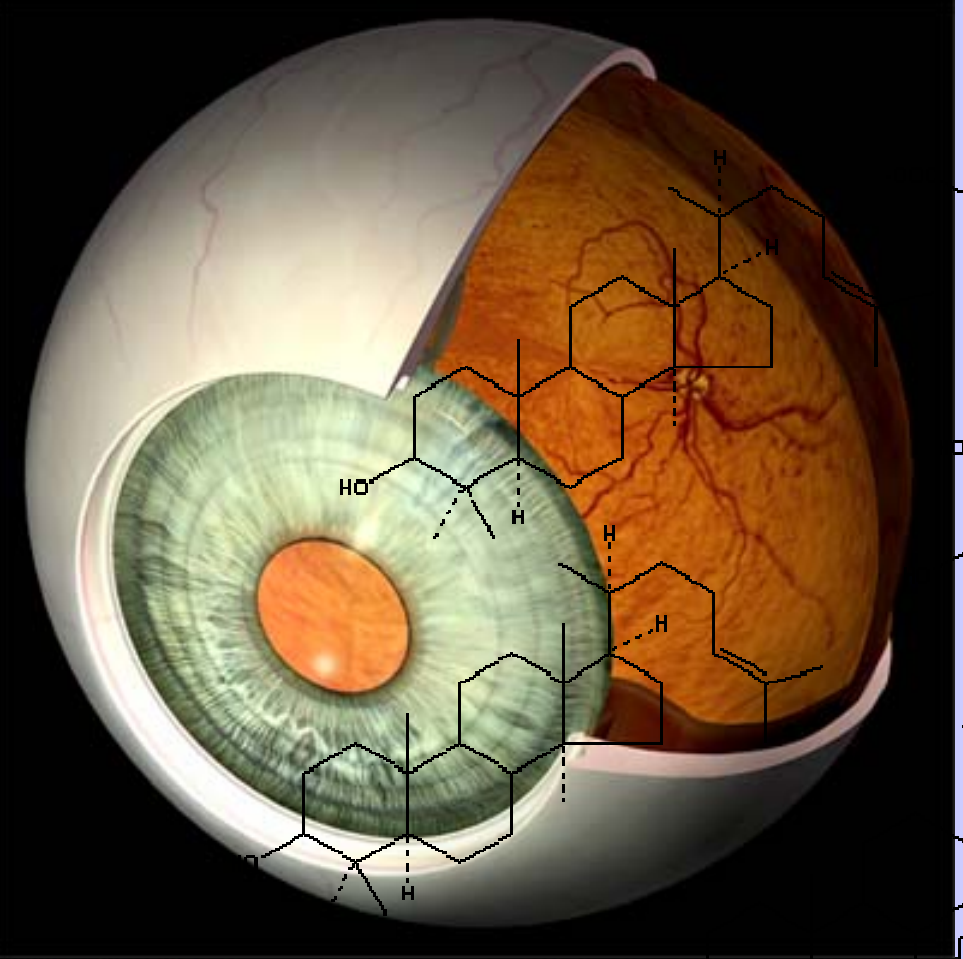
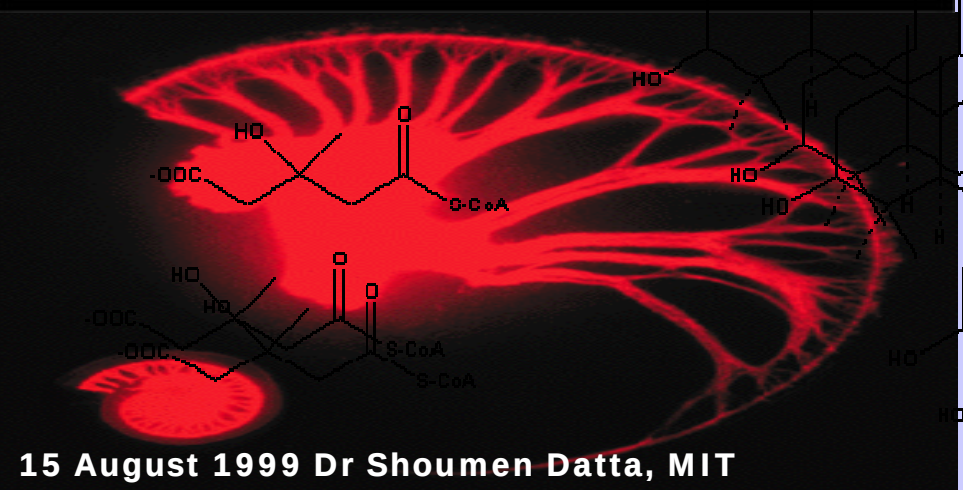
MICROARRAY for DUMMIES

An old dog learns new tricks!

Proteins in the Post-genomic Era

Dr. Shoumen Datta





Focus on Technology

Microarray & DNA Chips

- Who needs it?
- Scope
- Principle
- Materials & Methods
- Observations & Applications
- Limitations
- Summary

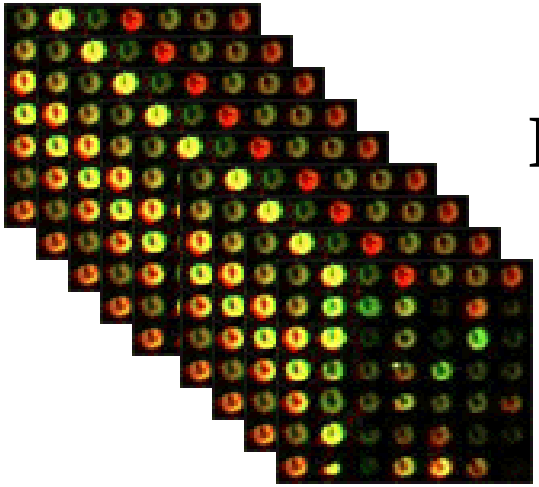
Focus on Technology

Microarray & DNA Chips: A Public View

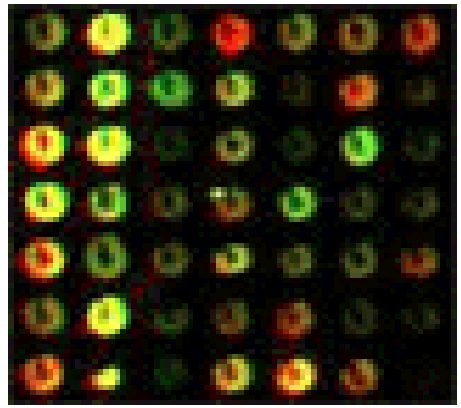
What ***is*** this new “fangled” technology?

It is an improved and highly scaled-up version of a 25 year old method to reveal very small changes in several hundreds or even thousands of genes in **one** step rather than searching one gene at a time. This technique can also estimate the activity of many genes, at a time. Have you ever stopped to wonder what your genes are doing inside your cells? Did you consider taking a ‘photograph’ of your genes in action? This technology allows us to take a ‘photograph’ of genes and catch them in action. A photo generally shows who is wearing what type of clothes or shoes [normal DNA versus mutations] and captures the activity of people, such as sitting, sleeping or running [activity of your genes, who is dormant, who is working hard or hardly working!].

WHAT THE HECK IS A GENE ??



Microarray & DNA Chips



Massively parallel post-genomic comparative hybridization using robot-fabricated immobilized cDNA probes on glass [microarray] representing all identified genes [ORFs] or photolithographic solid phase synthetic deoxyoligonucleotide probes [DNA Chip] based on ESTs and identified genes. Differentially [eg: Cye-3 dUTP and Cye-5 dUTP] labelled total or poly-dT purified RNA from normal and affected states are ‘targets’ for interrogation after RT/PCR. Fluorescence imaging coupled with robust bioinformatics analysis provides raw data on gene expression profiling, genotyping for polygenic traits, screening [SNPs/cSNPs] for allelic heterogeneity [eg: BRCA1], infectious [and other] disease patient management [HIV, TB] through genotyping [toxico- / pharmaco-genomics].

Microarray & DNA Chips

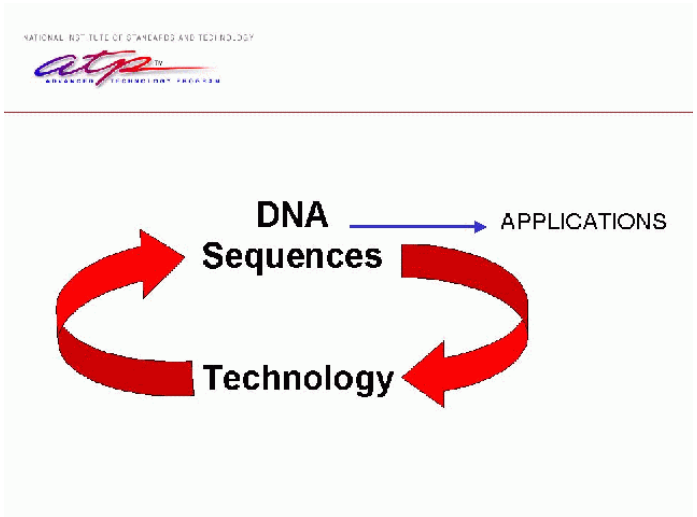
Who needs it?

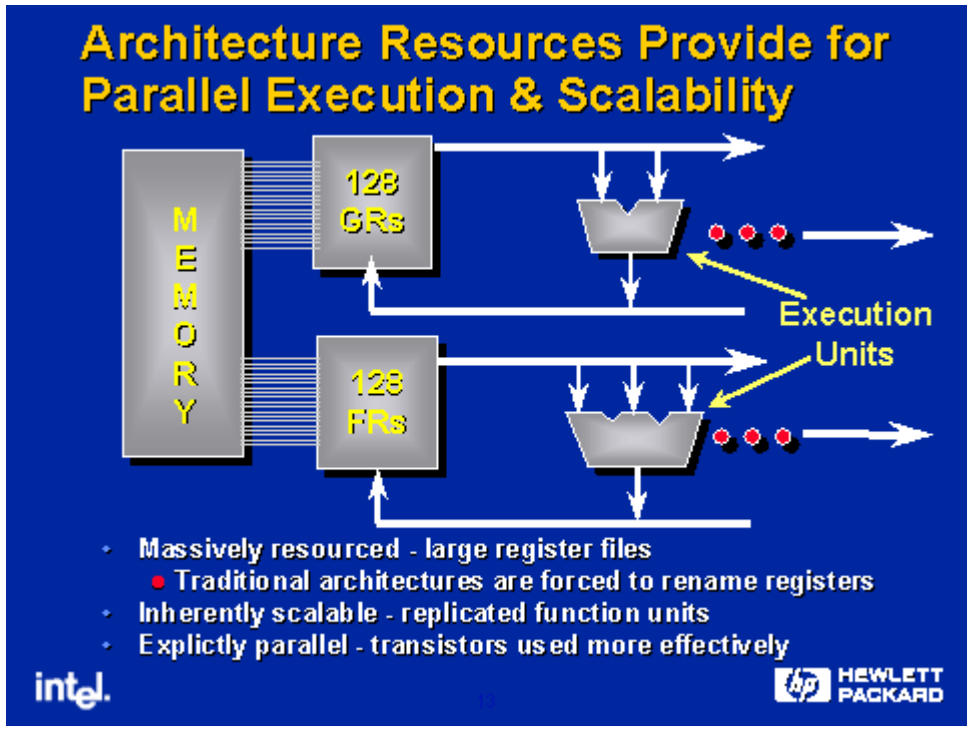
- Public
- Scientists
- Biotechnology
- Pharmaceuticals

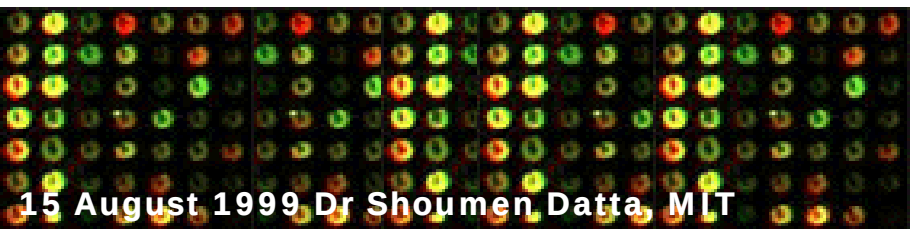
Why?

BROAD SCOPE

- Basic & Applied
- Exploratory Tool
- Massively Parallel

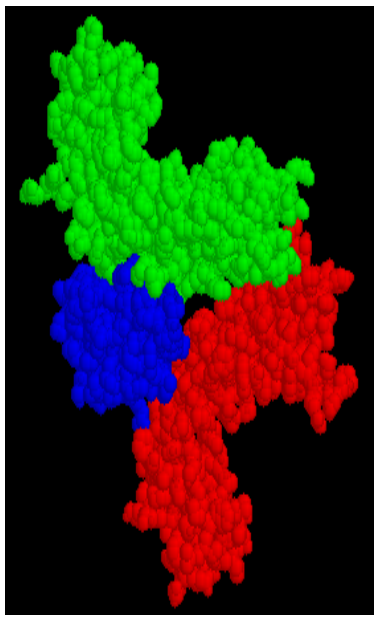






Microarray & DNA Chips

Broad Scope

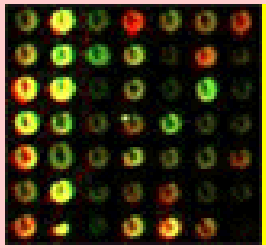


It can analyse the pattern and level of gene expression of 100,000 genes in 100 cell lines to test 1,000,000 synthetic molecules which may have pharmacological potential [NCI]. Instantly?

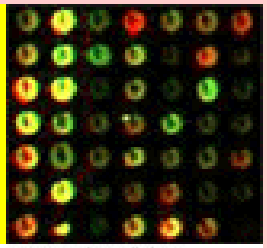
Zero calorie, “organic” potato chips from DNA Chips? Perhaps. Microarrays that may reveal how crops respond to infection by pathogens or adapts to stress [salt, drought, temperature]. Integrated understanding of coordinated multiple gene expression unlocks complex traits and allows engineering resistance to pathogens and natural stress.

Pre-clinical patient stratification may reduce failure of potential drugs [pharmacogenomics]. $20.6 billion was spent by US pharmaceuticals to discover and develop new drugs. 9 out of 10 drugs failed due to adverse metabolite side effect. Toxicogenomics may revive failed drugs!

Gene expression profiling leads to molecular disease management. Early profiles of at risk candidates may aid prognosis. Measures may be designed to transform a possible terminal ailment to a condition of life-long maintenance through genotypically tailored assortment of drugs.



Microarray & DNA Chips



Principle: HYBRIDIZATION
Same idea, better use

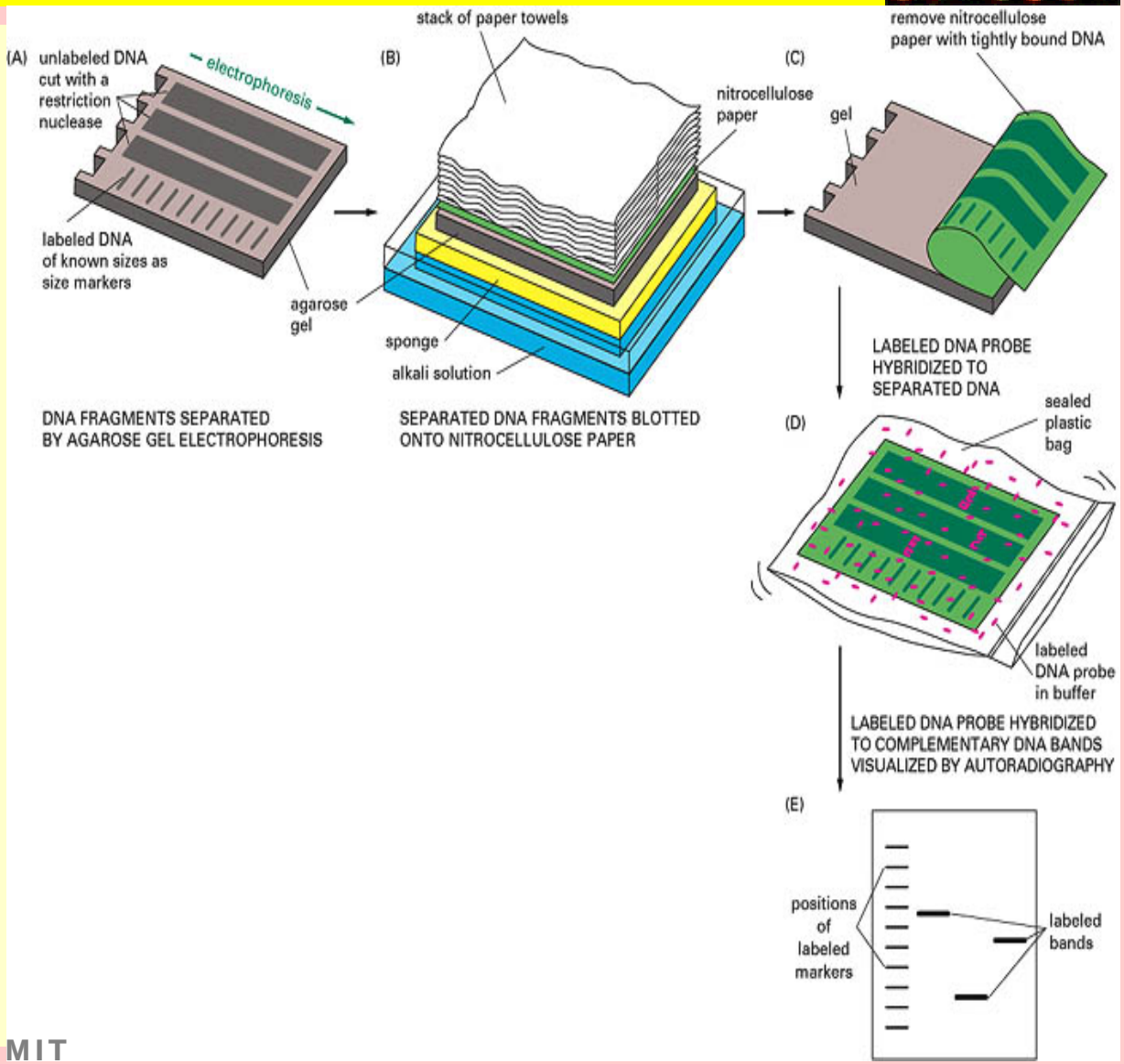
[Bardeen & Shockley]
Transistors to Computer Chips

[Edwin Southern]
Southern Blots to DNA Chips
One to One Correspondence: Clone to Hybridization Signal

History of Science
Other Landmark One to Ones

1941: **Beadle & Tatum**
One gene, one enzyme

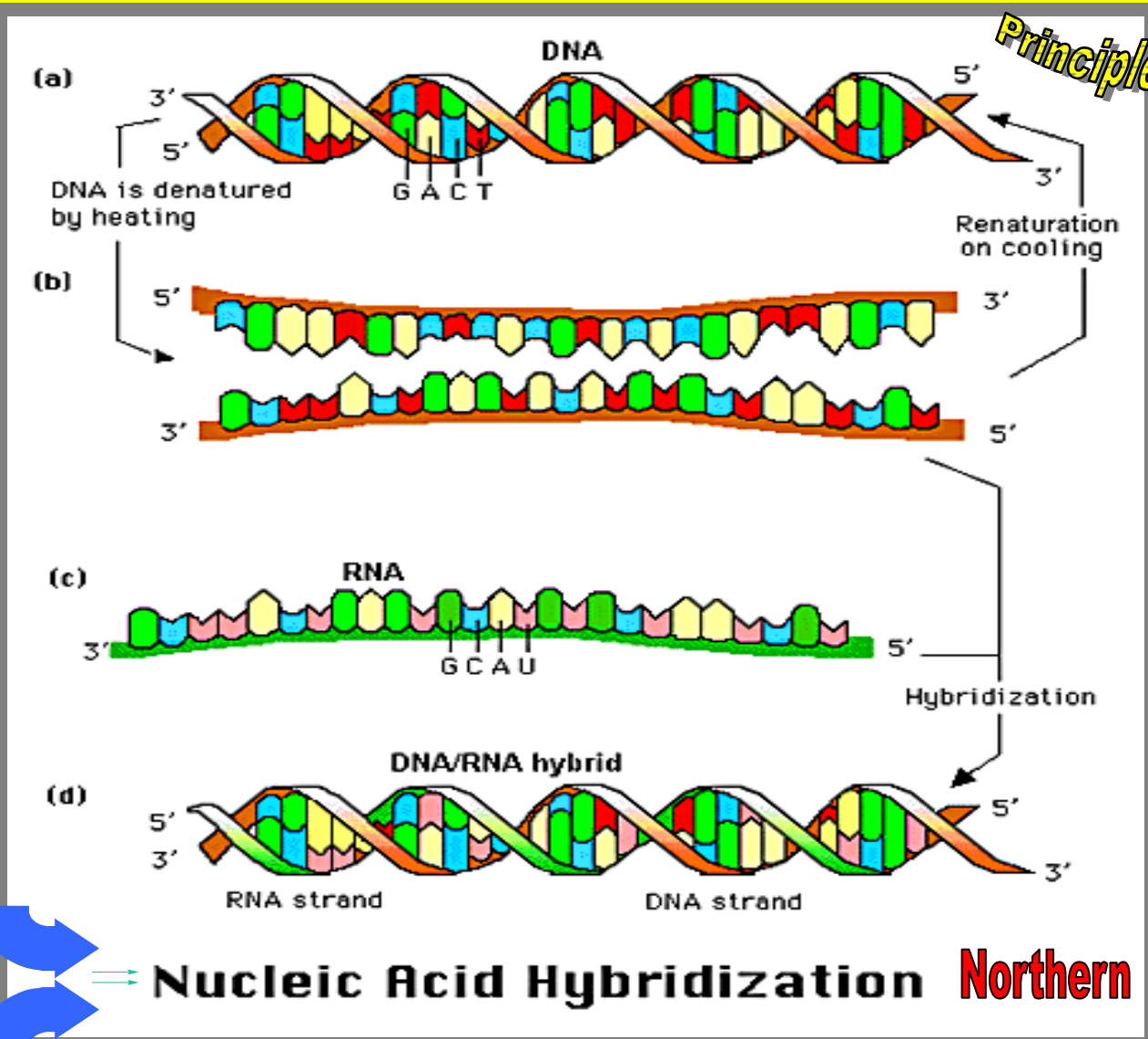
1964: **Charles Yanofsky**
DNA sequence colinear with Protein sequence



Microarray & DNA Chips

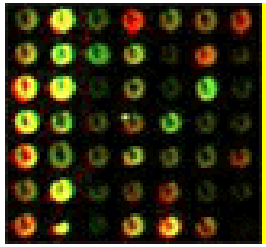
If you get this, then you have got it. That's all there's to it.

Really!

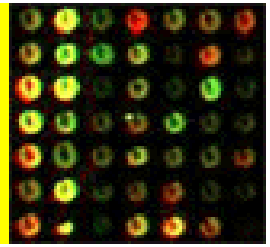


If you get this, then you have got it. That's all there's to it.

Really!



Microarray & DNA Chips



What are the steps?

[1] Choose cell population [or sample for diagnosis]

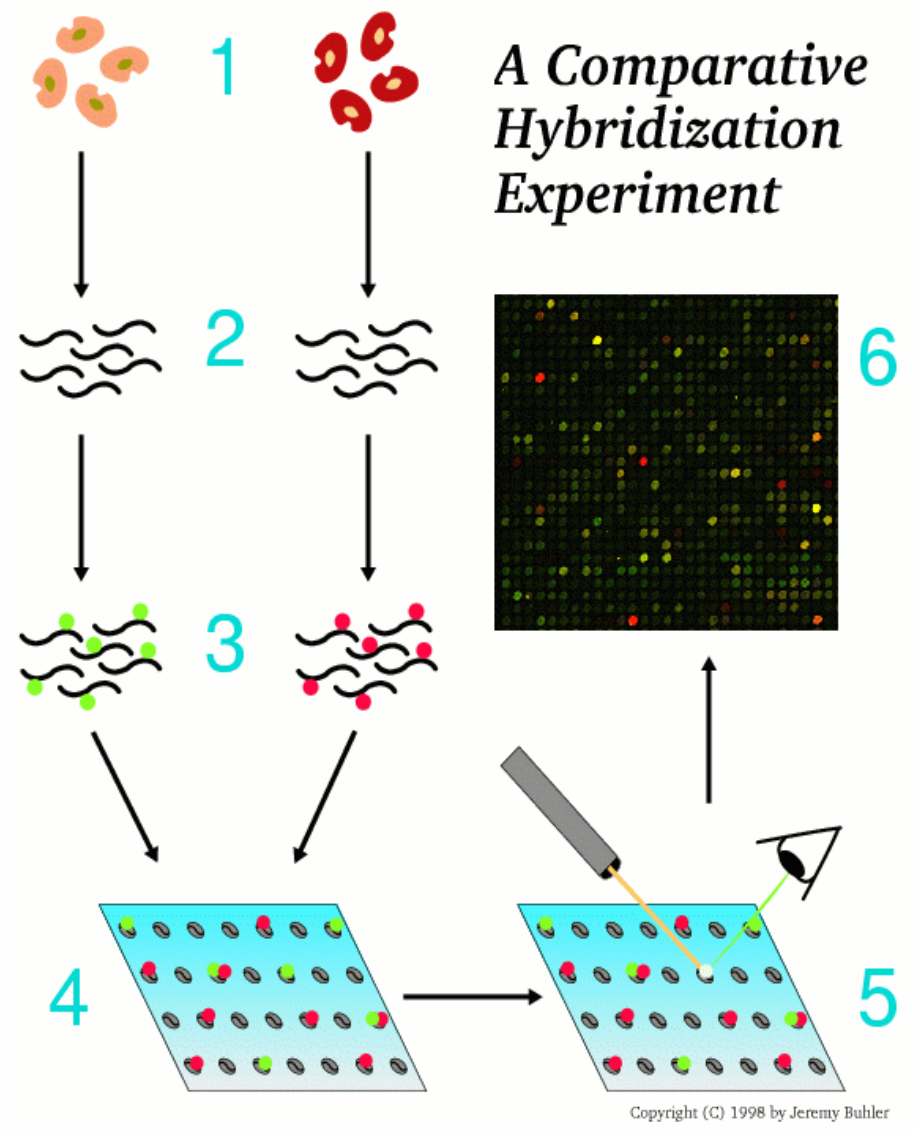
[2] RNA extraction, purify

[3] Fluorescent label cDNA

[4] Hybridize with **PROBE** on Microarray or DNA Chip

[5] Scan

[6] Interpret image



Microarray & DNA Chips

A Thumb Nail Version

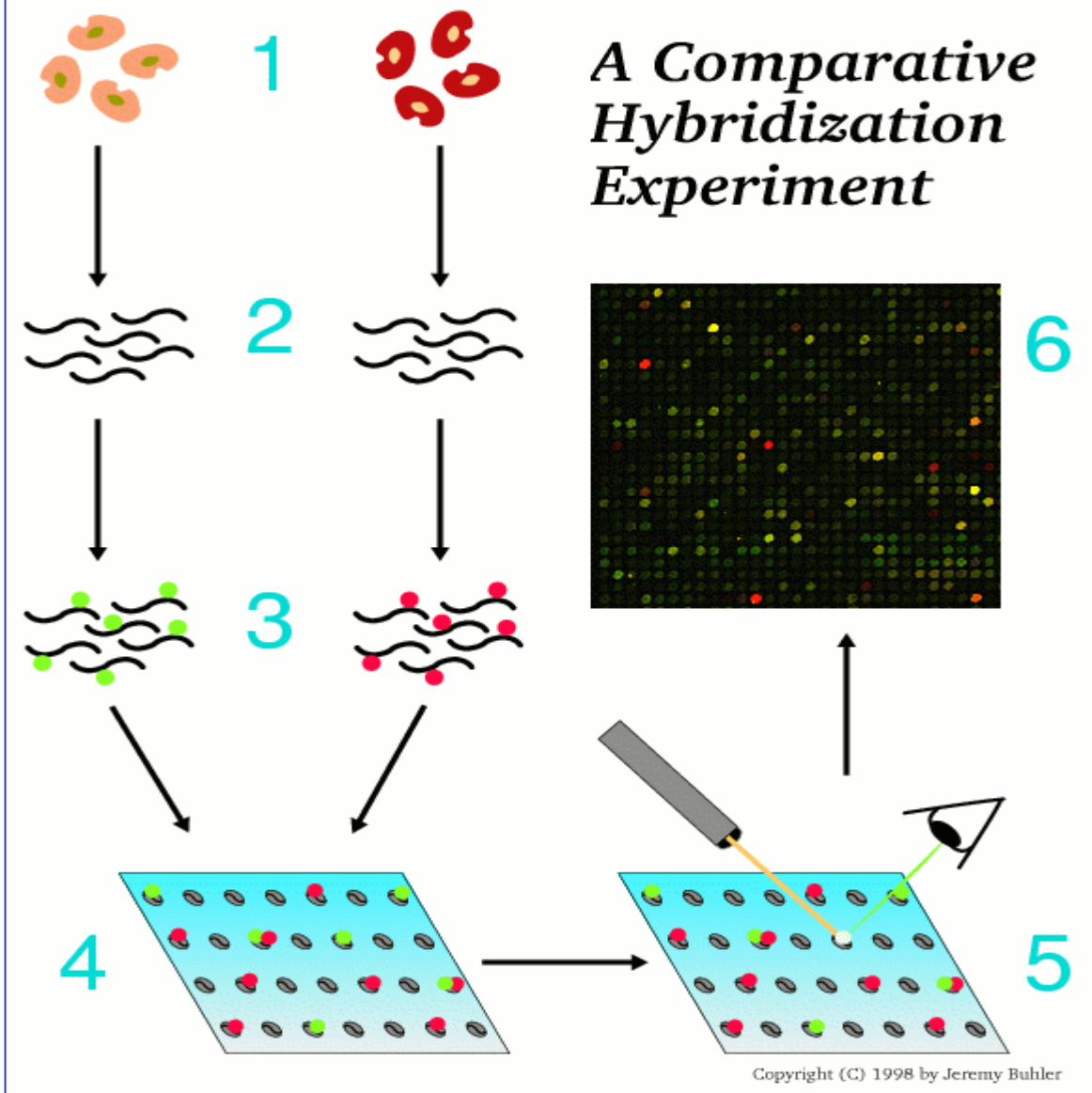
current issue

Are Microarrays and DNA Chips the same ?

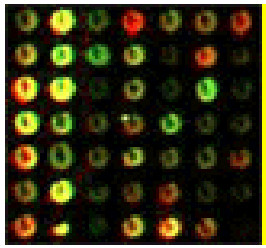
They share the same scientific principle

May be used with similar TARGETS

DIFFER in construction and type of **PROBE**

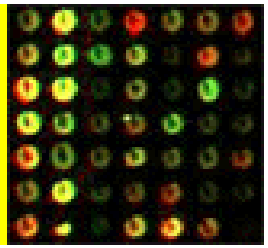


Nomenclature
nature genetics ARCHIVE



Focus on Technology

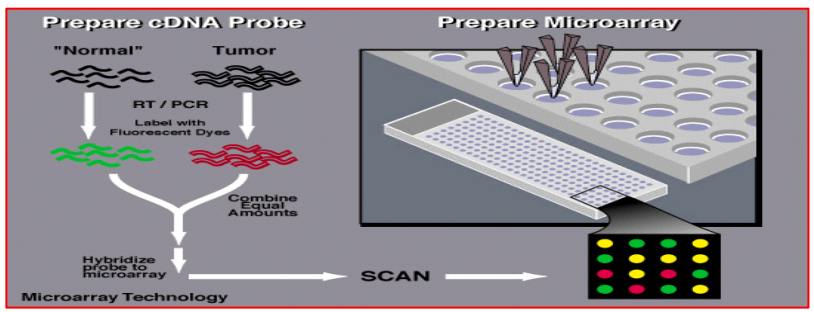
Microarray & DNA Chips



vive la difference

Microarray: SPOTTED

Probes [0.6 kb - 2.4 kb] are PCR amplified full-length cDNA or EST [expressed sequence tags] sequences. Spotted by ‘robo-arms’ on non-porous, solid support. About 10,000 ‘spots’ on a microscope glass slide.



DNA Chips: SYNTHESIZED

Probes are 20-25 deoxyoligo-nucleotides synthesized on glass by solid-phase DNA synthesis coupled with selectively masked light protection and deprotection [photolithography]. Commercial GeneChip have about 300,000 probes on 1.28x1.28cm surface. Experimental versions exceed 1,000,000 probes per array.

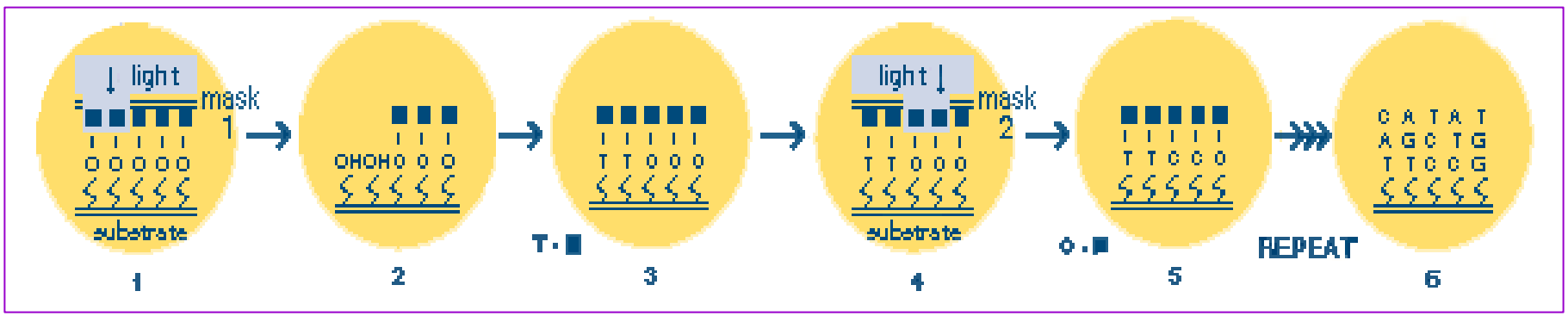
[Sound familiar? 286, 486, Pentium!]

Microarray & DNA Chips

Zorro [?] associated with DNA Chips ? Well…..

PHOTO-LITHOGRAPHY

Used by the semi-conductor industry to micro-fabricate computer chips by using a **mask**ing process [algorithm] to selectively expose specific areas of the silicon wafer to a light source [eg: laser or ultra-violet]. The light energy is used to etch micro-circuitry [integrated circuits] to produce computer chips [Pentium]. Using this principle, but adding nucleosides [DNA building blocks], chemical coupling occurs at sites illuminated. Sequential addition of A, T, G and C, controls sequence specificity. The steps are repeated.



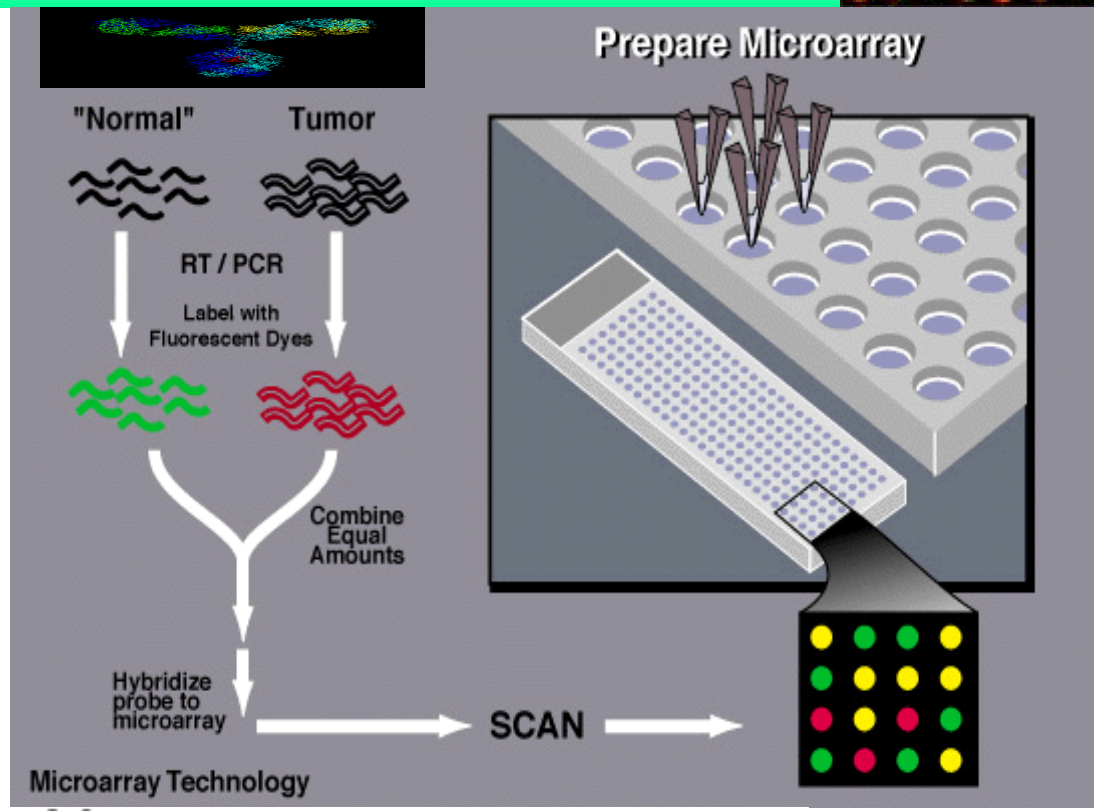
Microarray & DNA Chips

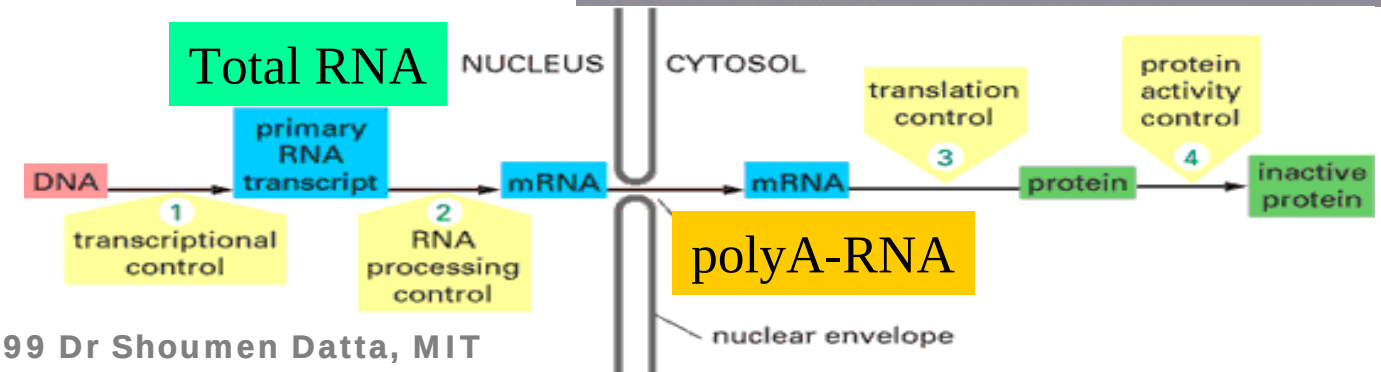
Depending on the question, most important determinant:

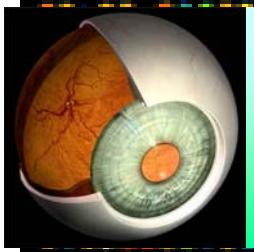
RNA [or Material] target
- Rare or limited source
- Purified vs whole cell
- Reliable amplification

Other Protocols:

1. Differential labeling to prevent cross-excitation
2. Quality of immobilized probe
3. Scanning and bioinformatics







Focus on Technology

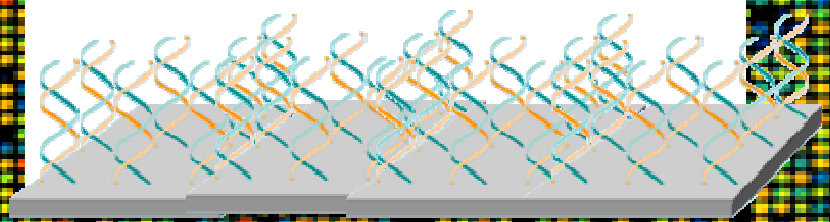
Microarray & DNA Chips

OBSERVATION

Two fluorescent reporters:

[G] Green [false coloured]
[R] Red [false coloured]

[Y] Yellow
Equal amounts of bound cDNA from each group of cells. Produces equal intensity in red and green channels [R + G = Y]



APPLICATION

What to use?

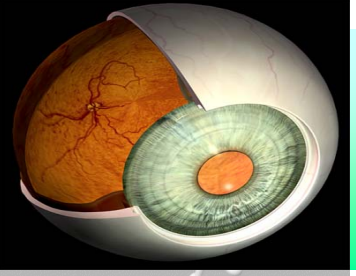
- Microarray
- SNP Chip, GeneChip

When to use?

- Preventive Care
- Infectious Disease
- Clinical Symptoms
- Genetic Predisposition

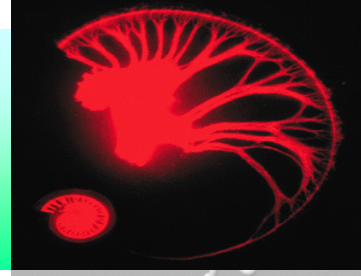
Where to use?

- Point of Care [POC]
- Clinic / Out-patient
- Hospital / Laboratory



Focus on Technology

Microarray & DNA Chips

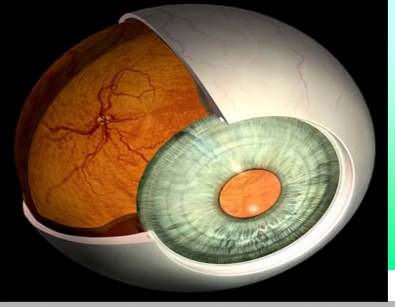


Few Examples of Applications

- Identifying drug targets and validation of new drugs
- Gene expression in pathogens [virulence determinants]
- Viral gene expression during latency and infection
- Expression profiling during cell cycle or apoptosis
- Population genetics: study of species diversity
- Homogeneous [HbS] v heterogeneous [MS] diseases
- Single nucleotide polymorphism map [SNP map]
- Prognosis and preventive measures
- Agro-biotechnology and animal husbandry

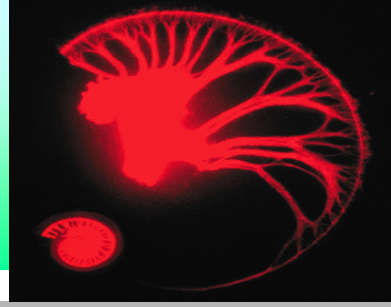
"Academics should concentrate on diseases that are rare or are predominantly in developing countries and thus hold little interest for the for-profit industry."

Daniel Cohen, Genset [*Science* 275 772 (7 February 1997)]



Focus on Technology

Microarray & DNA Chips



LIMITATIONS: CONCEPTUAL, SCIENTIFIC, TECHNICAL

Conceptual

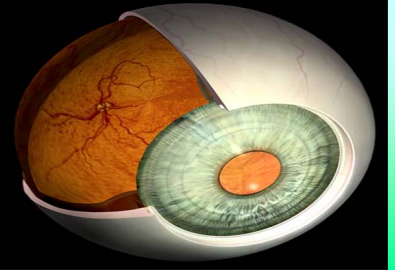
[1] How well can causation be inferred from correlation?
[2] RNA expression may not correlate with in vivo protein levels
[3] Are “housekeeping” genes reliable controls?

Scientific

[1] Use of RNA related issues and state of immobilized DNA
[2] Genomic DNA cannot be hybridized [100 fold complexity]

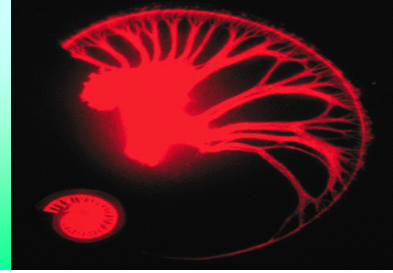
Technical

[1] Performance factors eg: arrayer pens [tips], spray-jets
[2] Bioinformatics: cluster analysis & subsignature profiling



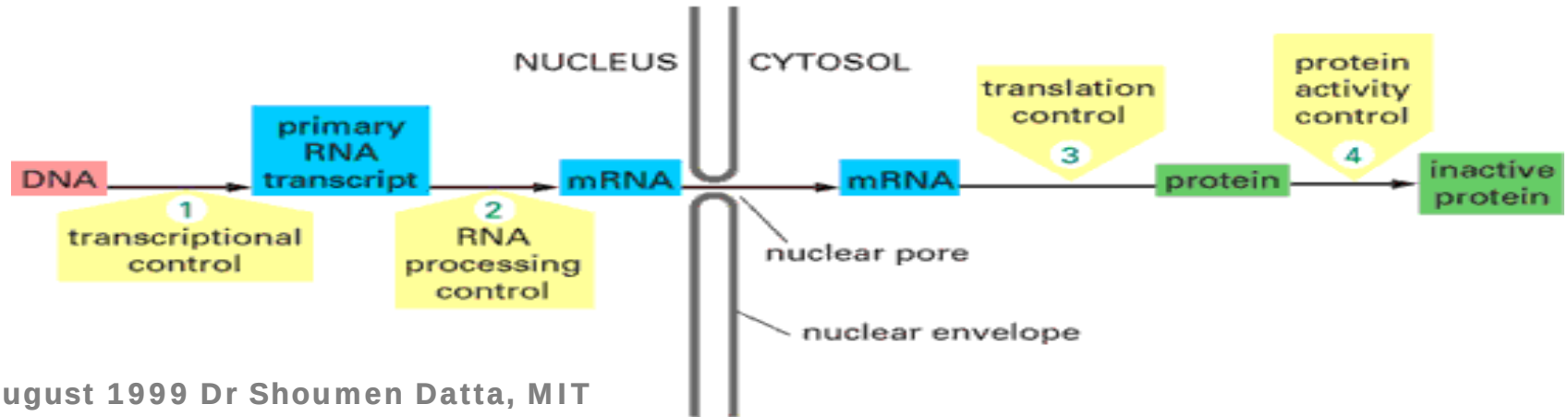
Focus on Technology

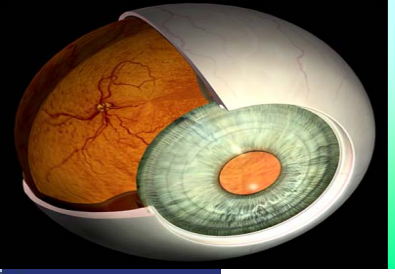
Microarray & DNA Chips



Some details of LIMITATIONS

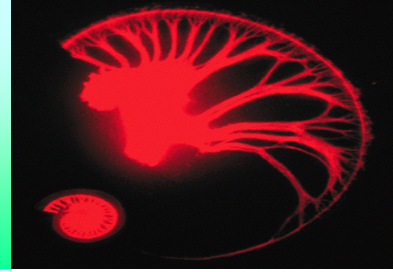
- RNA expression may be non-correlative with protein activity and physiological state
- Requires large amount of RNA for microarrays [but less for DNA Chips and GeneChip]
- Oligo-dT priming often aborts; random priming adds sample noise due to rRNA/tRNA
- Reverse transcription efficiency of RNA varies; affects fluorescent label incorporation
- Fluor depends on cDNA length and nucleotide composition [both may be unknown]
- Quantitative comparison of fluor intensities possible with same cDNA between groups
- Adjustment to same overall intensity assumes identical amount of RNA in different cell
- Quantitation subject to noise from RNA, shape of spot, dust, non-specific hybridization
- Unequal detection [PMT, CCD, Confocal Microscopy] efficiency across surface of slide
- Unchecked quality of immobilized DNA probe after synthesis, crosslinking & denaturation
- Oligonucleotide packing [10 picomoles / square mm] may introduce steric hindrance
- Probe redundancy [redundant PM/MM probes] critical to prevent cross-homology matches

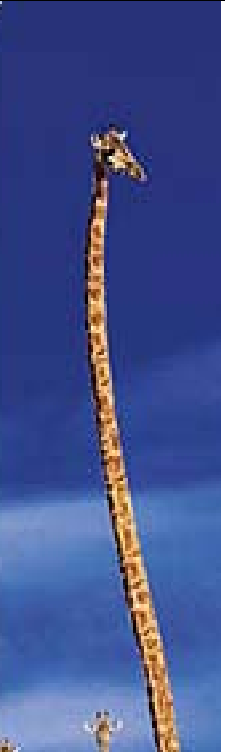




Focus on Technology

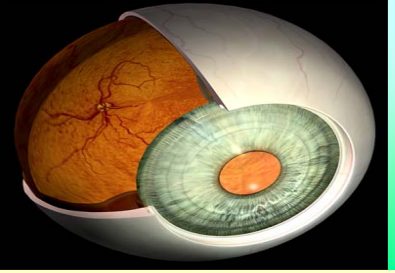
Microarray & DNA Chips





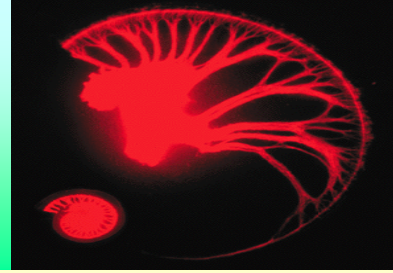
Choice of non-porous substrate is mission critical for high throughput

- Impenetrability allows immediate access & enhances kinetics of hybridization
- Washing steps unimpeded by diffusion; adds speed and reproducibility
- Flat, rigid, transparency improves probe location, image acquisition & processing
- Rigidity permits high throughput flow cell processing and automated scanning
- Inert nature assists microfluidics operation under variety of stringent conditions
- Oligoethylene glycols, poly-lysine or amino-reactive silanes functionalise surface
- In situ synthesis yields are high & permit combinatorial strategies for fabrication
- Photolithography increases density of arrays to make it massively parallel

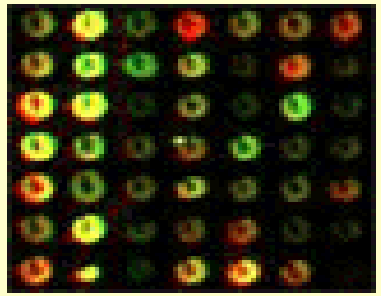


Focus on Technology

Microarray & DNA Chips



SUMMARY



Starting Principle

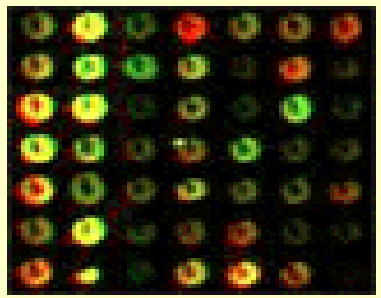
•Single strand DNA binds strongly to nitrocellulose membrane
•Permits hybridization to complementary RNA

And now...

•Scale-up causes paradigm shift: ***one gene at a time to one genome at a time***
•Thousands of such hybridization reactions performed on a glass slide
•Probes are based on sequencing of various organisms and human genome
•RNA from cell populations or diagnostic material are the usual **targets**
•Differential fluorescent labeling allows intensity dependent scanning
•Image is processed and data analysed with the help of bioinformatics

What it may be useful for ...

•Screen mutations, detect pathogens, gene expression, disease management
•Change biochemistry based drug discovery to genomics base



What it is not (yet)…

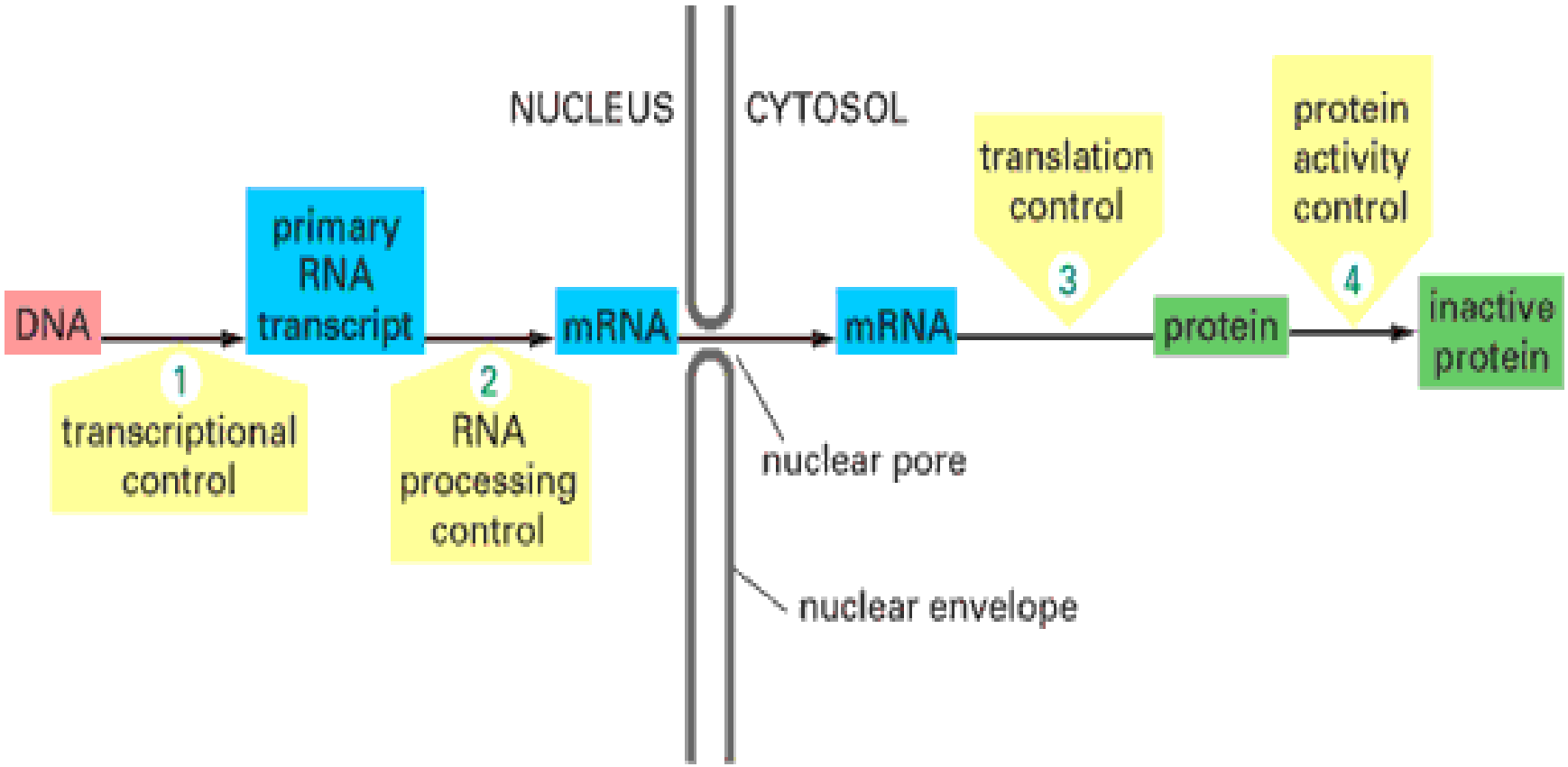
Panacea for gene expression studies or *de novo* gene discovery
POC infectious disease diagnostic tool

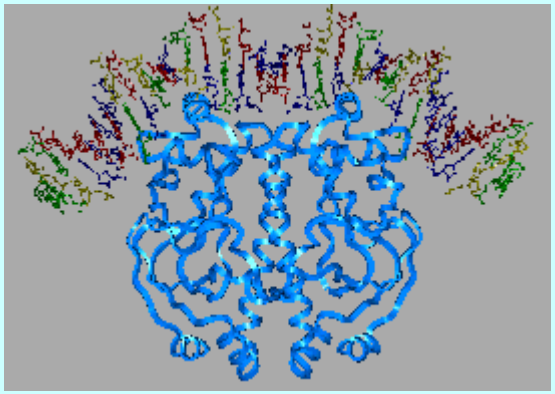
GENE EXPRESSION

VS

PROTEIN EXPRESSION

The physiological state of an organism may not be reflected by gene expression or RNA levels. Amount of mRNA may not correlate with amount of active protein or actual protein activity. Expression of a protein may not always produce a detectable physiological activity or response.



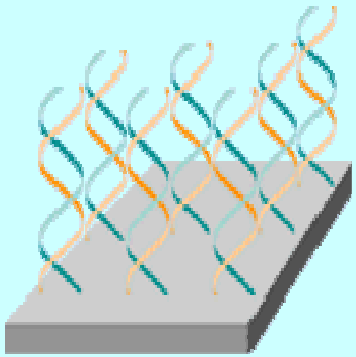


PROTEOMICS

- One Gene, One Enzyme: Synergy with Genomics
- High Throughput Analysis of Proteins
- Protein Profiling

Bioinformatics

- Experiments *in silico*
- Medical Genomics Information Management System



Remarks

DNA Chips in Diagnostics:
Too much gun for too little game?

Summary

Phosphorous
Nitrogen
Hydrogen
Oxygen
Carbon

PROTEOMICS

Protein Profiling

Signal

Protein Activity

Separation

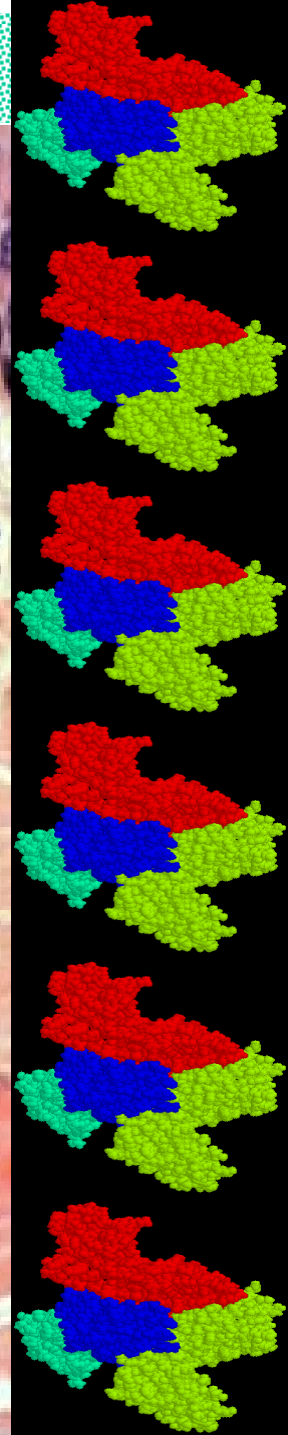
Analysis

Characterize

Protein Sequence

DNA Sequence

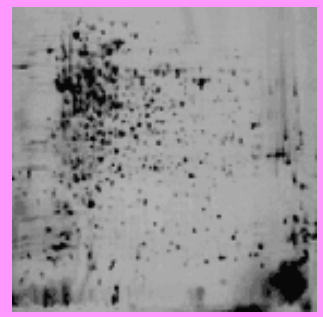
- Synergy with Genomics



PROTEOMICS

Protein Separation

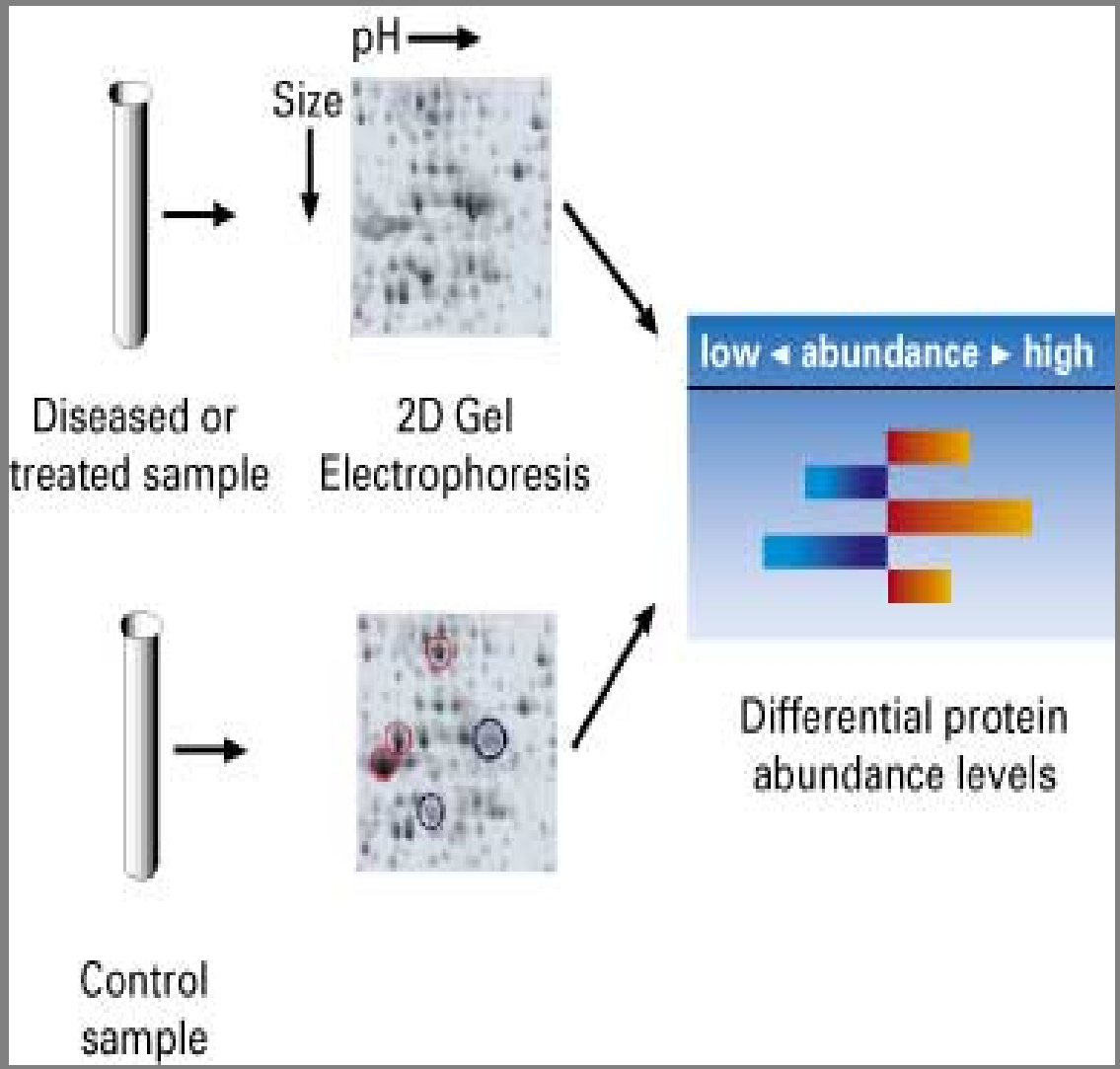
2D-GE



S. aureus
stationary

PROTEOMICS

Protein Analysis



STEPS

Sample prepared

Electrophoresis [2D]

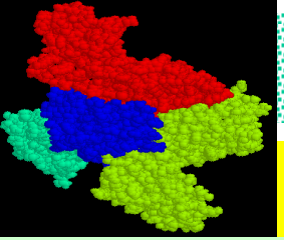
Fixed

Stained with fluor

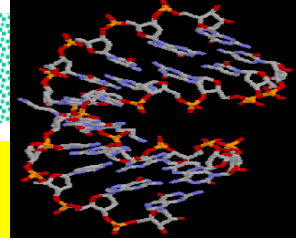
Expression scanned

Proteins selected

[Proteins sequenced]



PROTEOMICS



Protein Characterization

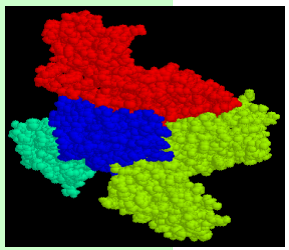
Steps:

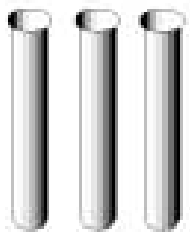
Extract target proteins

Sequence peptides

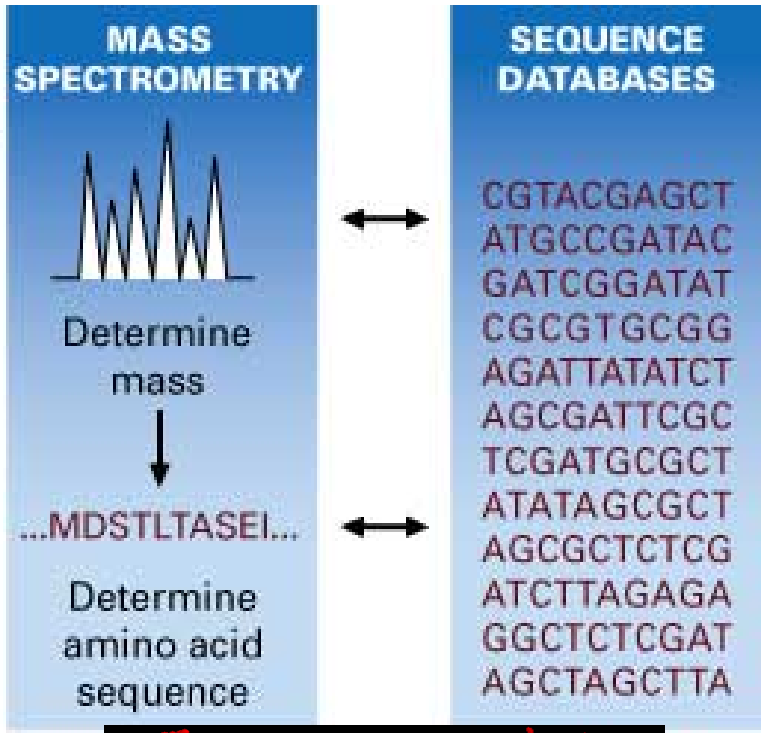
DNA Sequence match

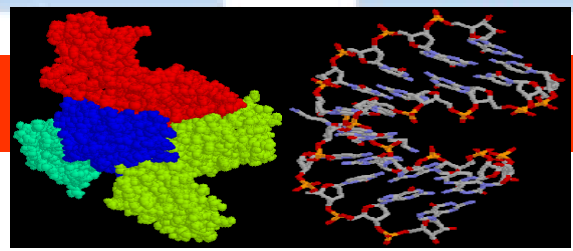
Protein Expression matched with Gene Expression

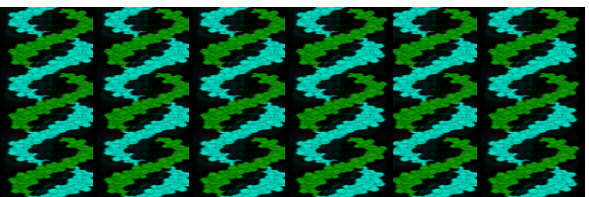




Extract proteins from gel and split into fragments of 5-10 amino acids







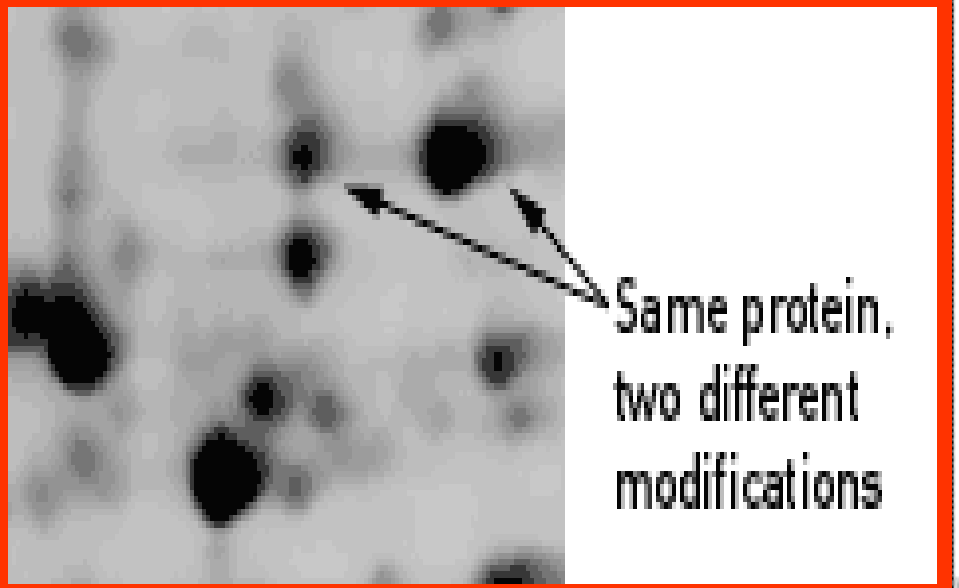
PROTEOMICS

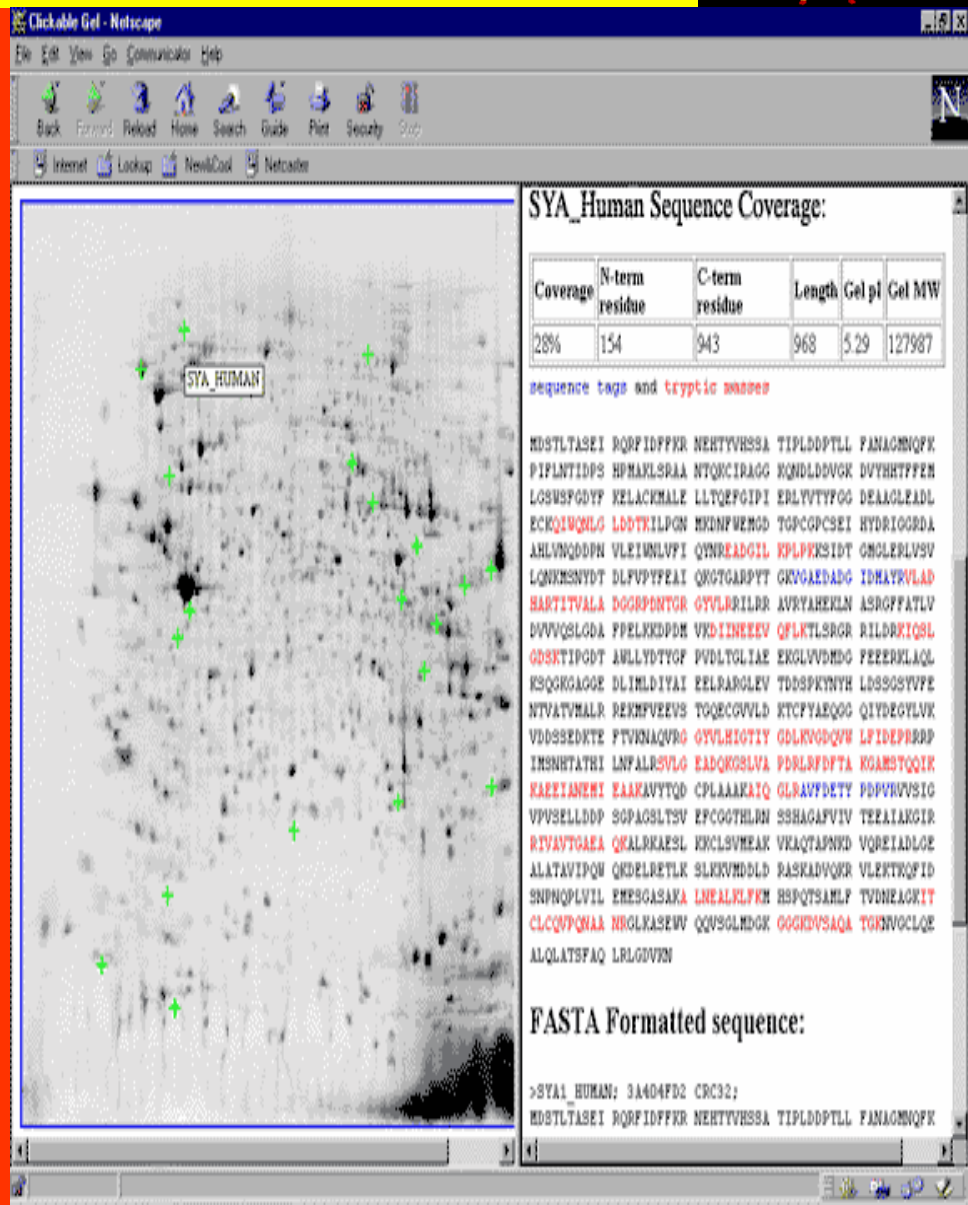
Protein Activity

Physiological Response

1. Same protein [no new gene expression] modified

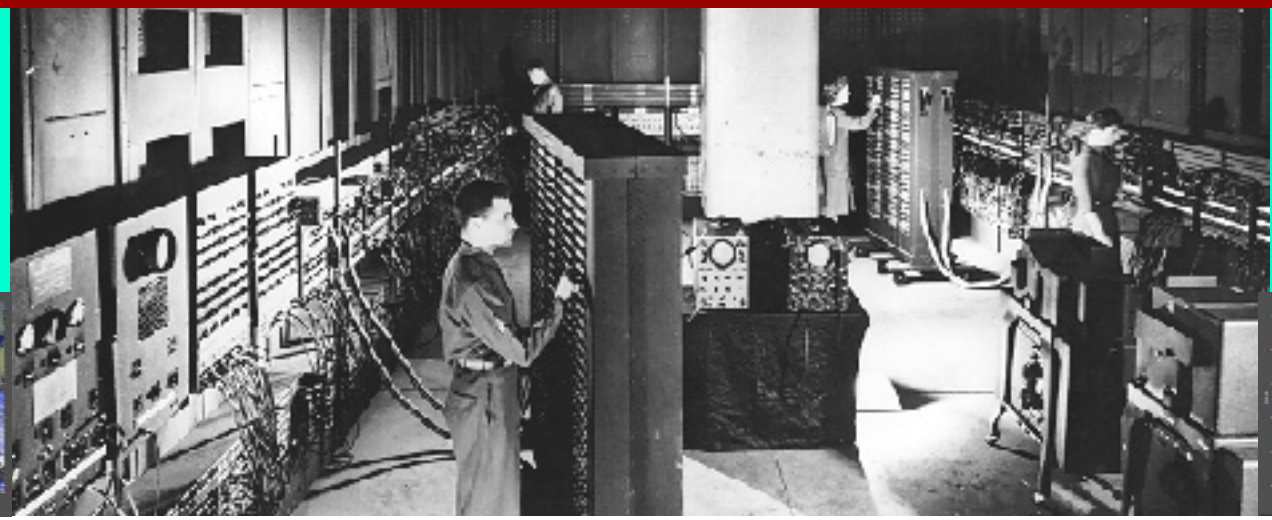
2. More or new protein (s) [gene expression detected]

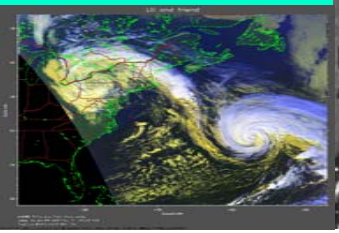


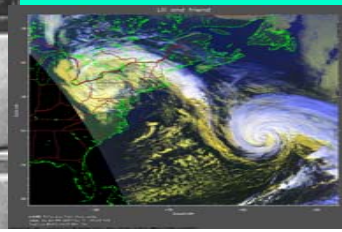


BIO-INFORMATICS

Experiments *in silico* and Medical Genomics Information Management System





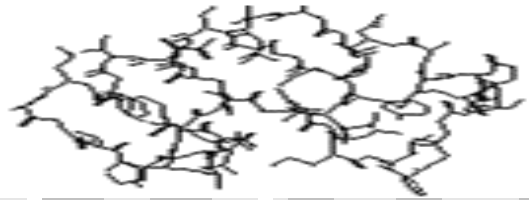


0101#01001010#10111010#
01010001#10010#1001#101
10010#1001001001011011#0

Residue	Φ	Ψ	ϖ
THR	0.0	147.7	172.9
THR	107.2	-125.3	187.4
CYS	123.4	63.6	103.7
PRO	60.3	83.9	-116.7

DNA → **Algorithm** → **Protein Model**

Natural Evolution →



BIO-INFORMATICS

Experiments in silico: Coupling 'tools' & 'analyses' to get information

1. Software acquires raw scan and performs image analysis from various on-line linked groups or database sites [uploaded with raw scan data].
2. Analyses [below] data from end-users and cross-links to databases

•Examples of experimental categories
Epidemiology, Infectious Disease, Molecular Pathology
Genomic Changes, Drug Sensitivity, Genetic Counseling

•Examples of biological categories [probe design]
Cell cycle phase, cell type, organism, small molecules, drug targets
protein motifs, DNA domains, signal sequence [EST, ORF, SNP, cDNA]

Use Statistical Cluster Analysis to sort, group and analyze data by:
Affinity grouping, rule induction, self-organization maps, decision tree, genetic algorithms, memory-based reasoning and other formats.

BIO-INFORMATICS

Medical Genomics Information Management System

M G I M S

•Not in place, yet. Resides in NIH and other databases.

•May not evolve till these scientific exploratory tools evolve to become general tools for hospital and clinics.

•MGIMS will allow micro-managing diseases with heterogeneous genetic risk [cancer, diabetes].

•May be helpful in HIV [not for detection] monitoring via mutation analysis of protease gene [database].

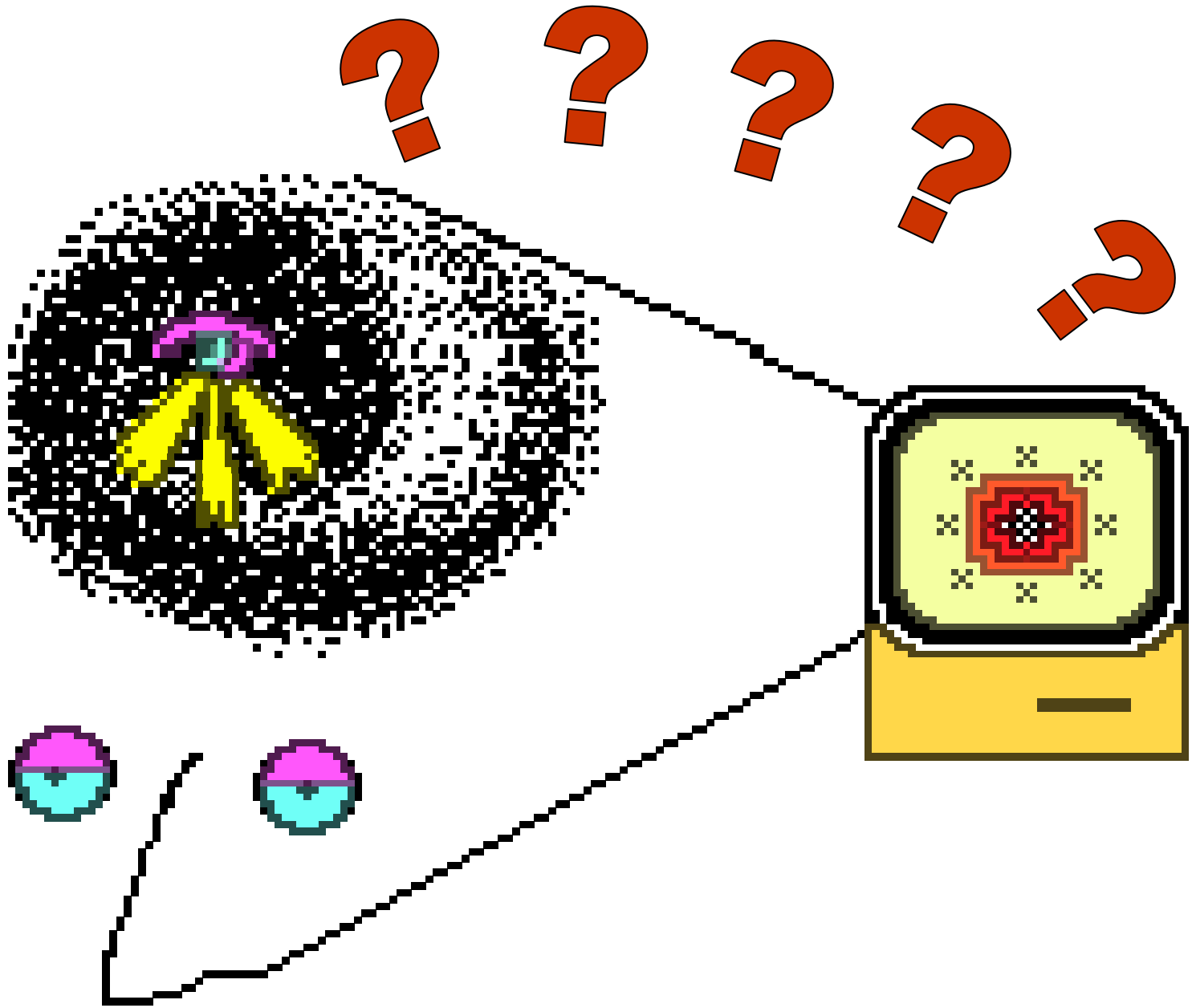
REMARKS

Infectious Disease [id] Diagnostics & the DNA Chip

Marriage of semiconductor microfabrication technology of photolithography with genomics has sired the GeneChip. Liquid phase reactions necessary for hybridization makes it less portable [than, for example, DuPont's *Riboprinter*] and microfluidics stations are still far from POC usage. High cost is another disadvantage for GeneChip in id-POC diagnostics compared to liquid phase nucleic acid probes and PCR tests [Roche's Cobas *Amplicor* or Gen-Probe's *Pace2C]*. The power of DNA Chips are 'wasted' even when used for patients with multiple HIV strains. Use in polygenic diseases may be suitable. For id-POC, *it is too much gun for too little game*, for now.

Summary

- DNA Chip is an exciting post-genomic exploratory tool
- May improve prognosis and care of patients with polygenic diseases or monitoring drug resistance
- Not yet optimized for point of care diagnostics
- Expensive [and unnecessary] for infectious disease diagnosis but may help in disease management
- Multiple uses in several fields still being defined



shoumen@mit.edu