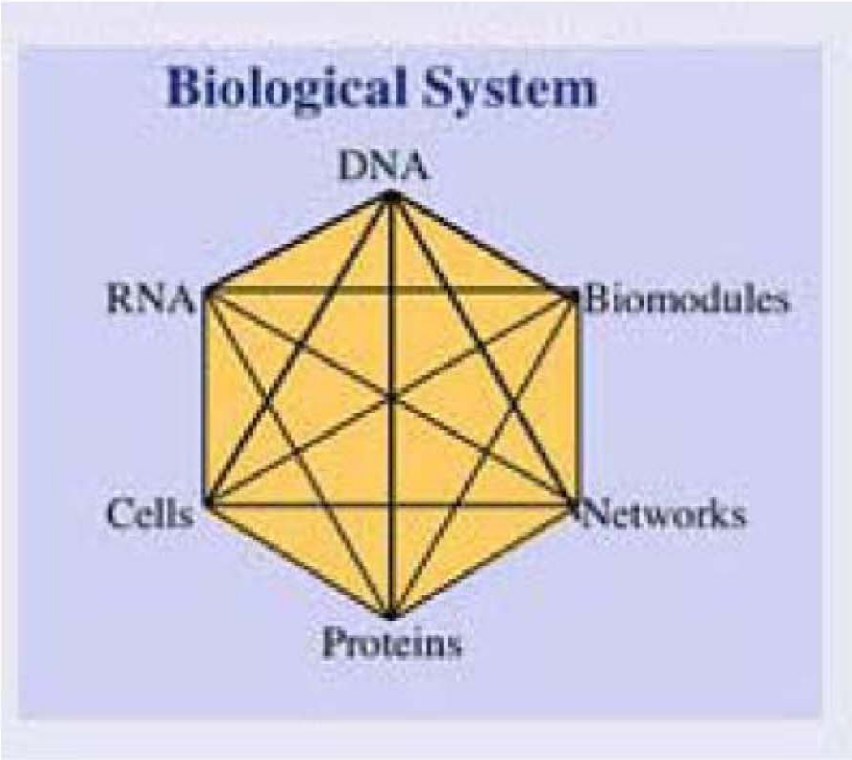
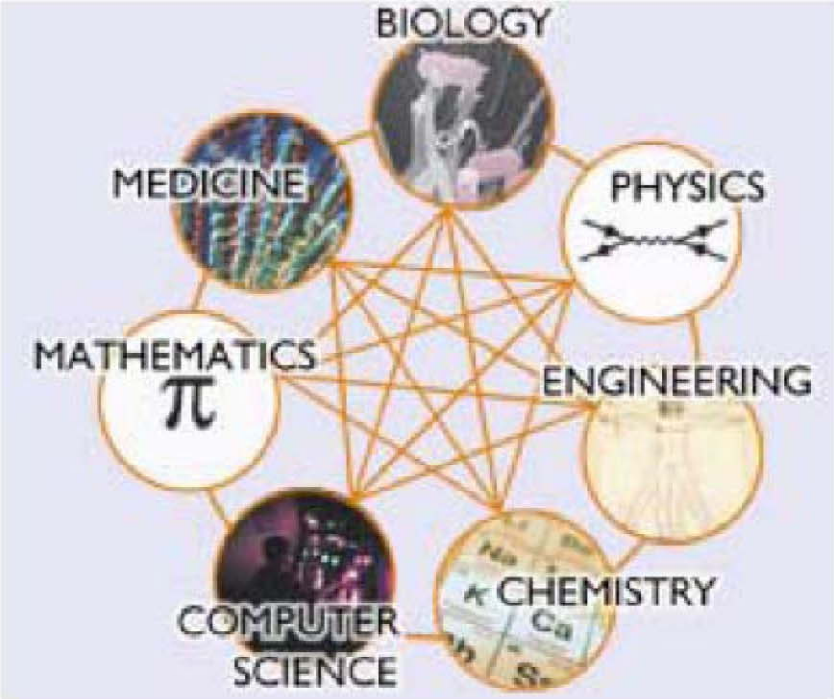
OMICS :

Systems Biology





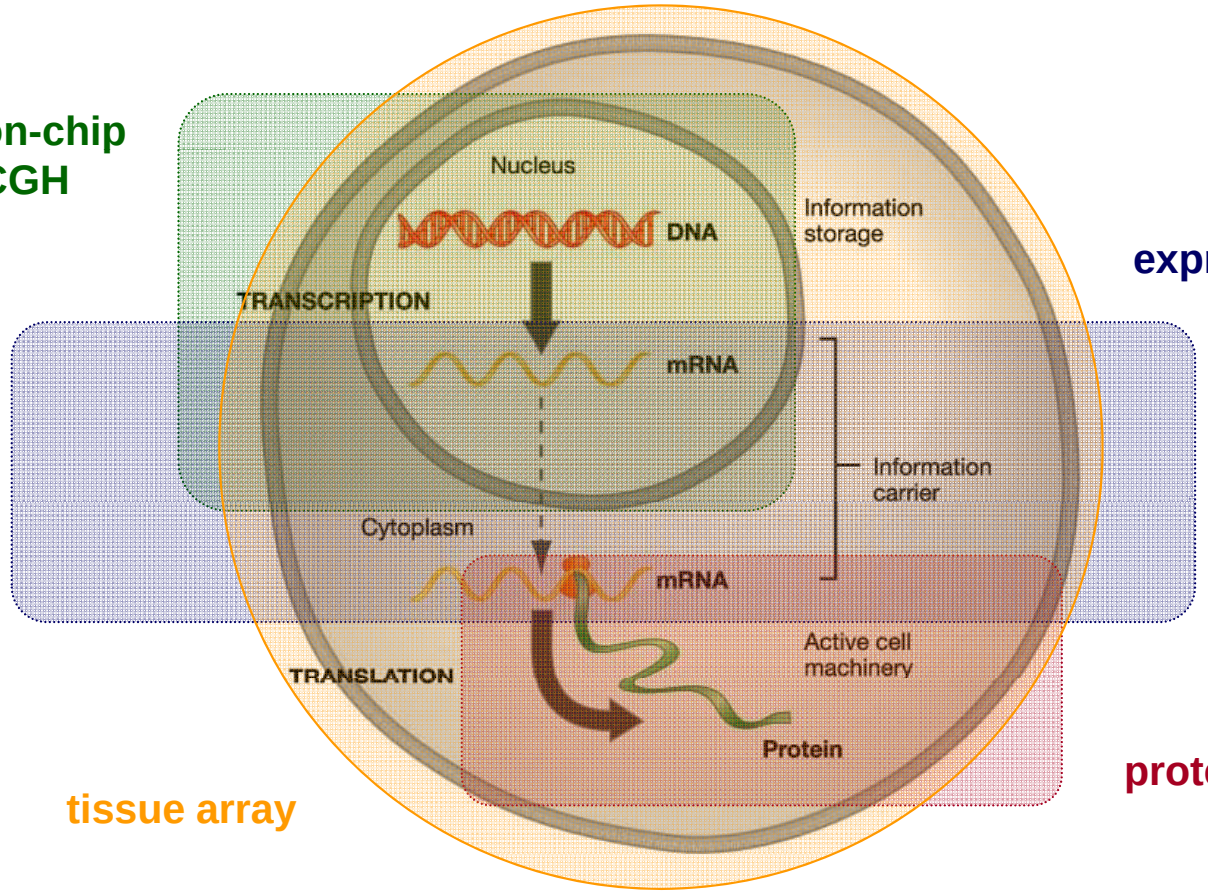
系統生物學 是一個試圖整合不同層次信息以理解生物系統如何行使功能的學術領域。通過研究某生物系統各不同部分之間的相互關係和相互作用（例如，與細胞信號傳導，代謝通路，細胞器，細胞，生理系統與生物等相關的基因和蛋白網路），系統生物學期望最終能夠建立整個系統的可理解模型。

2001年第二屆國際系統生物學會 (2nd International Conference on Systems Biology；ICSB 2001) 對「系統生物學」的解釋為：系統生物學是對生物體整個過程做一全面性的定量研究，並非以生物體的某一部分為對象。目的是要建立模式並以實驗來證實其可預測的生物體的表現。 簡單的說，這樣的研究方法就是利用資訊科學及微機電工程的技術來研究生物學的問題，最後並希望能夠利用電腦運算的結果，來預測細胞、器官、系統甚至完整生物體的表現。

系統生物學的研究可以包含四大部分，分別利用資訊科學 (computation)、分析(analysis)、技術 (technology)、基因組學 (genomics)、四者形成環型而連續的關係，建立出一個新的研究模式(Kitano, 2002)，並且利用這一模式所發展的一系列的工具來解決生物學家所面臨的研究問題

The Central Dogma

DNA → Transcription → RNA → Translation → Protein



CHIP-on-chip
array CGH

expression array

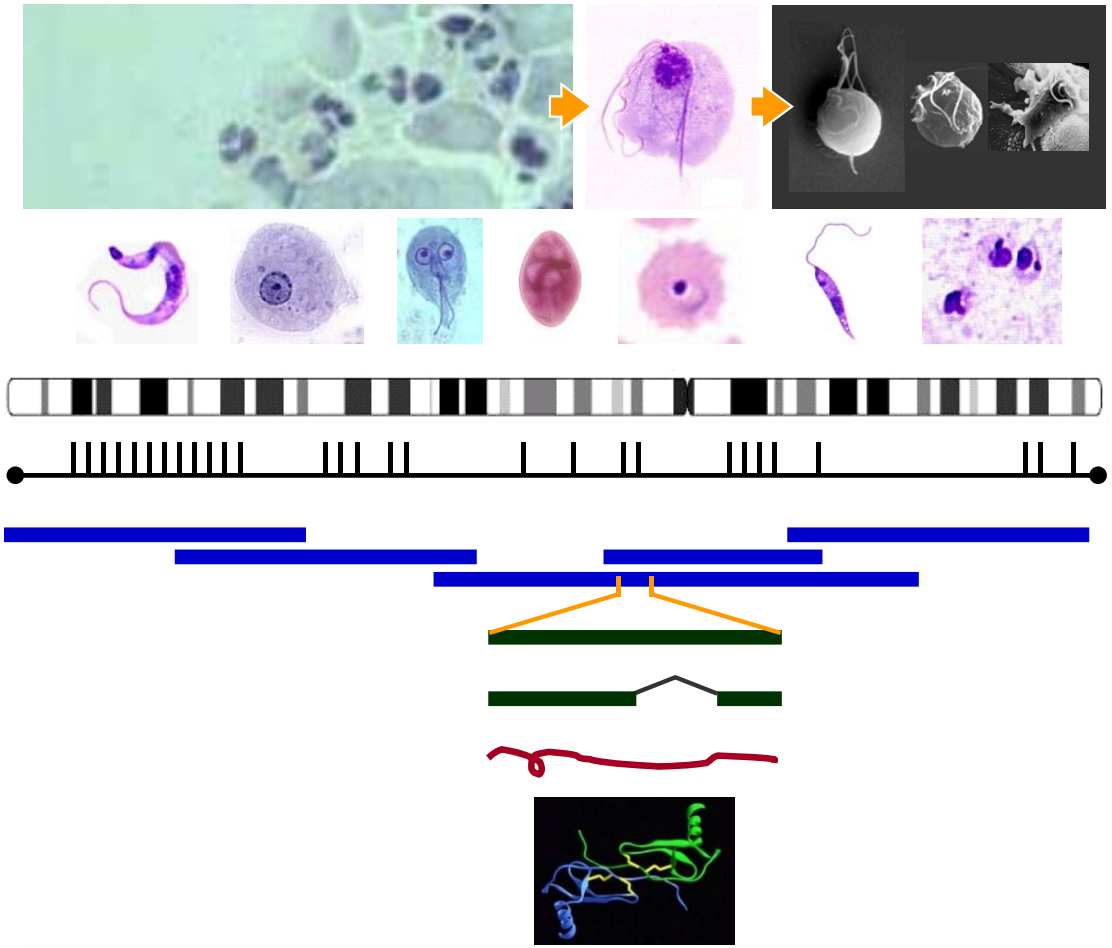
protein chip

tissue array

Data organization: Basic Model

Research

- Population
- Organism
- Development
- Phenotype
- Genetics
- Chromosomes
- Scaffolds
- Gene
- mRNA
- ProProtein
- Mature peptide
- Structure
- Domains
- Function
- Disease
- Phenotype

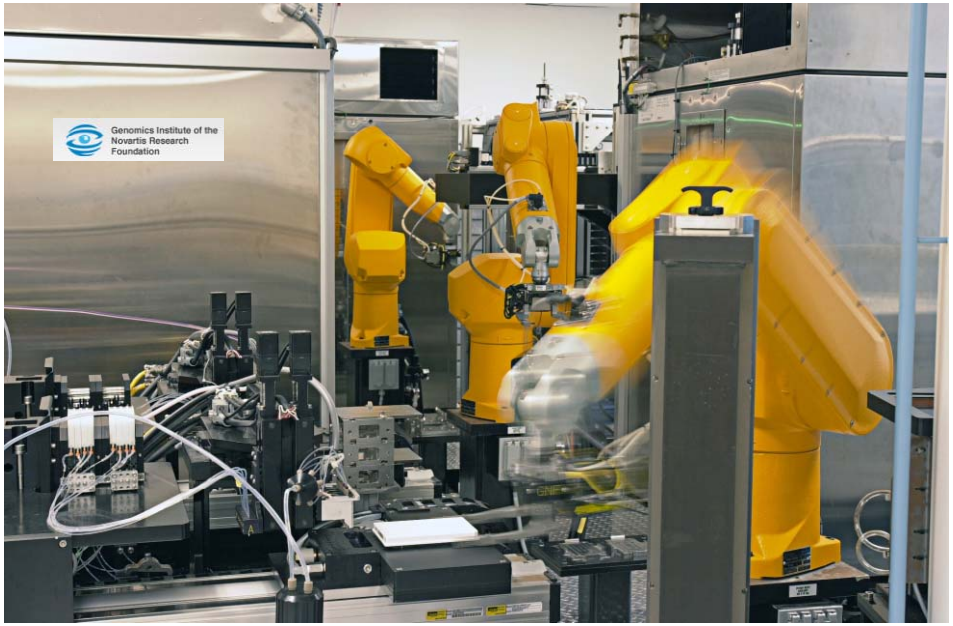


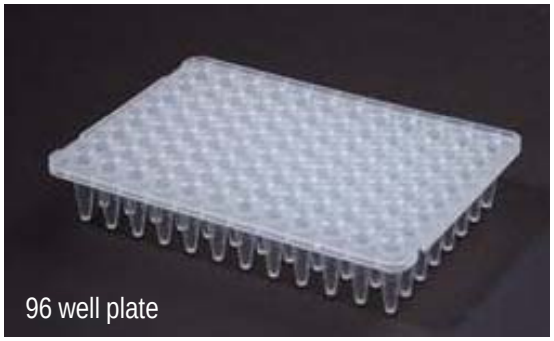
Resource

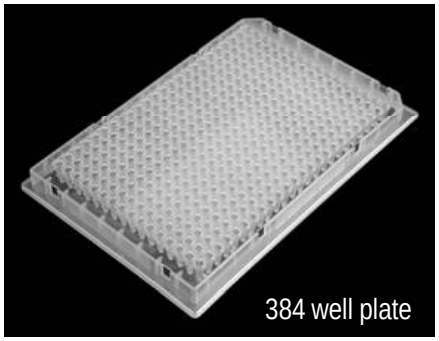
- Taxonomy
- Genome Project
- Genome Resources
- Map Viewer
- OMIM, OMIA
- Genome
- Trace
- Gene
- Nucleotide
- Protein
- RefSeq
- MMDB, Cn3D
- CDD
- Books
- OMIM, OMIA
- PubMed
- PubMed Central

高通量

High-throughput screening (HTS), is a *method* for scientific experimentation especially used in drug discovery and relevant to the fields of biology and chemistry.

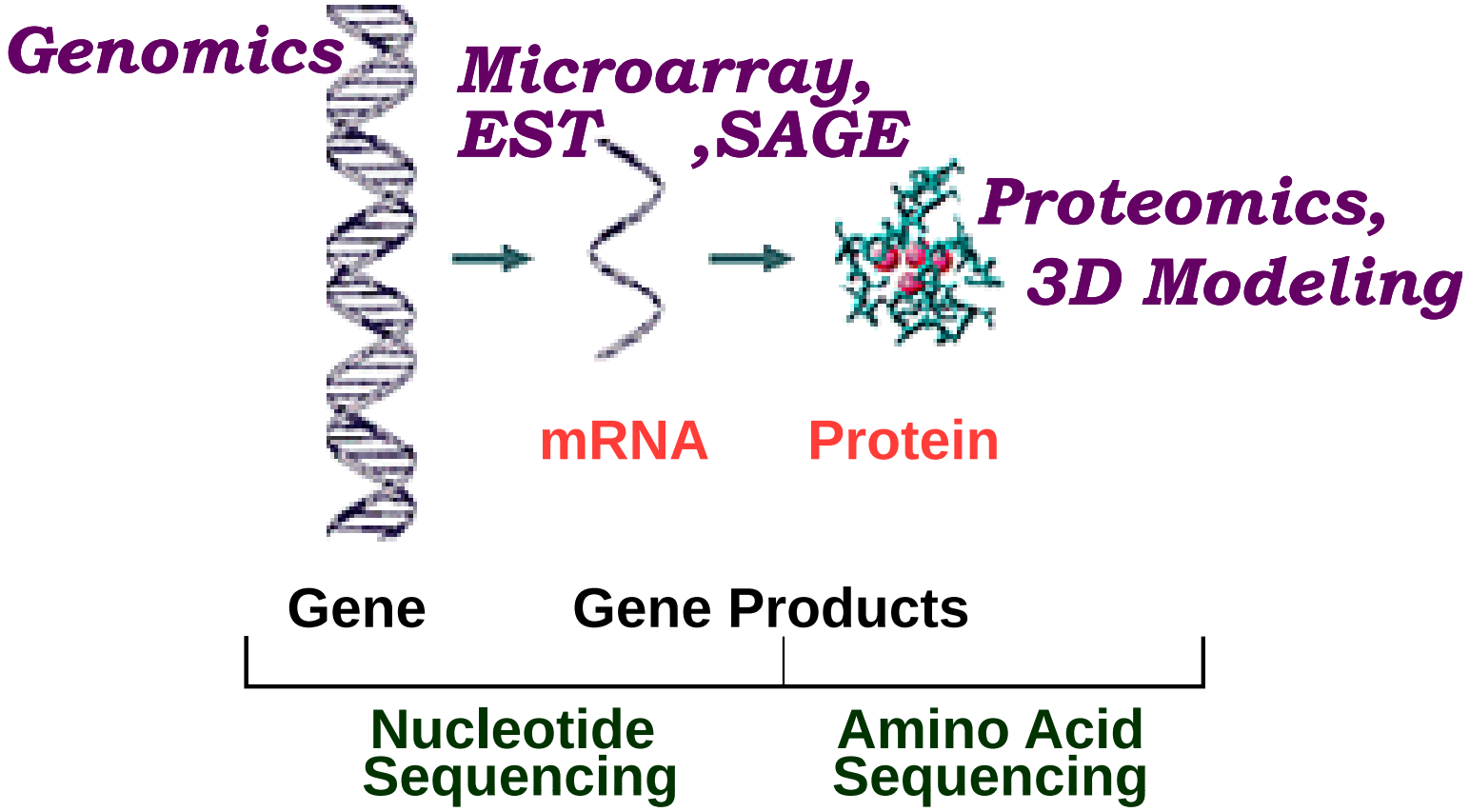


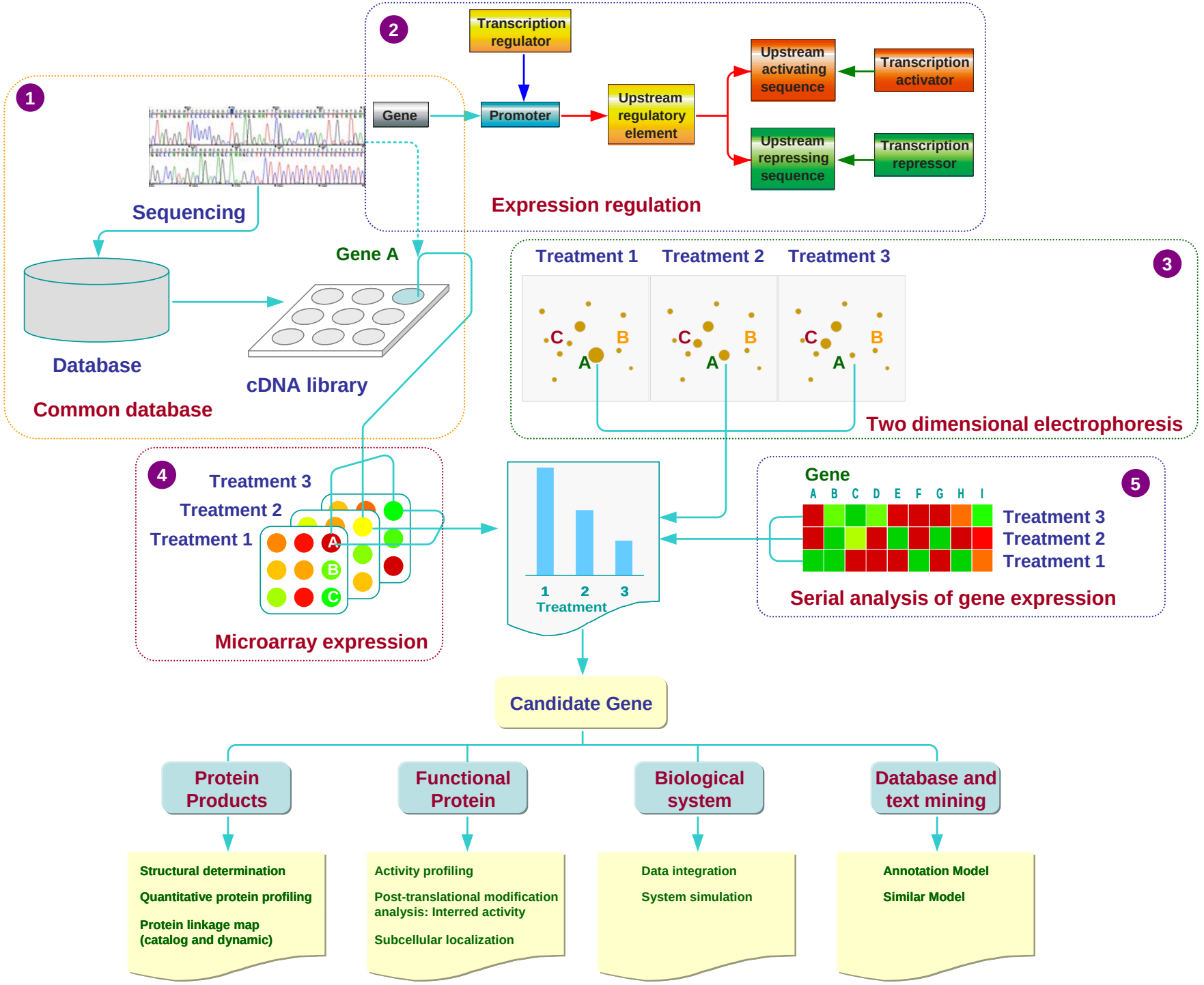




High Throughput Technologies: *The future of Molecular Diagnosis*

High Throughput Technologies (HTTs) are developed to produce huge amount of information from genome projects, but they have clear potential in mass screening and diagnostics of Infectious Diseases. The application of HTTs may revolutionize diagnostic techniques and replacing multiple individual assays.




1
2
Transcription regulator
Gene
Promoter
Upstream regulatory element
Upstream activating sequence
Transcription activator
Upstream repressing sequence
Transcription repressor
Expression regulation
Sequencing
Gene A
Database
cDNA library
Common database
3
Treatment 1
Treatment 2
Treatment 3
C
B
A
Two dimensional electrophoresis
4
Treatment 3
Treatment 2
Treatment 1
A
B
C
Microarray expression
1
2
3
Treatment
5
Gene
A B C D E F G H I
Treatment 3
Treatment 2
Treatment 1
Serial analysis of gene expression
Candidate Gene
Protein Products
Functional Protein
Biological system
Database and text mining
Structural determination
Quantitative protein profiling
Protein linkage map (catalog and dynamic)
Activity profiling
Post-translational modification analysis: Interred activity
Subcellular localization
Data integration
System simulation
Annotation Model
Similar Model

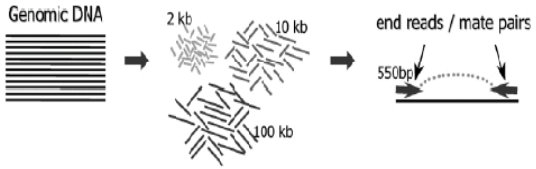
Going From Technology to Systems Biology

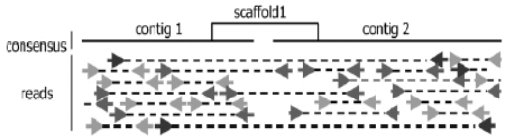
- Genomics → Genometrics
- Proteomics → Proteometrics
- Metabolomics → Metabometrics
- Phenomics → Phenometrics
- Bioinformatics → Biosimulation
- ***Quantify, quantify, quantify***

Genomics

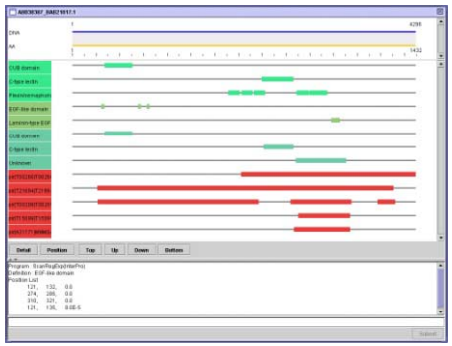
Genome Sequencing Project

Sequences





Identification



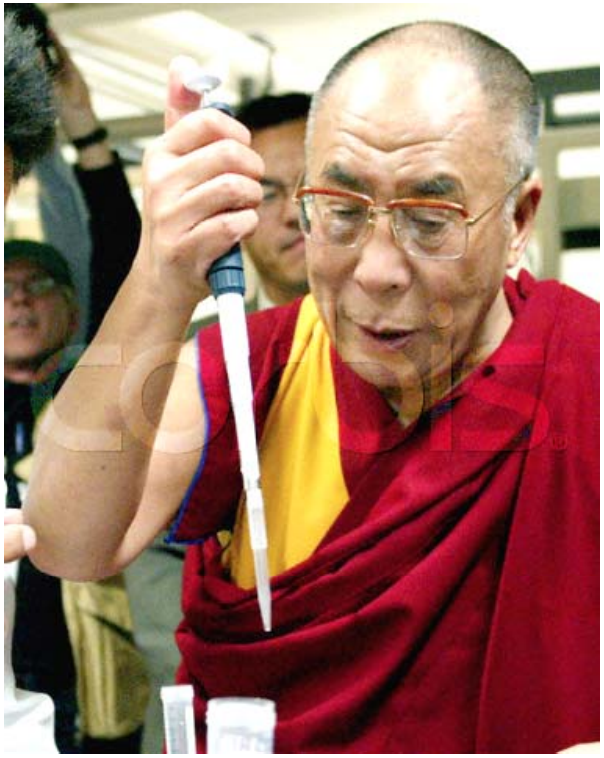
- Library Construction
- Genome Sequencing
- Genome Assembly
- Gap Closure
- Gene Prediction
- Gene Annotation
- Genome Analysis

Comparison

transcriptome - transcribed DNA sequence
proteome - peptide sequence
genome - related genomic sequence

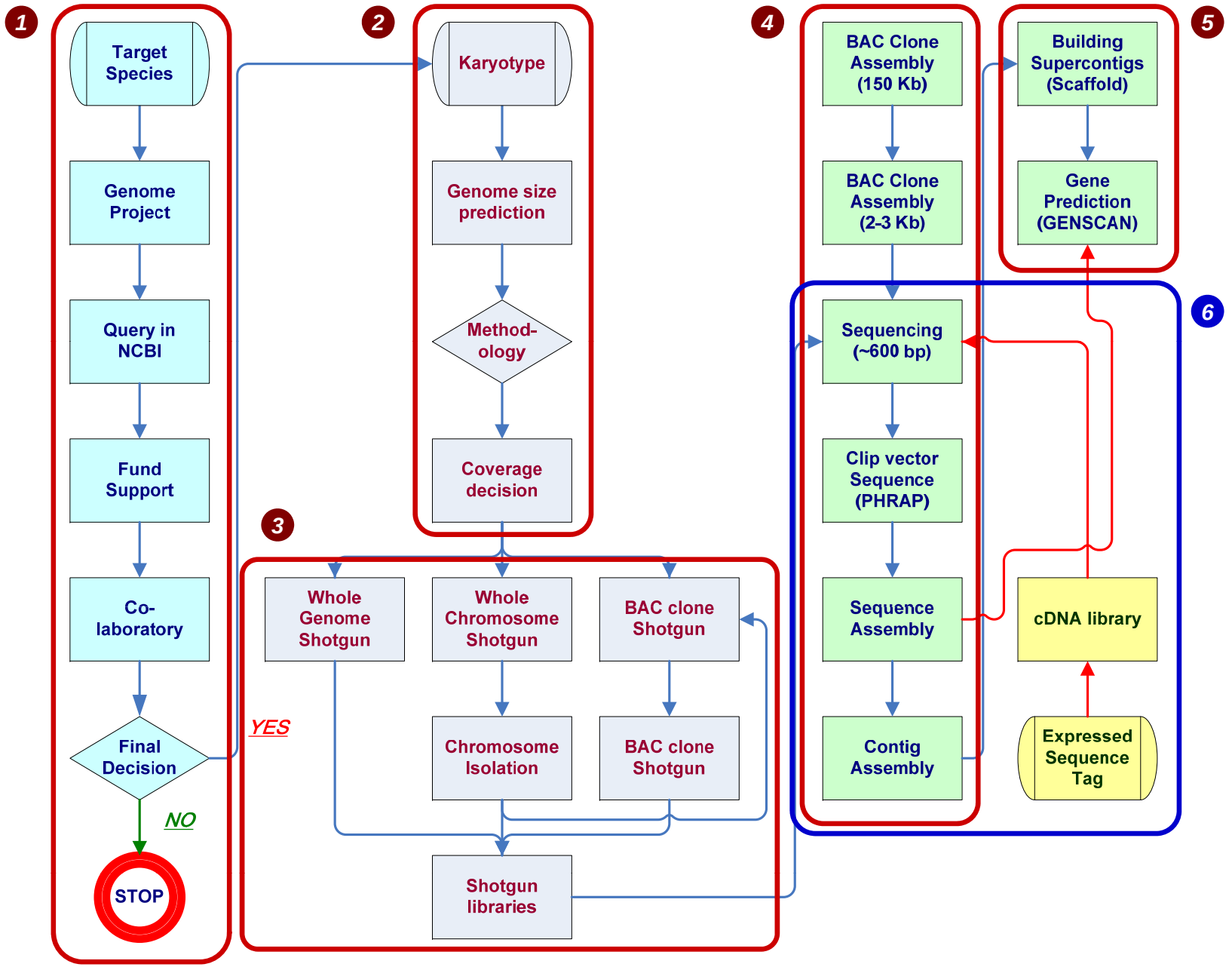
Functions

ORFeome based functional genomics
RNAi phenotypes
Gene Knockout
Expression Microarray



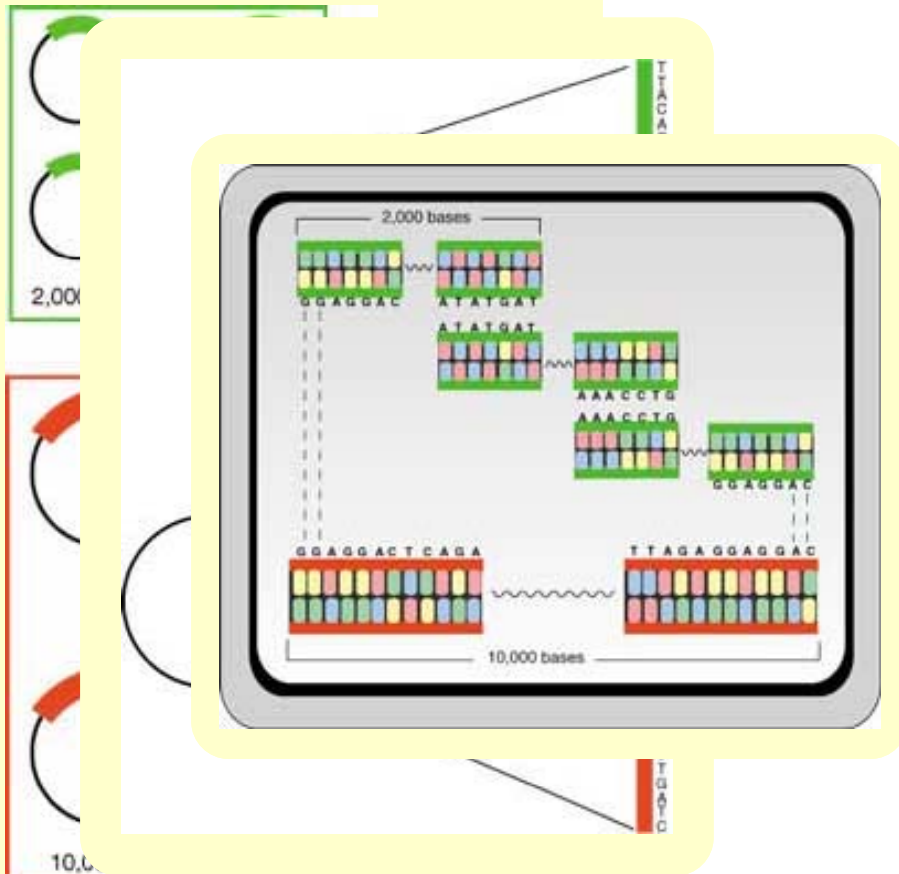
Let's go for a genome project

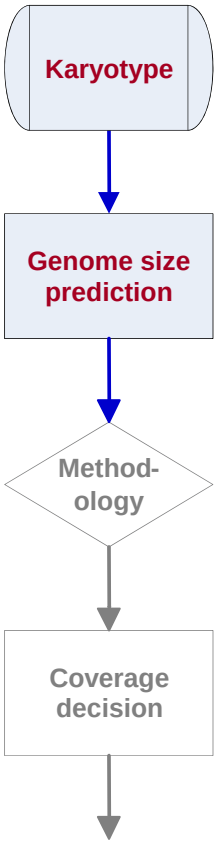
Genome Sequencing Project Flowchart



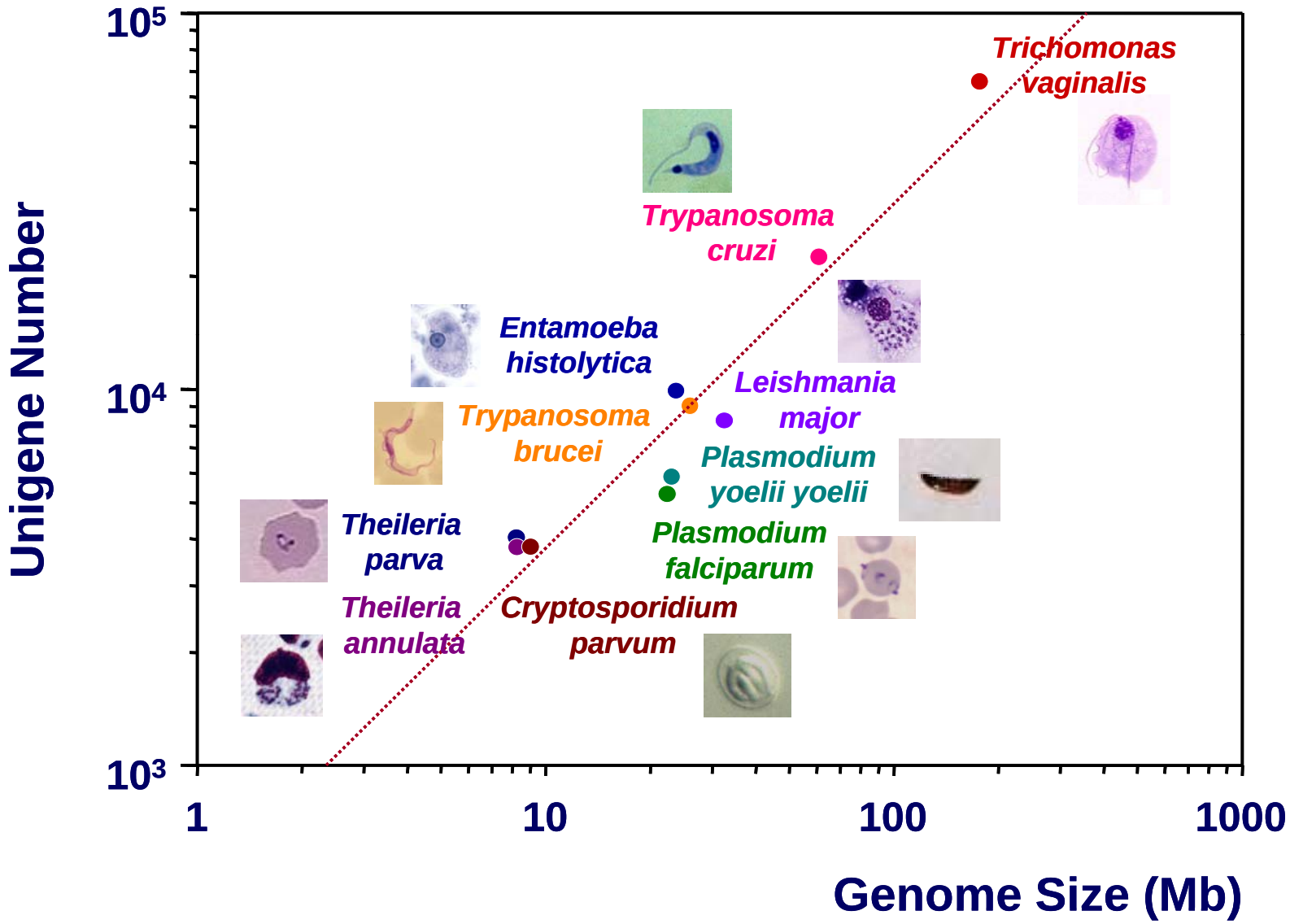
Sequencing the Plasmodium falciparum *genome*

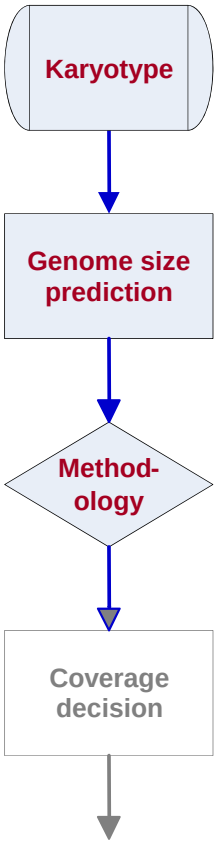
- Isolated individual chromosomes by PFGE
- Fragmented the DNA and cloned into *E. coli*
- Generated thousands of sequences from randomly selected fragments
- Assembled the random sequences to form “contigs” - Hundreds of fragments per chromosome obtained
- Ordered the fragments and closed the gaps
- Identified the genes, compared them to genes in other organisms, and annotated the genome sequence



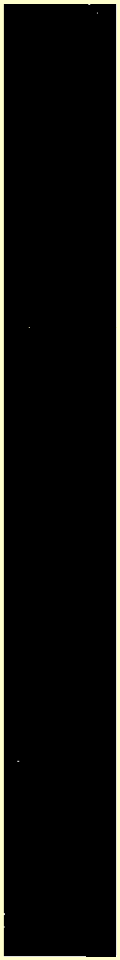


Protozoa Unigene Number





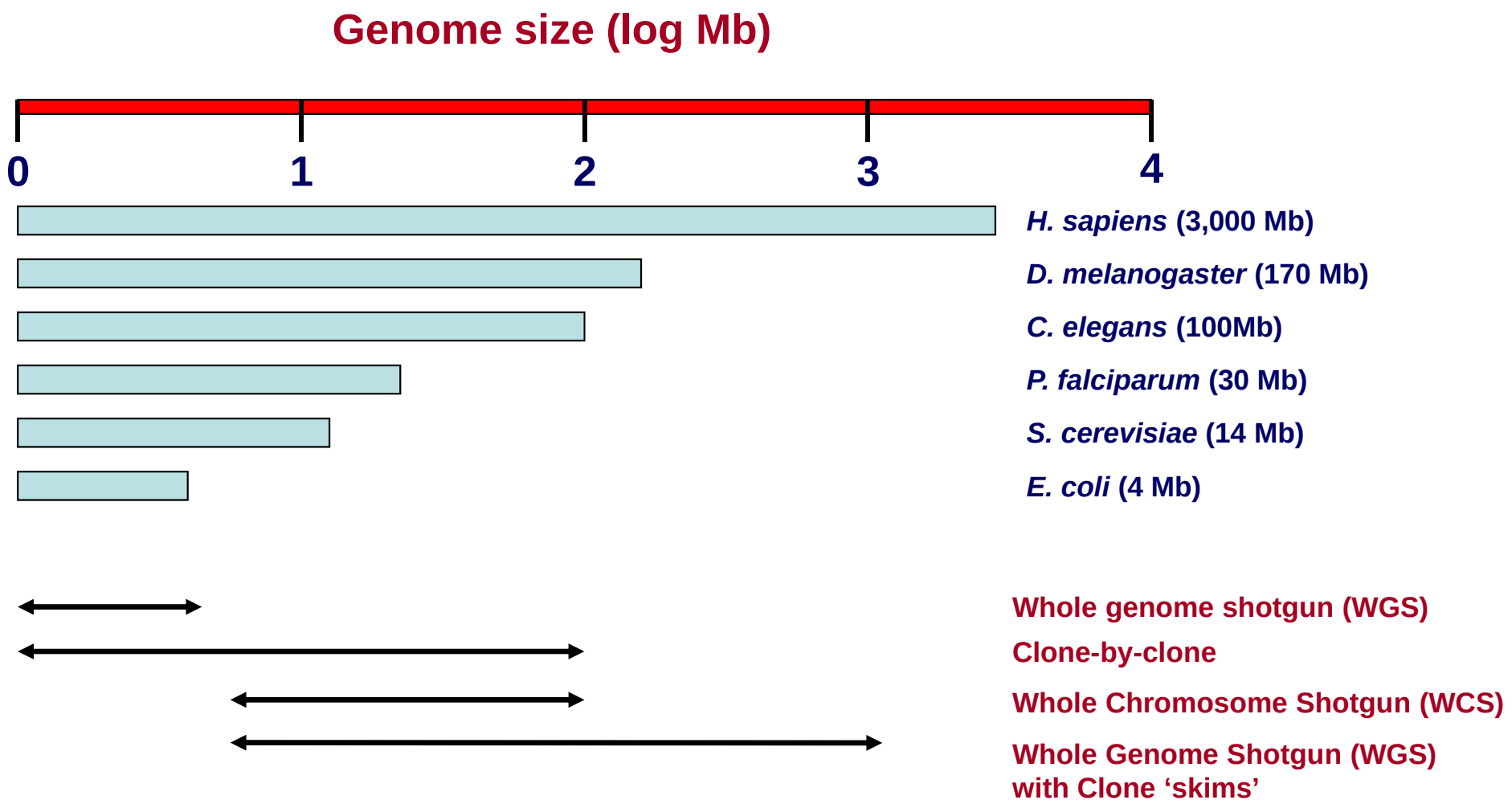
Plasmodium falciparum



13-14	3.2, 3.4 Mb
12	2.4 Mb
11	2.3 Mb
10	2.1 Mb
5-9	1.6 - 1.8 Mb
4	1.4 Mb
3	1.2 Mb
2	1.0 Mb
1	0.8 Mb

- ~30 million base pairs (Mb)
- 80% (A+T)
- 14 chromosomes
- DNA “unstable” in *E. coli*
- No large insert DNA clones suitable for sequencing
- Too large for whole genome shotgun
- Whole chromosome shotgun strategy was selected

Genome size and sequencing strategies



A History of Genome Sequencing

1981: Sanger et al. sequence Lambda (50Kbp) by the shotgun method.

Cloning:

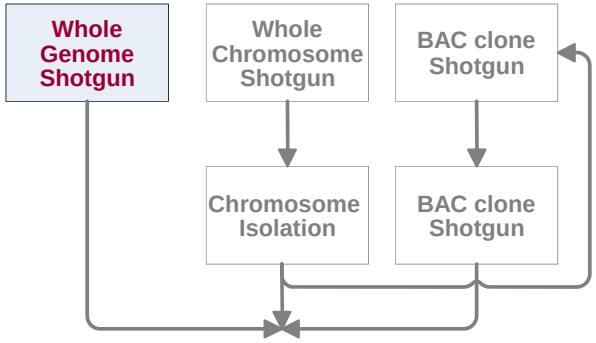
- **BACs permit 100-250 Kbp inserts**

Technology:

- **Cycle sequencing (linear PCR) permits efficient**
- **Sequencing of both insert ends**
- **Capillaries improve accuracy & efficiency**

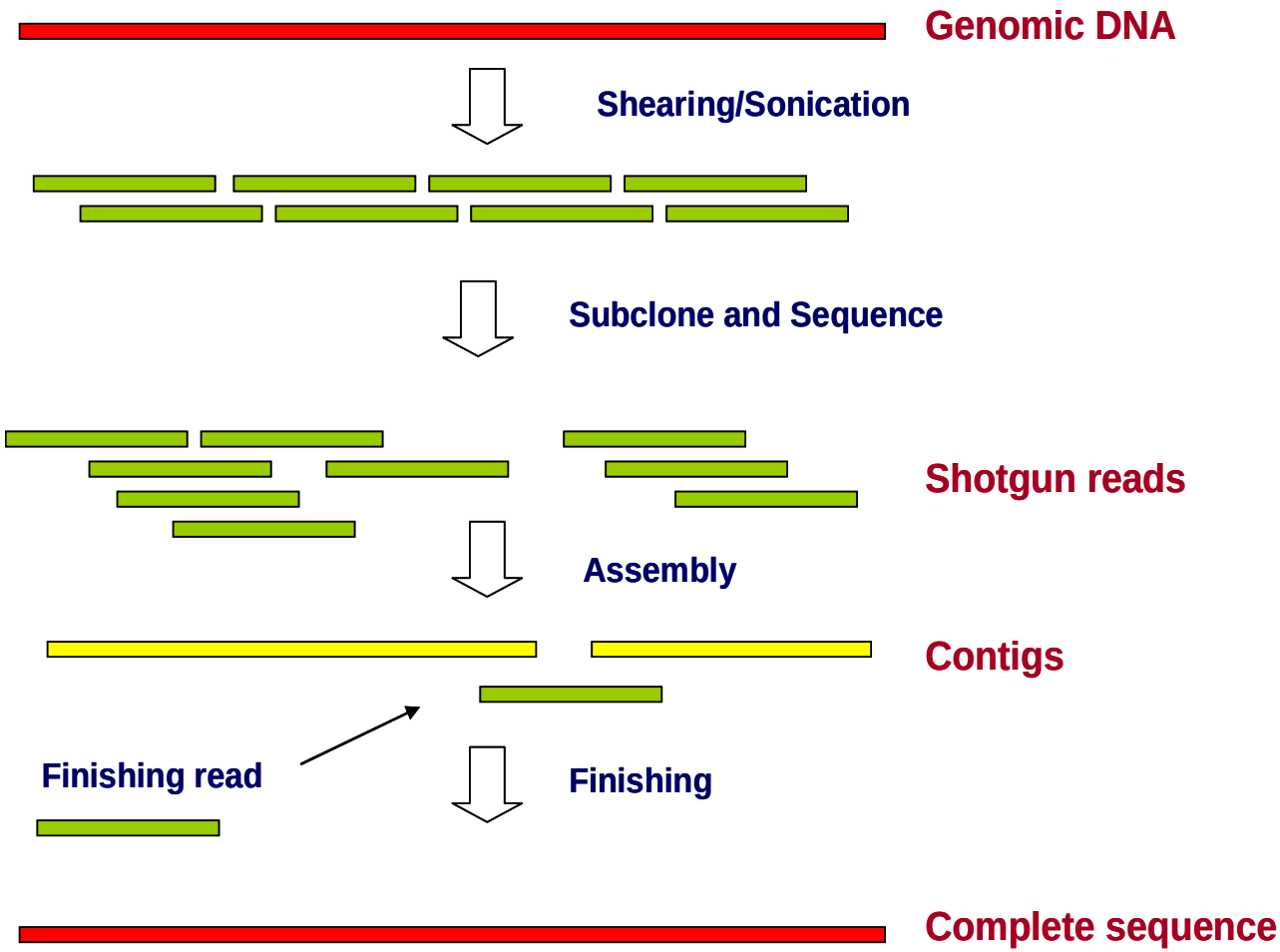
1998: 3% of the human genome has been sequenced using a BAC-based hierachical plan. Common wisdom is that shotgun approach does not scale beyond BACs save for simple bacterial sequences.

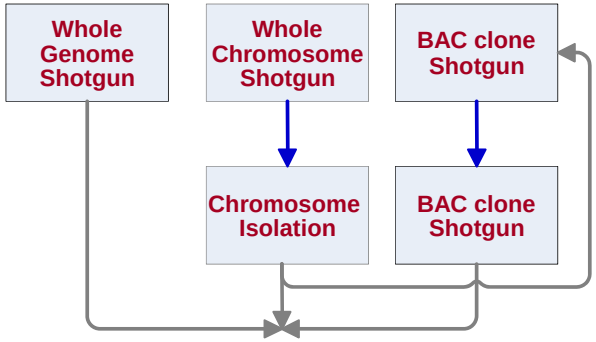
2001: 97% of the chromatin of the human genome has been determined. Mouse, Drosophila, Rice, Fugu, and Anopheles have all been sequenced with a whole genome shotgun approach.



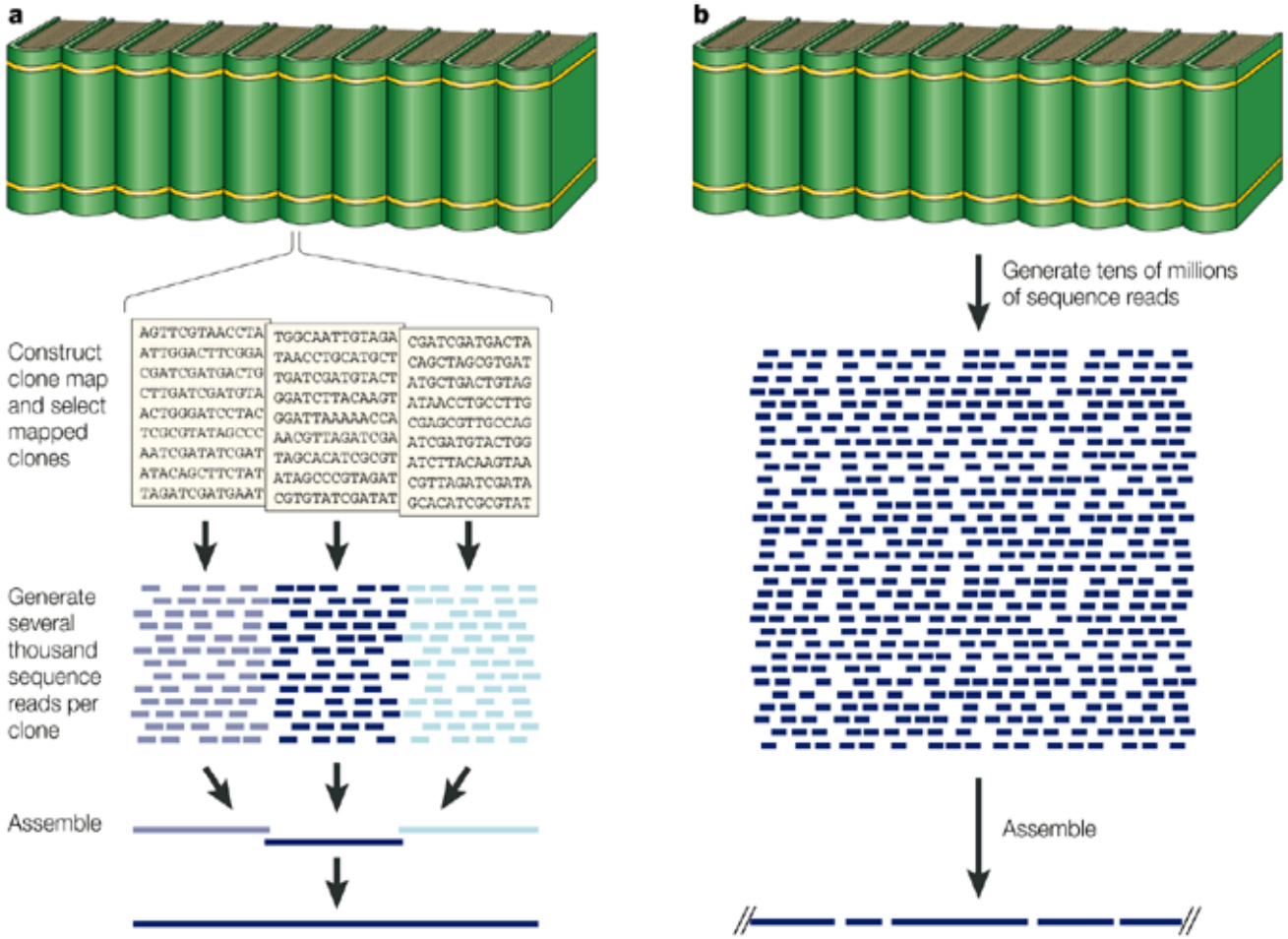
- Create clone library
- Sequencing the two ends of each clone
- Assembly the overlapped reads into contigs
- Assembly the contigs into super contigs
- Align the super contigs to the genome
- Genome Finishing

Full shotgun sequencing





Whole genome shotgun sequencing strategy



Building Supercontigs (Scaffold)

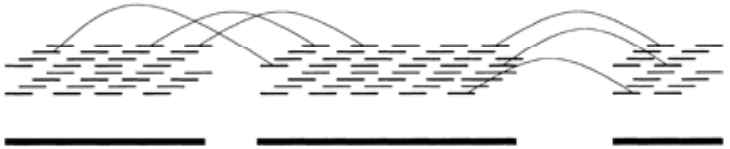
↓

Gene Prediction (GENSCAN)

Supercontig creation and gap filling

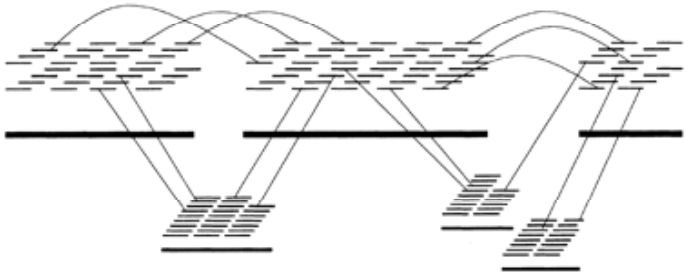
(A) A supercontig is constructed by successively linking pairs of contigs that share at least two forward-reverse links. Here, 3 contigs are joined into one supercontig.

A

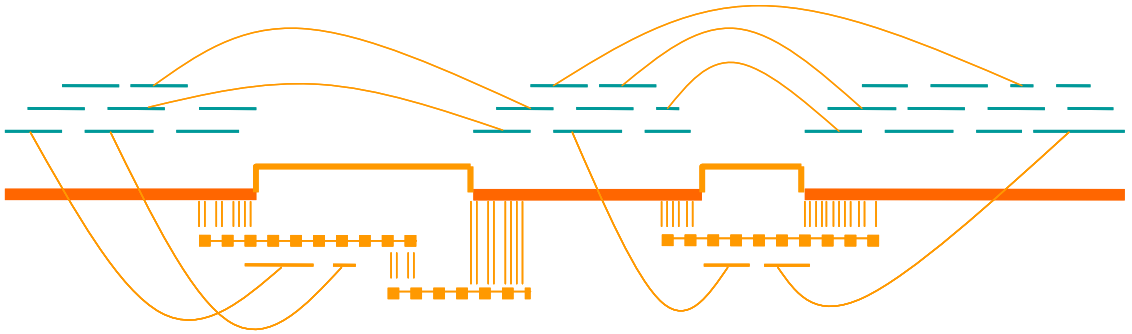


The layout now consists of a number of supercontigs with interleaved gaps. Most gaps belong to regions marked as repeat contigs, some correspond to regions of insufficient shotgun reads.

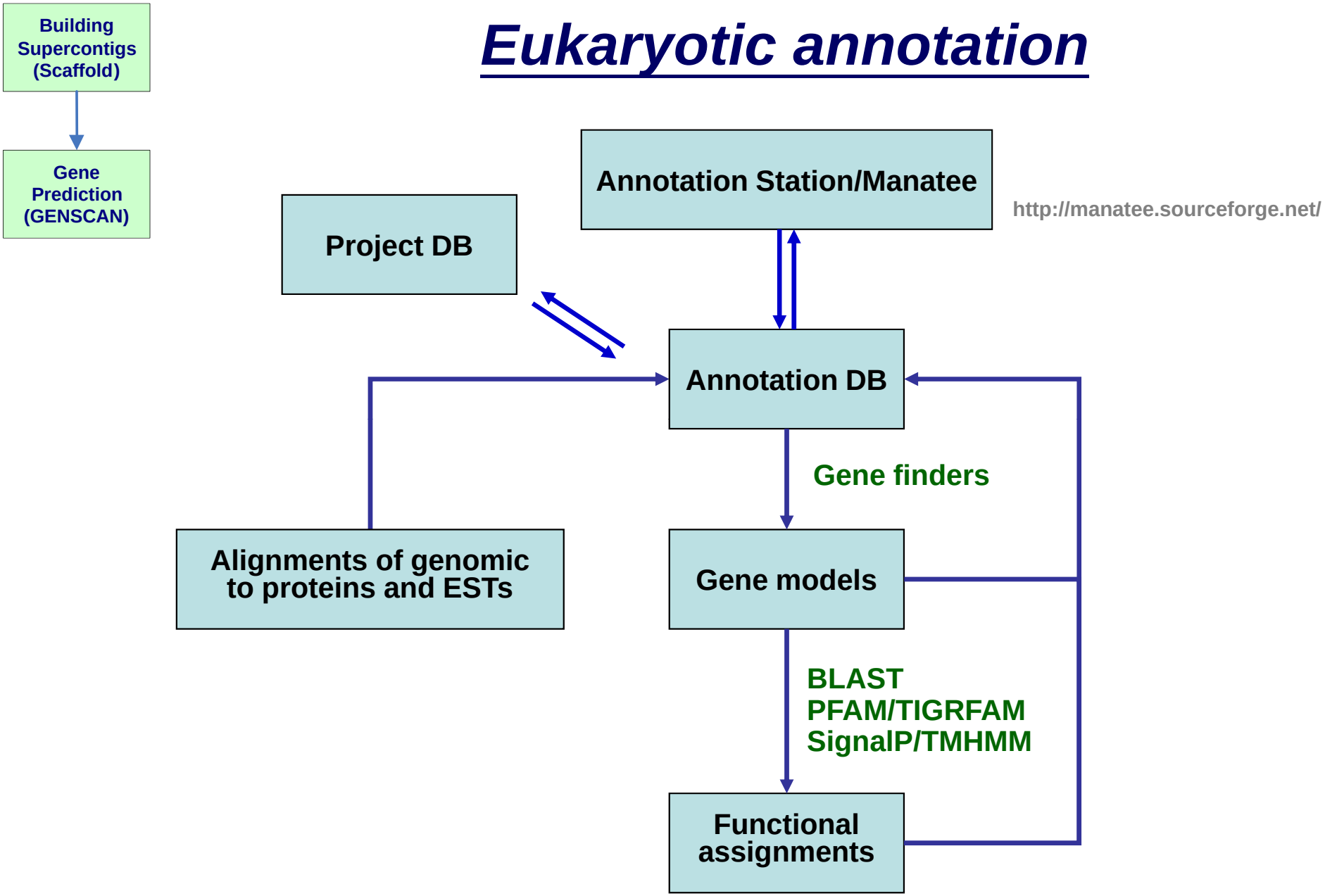
B



(B) Arachne attempts to fill gaps by using paths of contigs. The first gap in the supercontig shown here is filled with one contig, and the second gap is filled by a path consisting of two contigs.



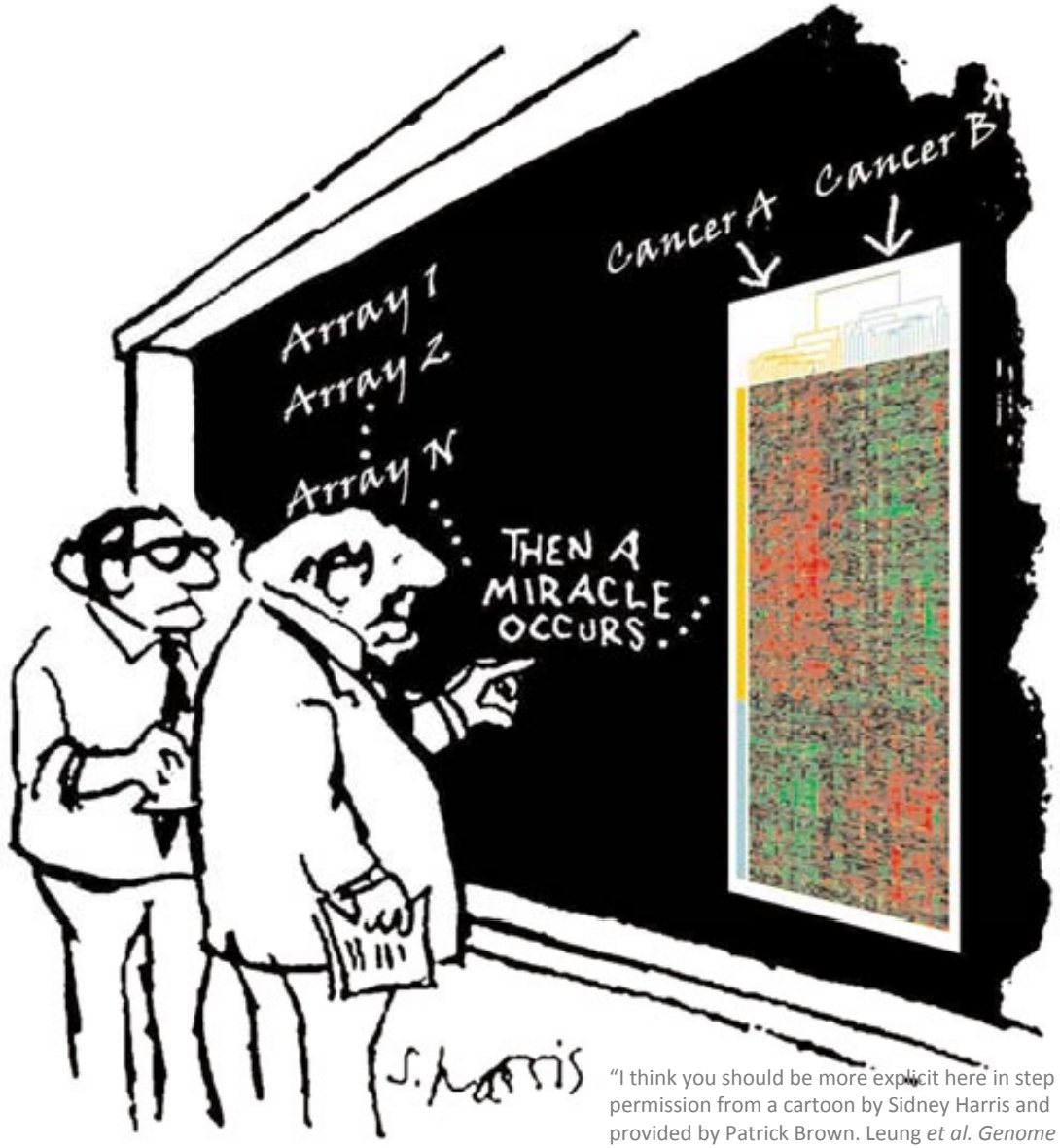
Eukaryotic annotation



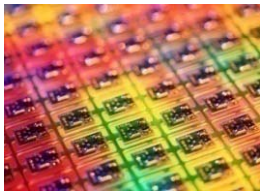
Transcriptomics

MICROARRAYS:

Chipping Away at the Mysteries of Science and Medicine



"I think you should be more explicit here in step two." Modified with permission from a cartoon by Sidney Harris and from an image provided by Patrick Brown. Leung *et al. Genome Biology* 2001 **2**:reports4021.1 doi:10.1186/gb-2001-2-9-reports4021



陣列

In computer science an **array** is a data ***structure*** consisting of a group of ***elements*** that are accessed by indexing. In most programming languages each element has the same data type and the array occupies a contiguous area of storage. Most programming languages have a built-in array data type.

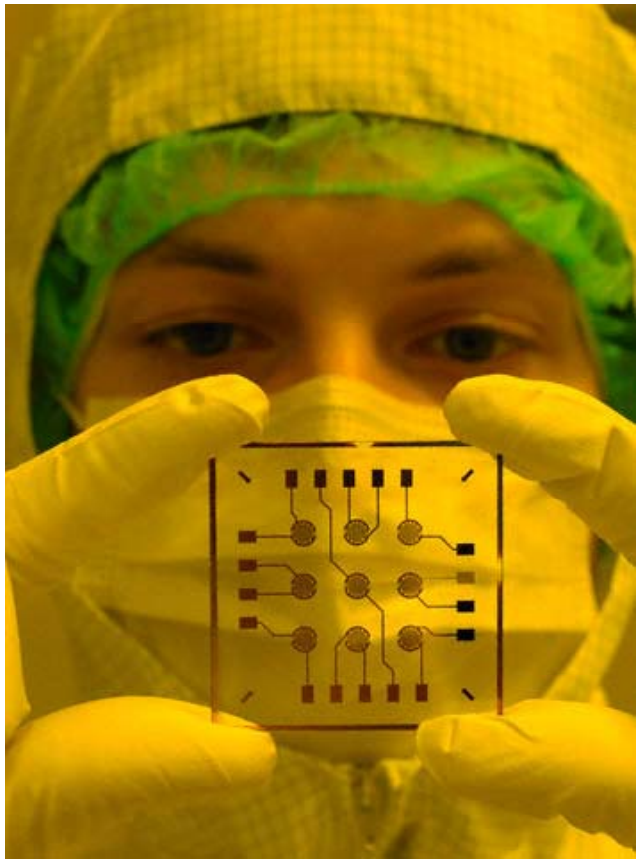
Some programming languages support array programming (e.g., APL, newer versions of Fortran) which generalises operations and functions to work transparently over arrays as they do with scalars, instead of requiring looping over array members.

Multi-dimensional arrays are accessed using more than one index: one for each dimension.

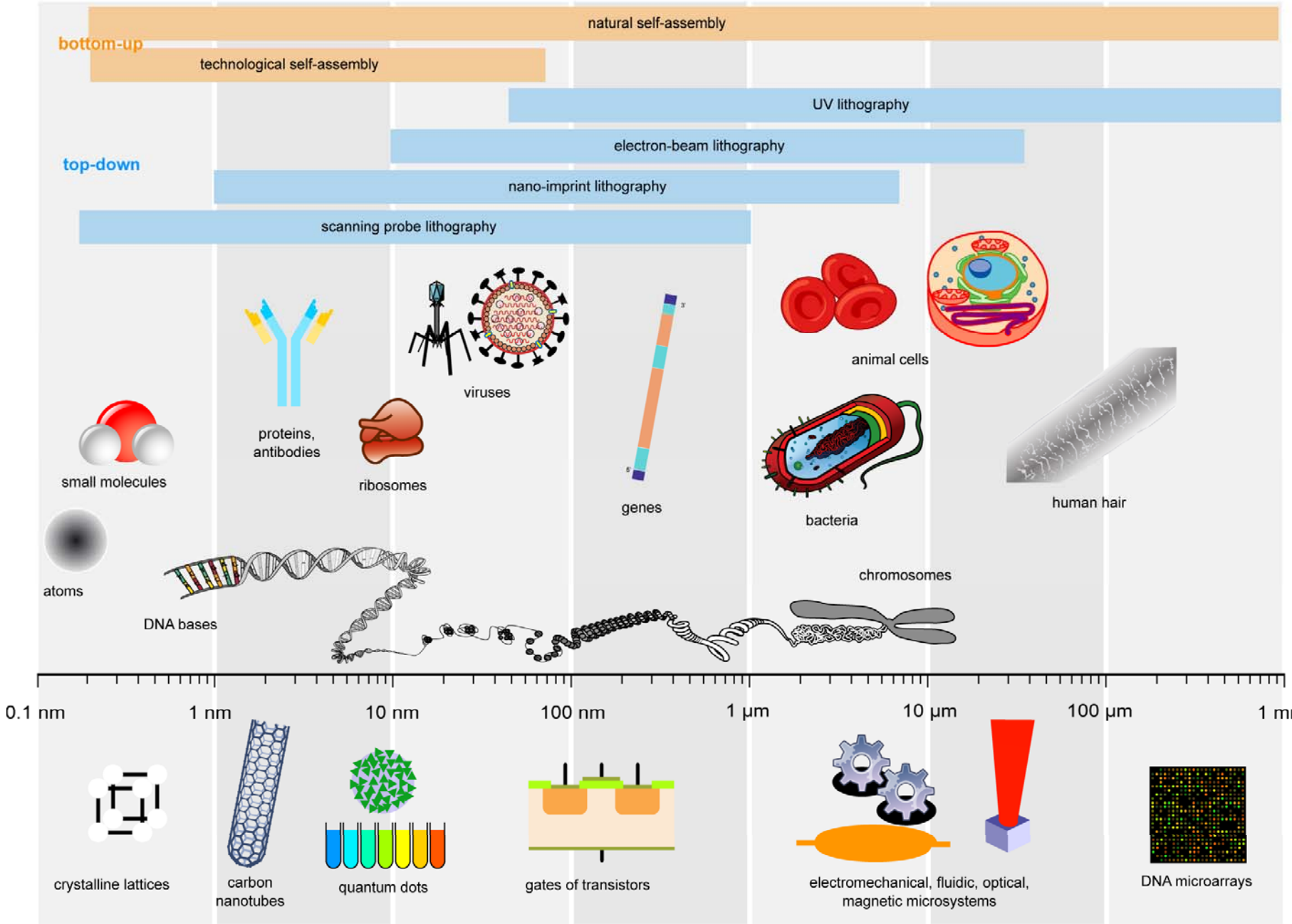
Arrays can be classified as fixed-sized arrays (sometimes known as static arrays) whose size cannot change once their storage has been allocated, or dynamic arrays, which can be resized.

微

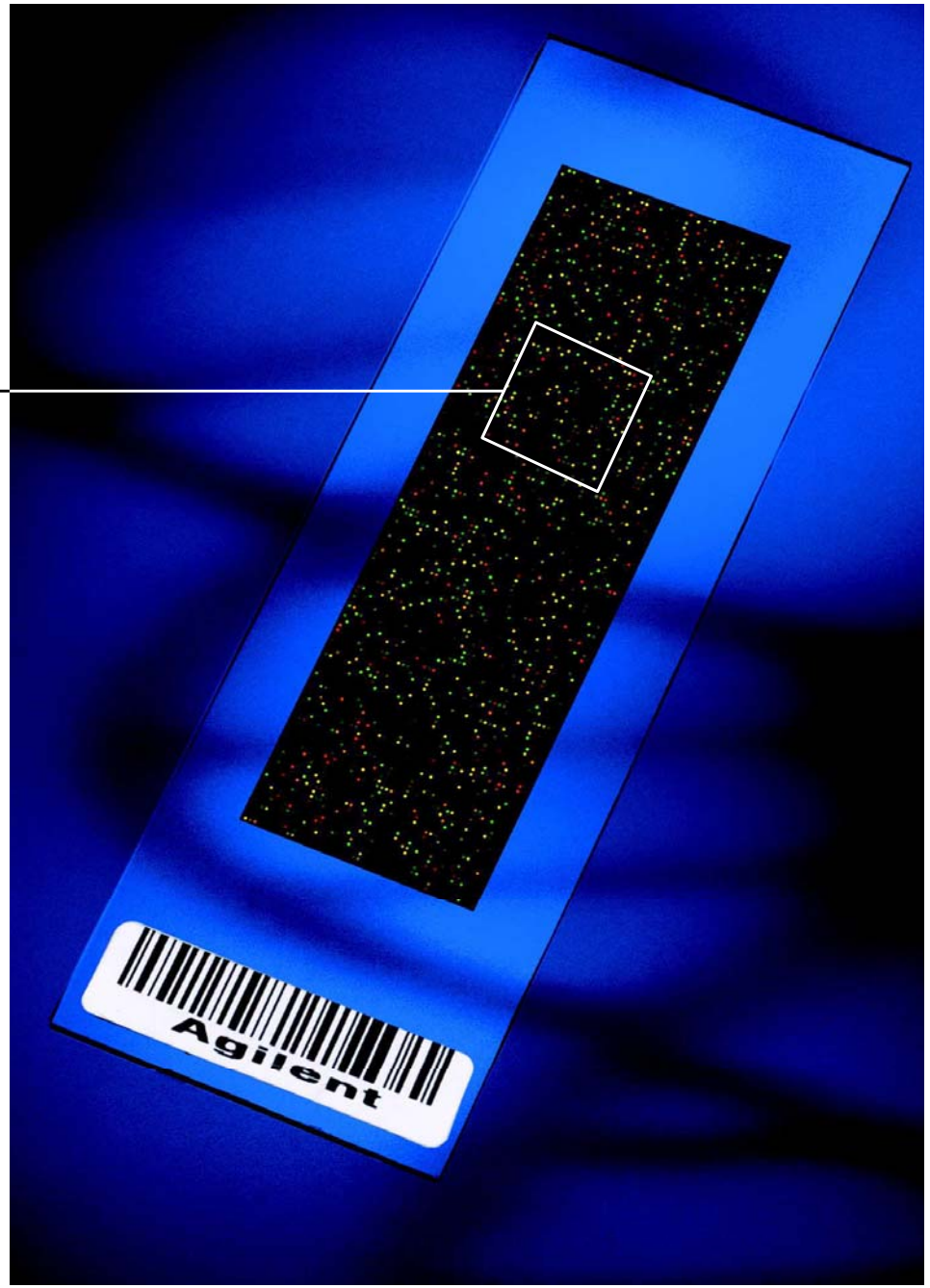
Micro is an English prefix of Greek origin that refers to an object as being ***smaller*** than an object or scale of focus, in contrast with macro.

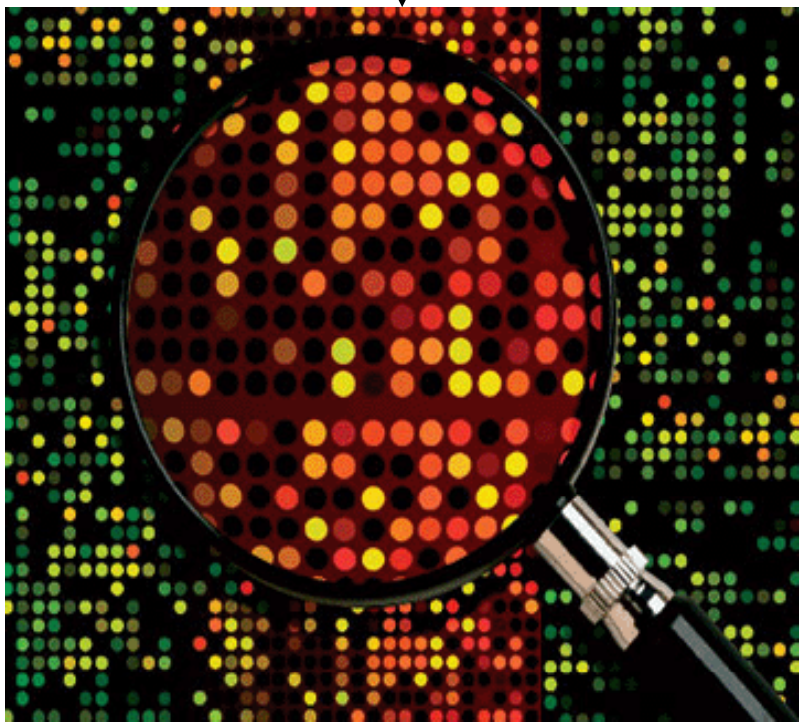


Comparison of various biological assemblies and technological devices



Microarrays are miniaturized biological devices consisting in molecules, for example DNA or protein, named the "probes", that are orderly arranged at a microscopic scale onto a solid support such as a membrane or a glass microscope slide.



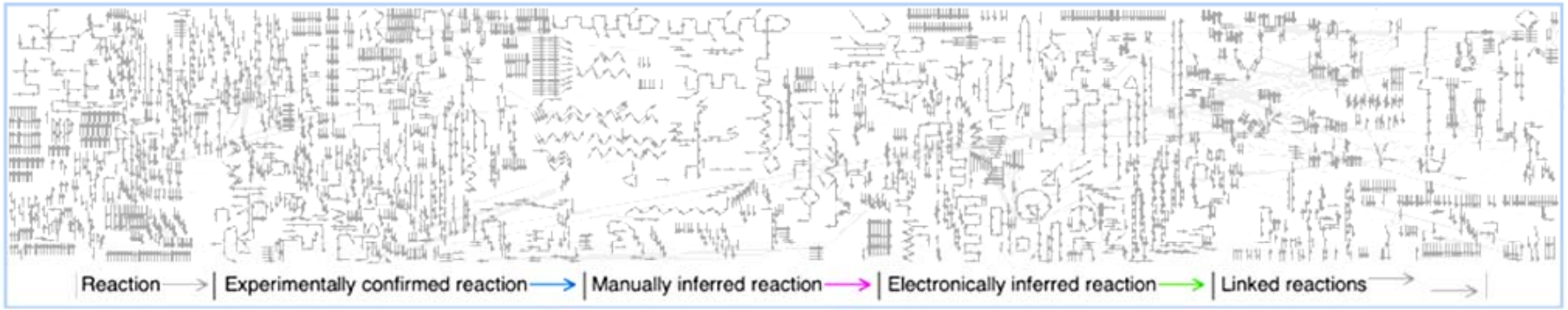


What problems can it solve?

1. Differing expression of genes over time, between tissues, and disease states
2. Identification of complex genetic diseases
3. Drug discovery and toxicology studies
4. Mutation/polymorphism detection (SNP's)
5. Pathogen analysis

Why Use DNA Microarrays for Expression Analysis?

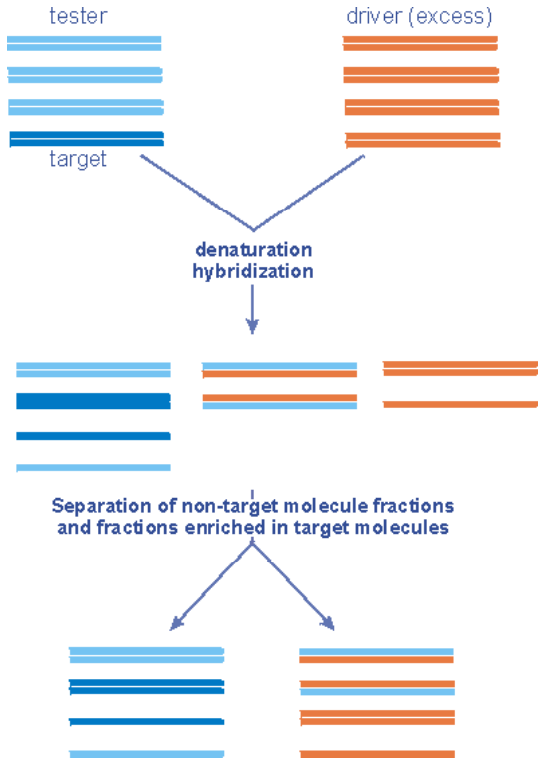
1. Conventional expression analysis only allows the study of the expression of a single gene in a single experiment.
2. The highly parallel nature of microarrays allows the simultaneous study of the expression of thousands or even tens of thousands of different genes in a single experiment.
3. Microarrays allow researchers to undertake global expression analysis that is not feasible with conventional techniques.

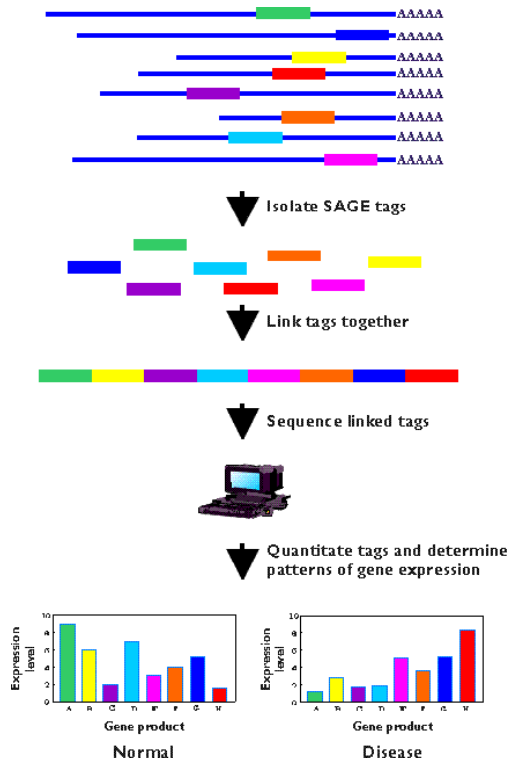


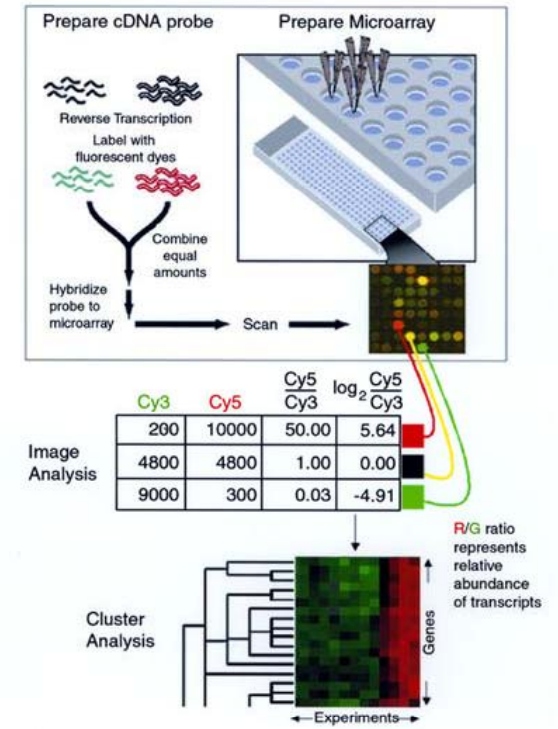
www.reactome.org

Evaluate global gene expression

1. Differential displaying
2. Suppressive subtractive hybridization (SSH)
3. Sequencing of expressed sequence tags (EST), serial analysis gene expression (SAGE & Long SAGE)
4. Hybridization to microarrays







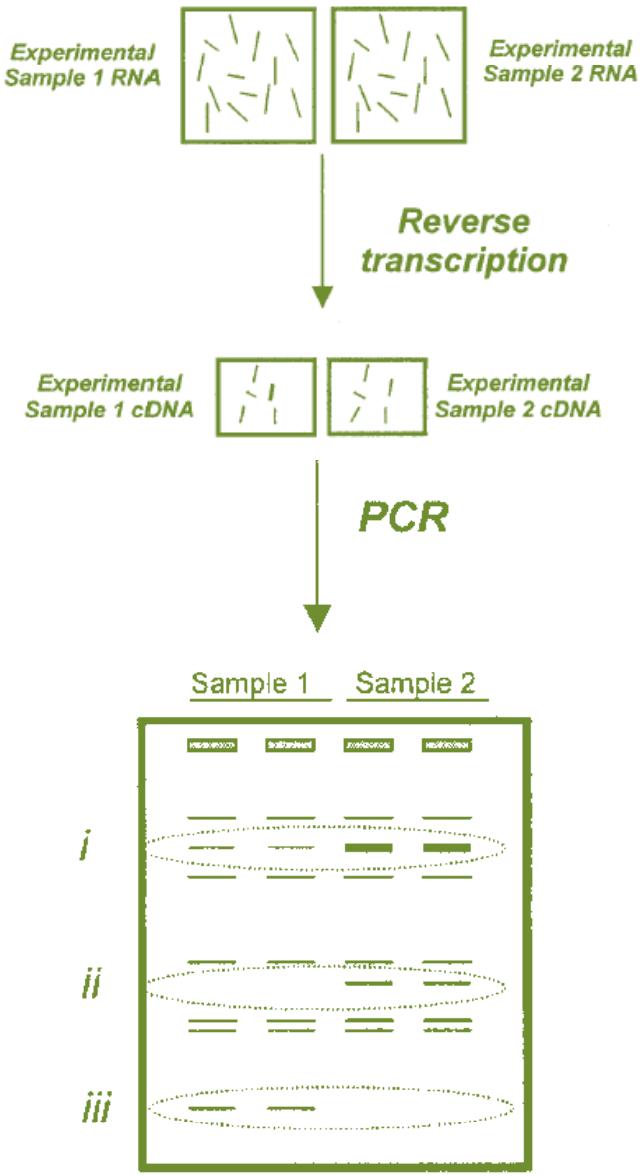
Differential displaying

Several significant limitations of the original protocol were:

- the requirement of large quantities of mRNA
- a bias against the identification of rare transcripts.
- not a quantitative method with a high rate of false positives, or
- gene fragments that seem to be differentially expressed as an artifact

Advantages

- ✓ to compare multiple experimental samples simultaneously
- ✓ to identify genes that are either up or down-regulated in one sample relative to another

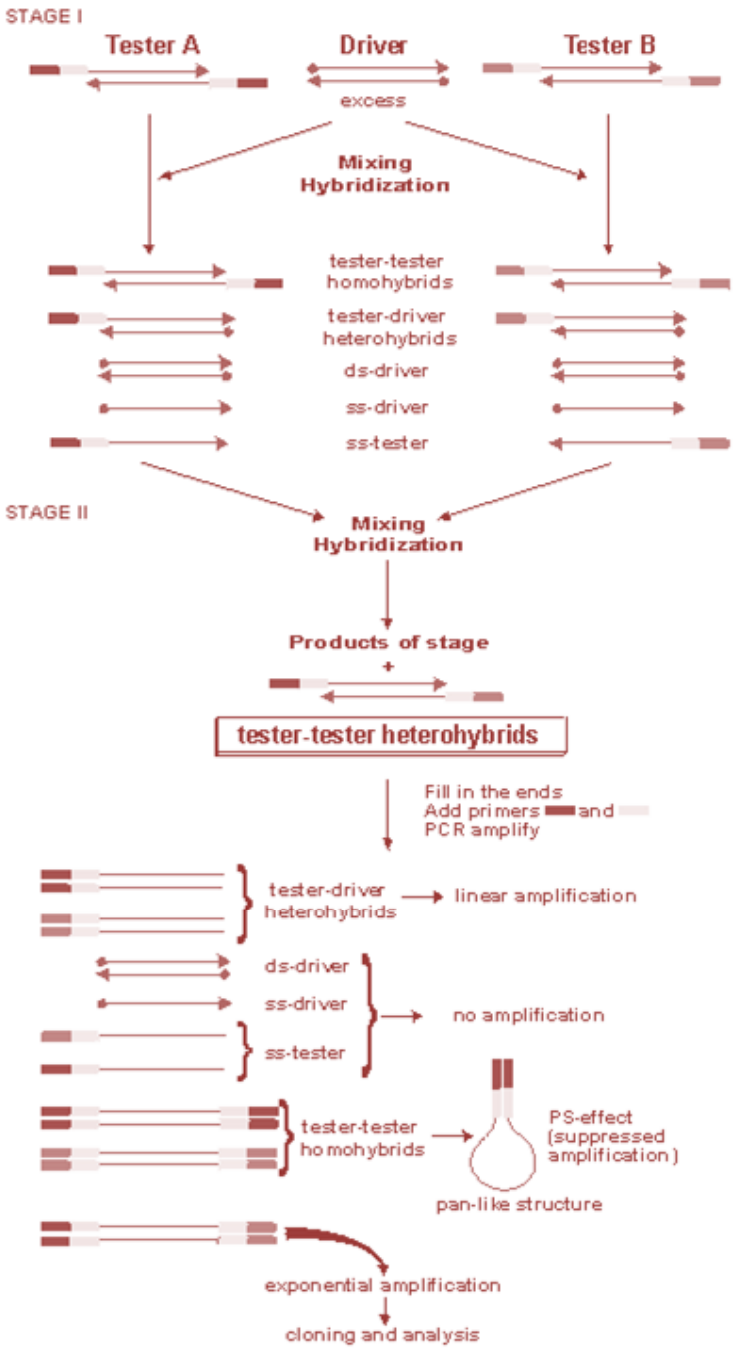


Subtractive hybridization

Subtractive hybridization (or subtraction) is a method for rapid isolation of differentially distributed nucleic acids (differentially expressed, differentially present, or differentially arranged in two or more different sources: different cells, cell populations or cell types, different tissues, disease, or development stages). The Suppressive subtractive hybridization (SSH) method is designed to selectively amplify differentially expressed transcripts while suppressing the amplification of abundant transcripts, thus eliminating the need to separate single- and double-stranded molecules.

Limitations

- only to pair-wise treatment comparisons
- does not provide a quantitative measure of expression differences

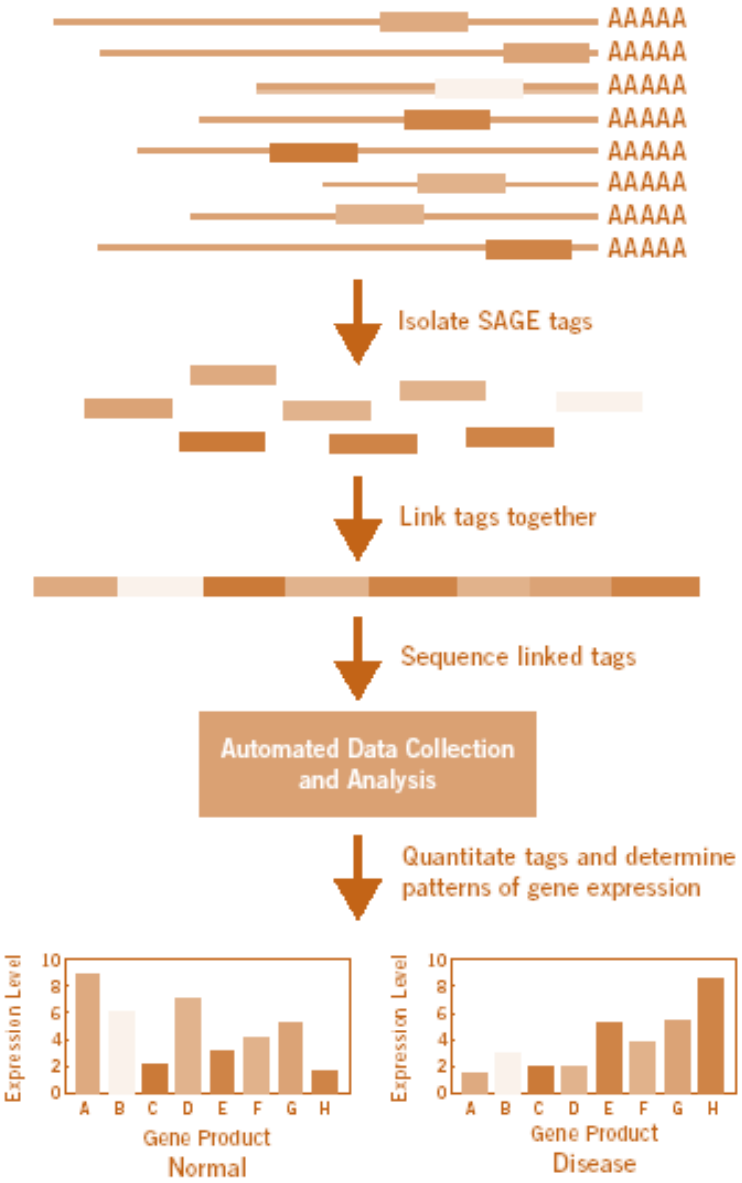


EST and SAGE

- ✓ Expressed sequence tag (EST) sequencing are generated by randomly picking clones from a cDNA library and performing a single sequencing reaction to produce 300 to 500 bps of sequence per clone. Differences in gene expression may be identified by counting the number of times a particular sequence appears in EST libraries of genes from different sources.
- ✓ Serial analysis of gene expression (SAGE) a technique for the analysis of gene expression that is essentially an accelerated version of EST sequencing.
  SAGE: 10-14 bps tags created by type IIS restriction endonuclease Nla III.
  Super SAGE: > 25 bps tag, created by type III restriction enzyme EcoP15I.

Advantages:

- ➢ SAGE data are quantitative and cumulative.
- ➢ Accurate, quantitative transcript profiles describing the abundance of all genes expressed in a cell or tissue are generated by SAGE, if sufficient sequencing is completed.

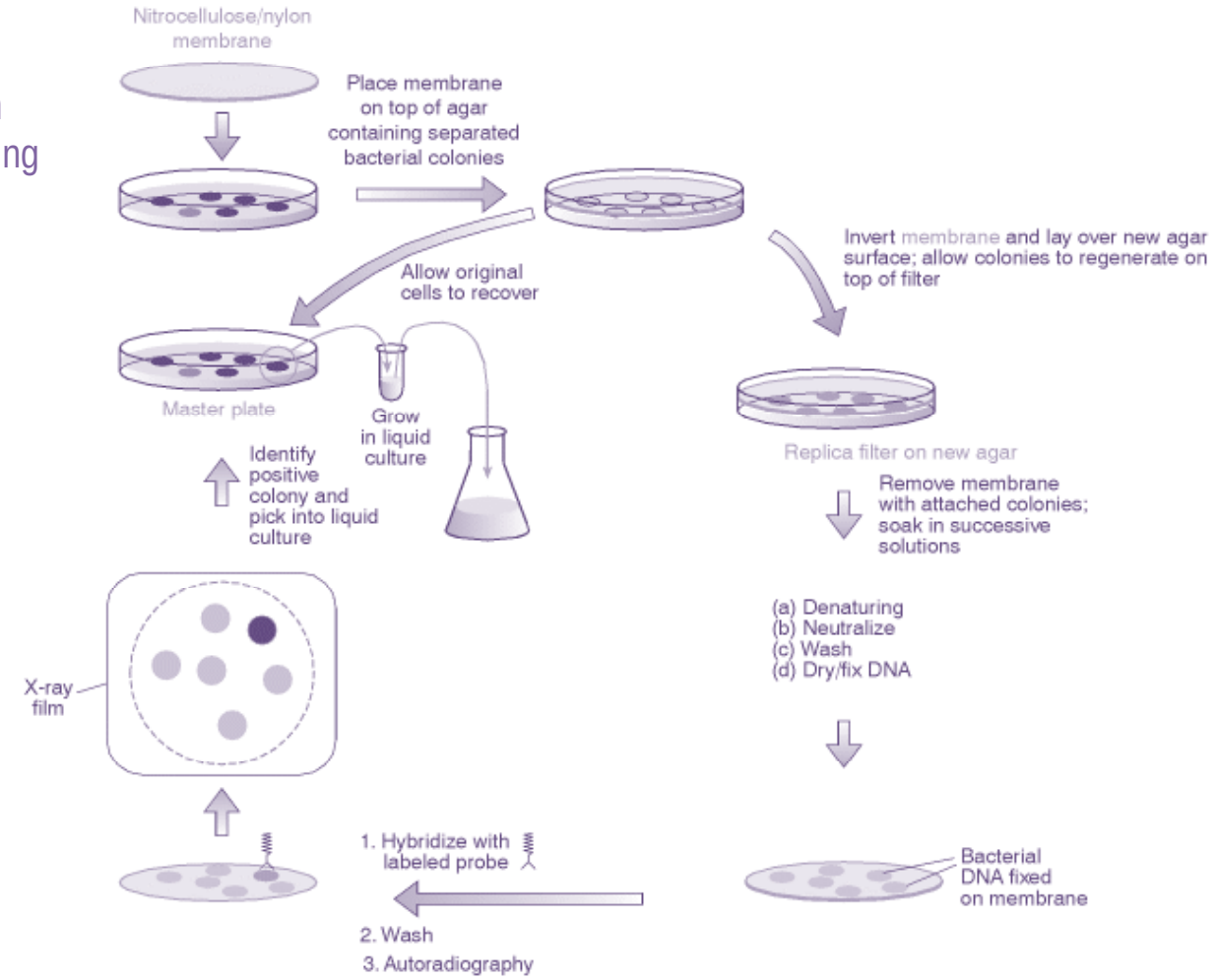


Microarray hybridization

The analysis of array data may be divided into three components:

1) identification and quantification of hybridization intensities,
2) visualization of data, and
3) clustering techniques

Based on the assumption that genes with related functions are coregulated, clustering of microarray data becomes a powerful method to assign putative functional classifications to novel genes.

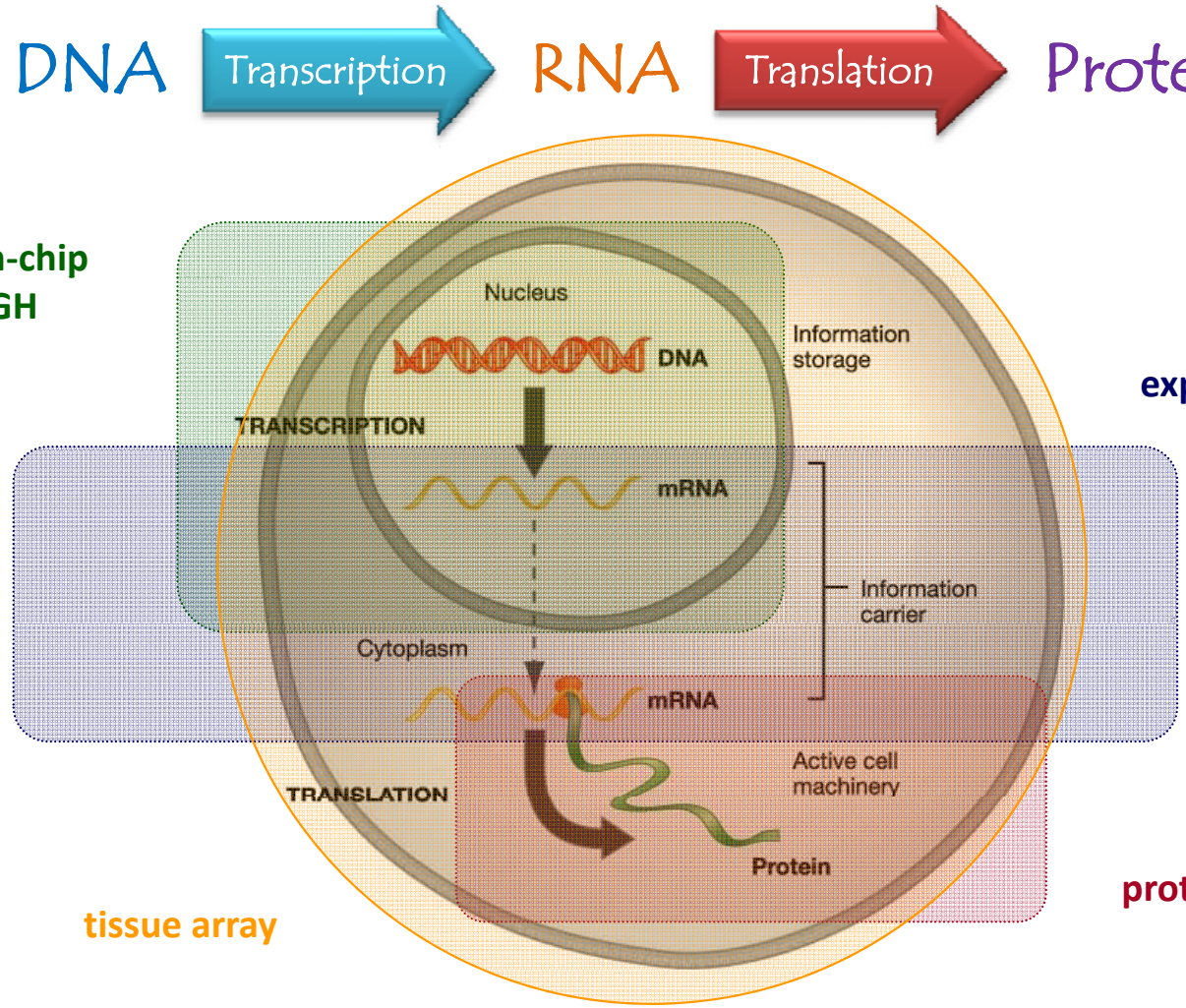


They are generally divided and differentiated by many ways

1. Type of target DNA (immobilized nucleic acid molecule) used in the array fabrication
2. Type of substrate to which the target DNA is printed or spotted
3. Methodology used to present the target DNA on the substrate
4. Density of probes on the array
5. Type of labeling for hybridization

1. **Type of target DNA (immobilized nucleic acid molecule) used in the array fabrication.**

cDNA, oligonucleotide, RNA, protein, antibody, tissue



CHIP-on-chip
array CGH

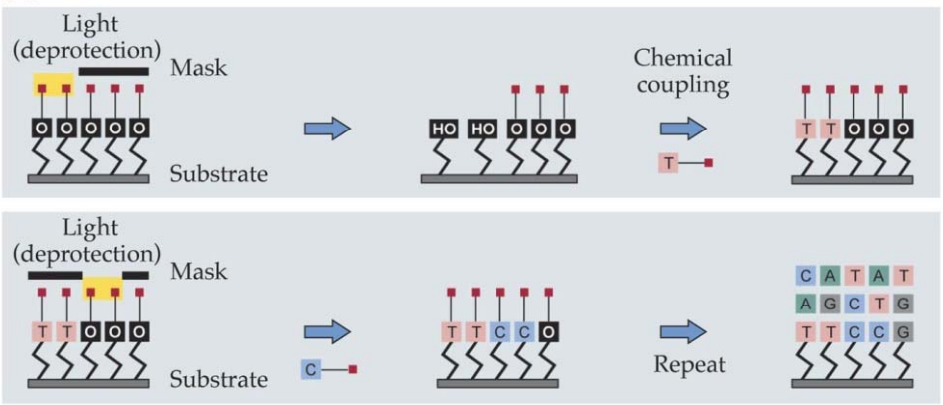
expression array

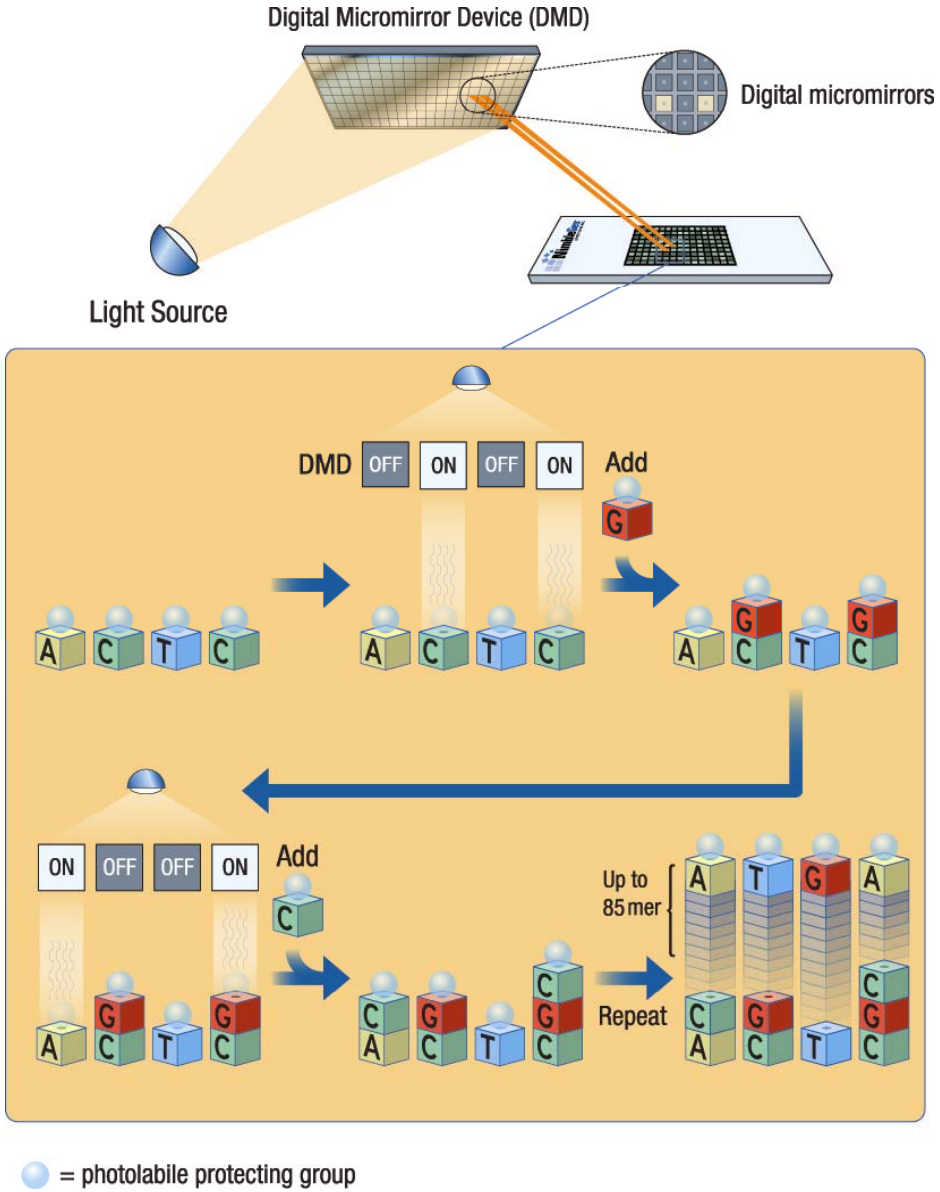
protein chip

tissue array

Construction of oligonucleotide arrays

- The light flash is produced by photolithography using a mask to allow light to strike only the required features on the surface of the chip.
- Oligonucleotide are synthesized *in situ* in the silicon chip. In each step, a flash of light "deprotects" the oligonucleotides at the desired location on the chip; then "protected" nucleotides of one of the four types (A, C, G or T) are added so that a single nucleotide can add to the desired chains. There are four types of masks according to the added nucleotide.



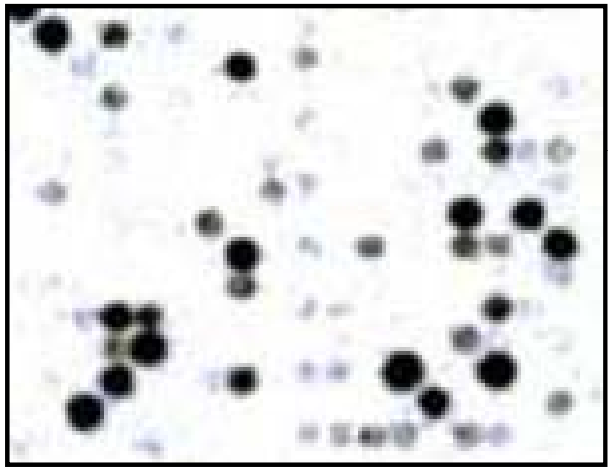


4. Density of probes on the array.

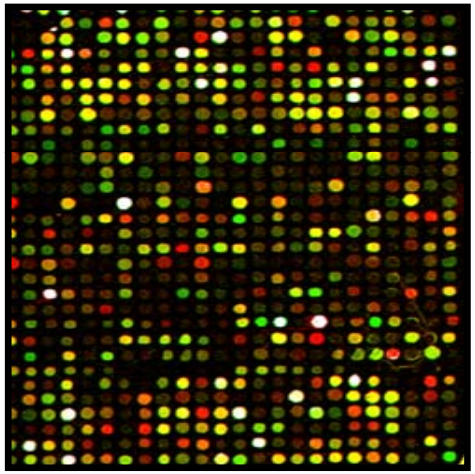
low density: 100 s (example: protein antibodies)

medium density: 1,000s to 10s of 1,000s and more (example: cDNA)

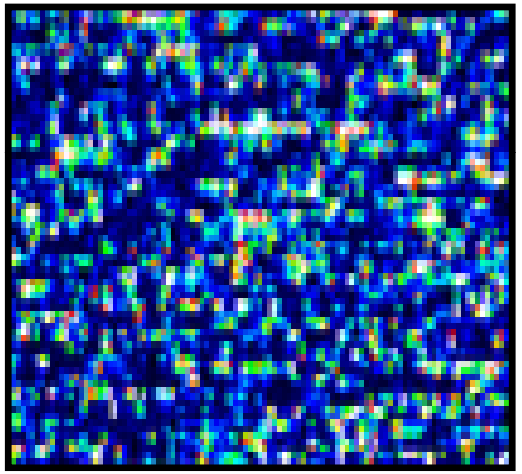
high density 100s to 1000s of 1000s (example: short oligonucloetide)



Isotope
Nylon – cDNA (300-900 nt)

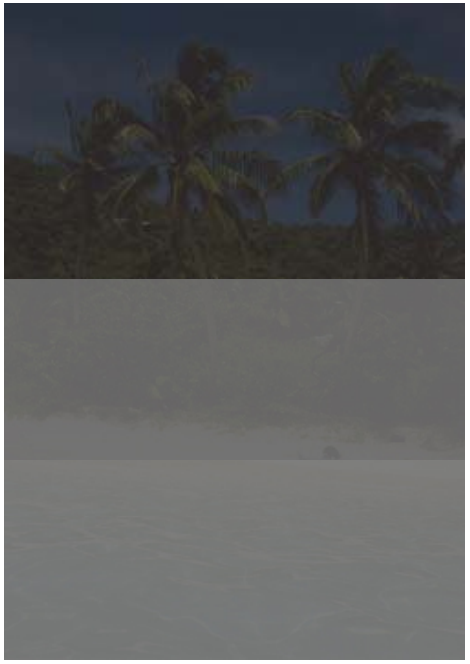
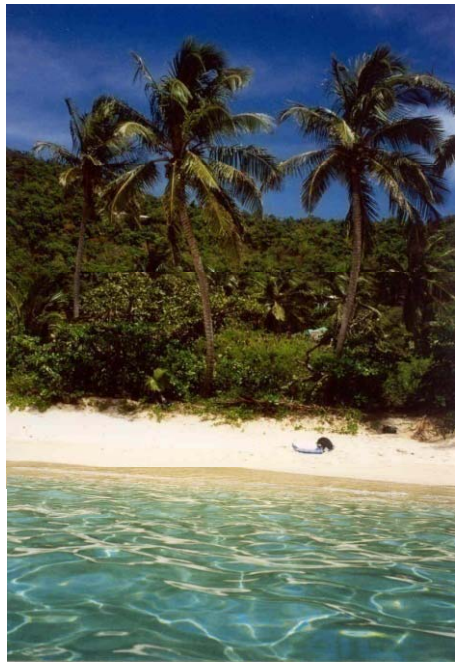


Two-colour
cDNA or Oligo (80 nt)
500 – 11,000 elements



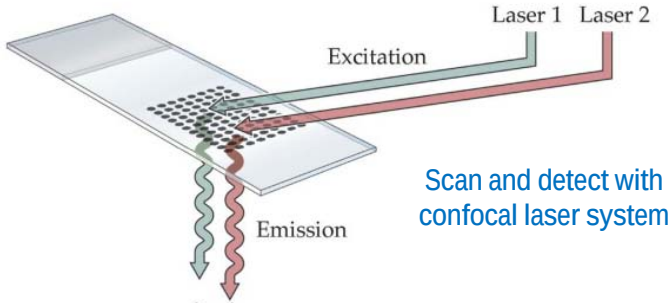
Affymetrix
Silicone – oligo (20 nt)
22 ,000 elements

Digital Image Analogy

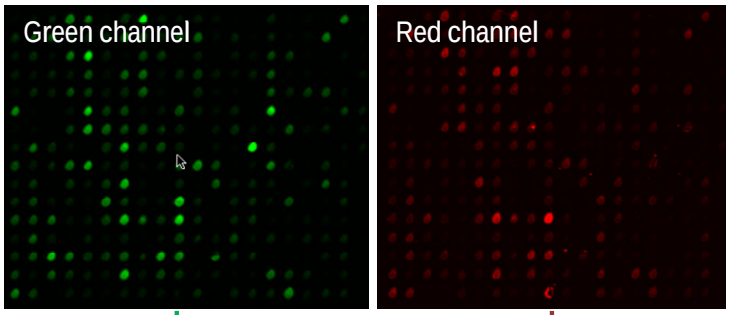
	Northern blot	Customized	Commercial
Expression			
CGH (LOH)			
ChIP, CGR, Exp Tiling, Methylation, …			
	Low density ▪ low content	High density ▪ high content Short probes ▪ limited contrast ▪ limited channels	High density ▪ high content Long, isothermal probes ▪ high contrast ▪ multiple channels

5. Type of labeling for hybridization.

single channel using one color and double channel using two color



Scan and detect with confocal laser system



Overlay images and normalize

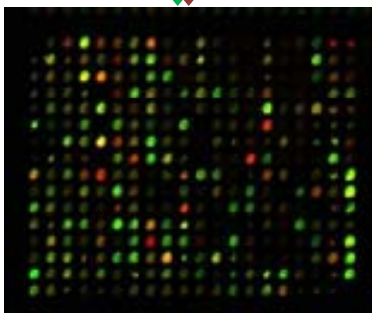
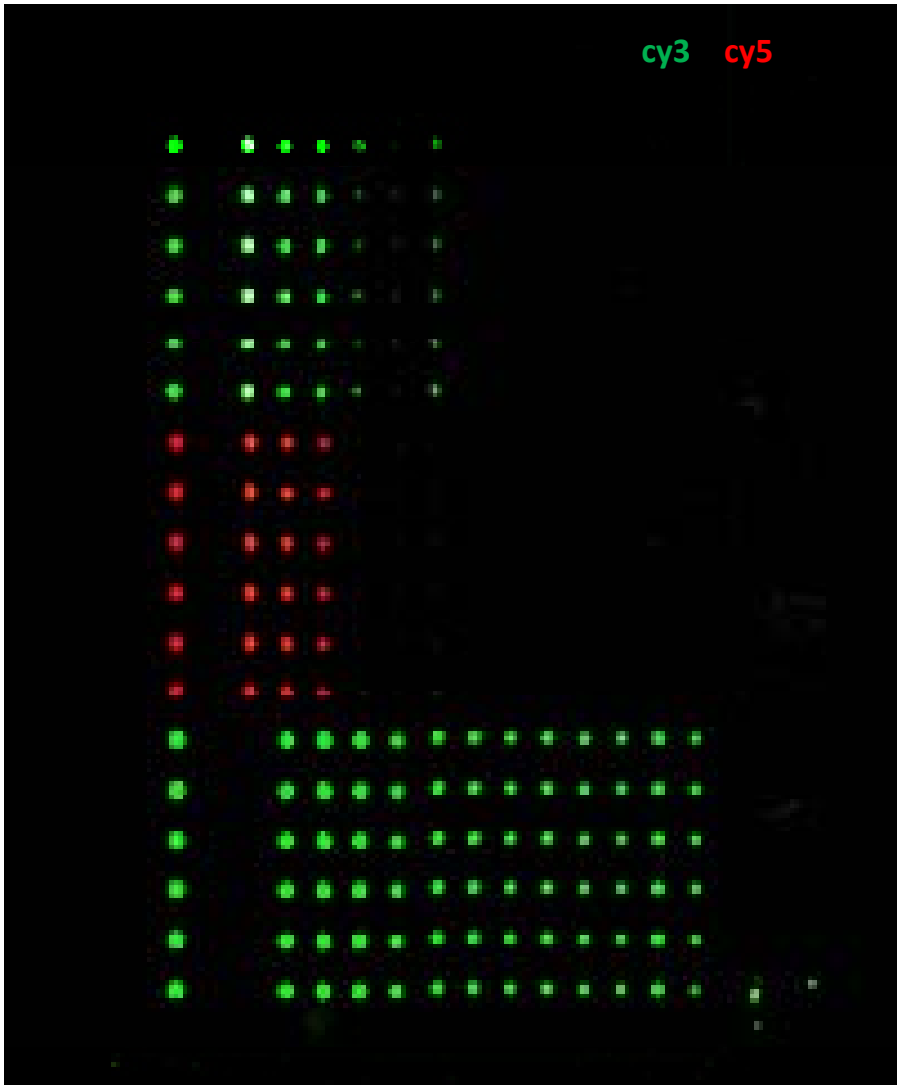
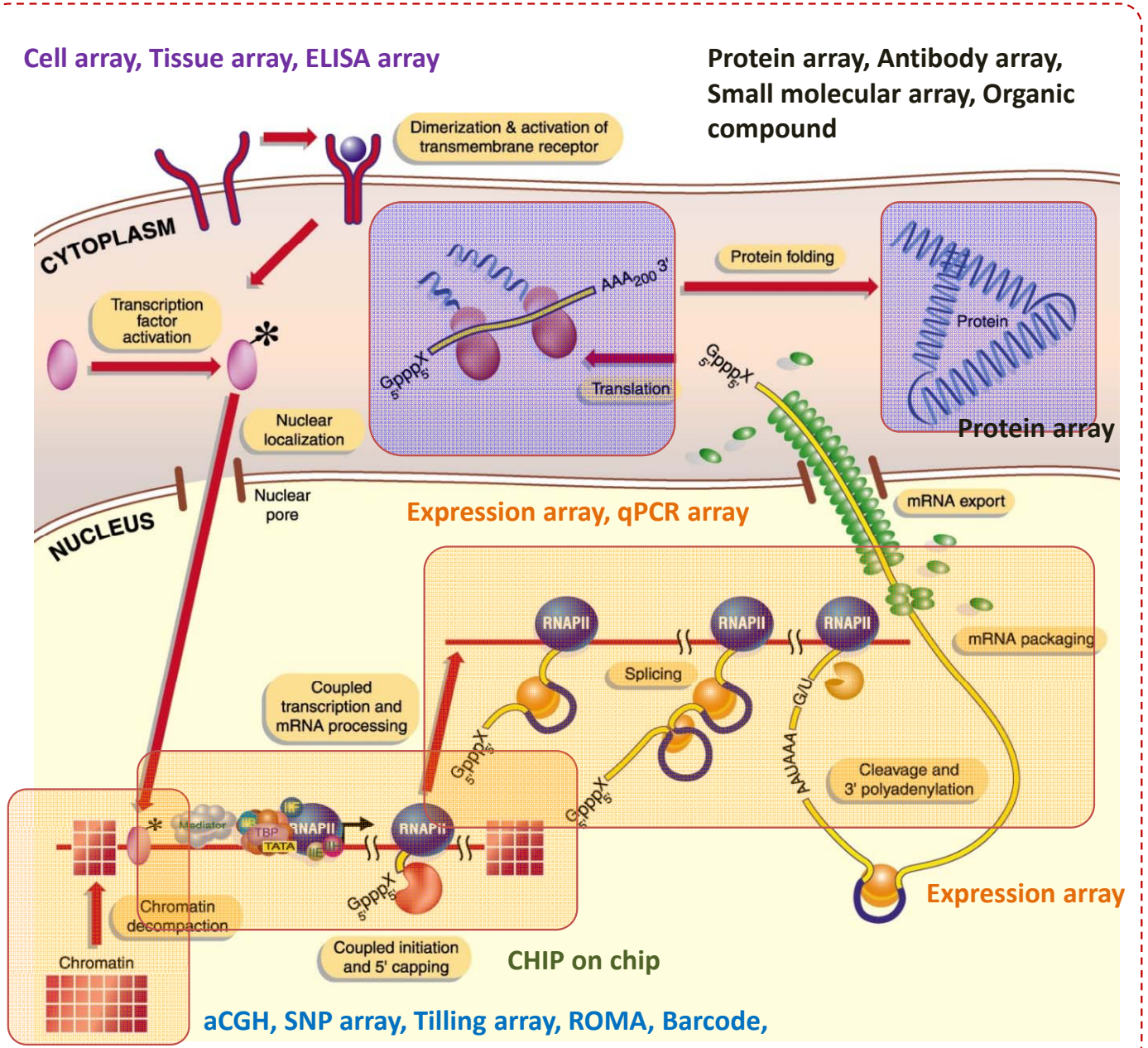


Image process and analyze



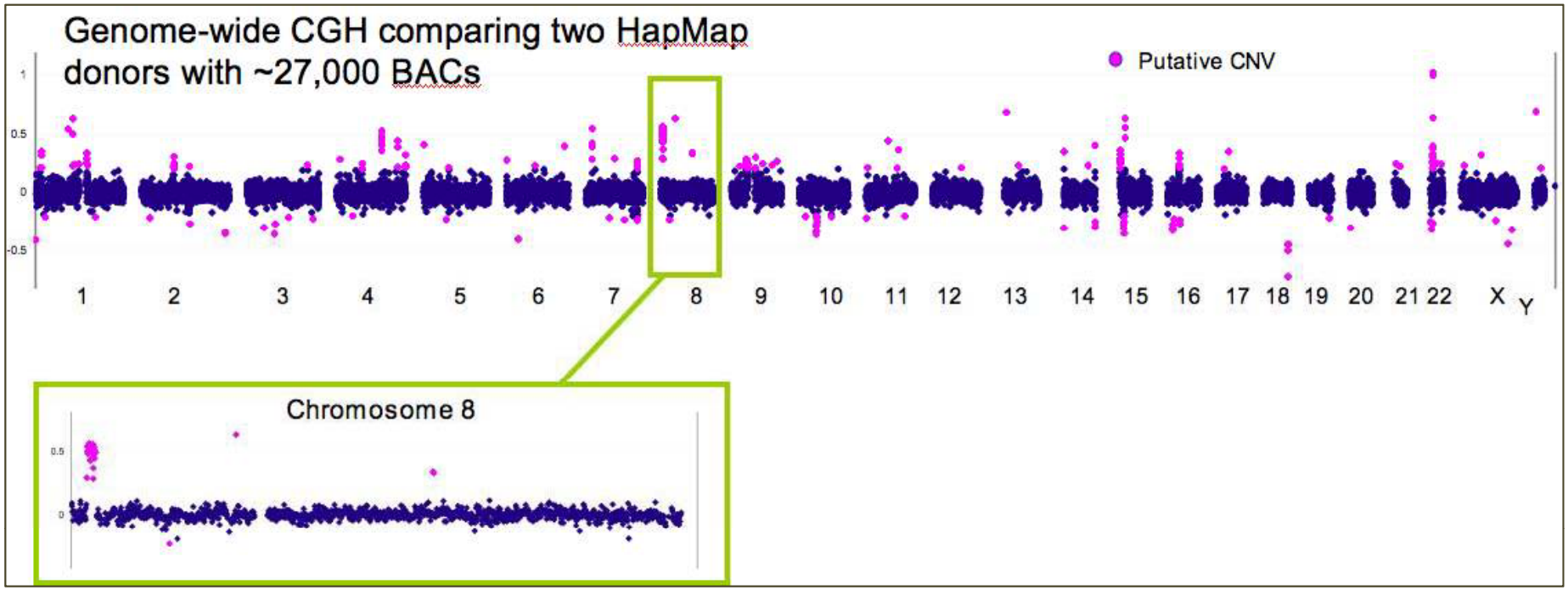
Central Dogma and Existing Microarrays Types

1. aCGH
2. SNP array
3. Tilling array
4. ROMA
5. Barcode array
6. cDNA array
7. qPCR array
8. RNA array
9. Protein array
10. CHIP on chip
11. Antibody array
12. Organic compound
13. Small molecule
14. ELISA array
15. Cell array
16. Tissue array
17. Lab-on-a-chip



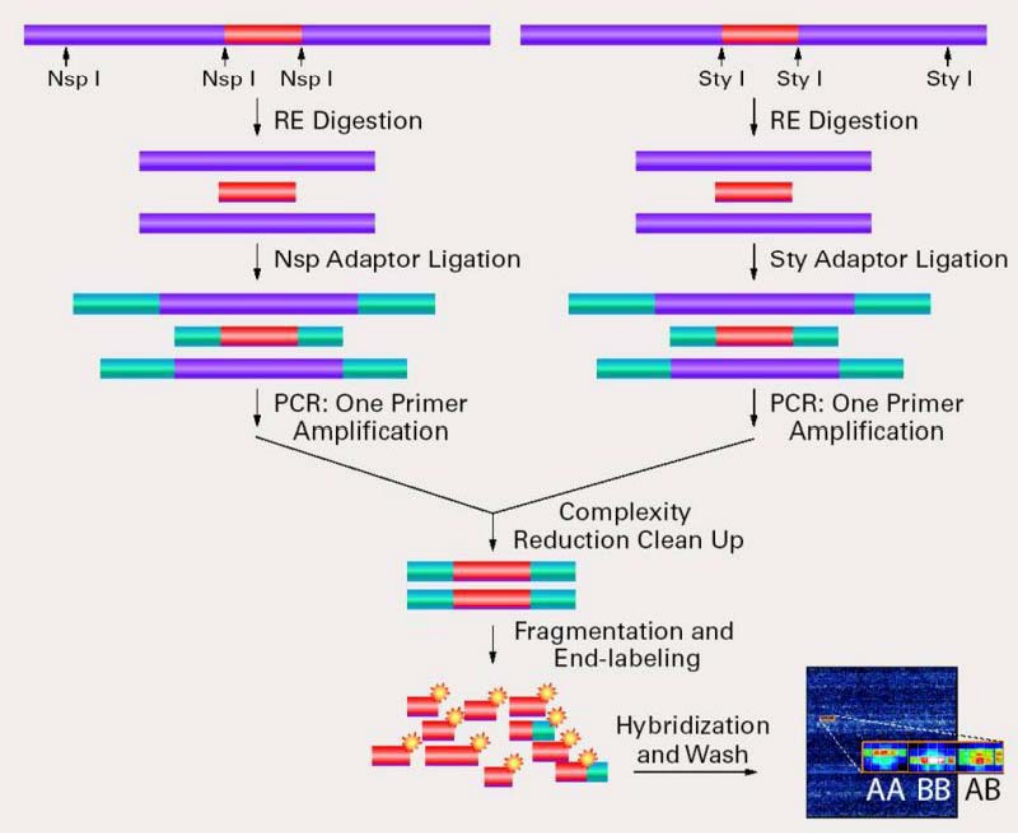
1. **aCGH**
2. SNP array
3. Tilling array
4. cDNA array
5. qPCR array
6. RNA array
7. Protein array
8. CHIP on chip
9. Antibody array
10. Organic compound
11. ELISA array
12. Cell array
13. Tissue array
14. Lab-on-a-chip

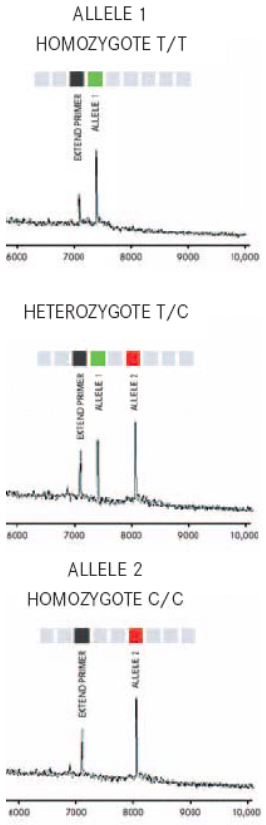
Array comparative genomic hybridization (also CMA, Chromosomal Microarray Analysis, Microarray-based comparative genomic hybridization, array CGH, a-CGH, aCGH, or virtual karyotype) genomic copy number variations at a higher resolution level than chromosome-based comparative genomic hybridization (CGH). This is a molecular-cytogenetic method for the analysis of copy number changes (gains/losses) in the DNA content of a given subject's DNA and often in tumor cells.



1. aCGH
2. **SNP array**
3. Tilling array
4. cDNA array
5. qPCR array
6. RNA array
7. Protein array
8. CHIP on chip
9. Antibody array
10. Organic compound
11. ELISA array
12. Cell array
13. Tissue array
14. Lab-on-a-chip

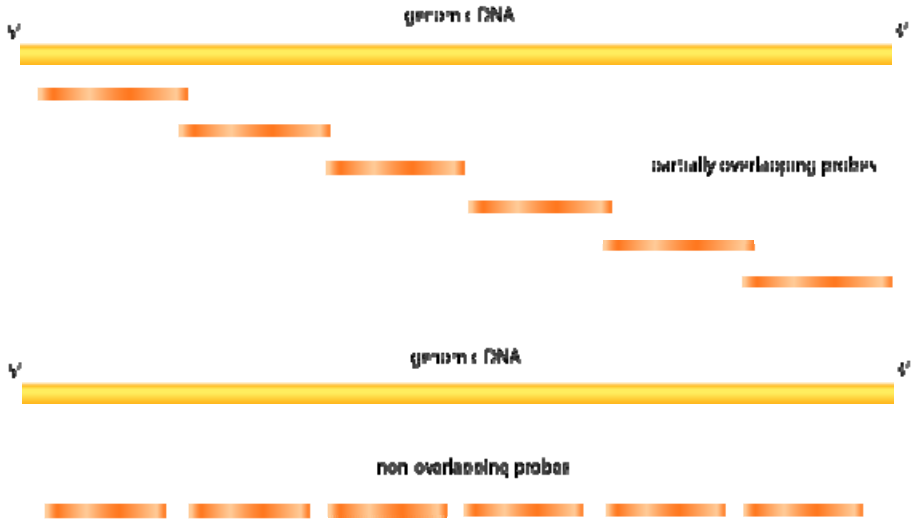
A **single nucleotide polymorphism** (**SNP**, pronounced snip) is a DNA sequence variation occurring when a single nucleotide — A, T, C, or G — in the genome (or other shared sequence) differs between members of a species (or between paired chromosomes in an individual). For example, two sequenced DNA fragments from different individuals, AAGCCTA to AAGCTTA, contain a difference in a single nucleotide. In molecular biology and bioinformatics, a SNP array is a type of DNA microarray which is used to **detect polymorphisms within a population**.

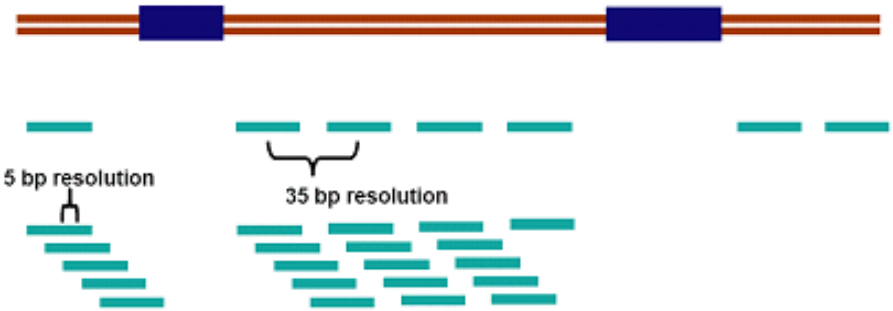




1. aCGH
2. SNP array
3. **Tilling array**
4. cDNA array
5. qPCR array
6. RNA array
7. Protein array
8. CHIP on chip
9. Antibody array
10. Organic compound
11. ELISA array
12. Cell array
13. Tissue array
14. Lab-on-a-chip

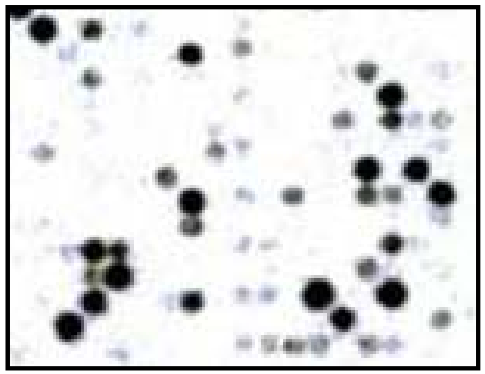
Tiling arrays differ in the nature of the probes. Short fragments are designed to **cover the entire genome or contigs of the genome**. Depending on the probe lengths and spacing different degrees of resolution can be achieved. Number of features on a single array can range from 10,000 to greater than 6,000,000, with each feature containing millions of copies of one probe. Tiling arrays can produce an unbiased look at gene expression because previously unidentified genes can still be incorporated. On top of individual gene expression analysis, other tiling arrays can be used in transcriptome mapping, ChIP-chip, MeDIP-chip and DNase Chip studies, Array CGH among others.



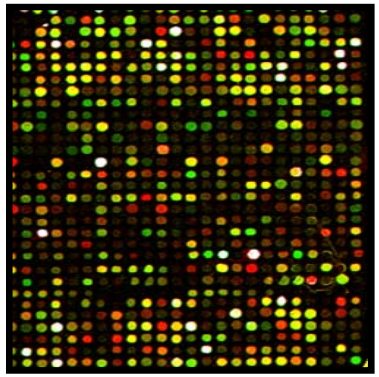


1. aCGH
2. SNP array
3. Tilling array
4. **cDNA array**
5. qPCR array
6. RNA array
7. Protein array
8. CHIP on chip
9. Antibody array
10. Organic compound
11. ELISA array
12. Cell array
13. Tissue array
14. Lab-on-a-chip

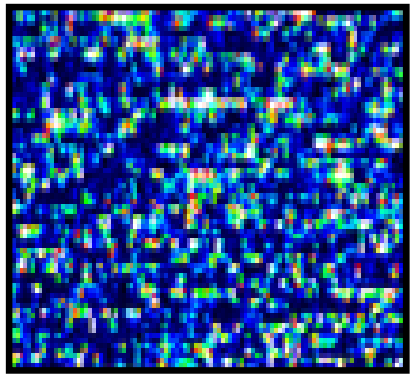
A **DNA microarray** (also commonly known as gene or genome chip, DNA chip, or gene array) is a collection of microscopic DNA spots, commonly representing single genes, arrayed on a solid surface by covalent attachment to a chemical matrix. DNA arrays are different from other types of microarray only in that they either **measure DNA or use DNA as part of its detection system**. Qualitative or quantitative measurements with DNA microarrays utilize the selective nature of DNA-DNA or DNA-RNA hybridization under high-stringency conditions and fluorophore-based detection. DNA arrays are commonly used for expression profiling, i.e., monitoring expression levels of thousands of genes simultaneously, or for comparative genomic hybridization.



Isotope
Nylon – cDNA (300-900 nt)



Two-colour
cDNA or Oligo (80 nt)
500 – 11,000 elements

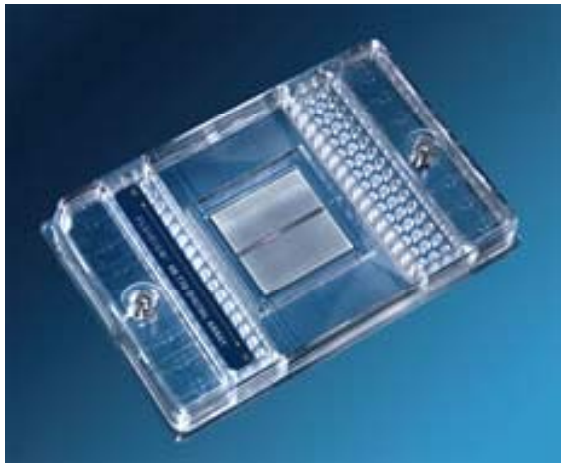


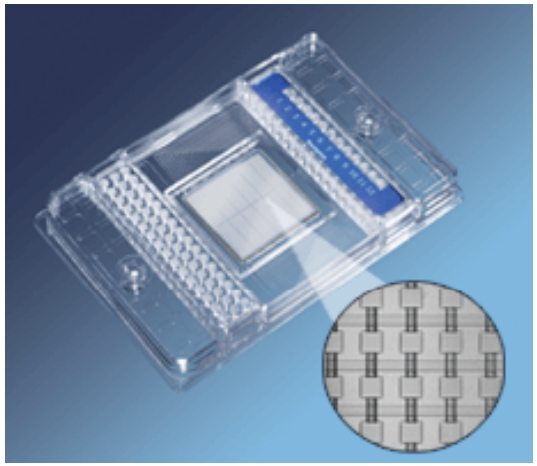
Affymetrix
Silicone – oligo (20 nt)
22 ,000 elements

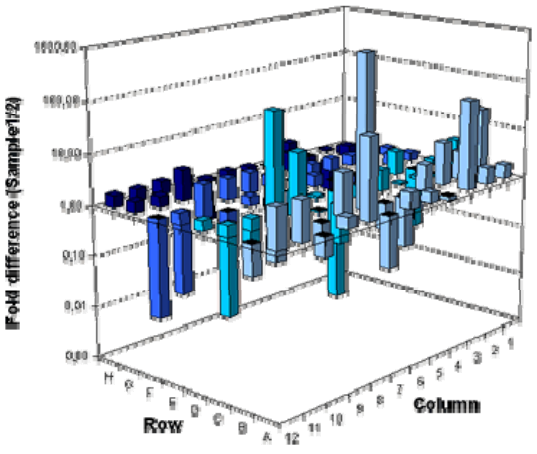
1. aCGH
2. SNP array
3. Tilling array
4. cDNA array
5. **qPCR array**
6. RNA array
7. Protein array
8. CHIP on chip
9. Antibody array
10. Organic compound
11. ELISA array
12. Cell array
13. Tissue array
14. Lab-on-a-chip

Tiling arrays

Quantitative polymerase chain reaction (qPCR) is a modification of the polymerase chain reaction used to rapidly measure the quantity of DNA, complementary DNA or ribonucleic acid present in a sample. Like other forms of polymerase chain reaction, the process is used to amplify DNA samples, via the temperature-mediated enzyme DNA polymerase. qPCR Arrays are designed to profile the expression of a panel of genes relevant to a specific pathway or disease state.

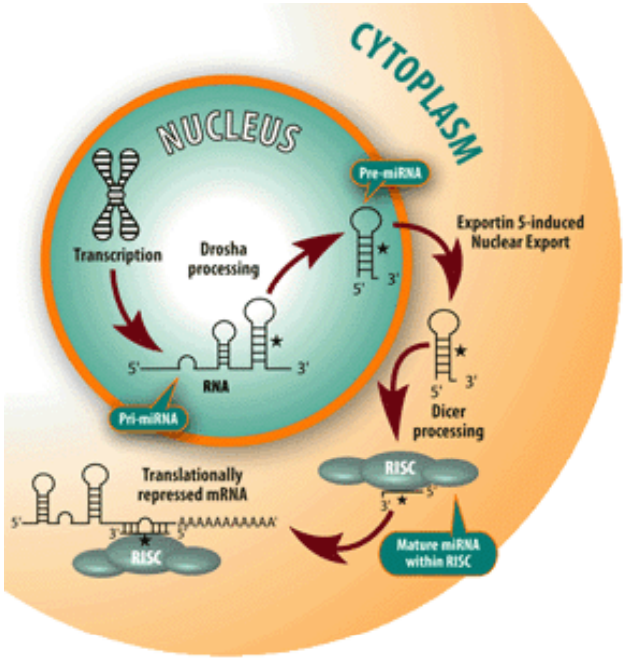


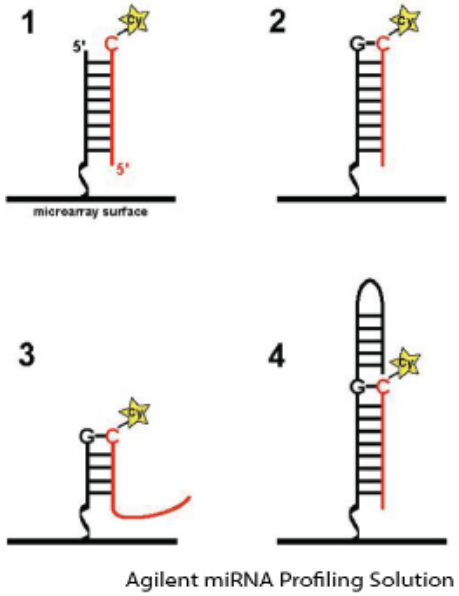




1. aCGH
2. SNP array
3. Tilling array
4. cDNA array
5. qPCR array
6. **RNA array**
7. Protein array
8. CHIP on chip
9. Antibody array
10. Organic compound
11. ELISA array
12. Cell array
13. Tissue array
14. Lab-on-a-chip

In genetics**, microRNAs** (miRNA) are single-stranded RNA molecules of 21-23 nucleotides in length, which regulate gene expression. miRNAs are encoded by genes from whose DNA they are transcribed but miRNAs are not translated into protein (non-coding RNA); instead each primary transcript (a pri-miRNA) is processed into a short stem-loop structure called a pre-miRNA and finally into a functional miRNA. Mature miRNA molecules are partially complementary to one or more messenger RNA (mRNA) molecules, and their main function is to down-regulate gene expression. MicroRNA array is a kind of arrays which is designed for the detection of **miRNA expression profile**.

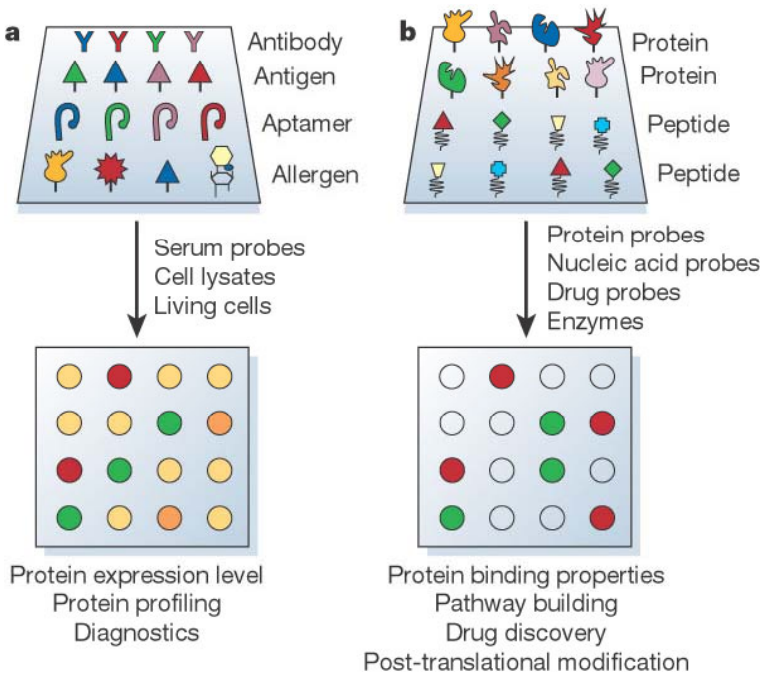




Agilent miRNA Profiling Solution

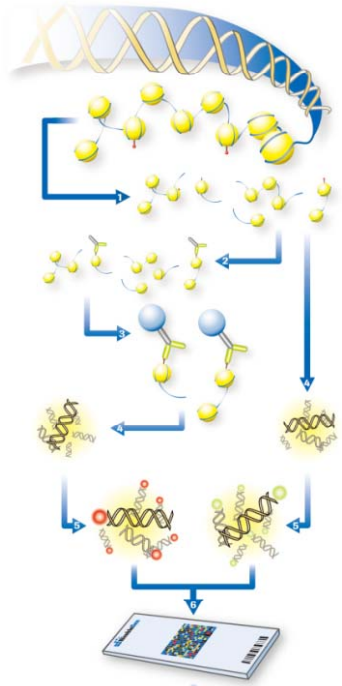
1. aCGH
2. SNP array
3. Tilling array
4. cDNA array
5. qPCR array
6. RNA array
7. **Protein array**
8. CHIP on chip
9. Antibody array
10. Organic compound
11. ELISA array
12. Cell array
13. Tissue array
14. Lab-on-a-chip

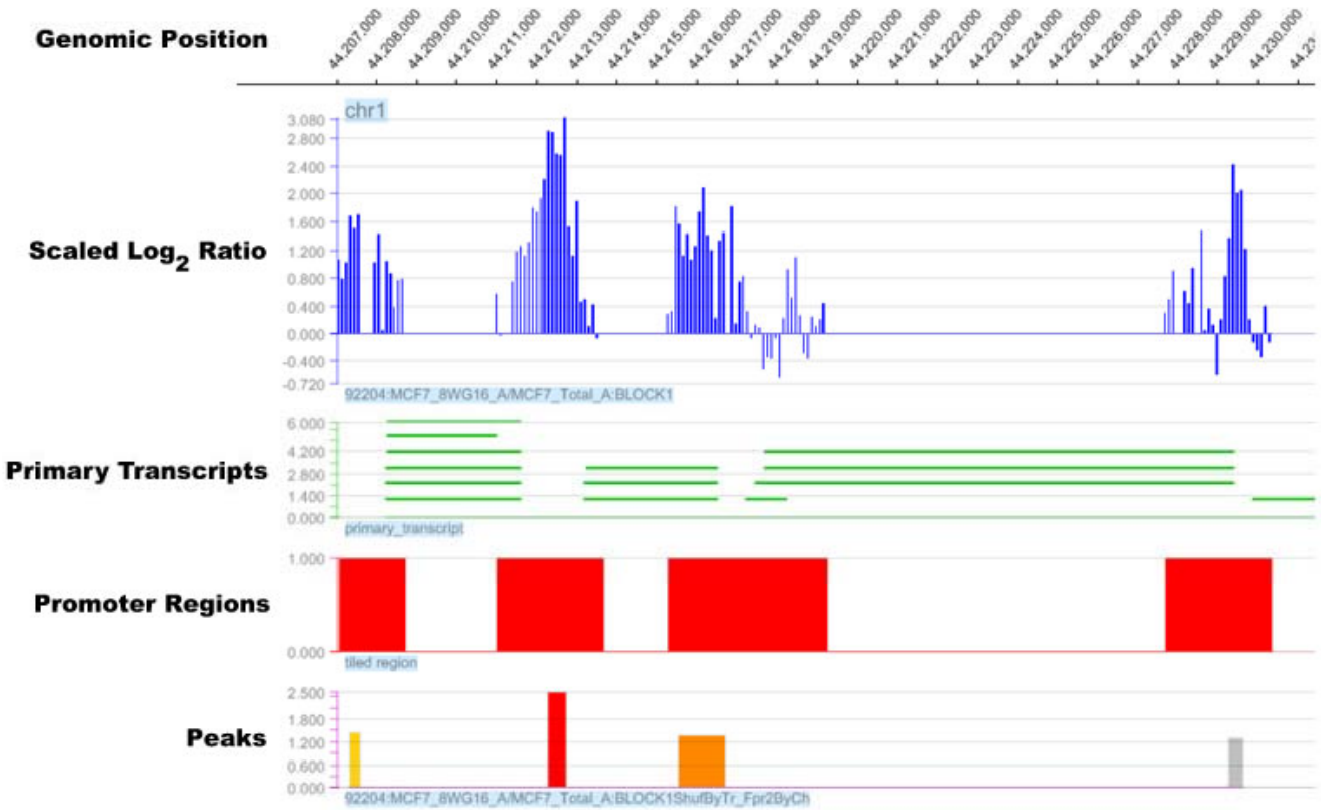
A **protein microarray** is a piece of glass on which different molecules of protein have been affixed at separate locations **in an ordered manner** thus forming a microscopic array. These are used to identify protein-protein interactions, to identify the substrates of protein kinases, or to identify the targets of biologically active small molecules. The most common protein microarray is the antibody microarray, where antibodies are spotted onto the protein chip and are used as capture molecules to detect proteins from cell lysate solutions. Related microarray technologies also include DNA microarrays, Antibody microarrays, Tissue microarrays and Chemical Compound Microarrays.



1. aCGH
2. SNP array
3. Tilling array
4. cDNA array
5. qPCR array
6. RNA array
7. Protein array
8. **CHIP on chip**
9. Antibody array
10. Organic compound
11. ELISA array
12. Cell array
13. Tissue array
14. Lab-on-a-chip

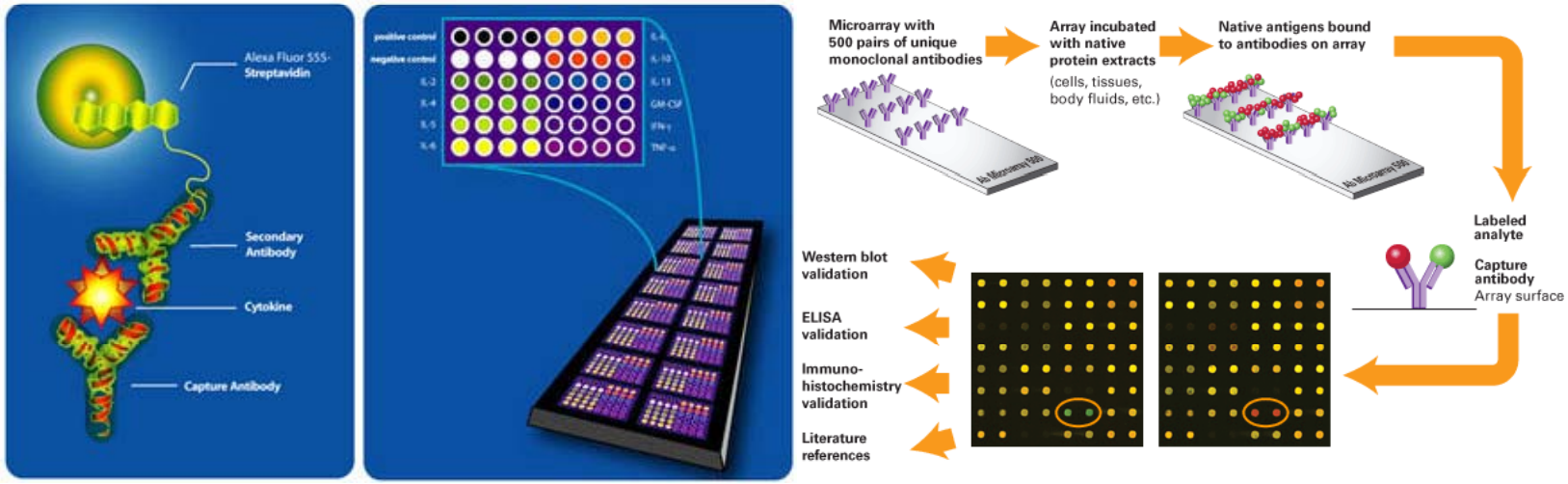
ChIP-on-chip (also known as ChIP-chip) is a technique that combines chromatin immunoprecipitation ("ChIP") with microarray technology ("chip"). Like regular ChIP, ChIP-on-chip is used to investigate **interactions between proteins and DNA** *in vivo*. Specifically, it allows the identification of binding sites of DNA-binding proteins on a genome-wide basis. One of the long-term goals ChIP-on-chip was designed for is to establish a catalogue of (selected) organisms that lists all protein-DNA interactions under various physiological conditions.





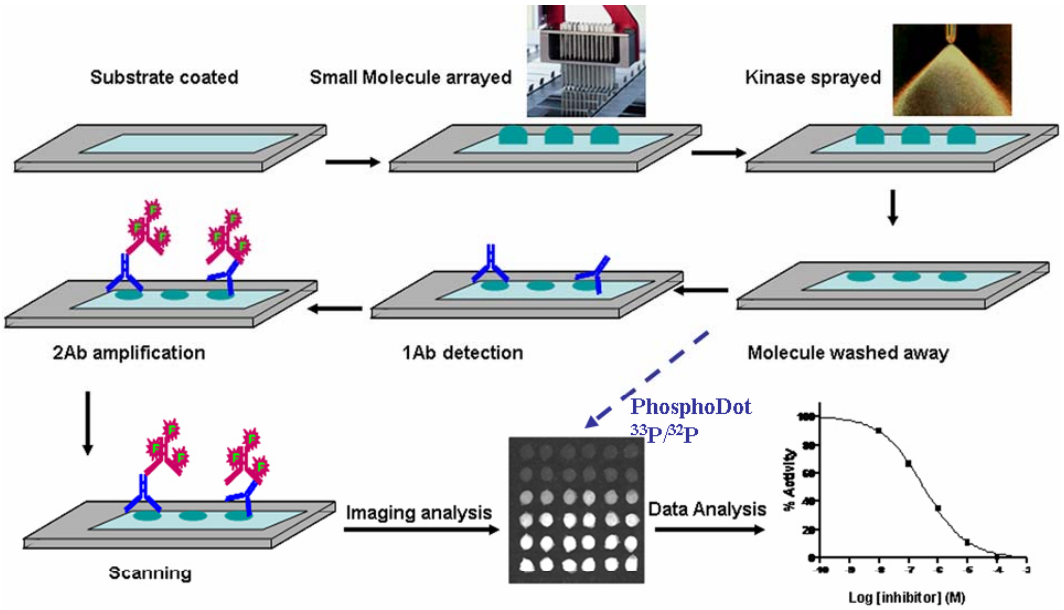
1. aCGH
2. SNP array
3. Tilling array
4. cDNA array
5. qPCR array
6. RNA array
7. Protein array
8. CHIP on chip
9. **aAntibody**
10. Organic compound
11. ELISA array
12. Cell array
13. Tissue array
14. Lab-on-a-chip

An **antibody microarray** is a specific form of protein microarrays, a collection of capture antibodies are spotted and fixed on a solid surface, such as glass, plastic and silicon chip for the purpose of detecting antigens. Antibody microarray is often used for **detecting protein expressions** from cell lysates in general research and special biomarkers from serum or urine for diagnostic applications.



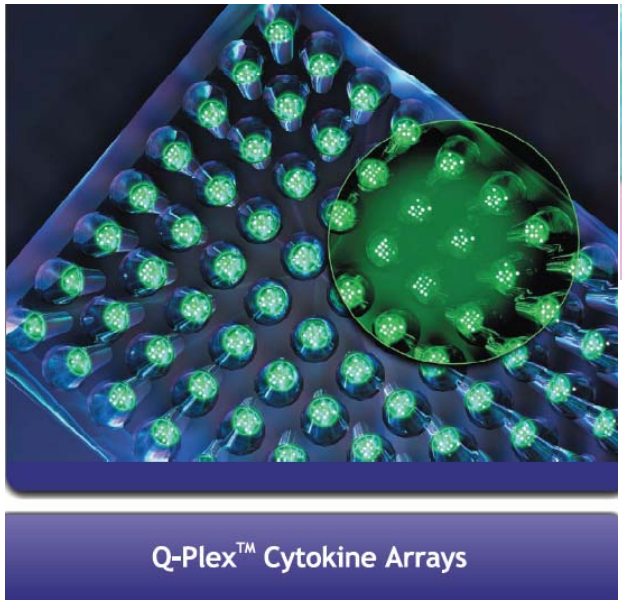
1. aCGH
2. SNP array
3. Tilling array
4. cDNA array
5. qPCR array
6. RNA array
7. Protein array
8. CHIP on chip
9. Antibody array
10. **aOrganic cpd**
11. ELISA array
12. Cell array
13. Tissue array
14. Lab-on-a-chip

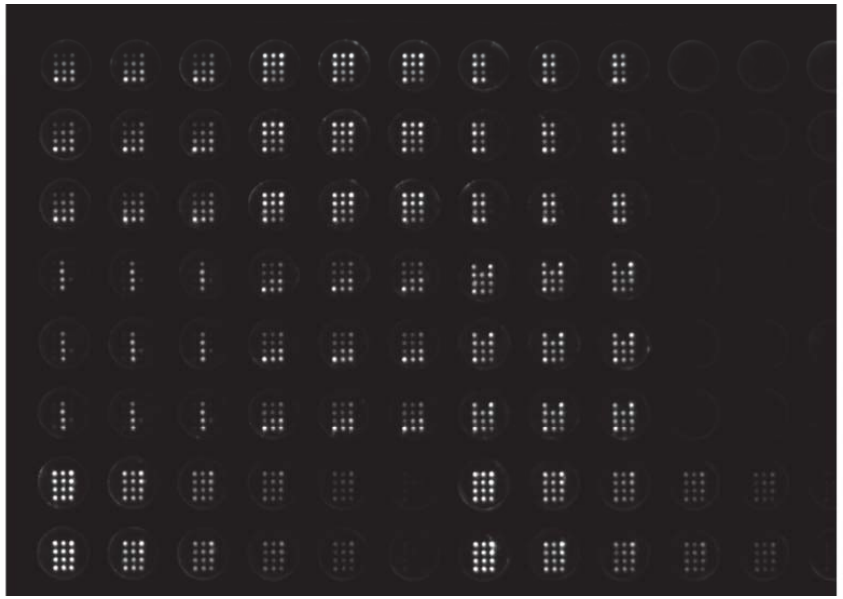
A **chemical compound microarray** is a collection of organic chemical compounds spotted on a solid surface, such as glass and plastic. In chemical genetics research, they are routinely used for searching proteins that bind with specific chemical compounds, and in general drug discovery research, they provide a multiplex way to search potential drugs for therapeutic targets. There are three different forms of chemical compound microarrays based on the fabrication method, **to covalently immobilize** (usually called Small Molecule Microarray), **to spot and dry** organic compounds on the solid surface without immobilization (commercial name as Micro Arrayed Compound Screening (μARCS)), and to spot organic compounds **in a homogenous solution** without immobilization and drying effect.



1. aCGH
2. SNP array
3. Tilling array
4. cDNA array
5. qPCR array
6. RNA array
7. Protein array
8. CHIP on chip
9. Antibody array
10. Organic compound
11. **ELISA array**
12. Cell array
13. Tissue array
14. Lab-on-a-chip

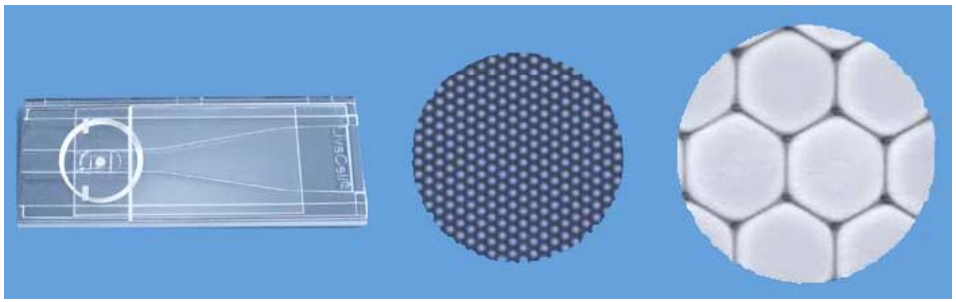
Enzyme-Linked ImmunoSorbent Assay, or ELISA, is a biochemical technique used mainly in immunology to detect the presence of an antibody or an antigen in a sample. The ELISA has been used as a diagnostic tool in medicine and plant pathology, as well as a quality control check in various industries. ELISA array (or array ELISA) is a new technology capable of simultaneously identifying multiple biomarkers to generate a biochemical profile. Once the ELISA protocol is completed the array is imaged using either a CCD imaging system or X-ray film to capture the chemiluminescent signal. The pixel intensity of each spot is directly correlated to the concentration.

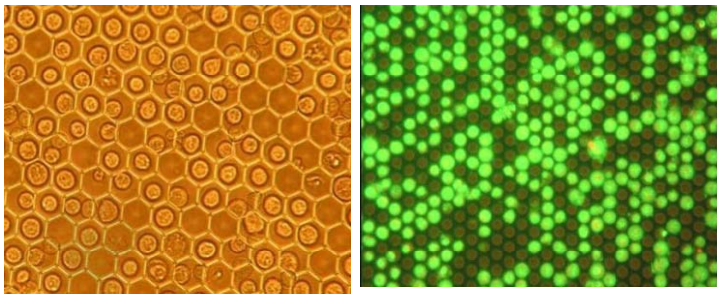




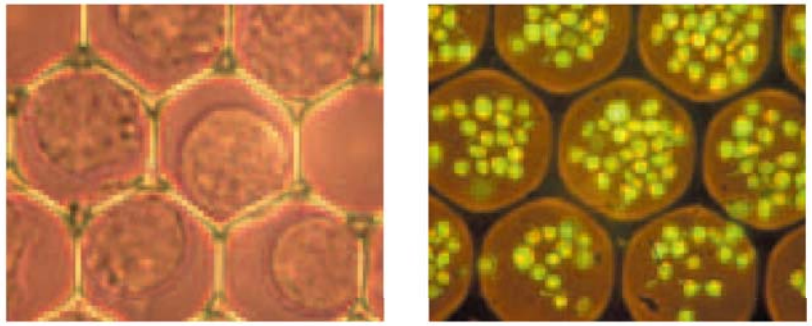
1. aCGH
2. SNP array
3. Tilling array
4. cDNA array
5. qPCR array
6. RNA array
7. Protein array
8. CHIP on chip
9. Antibody array
10. Organic compound
11. ELISA array
12. **Cell array**
13. Tissue array
14. Lab-on-a-chip

Cell array is a common term for different techniques, which are used in genomic-level cell biological testing. The live cell array, a microscope slide-based high content analysis tool, enables **multi-parametric imaging-based assays on thousands of intact individual cells**, including non-adhering blood and bone marrow cells. Cells can be observed in their own identified location, tracking individual, real-time responses to intervention. Multiple functional assays can be performed on a living cell, followed by post-fixation studies on the same cell to maximize usage of cell samples





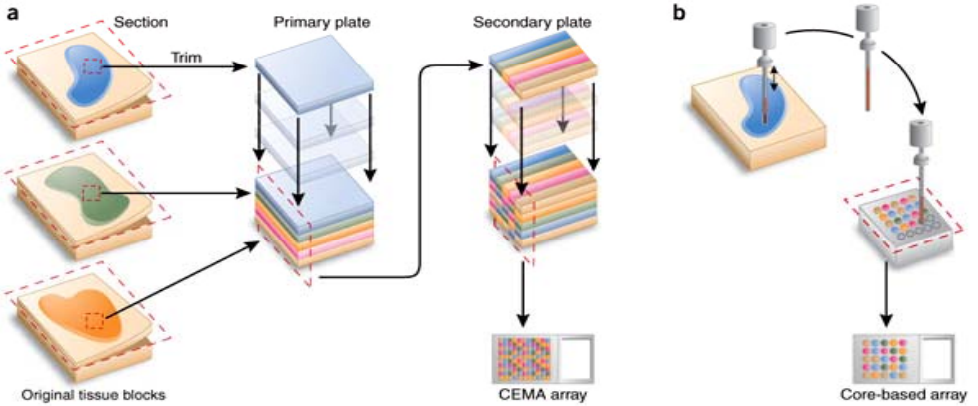
LIVECELL ARRAY microscope slide

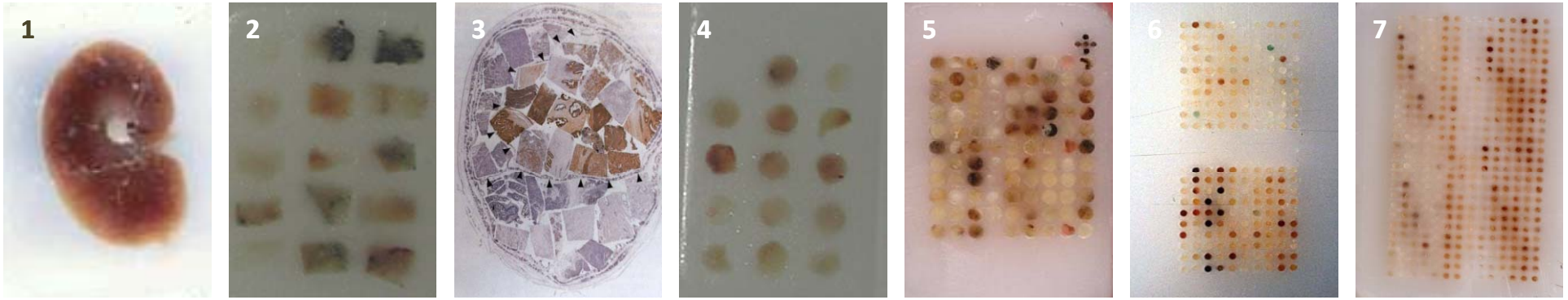


The Individual Cell Array with 15 (left) and 100-micron.

1. aCGH
2. SNP array
3. **Tilling array**
4. cDNA array
5. qPCR array
6. RNA array
7. Protein array
8. CHIP on chip
9. Antibody array
10. Organic compound
11. ELISA array
12. Cell array
13. **Tissue array**
14. Lab-on-a-chip

Tissue microarrays are produced by a method of **re-locating tissue** from conventional histologic paraffin blocks so that tissue from multiple patients or blocks can be seen on the same slide. This is done by using a needle to biopsy a standard histologic sections and placing the core into an array on a recipient paraffin block.



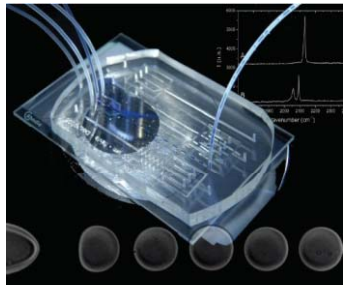


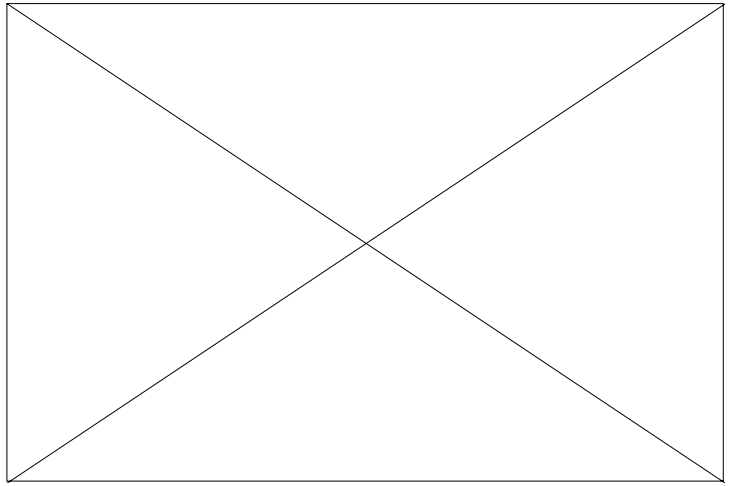
whole tissue | **tissue cocktail** | **tissue sausage** | **tissue array**

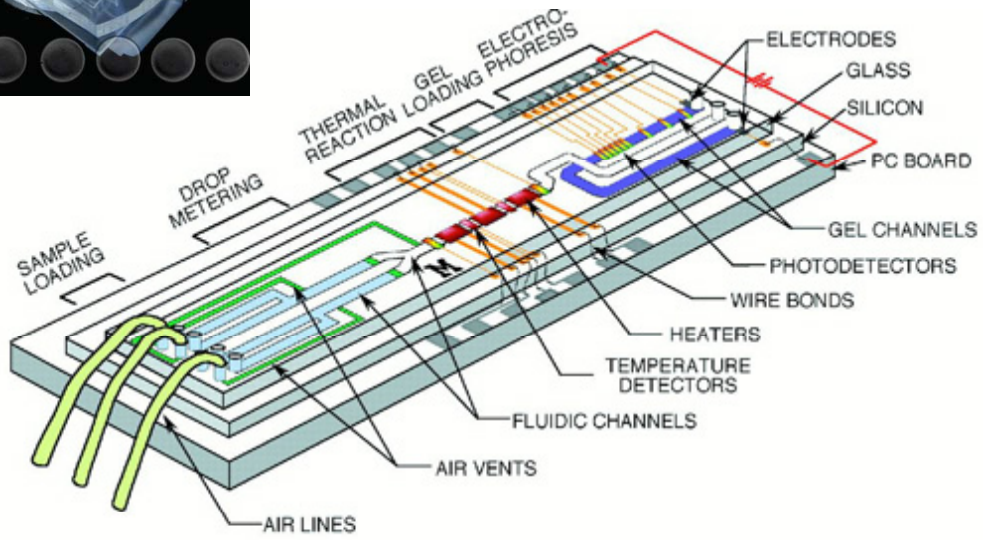
1. aCGH
2. SNP array
3. **Tilling array**
4. cDNA array
5. qPCR array
6. RNA array
7. Protein array
8. CHIP on chip
9. Antibody array
10. Organic compound
11. ELISA array
12. Cell array
13. Tissue array

14. Lab-on-a-chip

Lab-on-a-chip (LOC) is a term for devices that **integrate (multiple) laboratory functions on a single chip** of only millimeters to a few square centimeters in size and that are capable of handling extremely small fluid volumes down to less than pico liters.



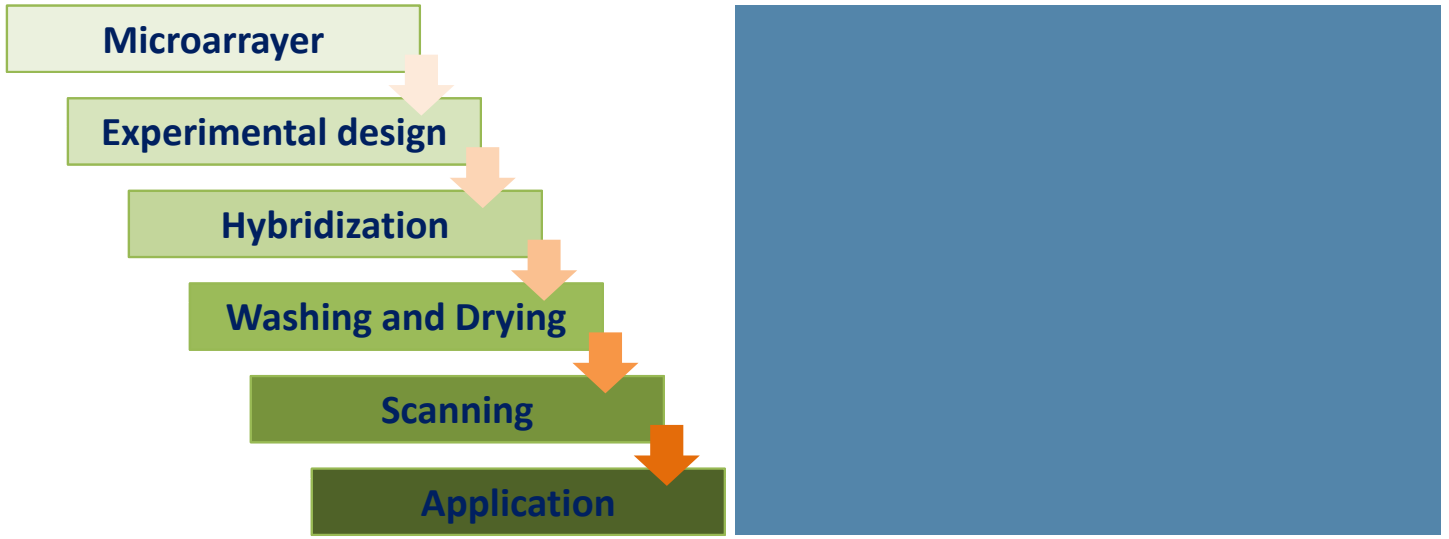





make a choice,
which one you like it and start on it

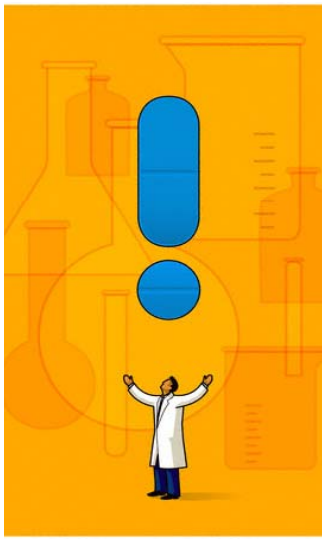
The 6 steps of a microarray experiment

1. Manufacturing of the microarray
2. Experimental design and choice of reference: what to compare to what?
3. Target preparation (labeling) and hybridization
4. Image acquisition (scanning) and quantification (signal intensity to numbers)
5. Database building, filtering and normalization
6. Statistical analysis and data mining



Experimental methodology

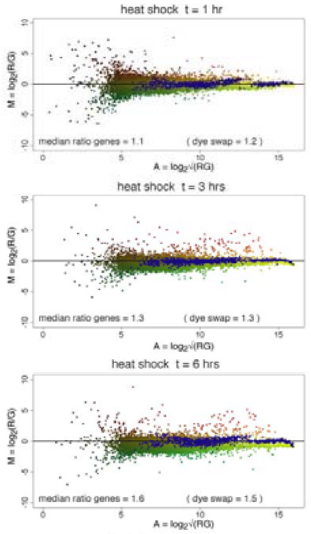
1 **Experimental Design**



2 **Microarray Experiments**

3 **Image Processing**

4 **Data Normalization**

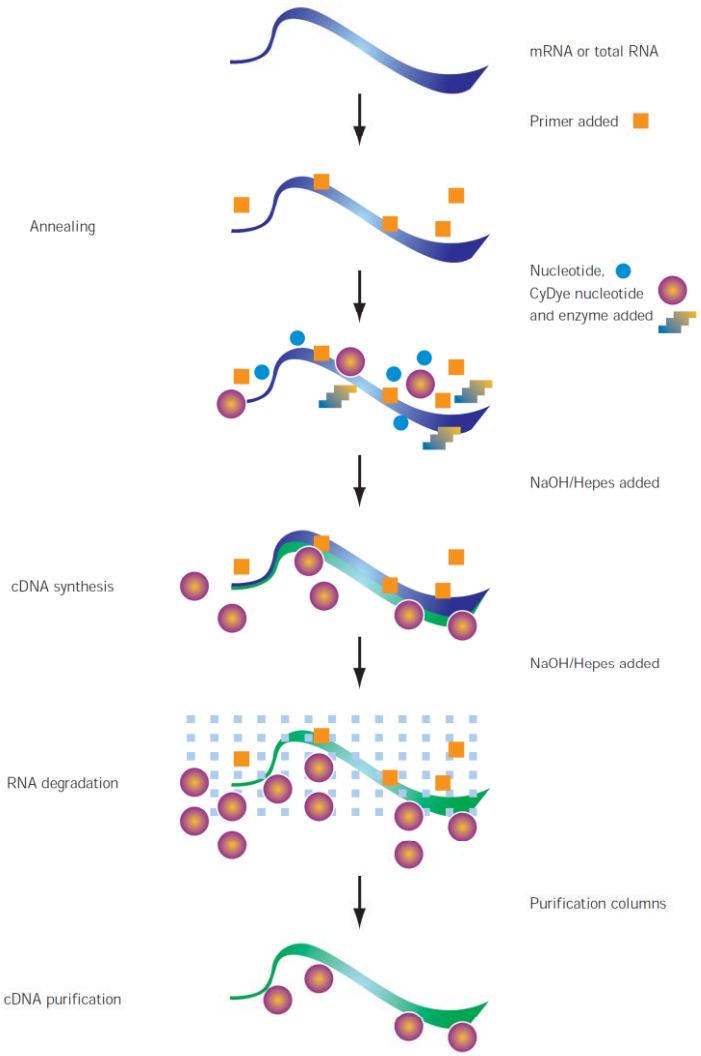


5 **Data Analysis**

1. Manufacturing of the microarray
2. Experimental design and choice of reference: what to compare to what?

3. Target preparation (labeling) and hybridization

4. Image acquisition (scanning) and quantification (signal intensity to numbers)
5. Database building, filtering and normalization
6. Statistical analysis and data mining



❶

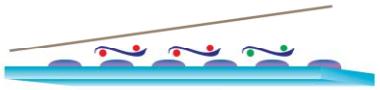
Attachment

Genes of interest are spotted onto a solid surface by the array spotter. These are known as the targets. Attachment chemistry will often be required to ensure that the DNA remains attached to the slide surface throughout the hybridization process.

❷

Hybridization

Hybridization buffer containing a known amount of labelled sample DNA—often referred to as probe— is then placed on the slide surface. A coverslip can then be carefully placed on top of the slide.



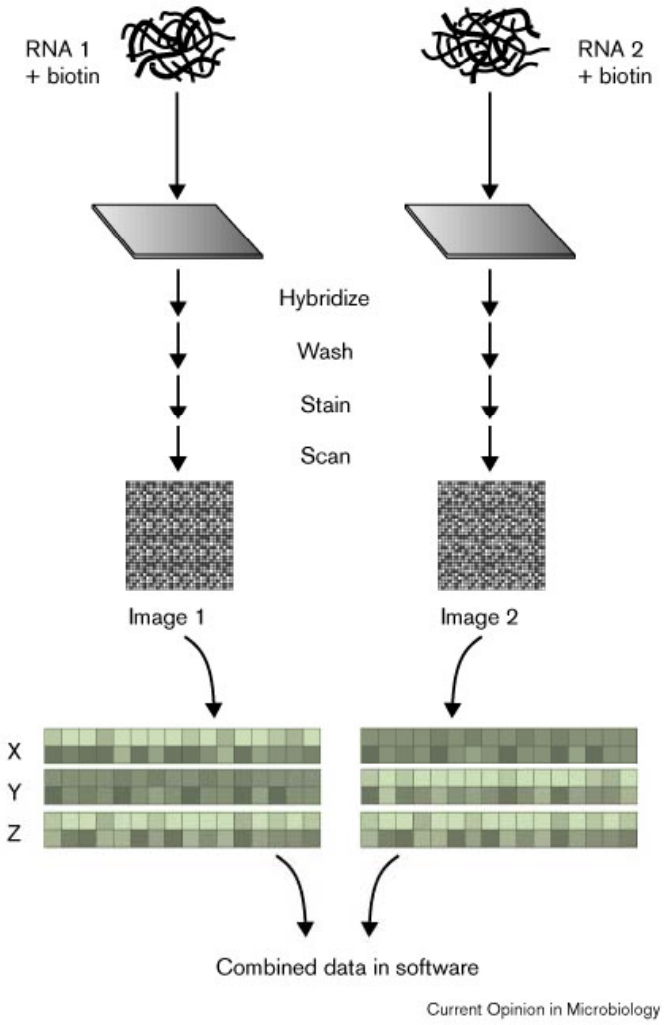
The slide is then incubated in a humid environment for up to 16 h. During this time the labelled probe is in contact with the targets on the slide. If the sequence homology is good then the probe will adhere to the target.

❸

Washing

Once the hybridization is complete, the slides are washed, and buffer and probe of little or no homology to the target will be washed away, leaving the labelled probe of high homology attached to the target and available for detection.

1. Manufacturing of the microarray
2. Experimental design and choice of reference: what to compare to what?
3. Target preparation (labeling) and hybridization
4. **Image acquisition (scanning) and quantification (signal intensity to numbers)**
5. Database building, filtering and normalization
6. Statistical analysis and data mining



Samples

Platform

Raw Data *(image, 16bit .tif file)*

(1) image acquisition
(2) spot location

Preprocessed Data *(.gal file)*

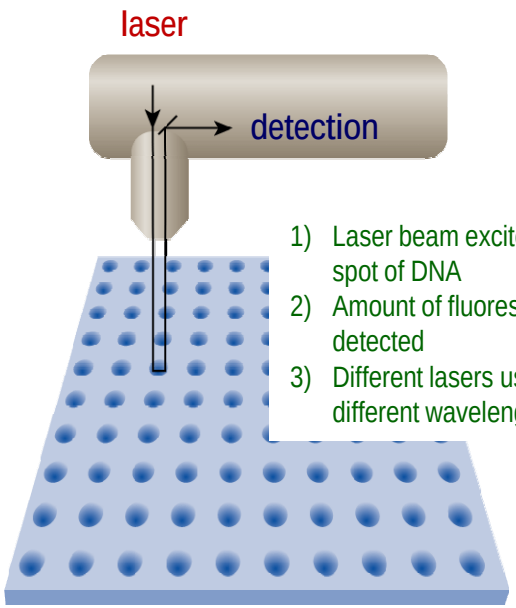
(3) computation of spot intensities

Data Analysis (different levels)

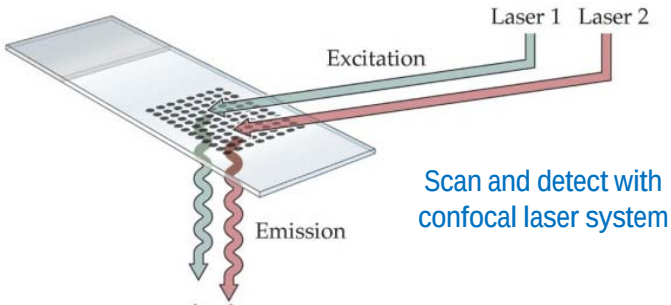
(4) data reporting

Image Processing

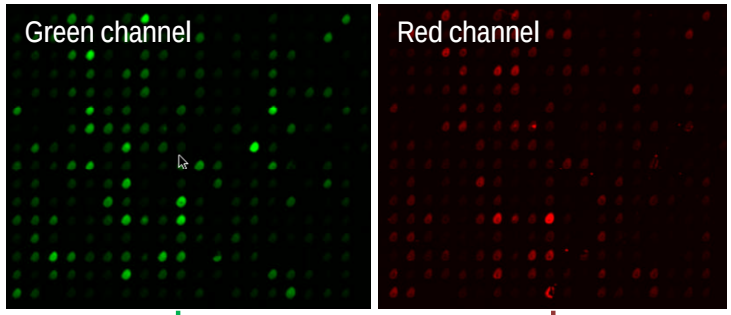
- Resolution
  - standard 10µm (100,000 atoms wide)
  - 100µm spot on chip = 10 pixels in diameter
- Image format
  - TIFF 16 bit (64K grey levels)
  - 1cm x 1cm image at 16 bit = 2Mb (uncompress
- Separate image for each fluorescent sample
  - channel 1, channel 2



1) Laser beam excites each spot of DNA
2) Amount of fluorescence detected
3) Different lasers used for different wavelengths



Scan and detect with confocal laser system



Overlay images and normalize

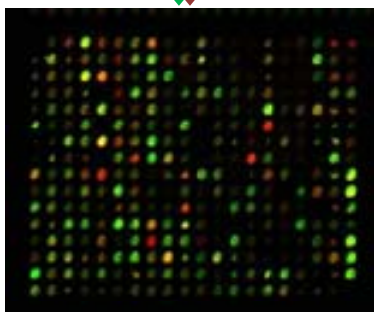
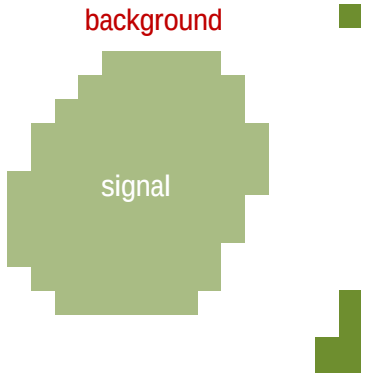
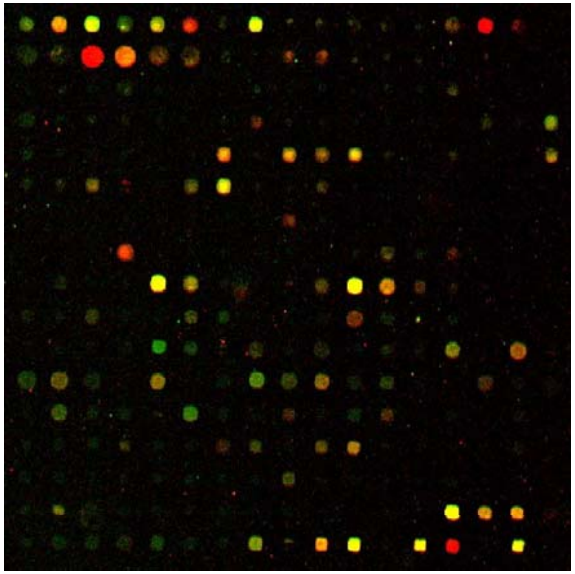


Image process and analyze

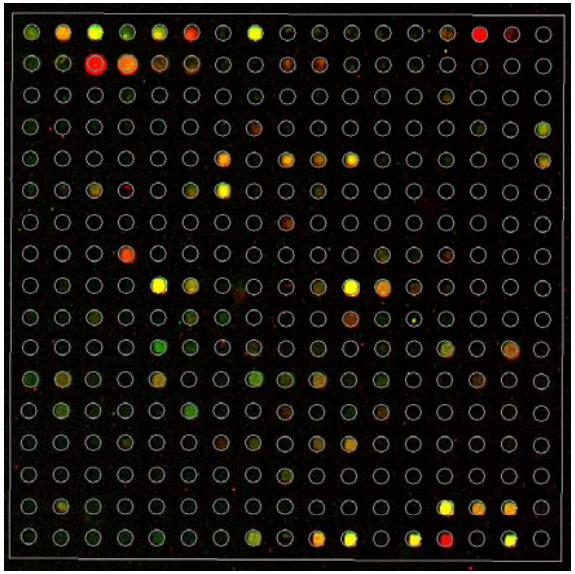
Image Processing

- **Addressing or gridding**
  - locate centers, assigning coordinates to each of the spots
- **Segmentation or spot picking**
  - classifying pixels either as foreground or as background
- **Intensity extraction** (for each spot)
  - Foreground fluorescence intensity pairs (R, G)
  - Background intensities
  - Quality measures
- **Information extraction**
  - for each spot of the array, calculates signal intensity pairs, background and quality measures

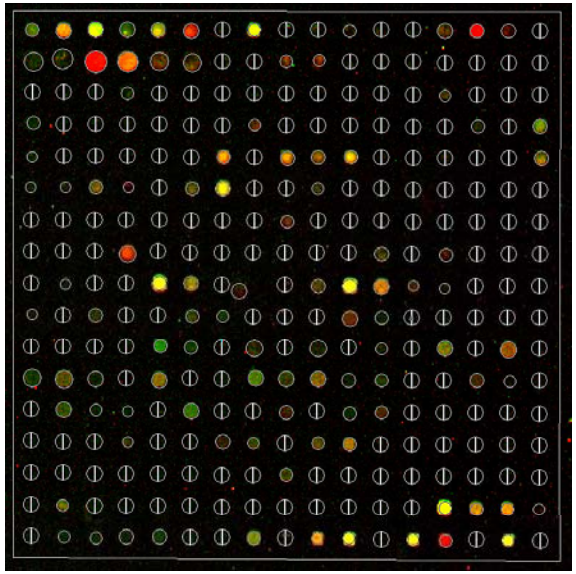




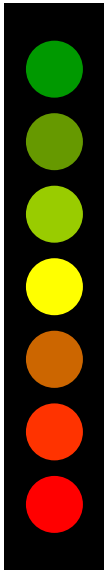
Raw (combined) image



Gridded



Spots picked & flagged

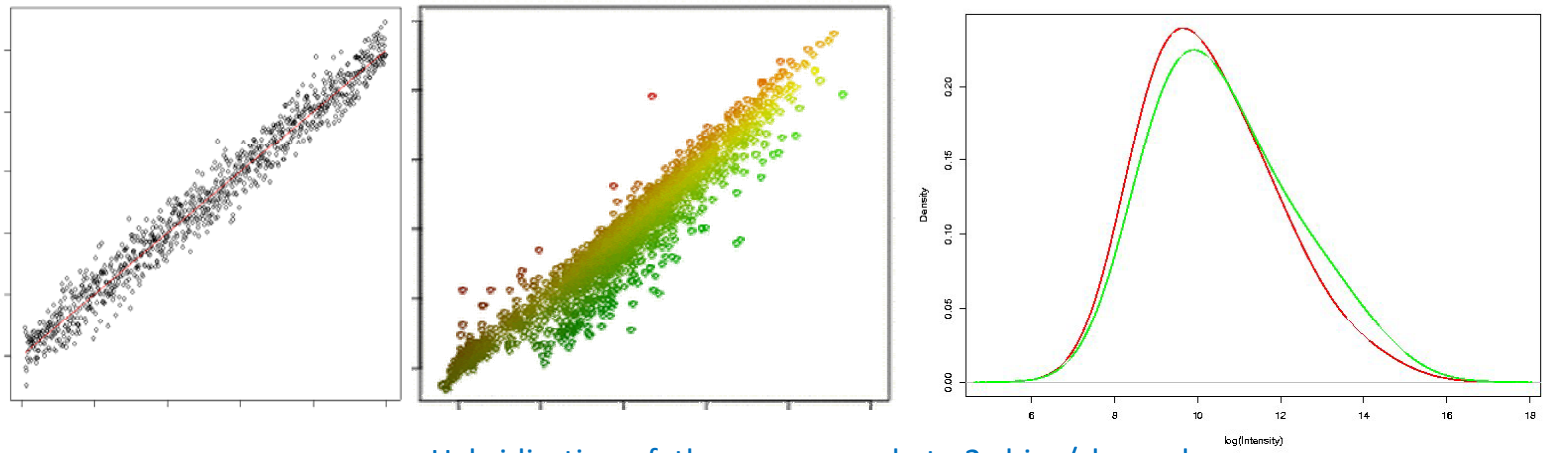


Intensity

Sources of Systematic Errors (*bias*)

- Different incorporation efficiency of dyes
- Different amounts of mRNA
- Experimenter/protocol issues (comparing chips processed by different labs)
- Different scanning parameters
- Batch bias

- Ideally: scatter plot coincides with the x=y diagonal
  - Due to Random errors: we expect to see a 'cloud' around the x=y diagonal.
- In practice: Both Random and Systematic measurement errors (Bias)
  - Due to Biases scatter plots are not centered around the x-y diagonal

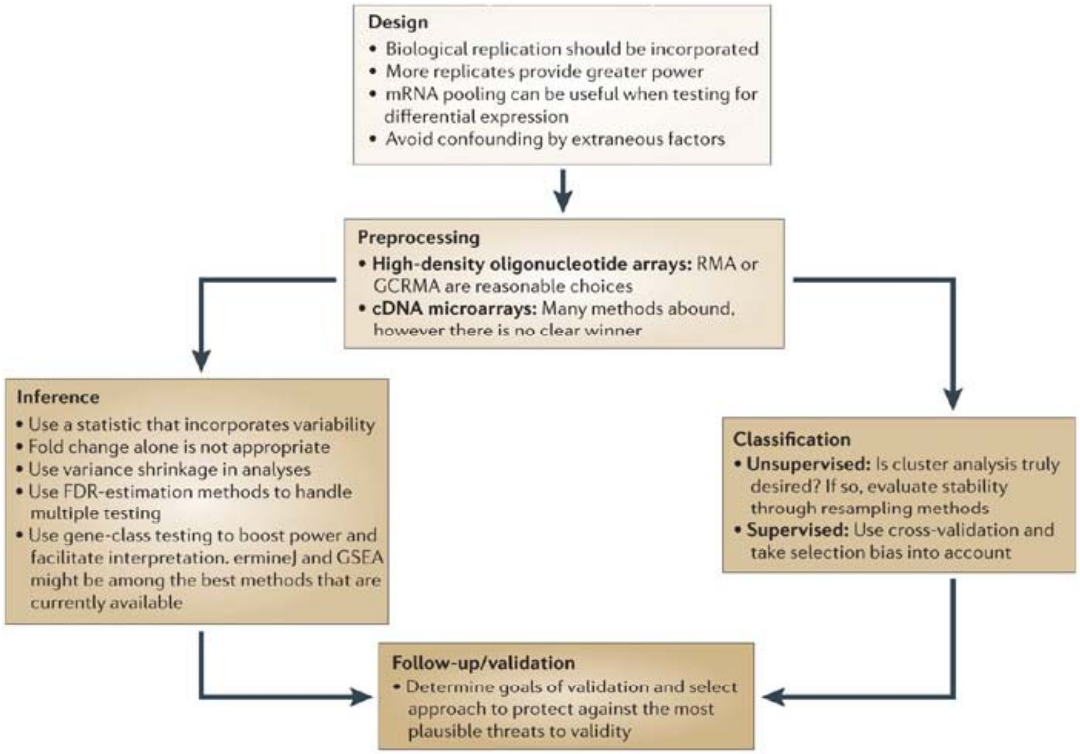


Hybridization of the same sample to 2 chips/channels

1. Manufacturing of the microarray
2. Experimental design and choice of reference: what to compare to what?
3. Target preparation (labeling) and hybridization
4. Image acquisition (scanning) and quantification (signal intensity to numbers)
5. Database building, filtering and normalization

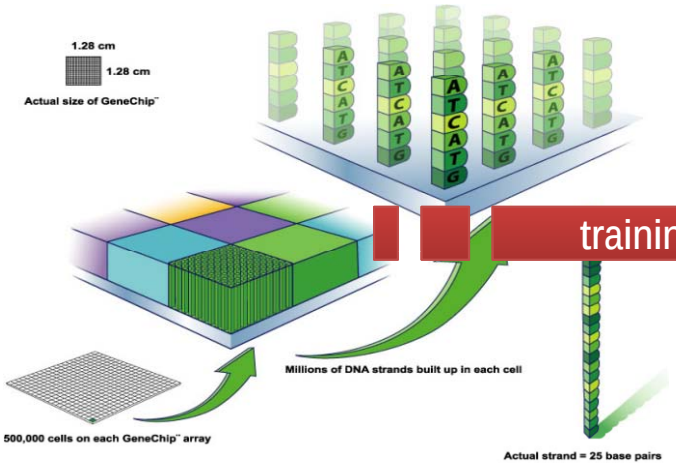
6. Statistical analysis and data mining

Guidelines for the statistical analysis of microarray experiments.

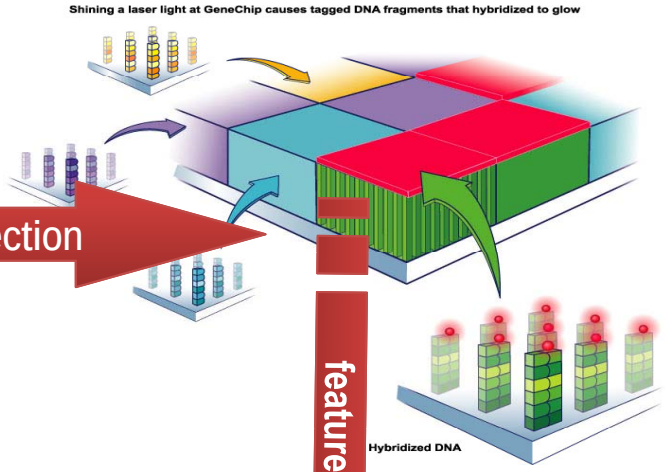


Allison DB *et al.* (2005) Microarray data analysis: from disarray to consolidation and consensus
Nat Rev gene. 7: 55–65 doi:10.1038/nri1749

Single-Test Platform of Microarray & Knowledge Discovery



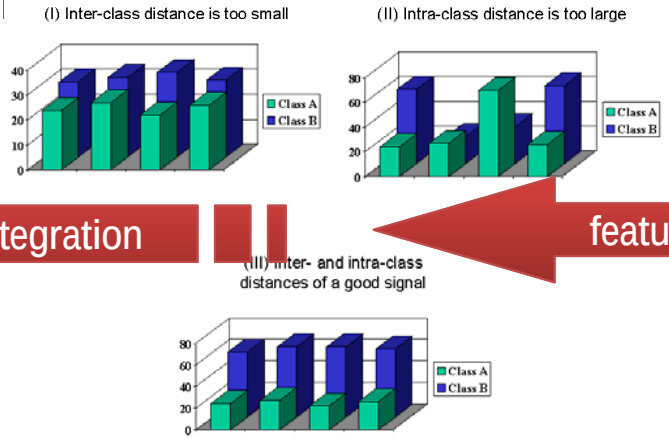
training data collection



feature generation

	00-0586-U	00-0586-U	00	U	00-0586-U	00-0586-U	Descriptions
	Positive	Negative			Avg Diff	Abs Call	
AFFX-Murl	5	2			297.5	A	M16762 Mouse int
AFFX-Murl	3	2			554.2	A	M37897 Mouse int
AFFX-Murl	4	2		9	308.6	A	M25892 Mus musc
AFFX-Murf	1	3		19	141	A	M83649 Mus musc
AFFX-BioE	13	1		19	9340.6	P	J04423 E coli bioB
AFFX-BioE	15	0		19	12862.4	P	J04423 E coli bioB
		0		19	8716.5	P	J04423 E coli bioB
		0		19	25942.5	P	J04423 E coli bioC
AFFX-BioC	16	0		20	28838.5	P	J04423 E coli bioC
AFFX-BioC	17	0		19	25765.2	P	J04423 E coli bioD
AFFX-BioC	19	0		20	140113.2	P	J04423 E coli bioD
AFFX-CreX	20	0		20	280036.6	P	X03453 Bacterioph
AFFX-CreX	20	0		20	401741.8	P	X03453 Bacterioph
AFFX-BioE	7	5		18	-483	A	J04423 E coli bioB
AFFX-BioE	5	4		18	313.7	A	J04423 E coli bioB
AFFX-BioE	7	6		20	-1016.2	A	J04423 E coli bioB

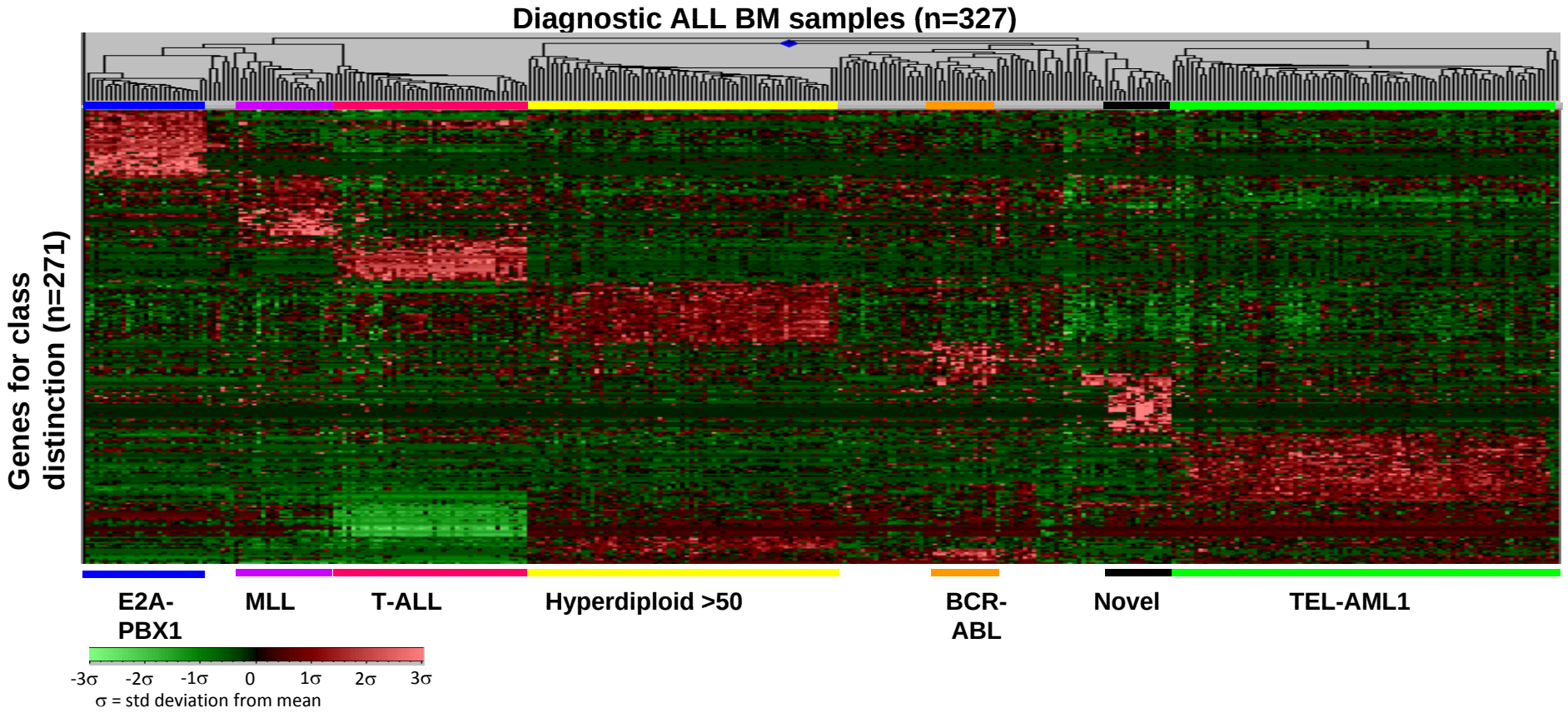
feature selection



feature integration

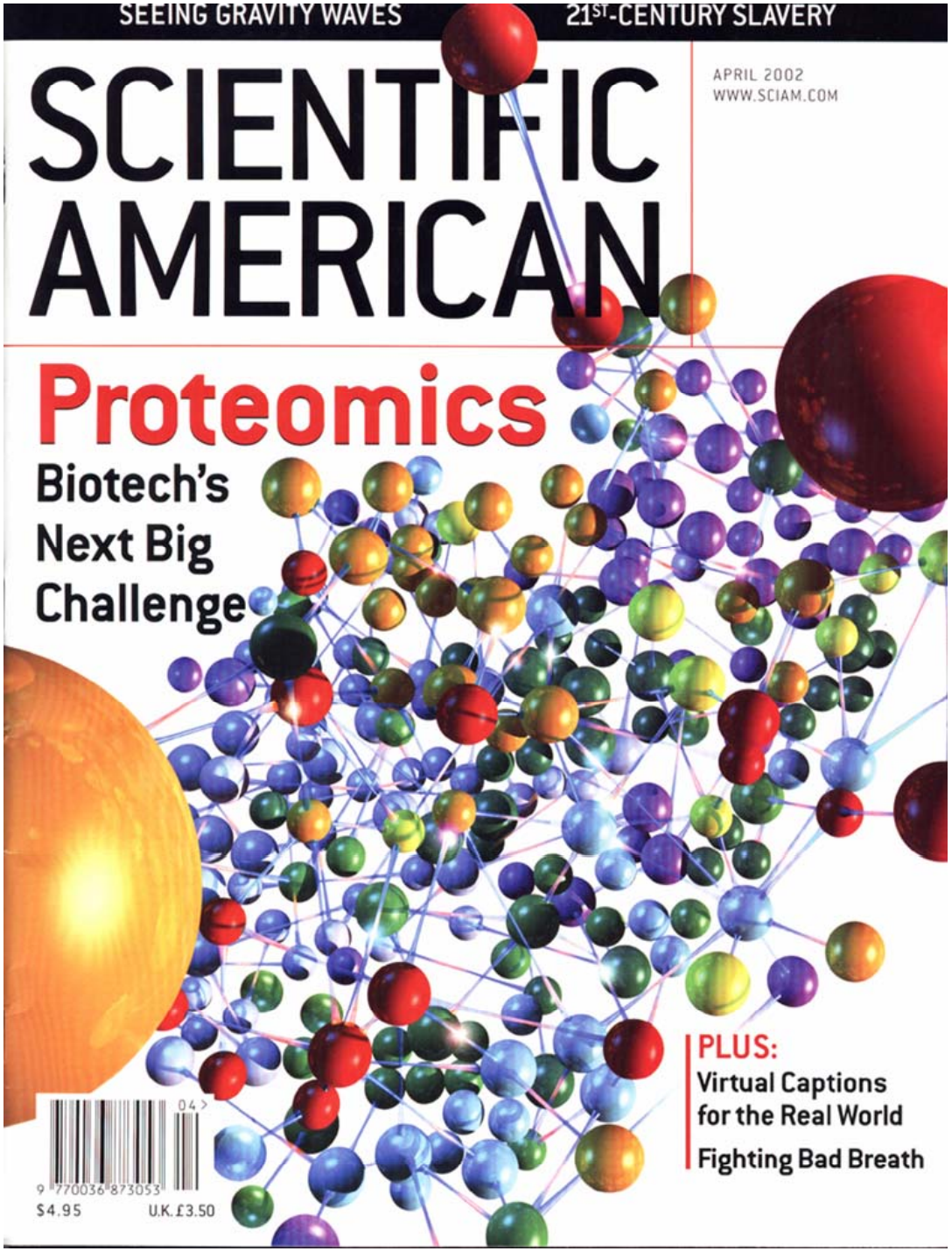
Higher Level Microarray data analysis

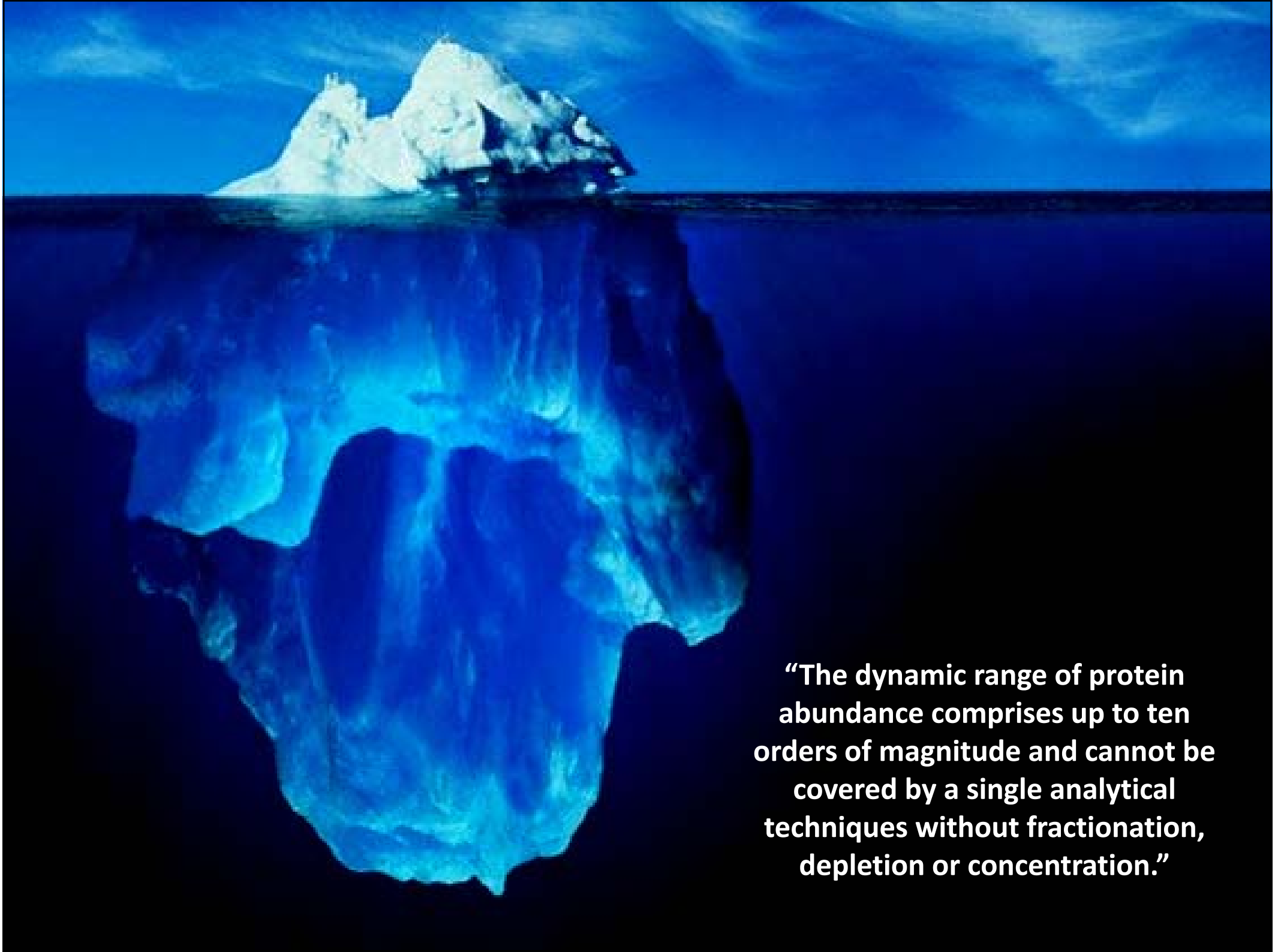
- Clustering and pattern detection
- Data mining and visualization
- Controls and normalization of results
- Statistical validatation
- Linkage between gene expression data and gene sequence/function/metabolic pathways databases
- Discovery of common sequences in co-regulated genes
- Meta-studies using data from multiple experiments



Proteomics

Proteomics and Its Application




“The dynamic range of protein abundance comprises up to ten orders of magnitude and cannot be covered by a single analytical techniques without fractionation, depletion or concentration.”

The Central Dogma of Life

SEMINAR

Seminar

Proteomics: new perspectives, new biomedical opportunities

Rosamonde E Banks, Michael J Dunn, Denis F Hochstrasser, Jean-Charles Sanchez, Walter Blackstock, Darryl J Pappin, Peter J Selby

Proteomics-based approaches, which examine the expressed proteins of a tissue or cell type, complement the genome initiatives and are increasingly being used to address biomedical questions. Proteins are the main functional output, and the genetic code cannot always indicate which proteins are expressed, in what quantity, and in what form. For example, post-translational modifications of proteins, such as phosphorylation or glycosylation, are very important in determining protein function. Similarly, the effects of environmental factors or multigenic processes such as ageing or disease cannot be assessed simply by examination of the genome alone. This review describes the underlying technology and illustrates several areas of biomedical research, ranging from pathogenesis of neurological disorders to drug and vaccine design, in which potential clinical applications are being explored.

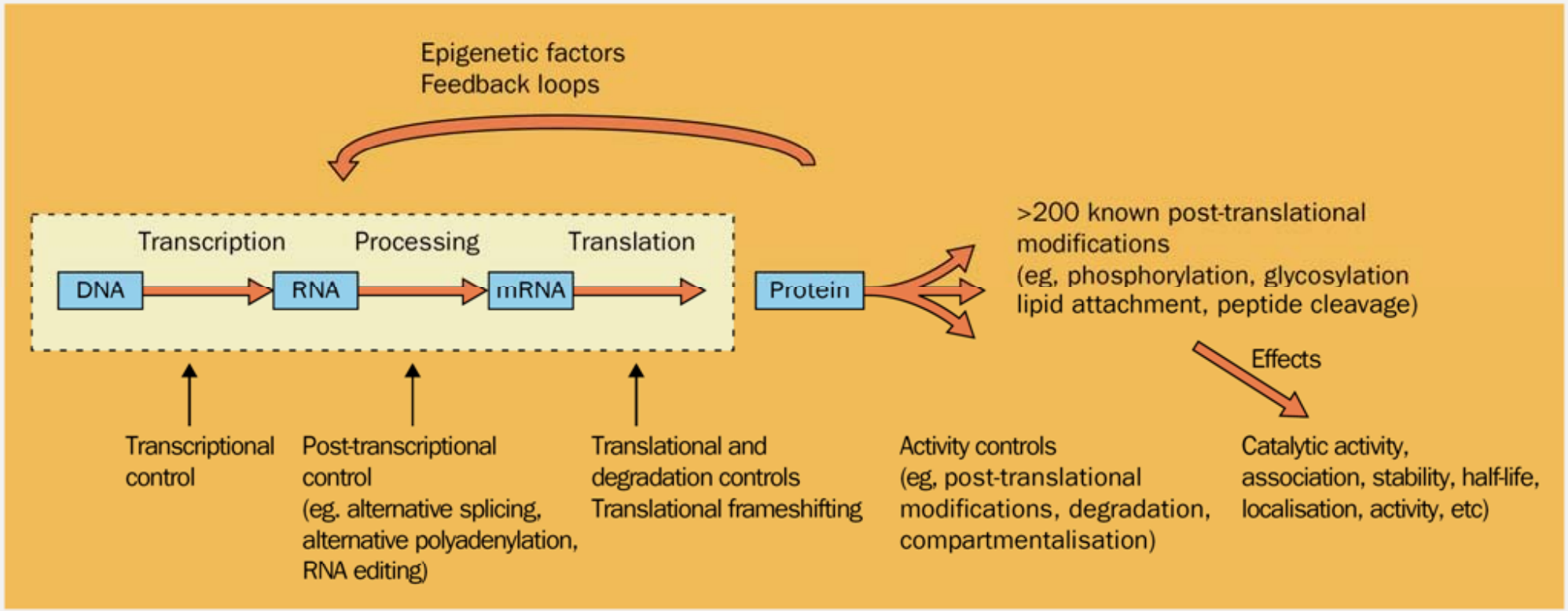
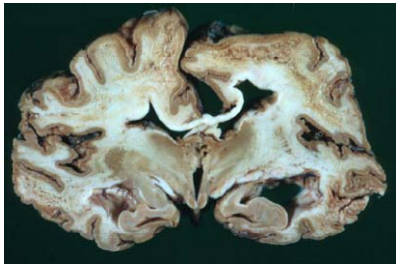
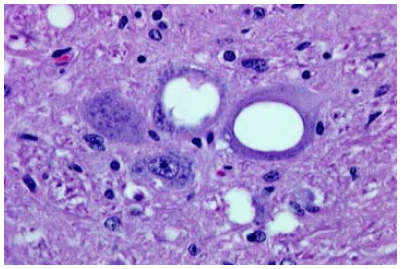


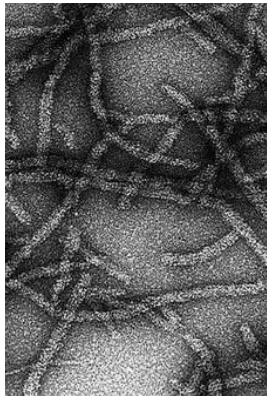
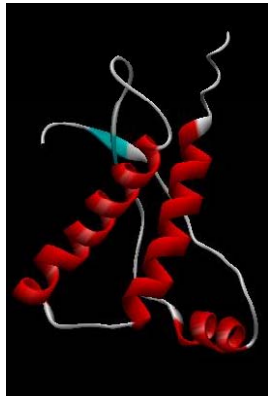
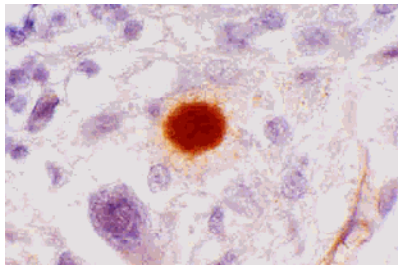
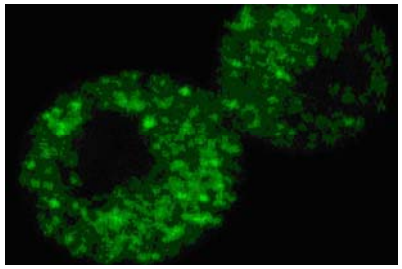
Figure 1: **The ways in which gene and protein expression can be regulated or modified from transcription to post translation**

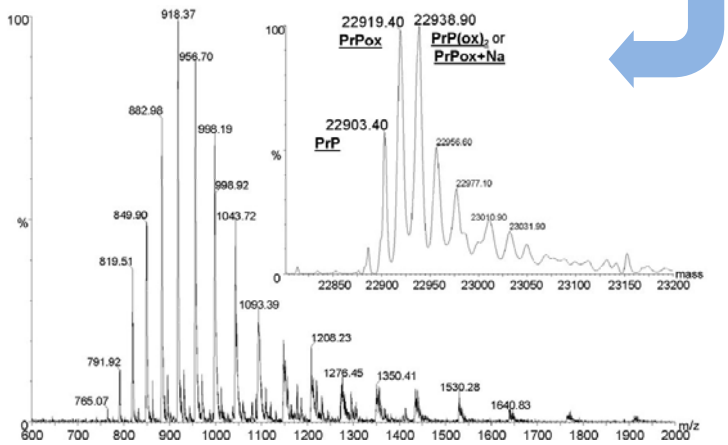
The Lancet (2000) 356:1749.

Prion







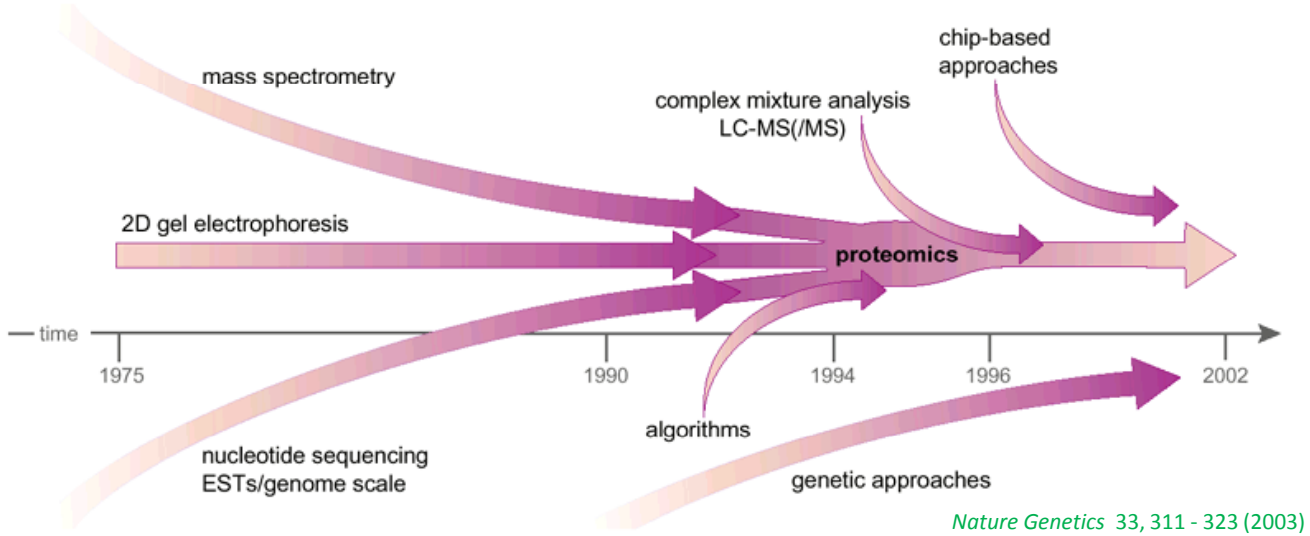
Complete human prion protein sequence:

MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNRYPPQGG
GGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGGGTHSQWNKPSKPKTN
MKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHFGSDYEDRYYRENMHRYPNQVY
YRPMDEYSNQNNFVHDCVNITIKQHTVTTTTKGENFTETDVKMMERVVEQMCIT
QYERESQAYYQRGSSMVLFSSPPVILLISFLIFLIVG

Complete human prion nucleotide sequence:

atggcgaaccttggctgctggatgctggttctctttgtggccacatggagtgac
ctgggcctctgcaagaagcgcccgaagcctggaggatggaacactgggggcagc
cgataccgggggcagggcagccctggaggcaaccgctacccacctcagggcggt
ggtggctgggggcagcctcatggtggtggctgggggcagcctcatggtggtggc
tgggggcagccccatggtggtggctggggacagcctcatggtggtggctggggt
caaggaggtggcacccacagtcagtggaacaagccgagtaagccaaaaaccaac
atgaagcacatggctggtgctgcagcagctggggcagtggtgggggggccttggc
ggctacatgctgggaagtgccatgagcaggcccatcatacatttcggcagtgac
tatgaggaccgttactatcgtgaaaacatgcaccgttaccccaaccaagtgtac
tacaggcccatggatgagtacagcaaccagaacaactttgtgcacgactgcgtc
aatatcacaatcaagcagcacacggtcaccacaaccaccaagggggagaacttc
accgagaccgacgttaagatgatggagcgcgtggttgagcagatgtgtatcacc
cagtacgagagggaatctcaggcctattaccagagaggatcgagcatggtcctc
ttctcctctccacctgtgatcctcctgatctctttcctcatcttcctgatagtg
ggatga

Current Proteomics Technologies



Nature Genetics 33, 311 - 323 (2003)

Figure 4. Time line indicating the convergence of different technologies and resources into a proteomic process.
Advances in mass spectrometry and the generation of large quantities of nucleotide sequence information, combined with computational algorithms that could correlate the two, led to the emergence of proteomics as a field.

- Proteome profiling/separation
  - 2D SDS PAGE (two-dimensional sodium dodecylsulphate polyacrylamide gel electro-phoresis)
  - 2-D LC/LC
    (LC = Liquid Chromatography)
  - 2-D LC/MS
    (MS= Mass spectrometry)
- Protein identification
  - Peptide mass fingerprint
  - Tandem Mass Spectrometry (MS/MS)
- Quantative proteomics
  - ICAT (isotope-coded affinity tag)


A NEW COMEDY BY ROB REINER
BILLY CRYSTAL MEG RYAN
Can two friends sleep together and still love each other in the morning?
When Harry Met Sally...
CASTLE ROCK ENTERTAINMENT NELSON ENTERTAINMENT ROB REINER BILLY CRYSTAL MEG RYAN
CARRIE FISHER BRUNO KIRBY ROBERT LEIGHTON JANE MUSKY
BARRY SONNENFELD MARC SHAIMAN ROB REINER AND ANDREW SCHEINMAN
NORA EPHRON ROB REINER
R RESTRICTED
A COLUMBIA PICTURES RELEASE
Includes The Music Video of Harry Connick Jr. Performing "It Had To Be You"
IN STORE DECEMBER 14
ORION HOME VIDEO

Proteomics - the classical definition

Two-dimensional gelelectrophoresis (*2D-PAGE*) of cell lysates

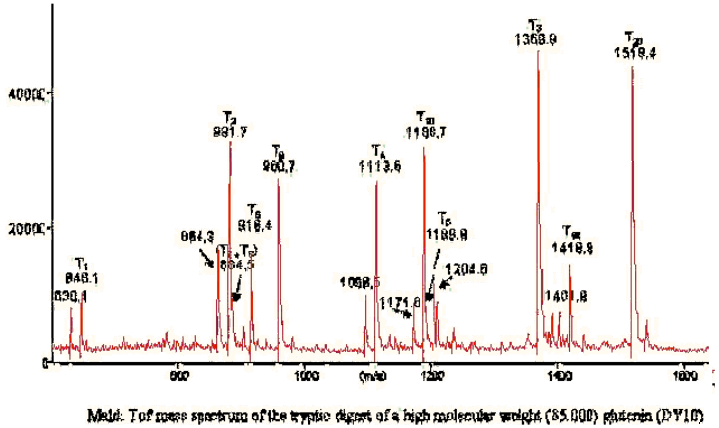
Mass spectrometry

generates global patterns of protein expression

Annotation

Large-scale visualization of differential protein expression

Peptide mass fingerprinting for protein identification



- High resolution 2D-PAGE first developed in 1975 (O'Farrell and Klose)
- Combination with biological mass spectrometry (1990s)
- Availability of genome sequences in databases

→ central role in proteomic studies

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 250, No. 10, Issue of May 25, pp. 4007–4021, 1975
Printed in U.S.A.

High Resolution Two-Dimensional Electrophoresis of Proteins*

(Received for publication, September 5, 1974)

PATRICK H. O'FARRELL‡
From the Department of Molecular, Cellular and Developmental Biology, University of Colorado, Boulder, Colorado 80302

P. H. O'Farrell
High resolution two-dimensional electrophoresis of proteins
J. Biol. Chem., May 1975; 250: 4007 - 4021

A technique has been developed for the separation of proteins by two-dimensional polyacrylamidegel electrophoresis. Due to its resolution and sensitivity, this technique is a powerful tool for the analysis and detection of proteins from complex biological sources.

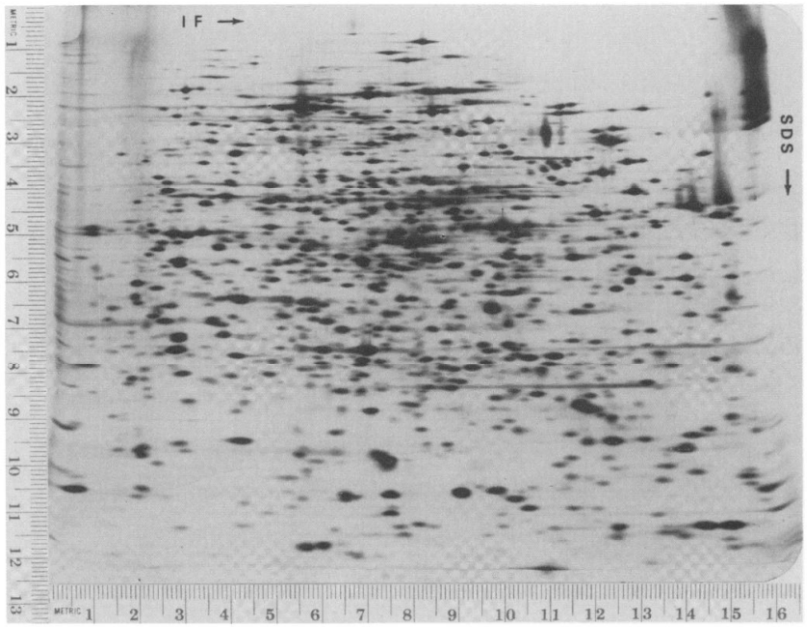


FIG. 2. Separation of *Escherichia coli* proteins. *E. coli* (1100) was labeled with ^{14}C-amino-acids as described under "Materials and Methods." The cells were lysed by sonication, treated with DNase and RNase and dissolved in lysis buffer. Twenty-five microliters of sample containing 180,000 cpm and approximately 10 μg of protein were loaded on the gel. The isoelectric focusing gel was equilibrated for 30 min. The gel in the SDS dimension was a 9.25 to 14.4% exponential acrylamide gradient. A volume of 10 ml of 14.4% acrylamide was used in the front chamber of the gradient mixer. The total volume of the gel was 16 ml. At this exposure, 825 hours, it is possible to count 1000 spots on the original autoradiogram. All autoradiograms of two-dimensional gels were photographed with a metric ruler along two edges of the autoradiogram. These rulers establish a coordinate system which is used to give spot positions. The verticle scale is given in units from top to bottom. The horizontal scale is given in units from left to right. The coordinates are given as horizontal × vertical.

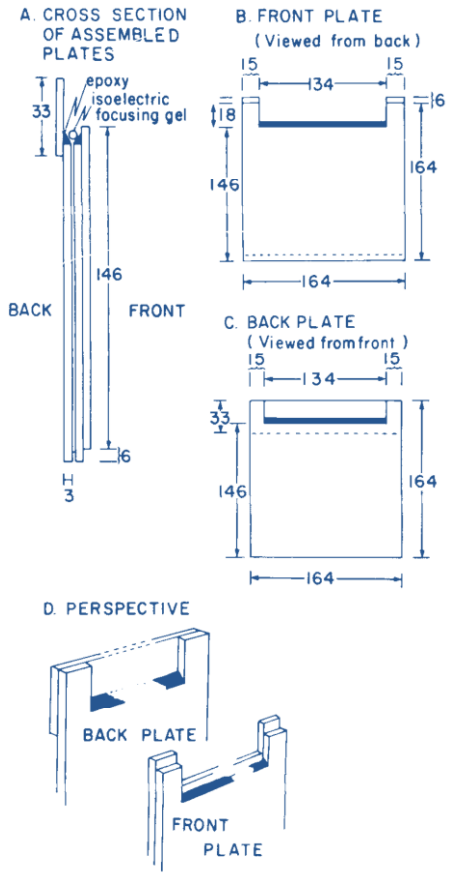


FIG. 1. Illustration of the slab gel plates. These plates are prepared as described in the text. All measurements are given in mm.

Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 1956–1960, April 1984
Biochemistry

Strategy for the mass spectrometric verification and correction of the primary structures of proteins deduced from their DNA sequences

(fast atom bombardment-mass spectrometry/tryptic peptides/molecular weight determination)

BRADFORD W. GIBSON AND KLAUS BIEMANN*

Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139

Communicated by John M. Buchanan, December 5, 1983

B.W. Gibson and K. Biemann
Strategy for the Mass Spectrometric Verification and Correction of the Primary Structures of Proteins Deduced from Their DNA Sequences
PNAS 1984 81: 1956-1960.

Fast atom bombardment mass spectrometry has been used to confirm and correct regions from the amino acid sequences of three large proteins, glutaminyl- and glycyl tRNA synthetase from Escherichia coli and methionyl-tRNA synthetase from yeast, whose primary structures had been deduced from the base sequences of their corresponding genes. The strategy is based on a comparison of the molecular weights of the tryptic peptides predicted from all three reading frames of the gene sequences with those determined mass spectrometrically.

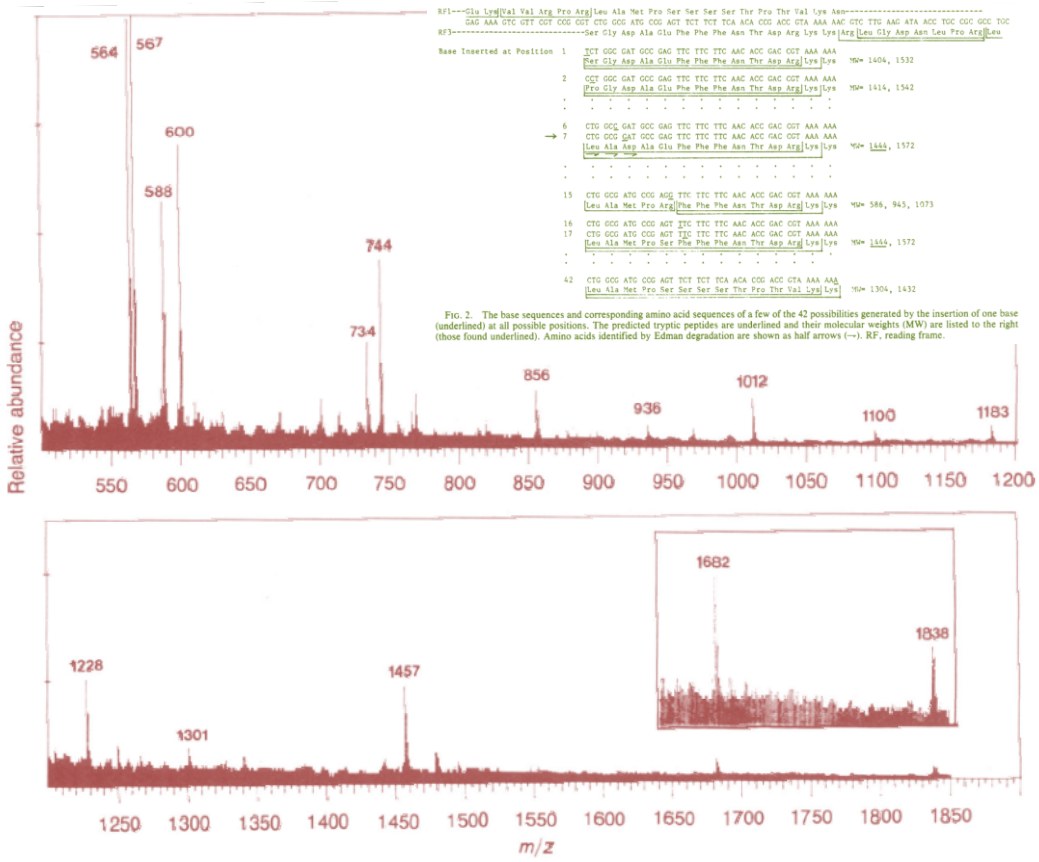


FIG. 1. FAB-MS of HPLC fraction 2 from the tryptic hydrolysate of Gly-tRNA synthetase. The MH^+ ions for each peptide are labeled.

Tools for Protein Identification

Enzyme Digestion

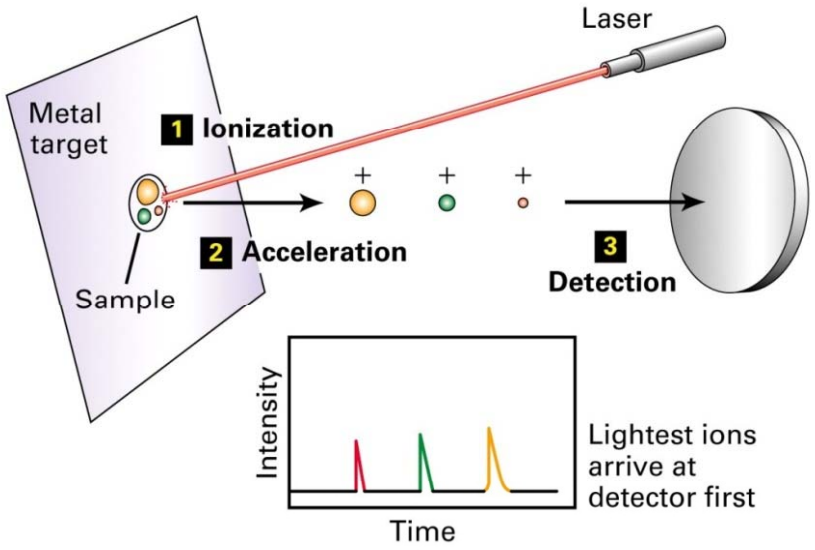
To get smaller fragments: Trypsin (99%), LysC, others

Mass Spectrometry

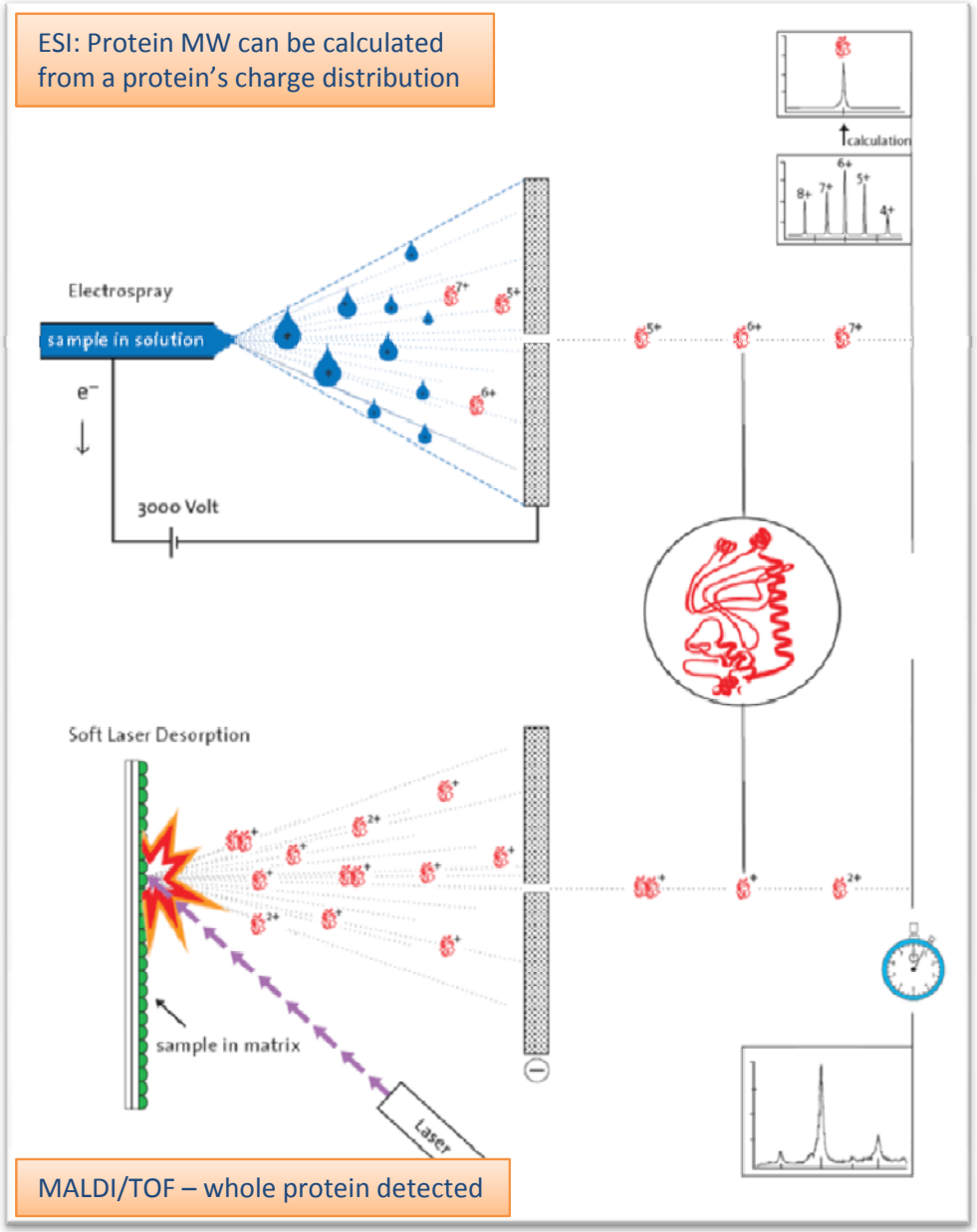
Two ionization techniques:

1. Matrix Assisted Laser Desorption/Ionization (MALDI)
2. Electrospray Ionization (ESI)

Both with many types of mass analyzers:
TOF, Quadrupole (Q), Q-TOF, FT ICR MS, Q-ITOF, etc.



ESI: Protein MW can be calculated from a protein's charge distribution



MALDI/TOF – whole protein detected

European J. Biochem. 1 (1967) 80—91

A Protein Sequenator

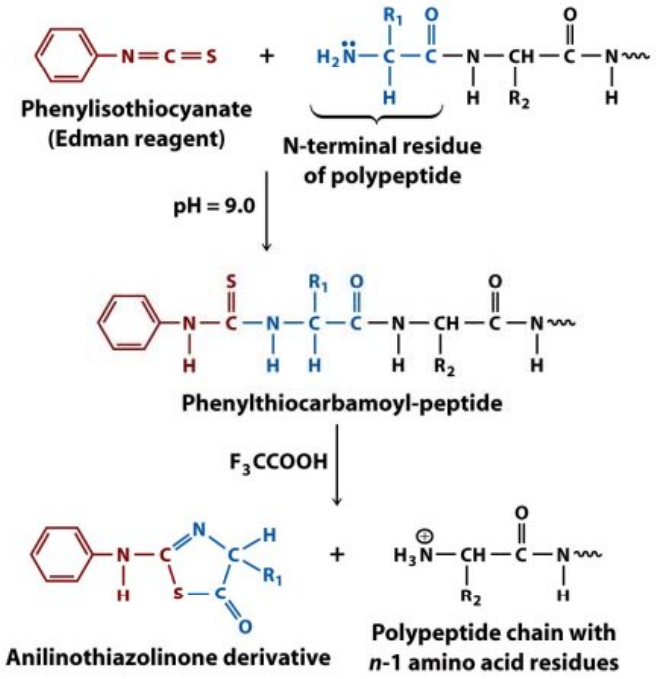
P. Edman and G. Begg

St. Vincent's School of Medical Research, Melbourne, N. 6, Victoria

(Received October 5, 1966)

Edman P, Begg G.
A protein sequenator.
Eur J Biochem. 1967
Mar;1(1):80-91.

Edman degradation, developed by Pehr Victor Edman, is a method of sequencing amino acids in a peptide. In this method, the amino-terminal residue is labeled and cleaved from the peptide without disrupting the peptide bonds between other amino acid residues.



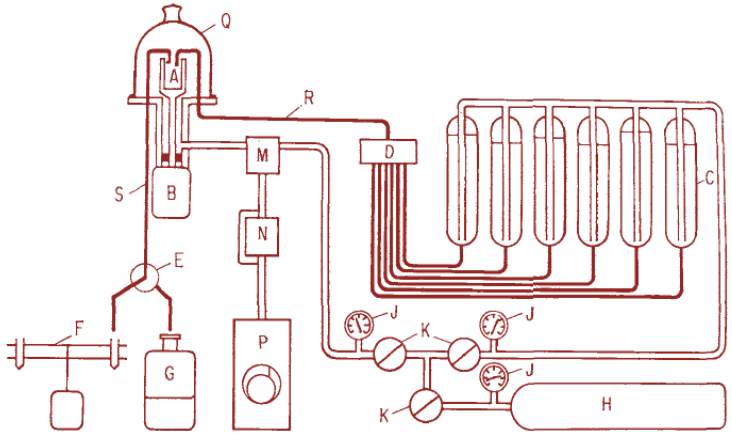
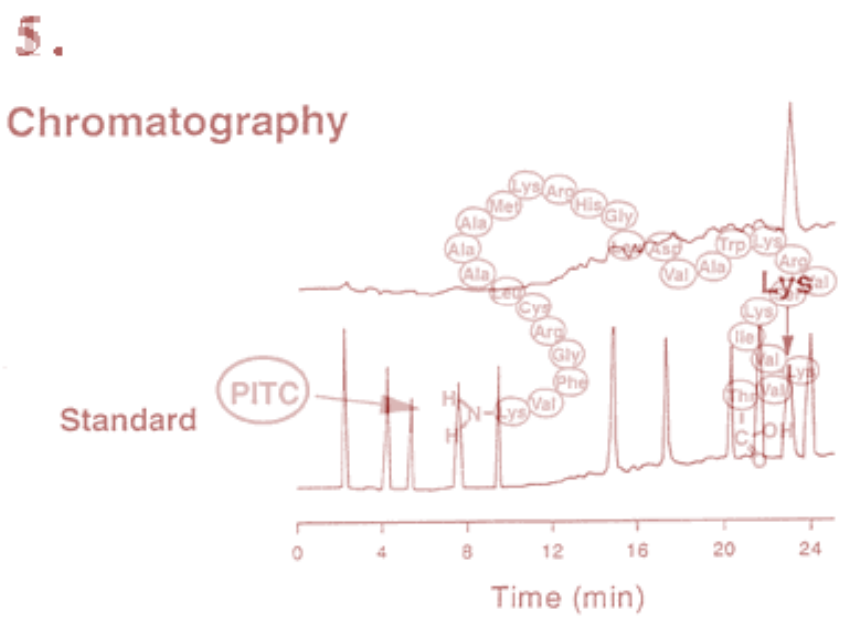


Fig. 1. *Diagram of sequenator.* *A*, spinning cup; *B*, electric motor; *C*, reagent (solvent) reservoir; *D*, valve assembly; *E*, outlet stopcock assembly; *F*, fraction collector; *G*, waste container; *H*, nitrogen cylinder; *J*, pressure gauges; *K*, pressure regulators; *M*, 3-way valve; *N*, 2-way valve with bypass; *P*, rotary vacuum pump; *Q*, bell jar; *R*, feed line; *S*, effluent line. Gas lines are doubly contoured, and liquid lines are filled

Two major techniques for sequencing a polypeptide

(Edman degradation versus Mass Spectrometry)



Edman degradation for N-terminal analysis peptides

- Involves treatment of a peptide with phenyl isothiocyanate (PITC), C_6H_5–N=C=S, followed by treatment with trifluoroacetic acid .
- The phenylthiohydantoin (PTH) is identified chromatographically by comparison of its elution times with the known elution times of PTH derivatives of all 20 common amino acids.

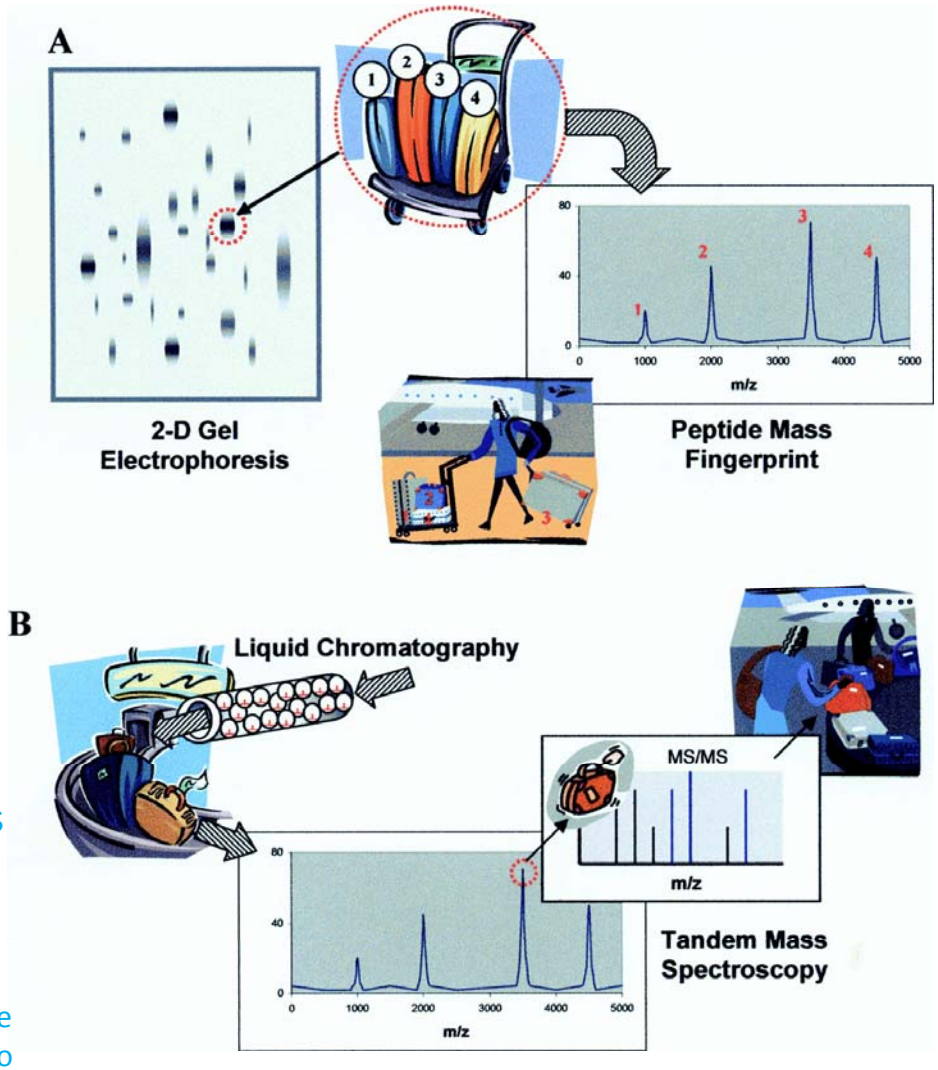
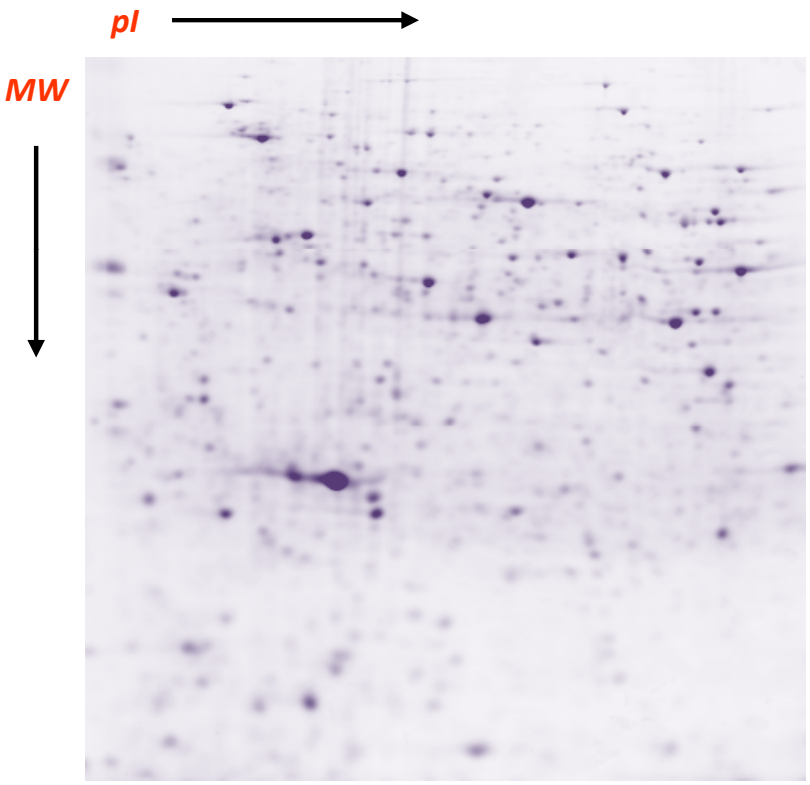


Figure 1 A simple analogy to illustrate the two most commonly used proteomic strategies

Clin. Sci. (2005) 109, 421-430

2D-PAGE based expression proteomics

Protein expression profiling: ~ 1000 proteins routinely detectable in a 2D-gel → global changes in the proteome readily detectable



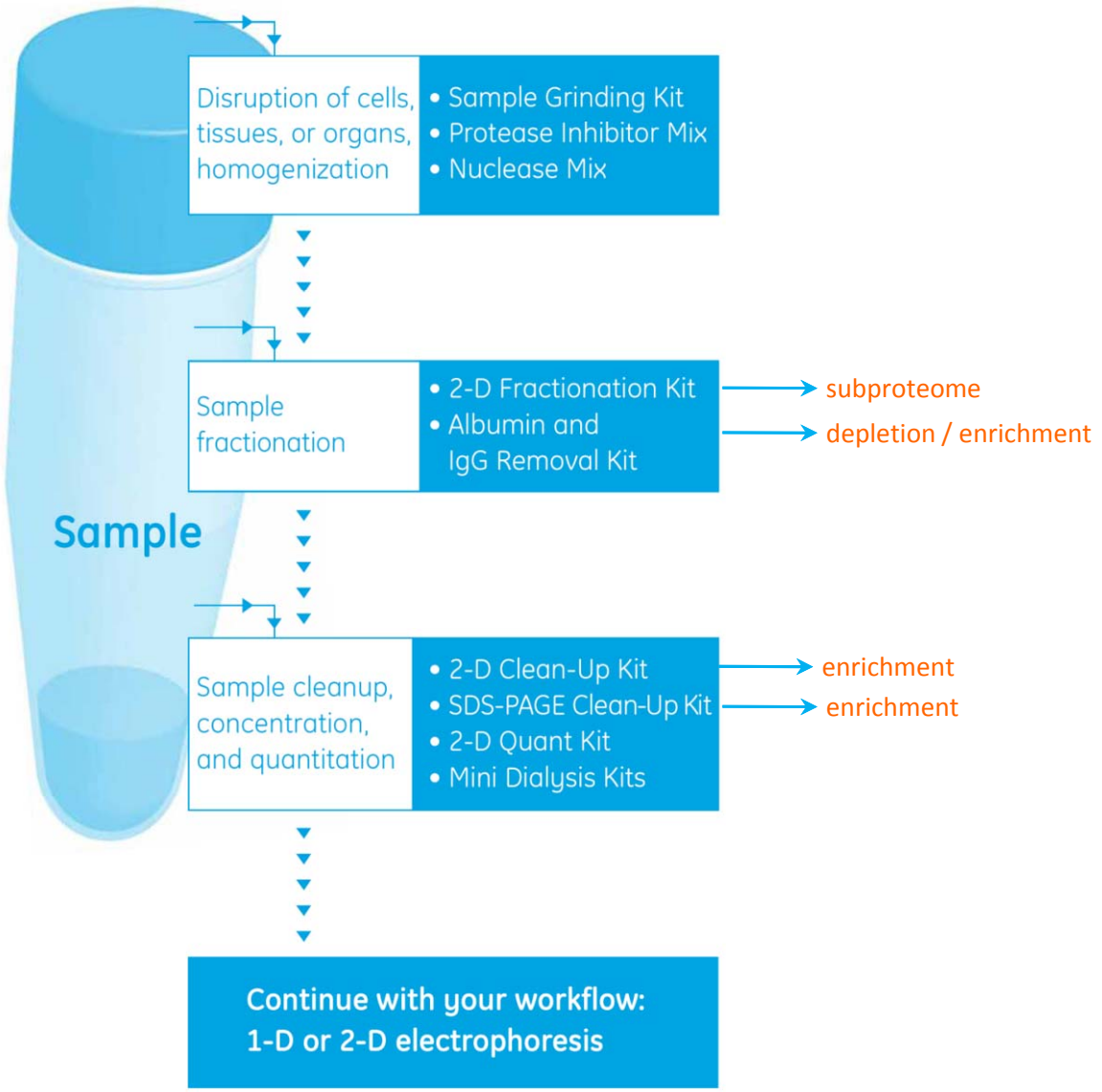
SYPRO Ruby stained gel

- Identify specific proteins in a cell that undergo changes in abundance, localization, or modification in response to a specific biological condition
- Often combined with complementary techniques (protein biochemistry, molecular biology and cell physiology)

- posttranscriptional control mechanisms can influence protein expression
- posttranslational modifications of a protein such as phosphorylation, glycosylation, processing of signal sequences or degradation can be visualized

- Sample preparation
- Isoelectrofocusing (1st dimension)
- Equilibration incl. reduction, alkylation
- SDS-PAGE (2nd dimension)
- Staining
- Imaging
- Spot detection and matching
- Normalization and quantification
- Analysis
- Cutting of selected spots
- Trypsin digestion
- Identification with mass spectroscopy
- Database comparison

- Detergents: solubilize membrane proteins-separation from lipids
- Reductants: Reduce S-S bonds
- Denaturing agents: Disrupt protein-protein interactions-unfold proteins
- Enzymes: Digest contaminating molecules (nucleic acids etc.)
- Protease inhibitors



Isoelectric point (pI):

Isoelectric Focusing is an electrophoretic method that separates proteins according to their **isoelectric points (pI).** Proteins are amphoteric molecules; they carry either positive, negative, or zero net charge, depending on the pH of their surroundings. The net charge of a protein is the sum of all the negative and positive charges of its amino acid side chains and amino- and carboxyl-termini. The isoelectric point (pI) is the specific pH at which the net charge of the protein is zero. Proteins are positively charged at pH values below their pI and negatively charged at pH values above their pI. If the net charge of a protein is plotted versus the pH of its environment, the resulting curve intersects the x-axis at the isoelectric point.

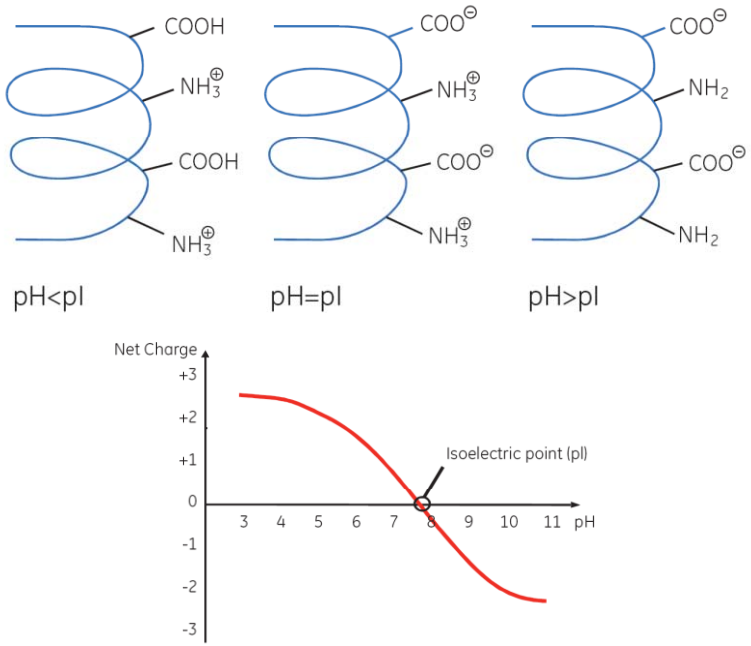


FIGURE: Plot of the net charge of a protein versus the pH of its environment. The point of intersection of the curve at the x-axis represents the isoelectric point of the protein.

Immobilized pH gradients

Immobilized pH gradients are formed using two solutions, one containing a relatively acidic mixture of acrylamido buffers and the other containing a relatively basic buffer mixture. The concentrations of the various buffers in the two solutions define the range and shape of the pH gradient produced. Both solutions contain acrylamide monomers and catalysts. During polymerization, the acrylamide portion of the buffers copolymerizes with the acrylamide and bisacrylamide monomers to form a polyacrylamide gel.

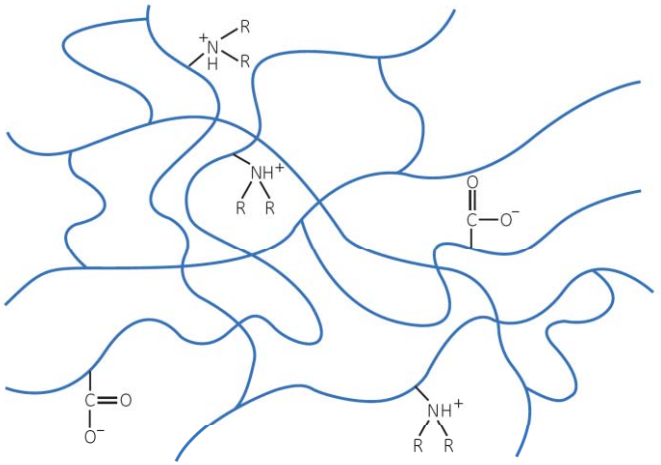
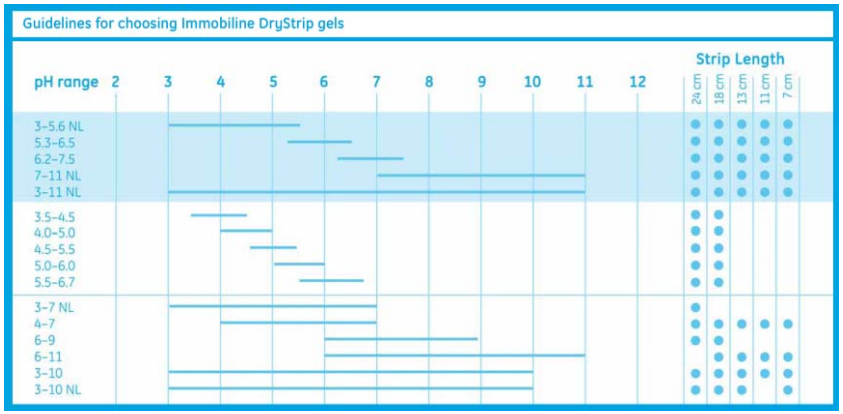
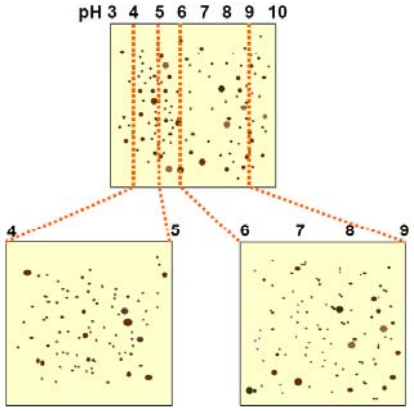
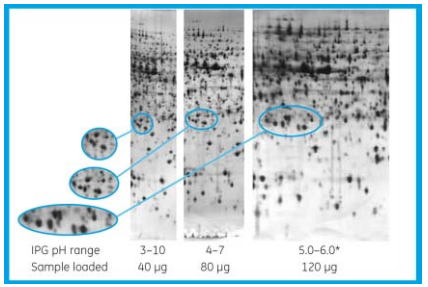


FIGURE: Immobilized pH gradient polyacrylamide gel matrix showing attached buffering groups.

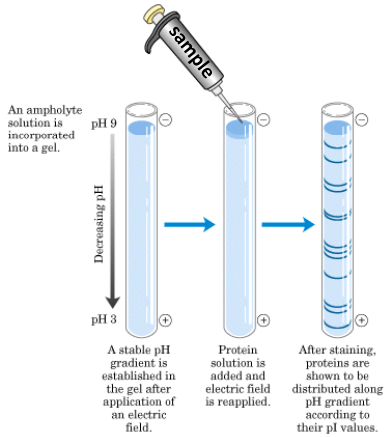
- Sample preparation
- Isoelectrofocusing (1st dimension)
- Equilibration incl. reduction, alkylation
- SDS-PAGE (2nd dimension)
- Staining
- Imaging
- Spot detection and matching
- Normalization and quantification
- Analysis
- Cutting of selected spots
- Trypsin digestion
- Identification with mass spectroscopy
- Database comparison

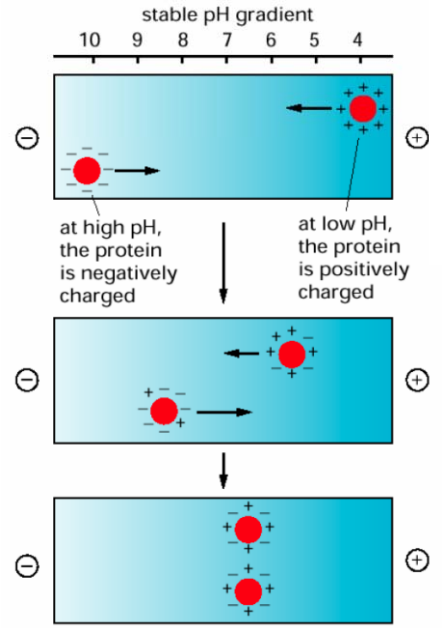






Ettan IPGphor 3 IEF System





The protein shown here has an isoelectric pH of 6.5.

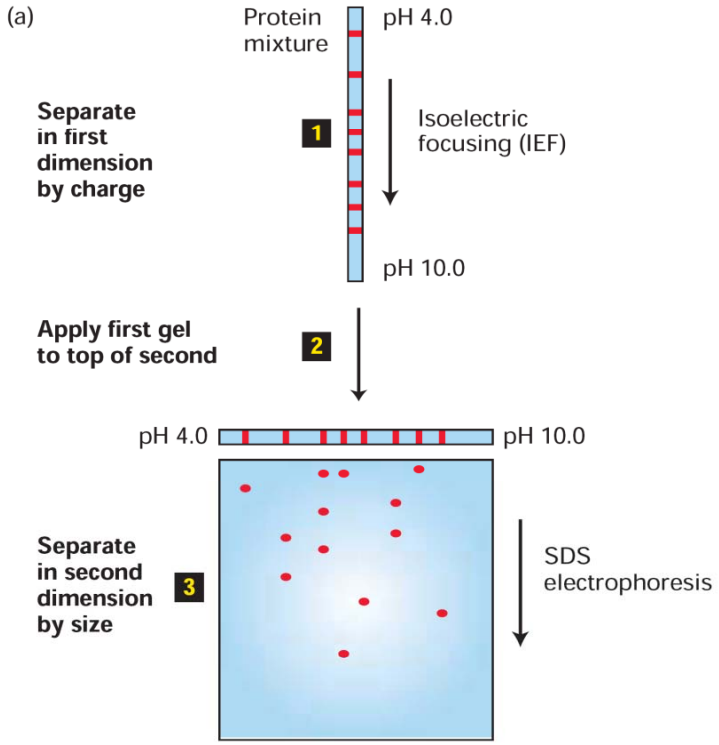
At the isolectric point the protein has no net charge and therefore no longer migrates in the electric field.

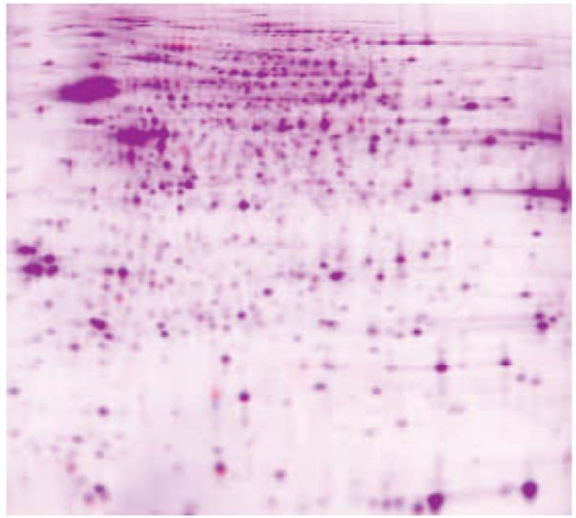
Wide and Narrow pH Gradients

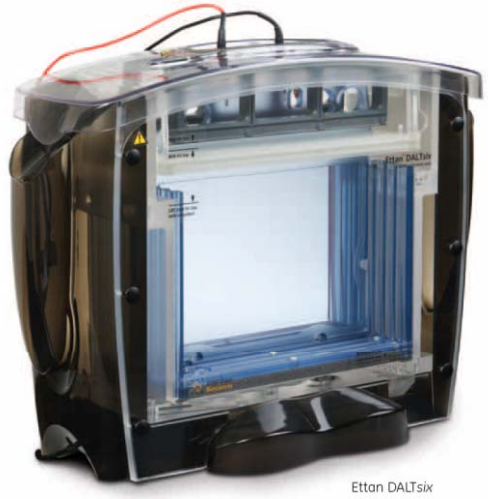
- Wide gradients are applied for:
  entire protein spectrum
- Narrow gradients are applied for
  increased resolution
  increased loading capacity to detect and analyze more proteins

- Sample preparation
- Isoelectrofocusing (1st dimension)
- Equilibration incl. reduction, alkylation
- SDS-PAGE (2nd dimension)
- Staining
- Imaging
- Spot detection and matching
- Normalization and quantification
- Analysis
- Cutting of selected spots
- Trypsin digestion
- Identification with mass spectroscopy
- Database comparison

- Proteins enter SDS-Polyacrylamide gel and are dissolved according to their molecular mass.
- Postelectrophoretic staining of the proteins with: Coomassie, Silver, Fluorescent stains (SYPRO Ruby)
- Protein expression profiling: ~ 1000 proteins routinely detectable in a 2D-gel → global changes in the proteome readily detectable.
- Posttranscriptional control mechanisms can influence protein expression.
- Posttranslational modifications of a protein such as phosphorylation, glycosylation, processing of signal sequences or degradation can be visualized.



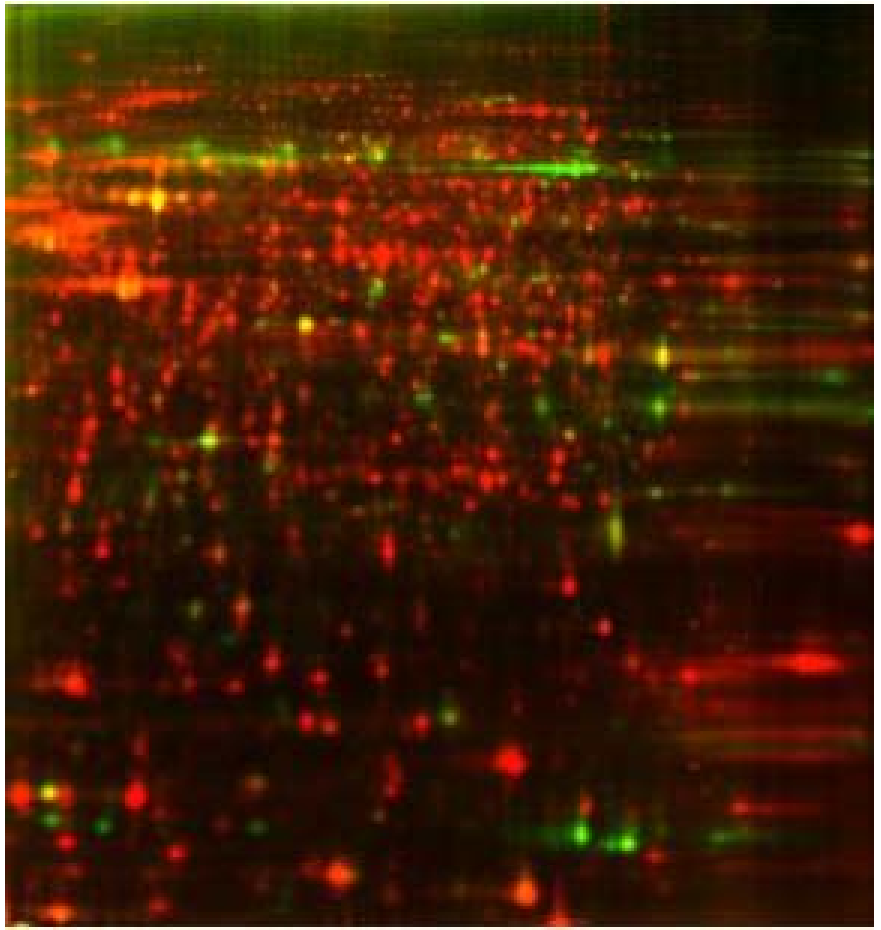




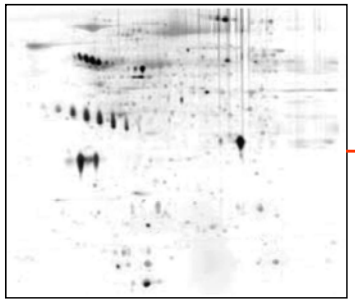
Second-dimension separation with Ettan DALT electrophoresis system. First-dimension separation: Ettan IPGphor IEF system, 24-cm Immobiline DryStrip gel, pH 3-10 NL. Second-dimension separation: Ettan DALT electrophoresis system, DALT Gel 12.5 (26 x 20 cm).

Limitations and challenges of gel-based approaches

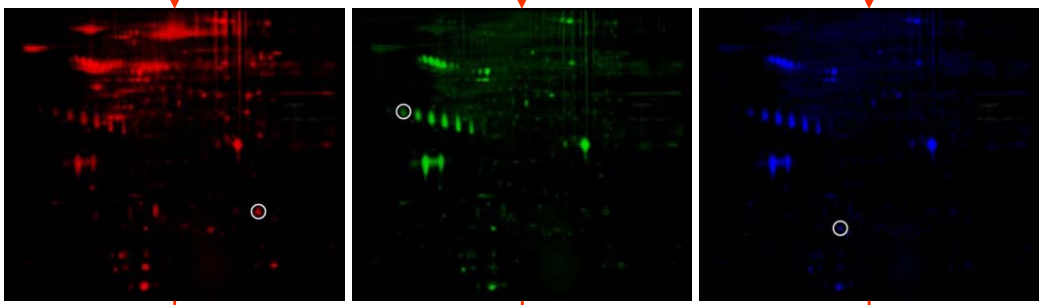
- Dynamic range detectable on 2D-gels: 10^4, protein expression levels of a cell can vary between 10^5 (yeast) and even 10^{10} (humans).
  → enrichment or pre-fractionation strategies needed to reach less abundant proteins
- Resolution of 2D-gels has its limits.
  → use narrow pH range gels and combine
- Protein extraction and solubility during IEF can be a problem for poorly water-soluble proteins e.g. membrane proteins or nuclear proteins.
- ***Challenges*** for further development in gel-based proteomics: improve sample preparation to be ***able to analyze extreme proteins*** (extremely basic or acidic, extremely small or big, extremely hydrophobic), ***sensitivity, dynamic range, automation.***

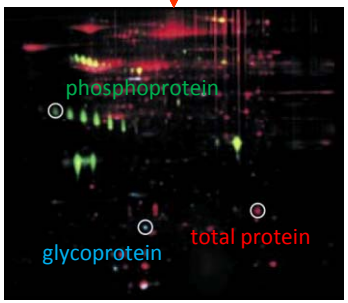


Multiplexed Proteomics Technology



SYPRO® Ruby
[total protein]
Pro-Q® Diamond
[phosphoprotein]
Pro-Q® Emerald
[glycoprotein]

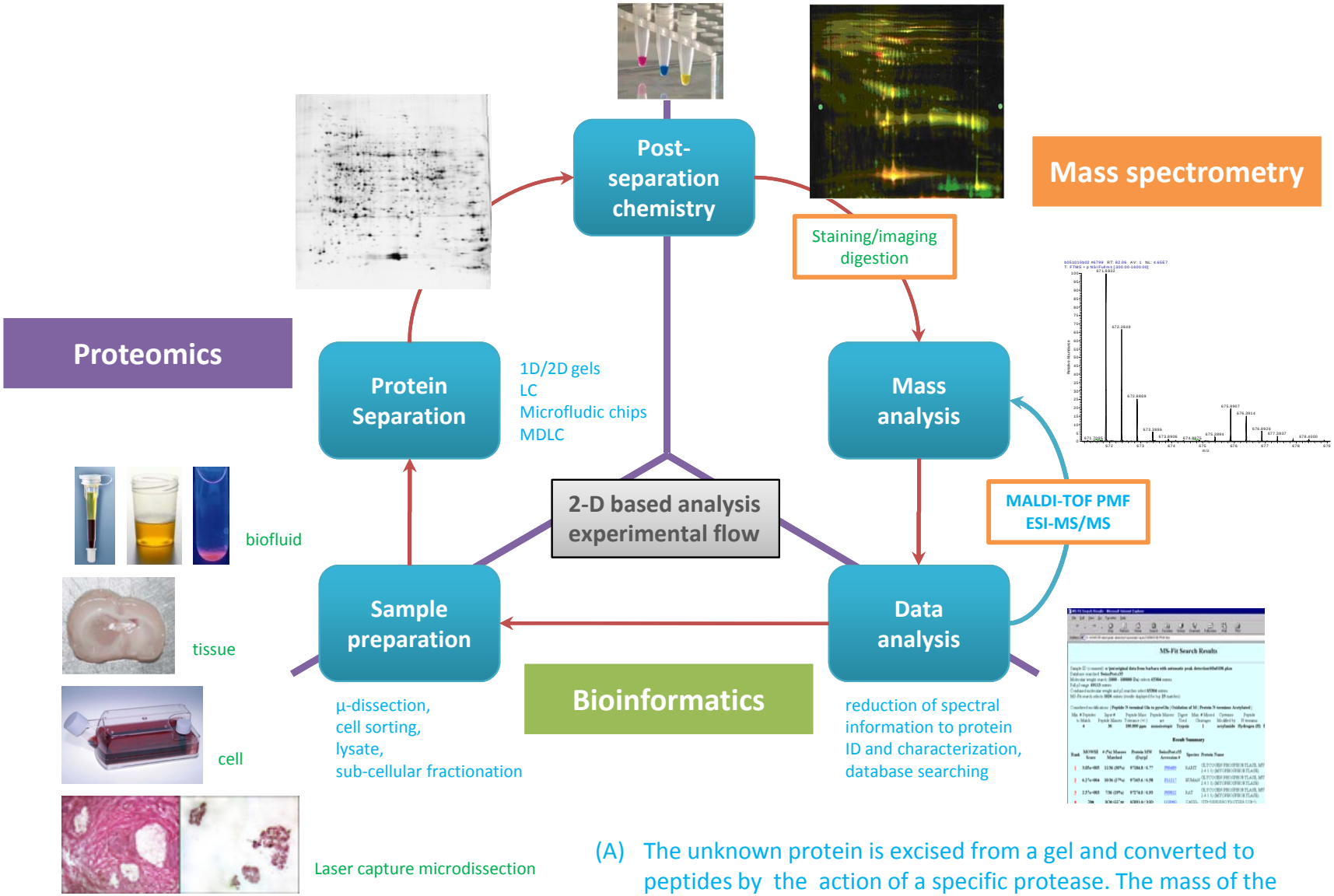




Three stains in one gel

- Identify changes in expression, phosphorylation and glycosylation
- Compare multiple samples with high accuracy and reproducibility
- Obtain more data from precious samples

General workflow of proteomics analysis (peptide mass fingerprinting method)

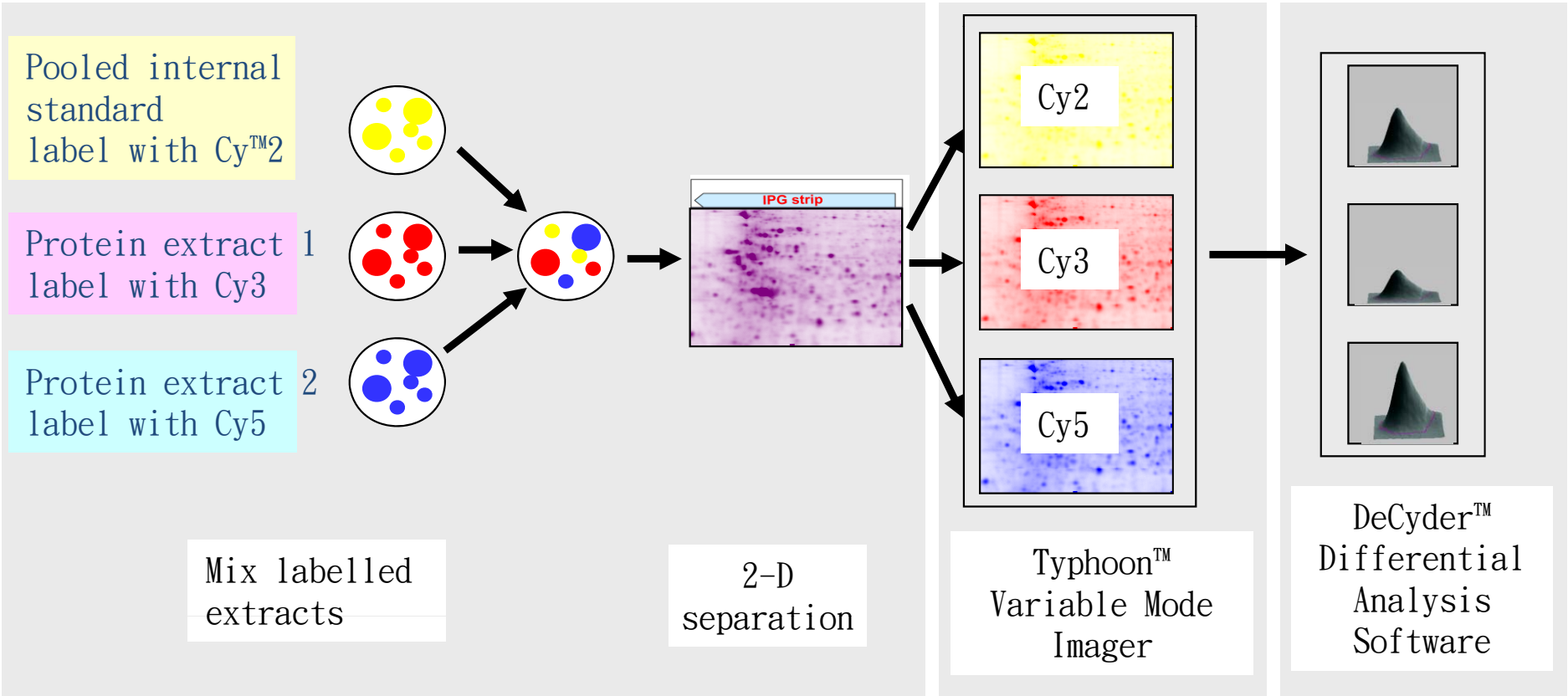


(A) The unknown protein is excised from a gel and converted to peptides by the action of a specific protease. The mass of the peptides produced is then measured in a mass spectrometer.

(B) The mass spectrum of the unknown protein is searched against theoretical mass spectra produced by computer-generated cleavage of proteins in the database.

Difference in-gel 2D-PAGE system (DIGE)

- Built upon the classical gel approach to protein quantification (gel densitometry)
- Separate samples are treated with unique fluorophore tags (binding covalently with lysine ε-amino groups)
- Samples are combined and run on the same 2D gel (ΔMW of proteins is negligible)
- Quantitative Analysis is based on relative intensities of fluorescing labels at specific spots (relative quantitation) or to labeled standard (absolute quantitation).
- Allows use of an internal standard in each gel which reduces gel-to-gel variation, reduces the number of gels to be run.

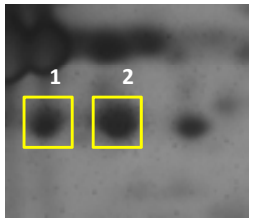


- Sample preparation
- Isoelectrofocusing (1st dimension)
- Equilibration incl. reduction, alkylation
- SDS-PAGE (2nd dimension)
- Staining
- Imaging
- Spot detection and matching
- Normalization and quantification
- Analysis
- Cutting of selected spots
- Trypsin digestion
- Identification with mass spectroscopy
- Database comparison

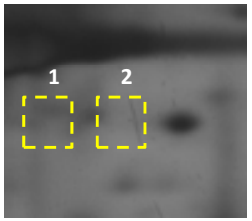
melanie

http://expasy.org/melanie/

ImageQuant ECL

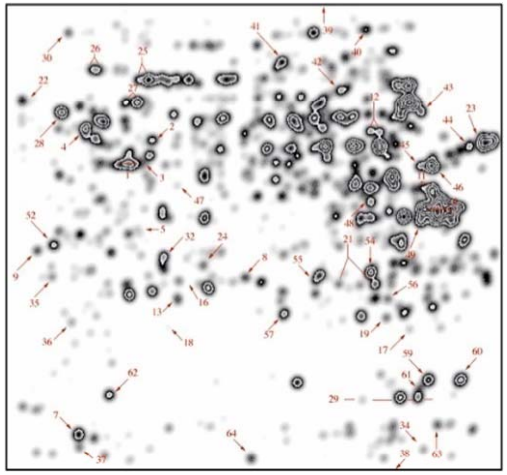


Pre-vaccine

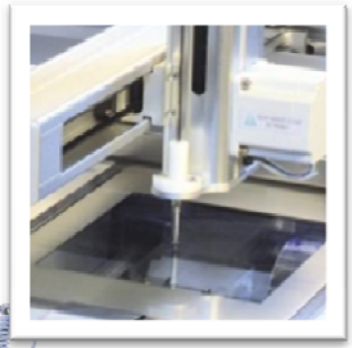


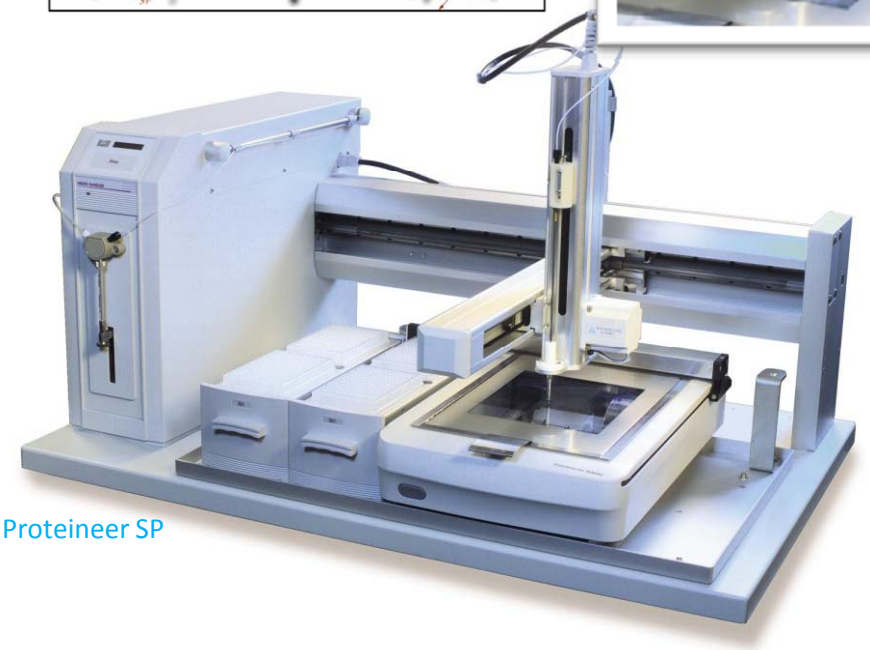
Post-vaccine

manual



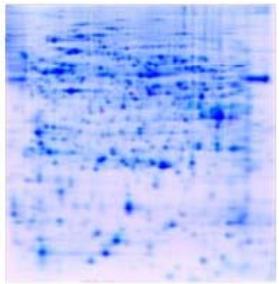
robotics



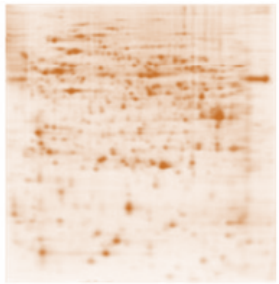


Proteineer SP

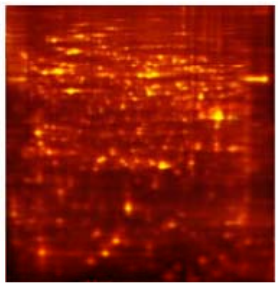
Commassie

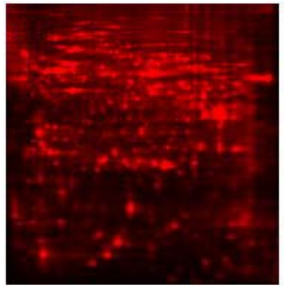


Silver

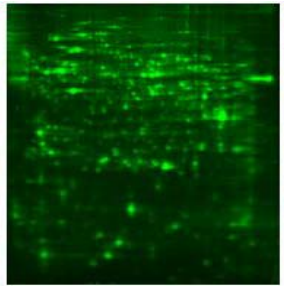


Sypro Ruby

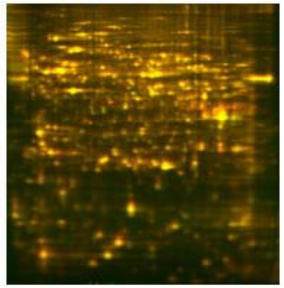




Cy5

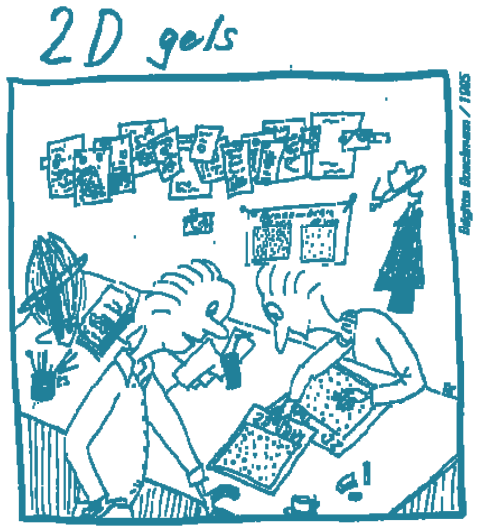


Cy3

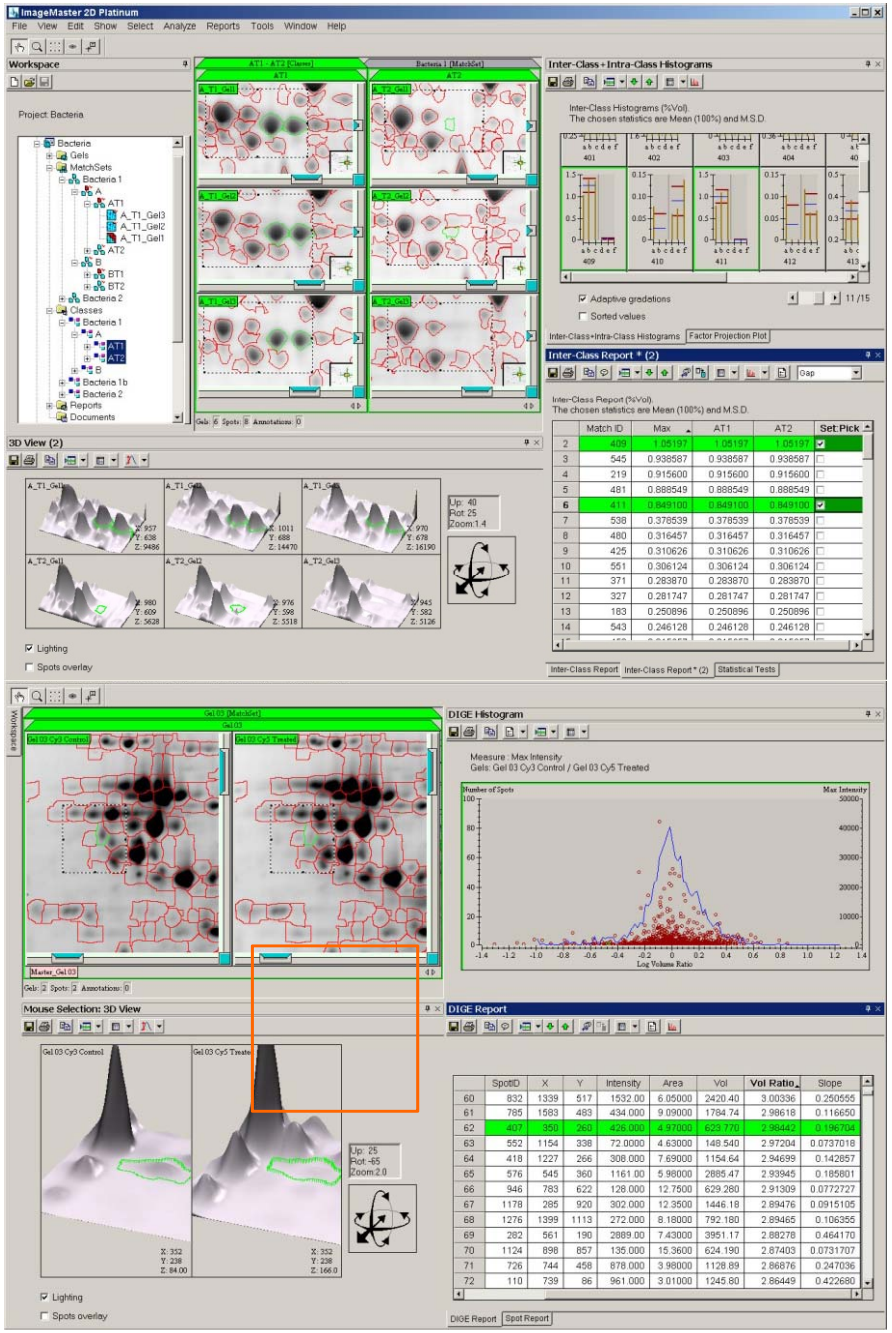


Cy5/Cy3

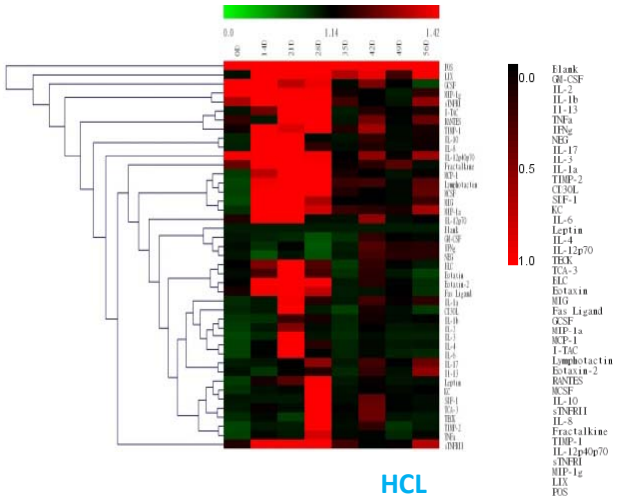
- Sample preparation
- Isoelectrofocusing (1st dimension)
- Equilibration incl. reduction, alkylation
- SDS-PAGE (2nd dimension)
- Staining
- Imaging
- Spot detection and matching
- Normalization and quantification
- Analysis
- Cutting of selected spots
- Trypsin digestion
- Identification with mass spectroscopy
- Database comparison



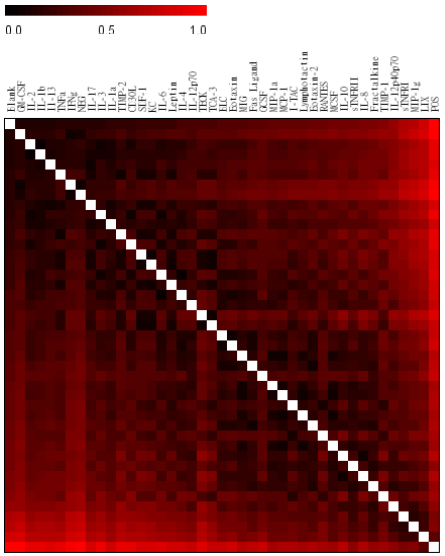
http://expasy.org/images/cartoon/2dgels.gif



Expression clustering



HCL

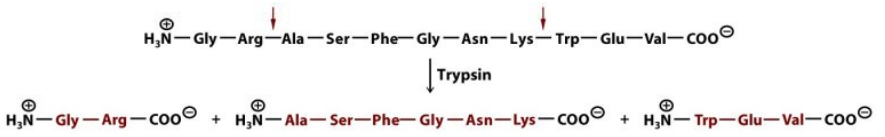


GDM

- Sample preparation
- Isoelectrofocusing (1st dimension)
- Equilibration incl. reduction, alkylation
- SDS-PAGE (2nd dimension)
- Staining
- Imaging
- Spot detection and matching
- Normalization and quantification
- Analysis
- Cutting of selected spots
- Trypsin digestion
- Identification with mass spectroscopy
- Database comparison

Trypsin

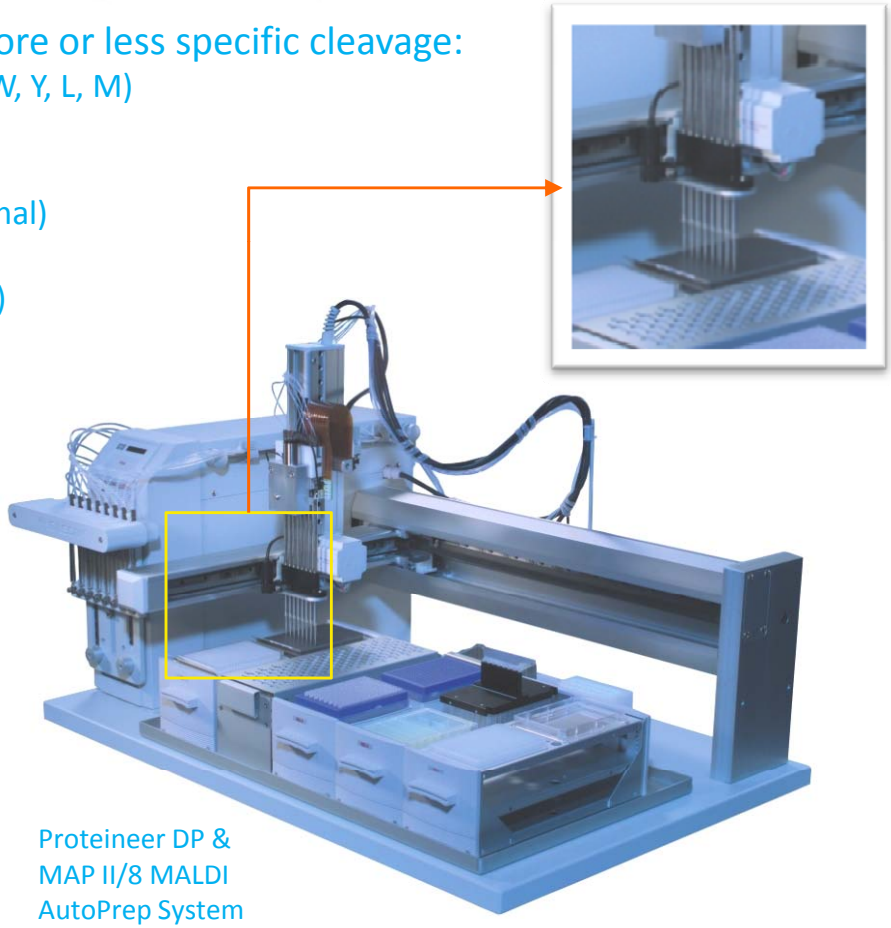
Good activity both in gel digestion and in solution.
(Cleaves at lysine [K] and arginine [R], unless either is followed by proline [P] in C-terminal direction)



Other enzymes with more or less specific cleavage:

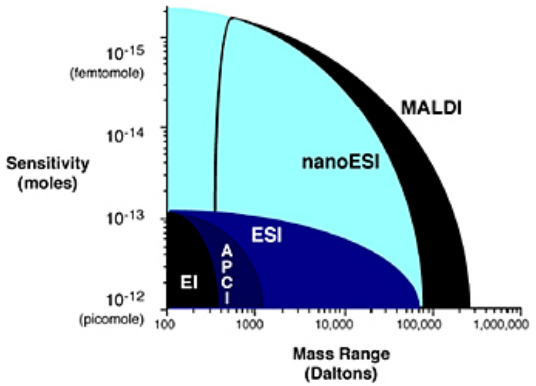
- Chymotrypsin (F, W, Y, L, M)
- Lys-C (K)
- Arg-C (R)
- Asp-N (D, N-terminal)
- V8-bicarb (E)
- V8-biphosph (E, D)
- {CNBr (M)}

1. Destain
2. Shrink with acetonitrile
3. Digest
4. Extract peptides and wash
5. Capture peptides on C_{18}
6. Elute in MS-compatible buffer

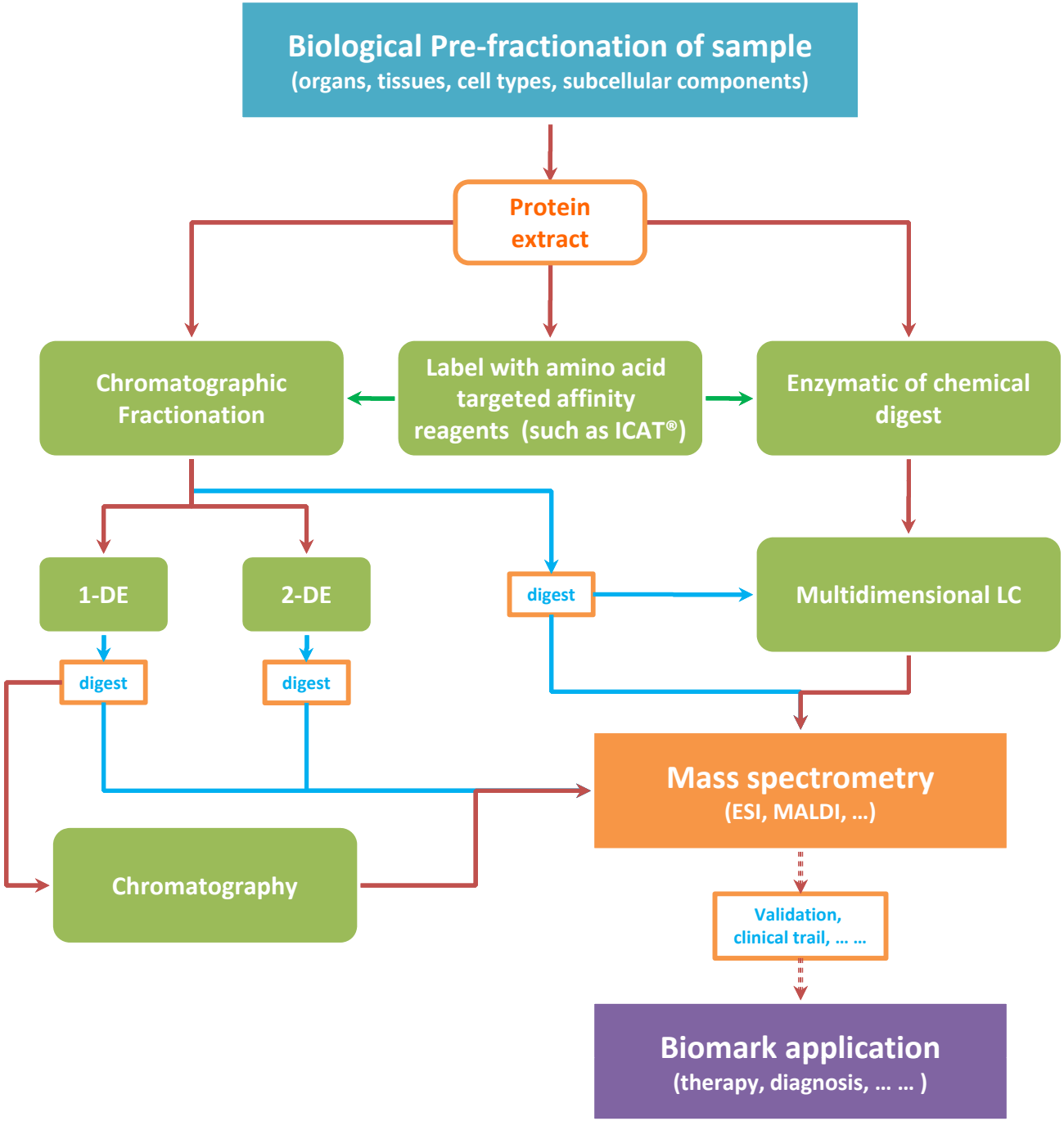


Proteineer DP &
MAP II/8 MALDI
AutoPrep System

- Sample preparation
- Isoelectrofocusing (1st dimension)
- Equilibration incl. reduction, alkylation
- SDS-PAGE (2nd dimension)
- Staining
- Imaging
- Spot detection and matching
- Normalization and quantification
- Analysis
- Cutting of selected spots
- Trypsin digestion
- Identification with mass spectroscopy
- Database comparison

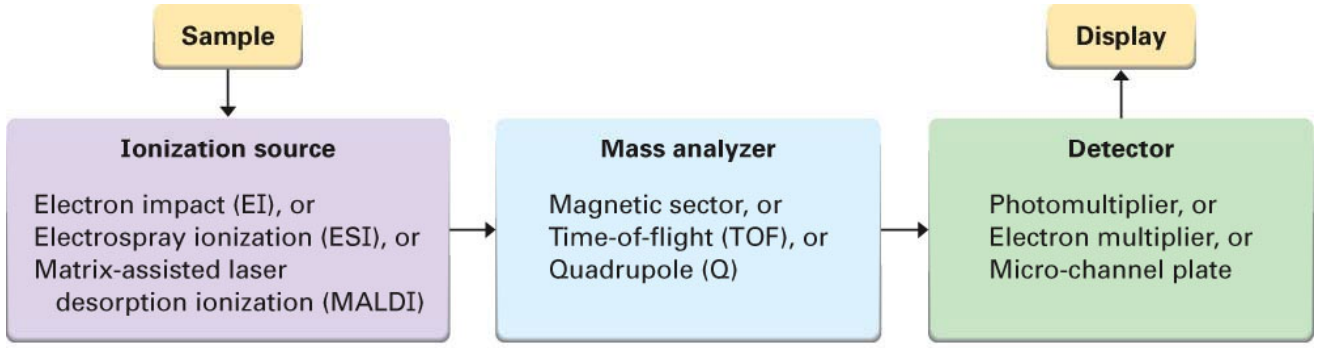


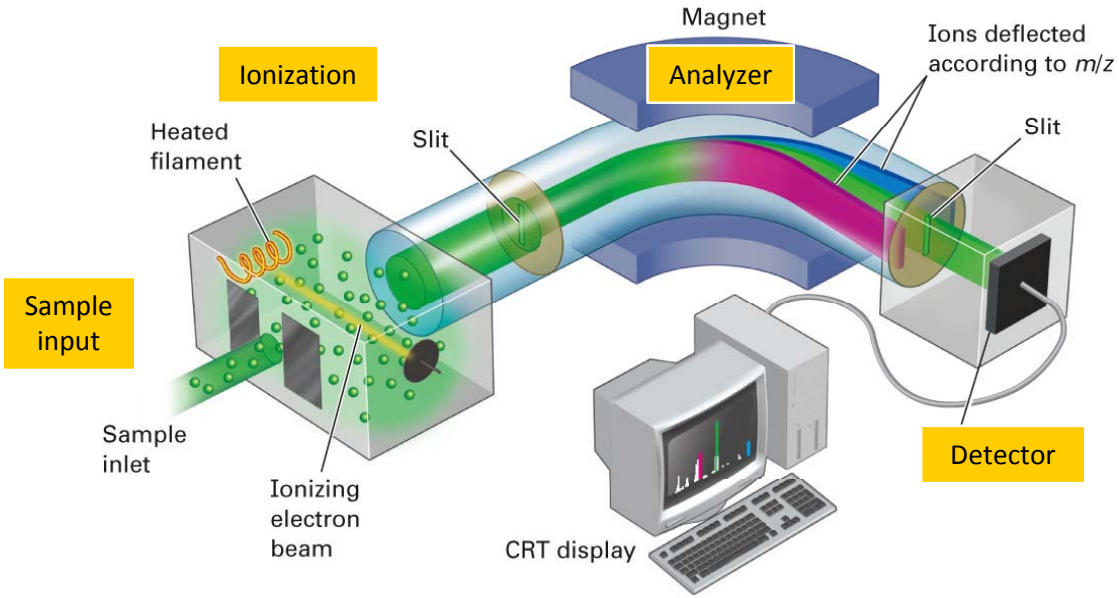
A glance at the typical sensitivity and mass ranges allowed by different ionization techniques provides a clear answer to the question of which are most useful; electron ionization (EI), atmospheric pressure chemical ionization (APCI) and desorption/ionization on silicon (DIOS) are somewhat limiting in terms of upper mass range, while electrospray ionization (ESI), nanoelectrospray ionization (nanoESI), and matrix-assisted laser desorption ionization (MALDI) have a high practical mass range.



Mass Spectrometry of Small Molecules: Magnetic-Sector Instruments

- Mass spectrometry (MS) measures the mass and molecular weight (MW) of a molecule.
- Provides structural information by finding the masses of fragments produced when molecules break apart.
- Three basic parts of mass spectrometers





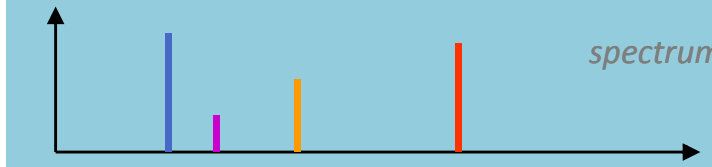
Bruker Autoflex
Atomatic MALDI-TOF MS

trypsin digestion

MPSESSYKVHRPAKSGGS *protein*

MPSESSYK VHR PAK SGGS *peptides*

spectrum



m/z : mass-to-charge ratio

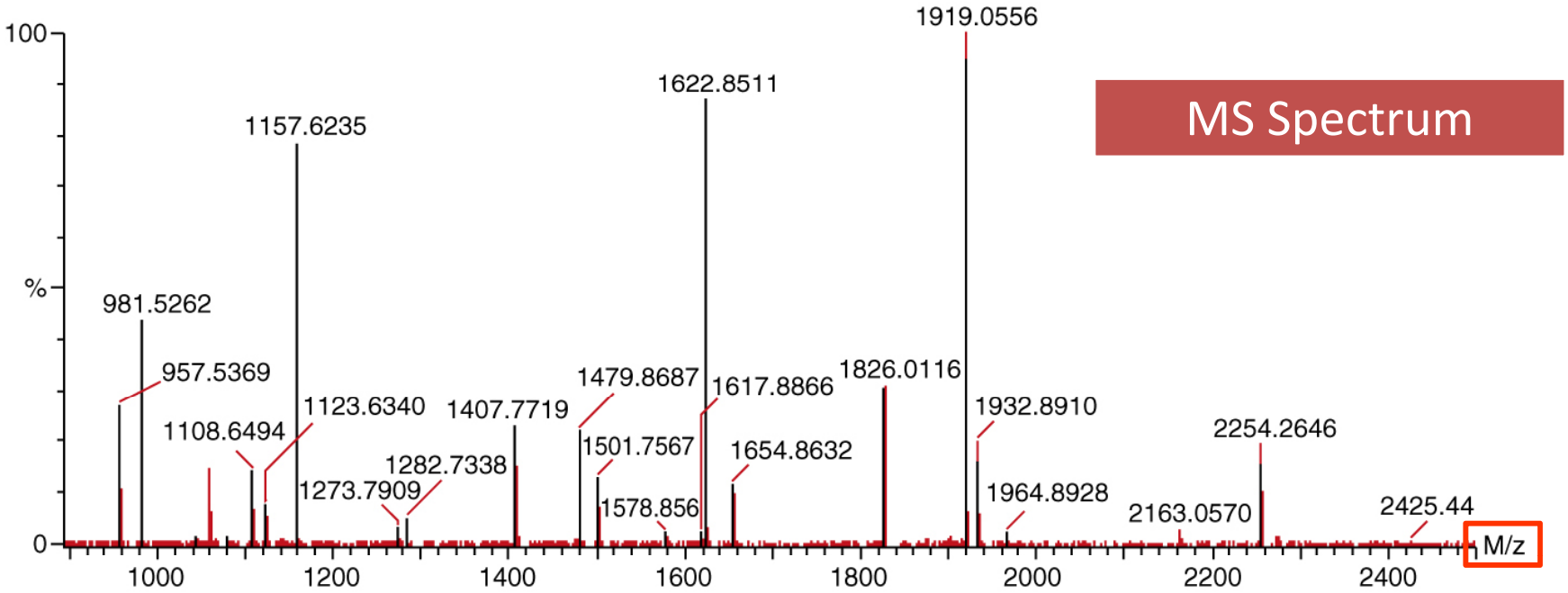
VALEACVQAR Mass = 1059.25 Da

VALEACVQAR
|
H^+

Mass = 1060.25 Da
m/z = 1060.25 / 1
= 1060.25

H^+ — VALEACVQAR
|
H^+

Mass = 1061.25 Da
m/z = 1061.25 / 2
= 530.63



The Nobel Prize in Chemistry 2002

http://nobelprize.org

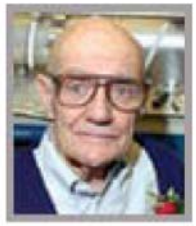
BACK

The Nobel Prize in Chemistry 2002

The Royal Swedish Academy of Sciences has decided to award the Nobel Prize in Chemistry for 2002 "for the development of methods for identification and structure analyses of biological macromolecules" with one half jointly to John B. Fenn and Koichi Tanaka "for their development of soft desorption ionisation methods for mass spectrometric analyses of biological macromolecules" and the other half to Kurt Wüthrich "for his development of nuclear magnetic resonance spectroscopy for determining the three-dimensional structure of biological macromolecules in solution".

Proteins in close-up

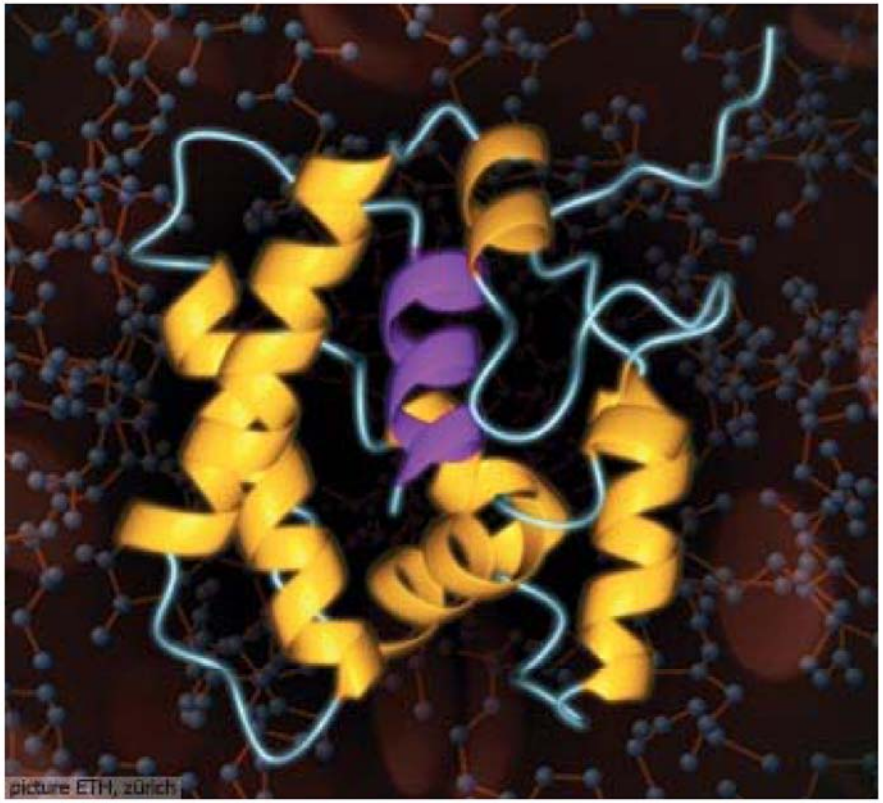
Living cells consist of myriads of molecules. The large molecules, which include the proteins, interact with one another and with other molecules in a never-resting molecular machinery. How can we understand what is happening inside the cell? One important step is to develop tools to "see" with, and this is what this year's Nobel Laureates in Chemistry have done.



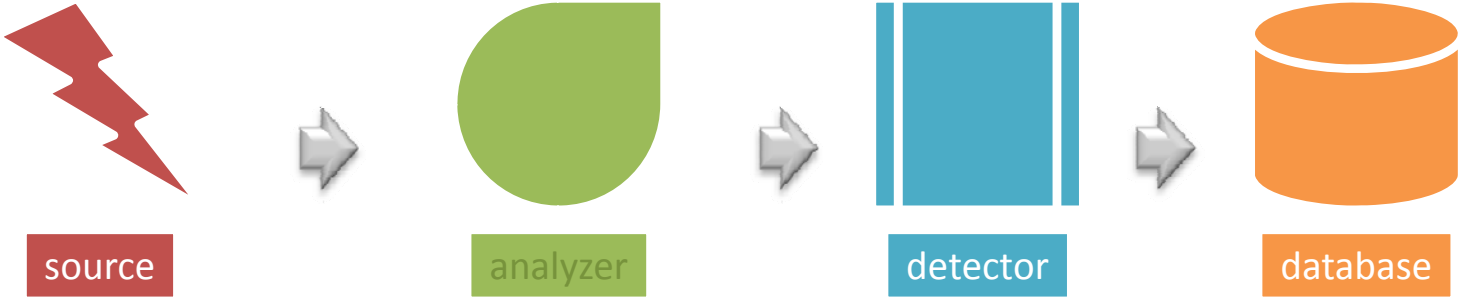
John B. Fenn

Koichi Tanaka

Kurt Wüthrich



Mass Spectrometry



source | analyzer | detector | database

produces the ions from the sample (vaporization /ionization)

resolves ions based on their mass/charge (m/z) ratio

detection of mass separated ions

identification and prediction

Ion source
ESI, MALDI

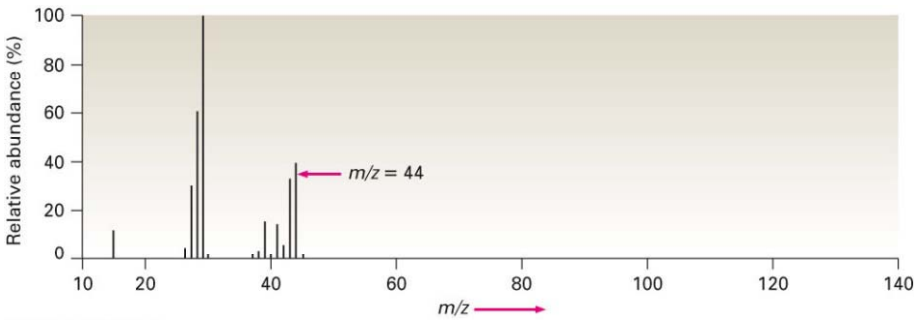
Analyzer
PSD, Ion trap, Q-TOF, TOF/TOF, FT-ICR

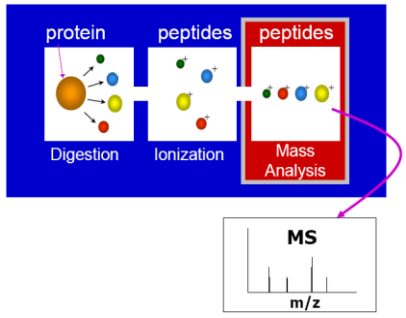
Strengths

- Precise molecular weight
- Fragmentation
- Automated

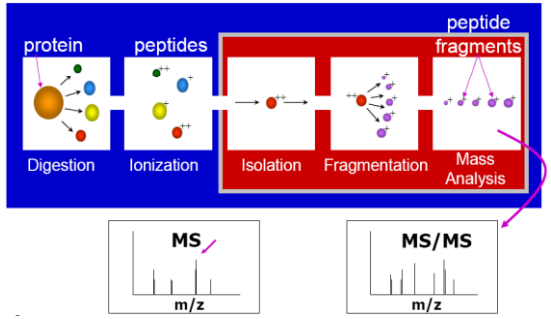
Weaknesses

- Best for a few molecules at a time
- Best for small molecules
- Mass-to-charge ratio, not mass
- Intensity ≠ Abundance





Single Stage MS

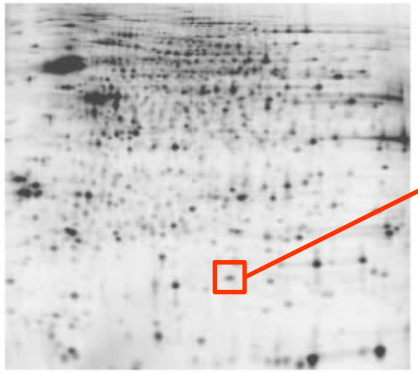


Tandem MS

Peptide Mass Fingerprinting

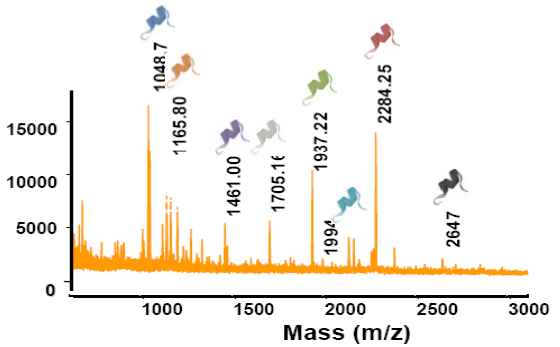
A protein identification technique, that correlates experimental data with theoretical data.

Protein



Proteolytic digestion

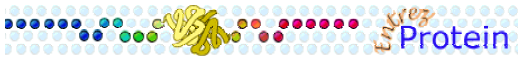
Experimental MS



1. Excise band from gel
2. Tryptic Digestion of gel fragment
3. Supernatant transferred to fresh eppendorf
4. Sample transferred to target plate

Computer search

Protein sequence from database



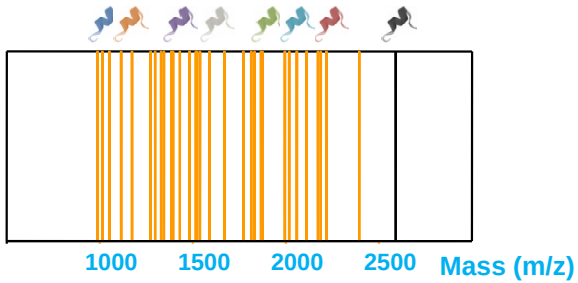
proteasome activator protein PA26
[*Trypanosoma brucei*]

```
  1 MPPKRAALIQ NLRDSYTETS SFAVIEEWAA
 31 GTLQEIEGIA KAAAEAHGTI RNSTYGRAQA
 61 EKSPEQLLGV LQRYQDLCHN VYCQAETIRT
 91 VIAIRIPEHK EEDNLGVAVQ HAVLKIIDEL
121 EIKTLGSGEK SGSGGAPTPI GMYALREYLS
151 ARSTVEDKLL GSVDAESGKT KGGSQSPSLL
181 LELRQIDADF MLKVELATTH LSTMVRAVIN
211 AYLLNWKKLI QPRTGSDHMV S
```

In silico digestion

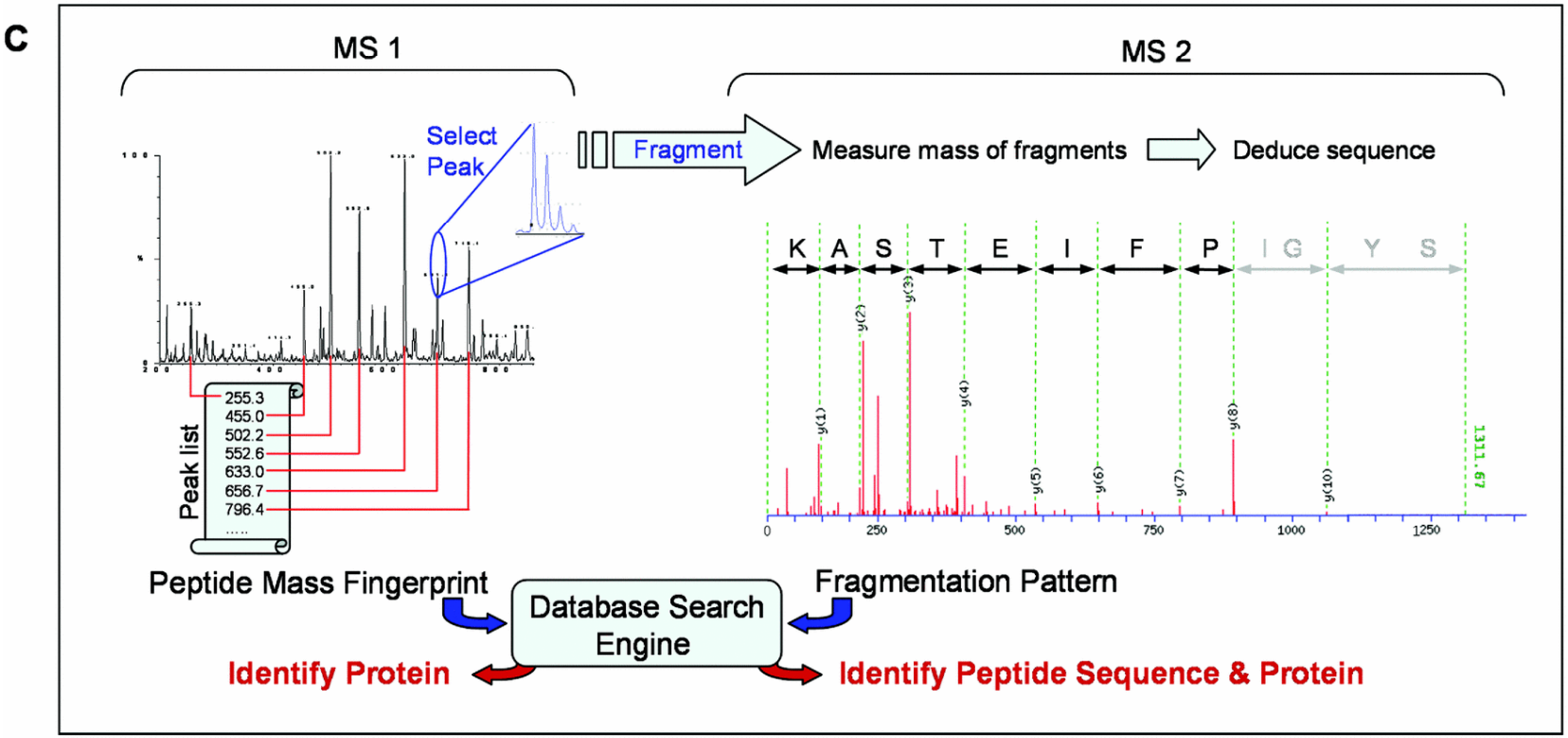
```
TGTDHMVS
AALIQNLR
IIDELEIK
AAAEAHGTIR
LLGSVDAESGK
QIDADFMLK
SPEQLLGVLQR
AVINAYLLNWK
GGSQSPSLLLELR
VELATTHLSTMVR
SGSGGAPTPIGMYALR
EEDNLGVAVQHAVLK
YQDLCHNVYCQAETIR
DSYTETSSFAVIEEWAAGTLQEIEGIAK
```

Theoretical MS



1. *In silico* digest
2. Mass computation
3. Compare computer generated masses with observed spectrum

Some Principles of Mass Spectrometry



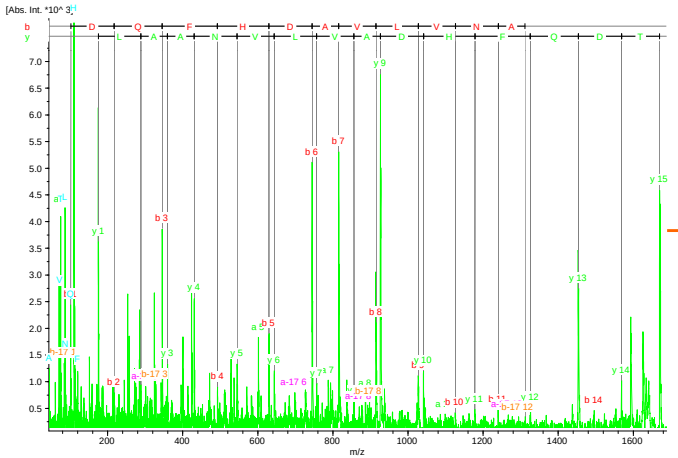
MS or MS/MS

Peptide Mass Fingerprint

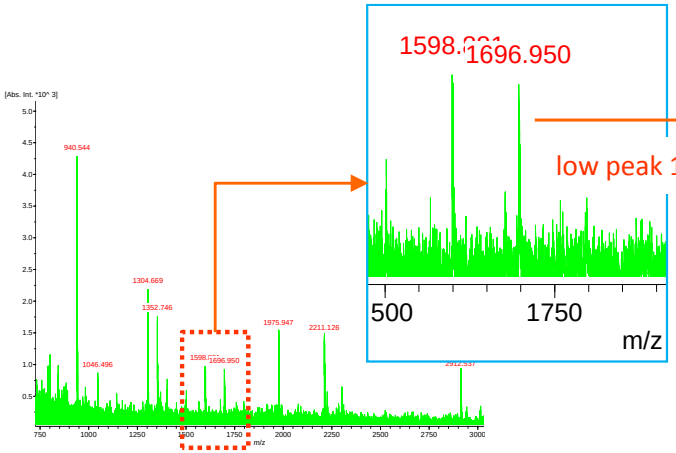
- Fast, simple analysis
- High sensitivity
- Need database of protein sequences
- Sequence must be present in database
- Not good for mixture

Peptide Fragmentation Search

- Easily automated for high throughput
- Can get matches from marginal data
- Can be slow
- MS/MS is peptide identification

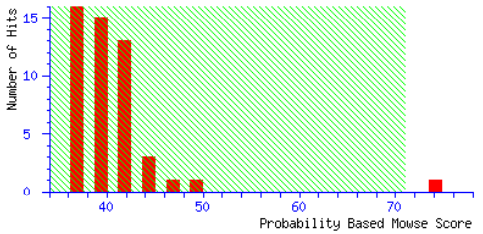


MS/MS from 1669 amu: TDQFHDAVLVNAALR



low peak 1697 amu selected

In-Gel (Spot 16) Digest MS Spectrum



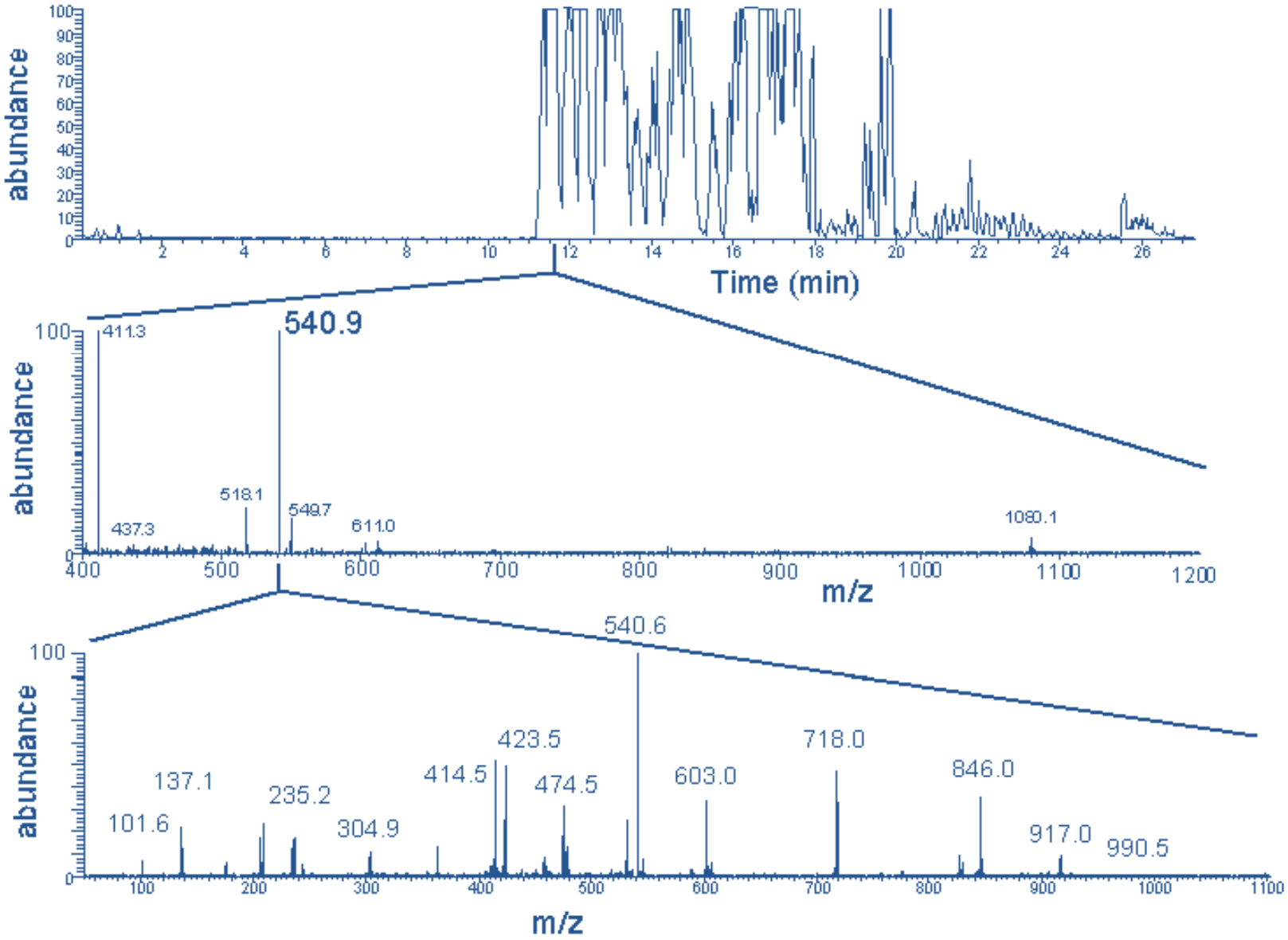
Mascot Score for UltraFlex TOF/TOF of 1669.93

LC/MS/MS of Peptide Mixtures

LC

MS
(MW Profile)

MS/MS
(AA Identity)

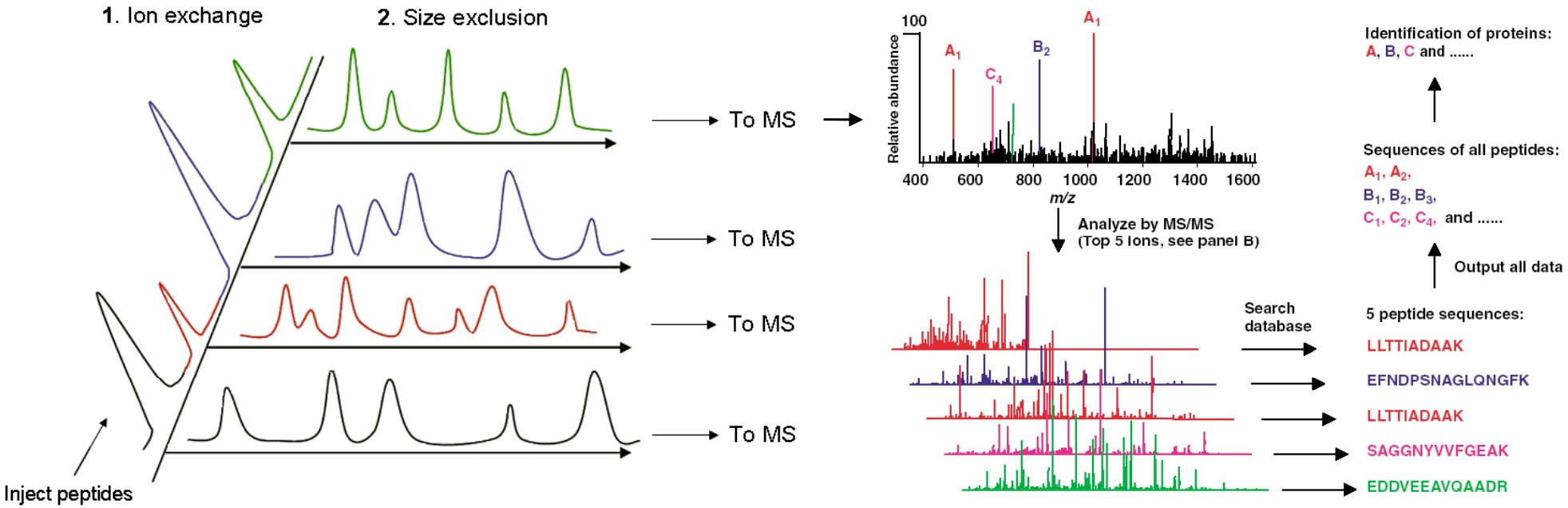


Advantages of LC/MS over 2D PAGE

- Easier automation
- no robots
- no gel handling
- Better separation power
- Monolithic columns
- Multi-dimensional chromatography
- Simpler coupling to MS

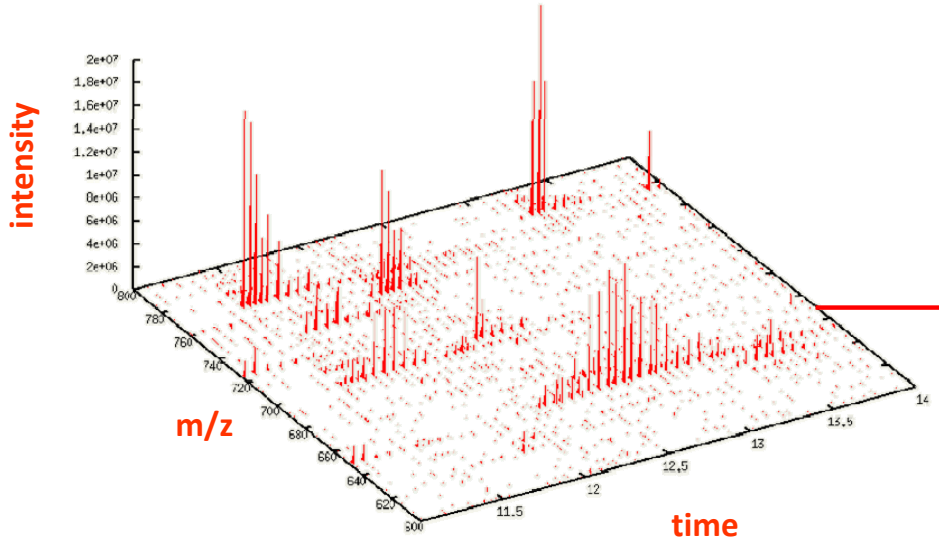
Disadvantages of LC/MS

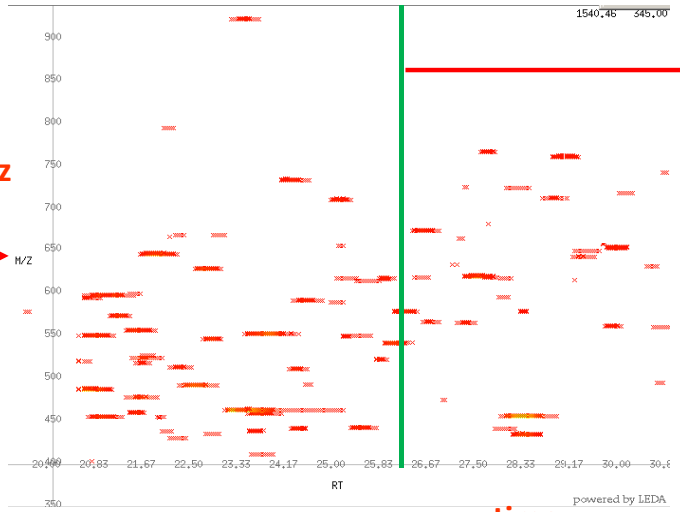
- Proteins are chopped up
- Quantization difficult
- Huge amount of data (~10 GB/run)
- Data hard to manage/interpret



LC/MS/TOF

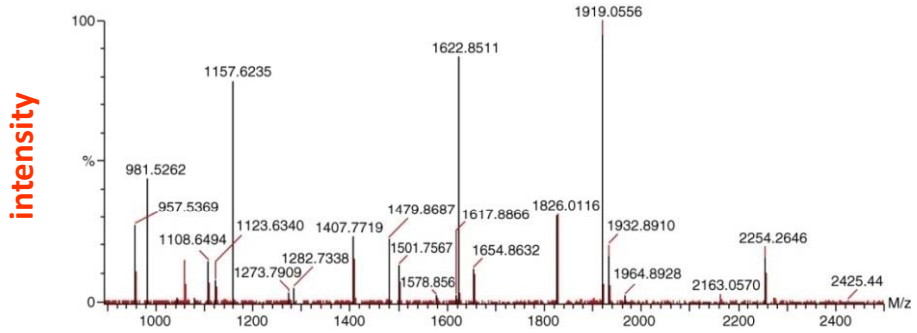
3D view: m/z, intensity, time

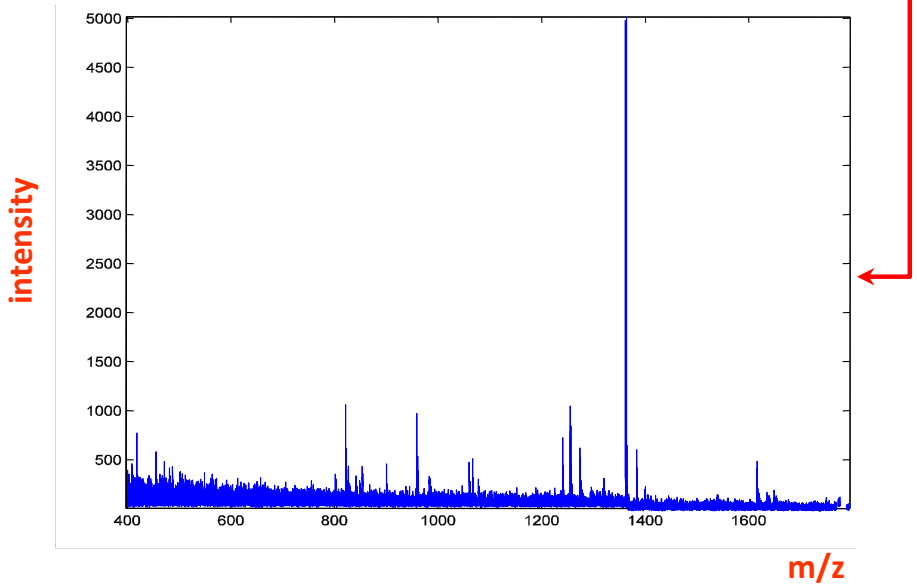




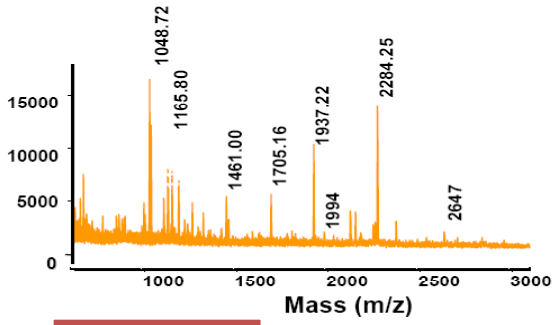
MALDI-TOF

2D view: m/z, intensity

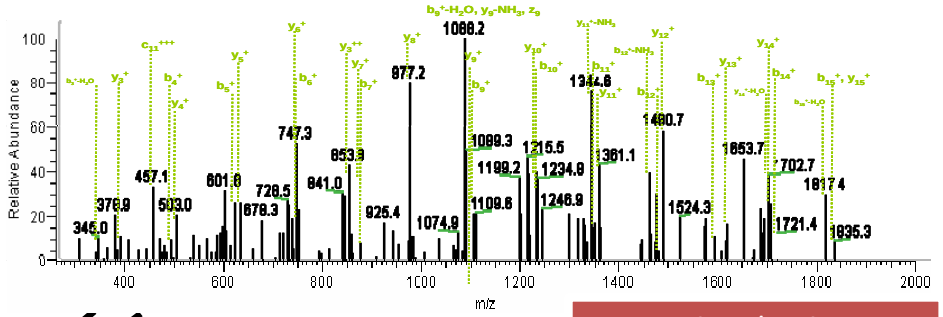




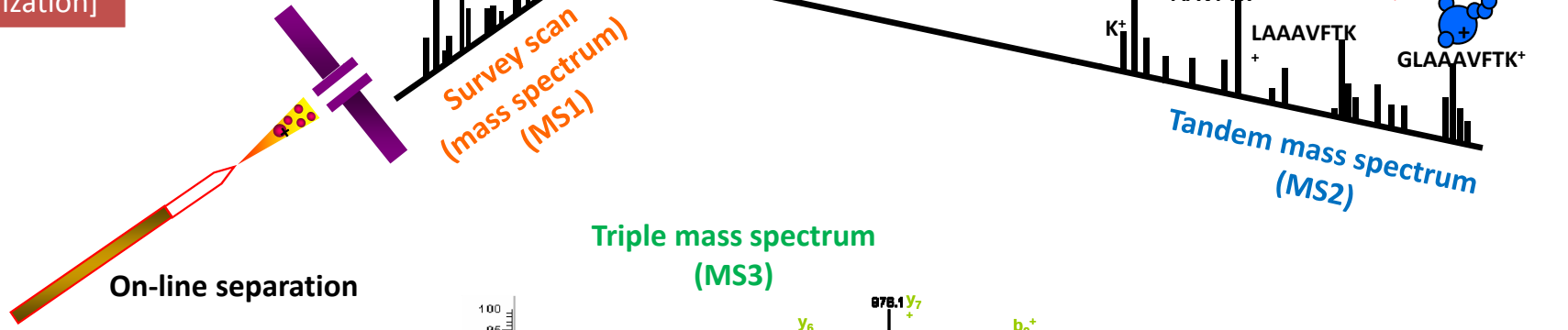
Peptide Sequencing by LC/MS/MS



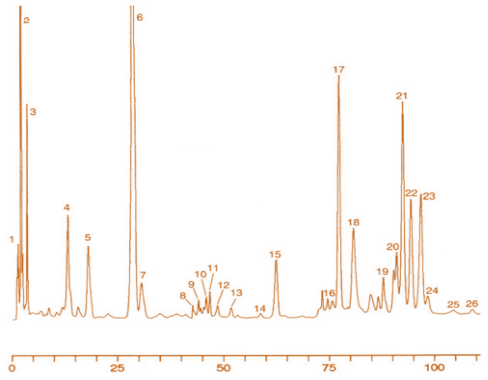
STEP 2 {MS} [ionization]



STEP 3 {MS/MS} [peptide identification]

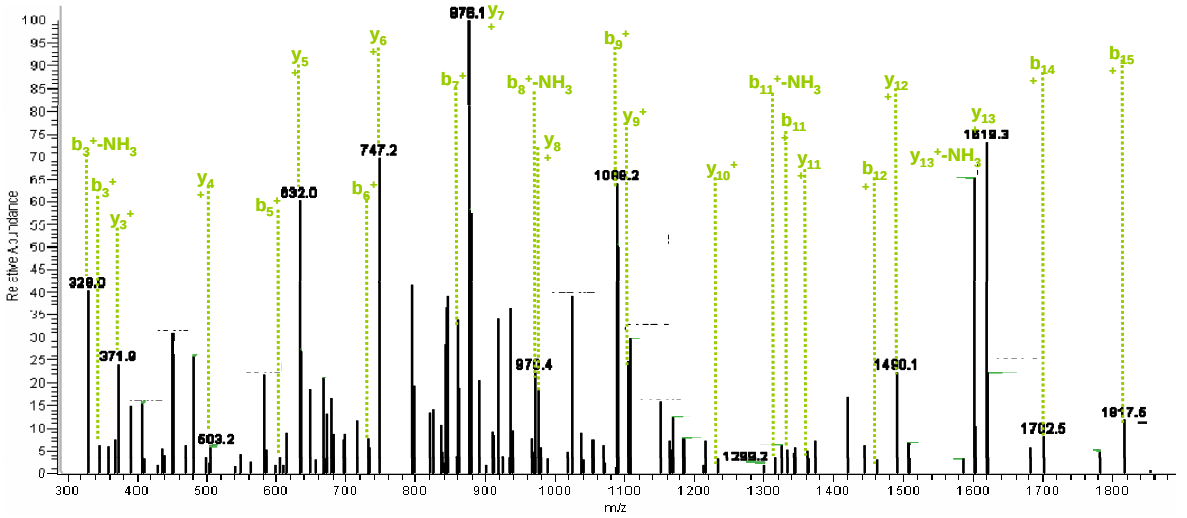


On-line separation



STEP 1 {LC} [separation 1]

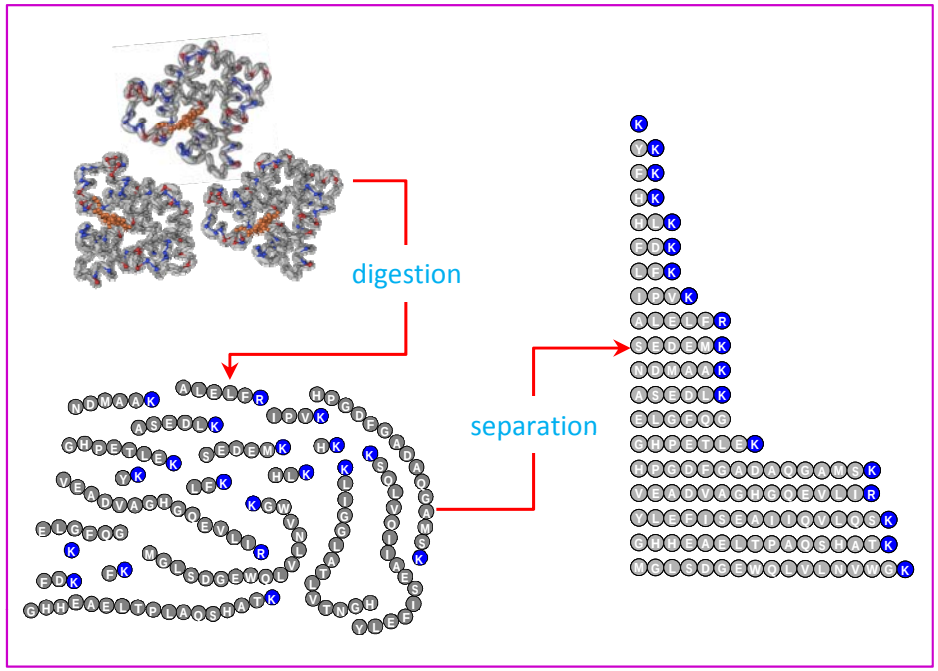
Triple mass spectrum
(MS3)

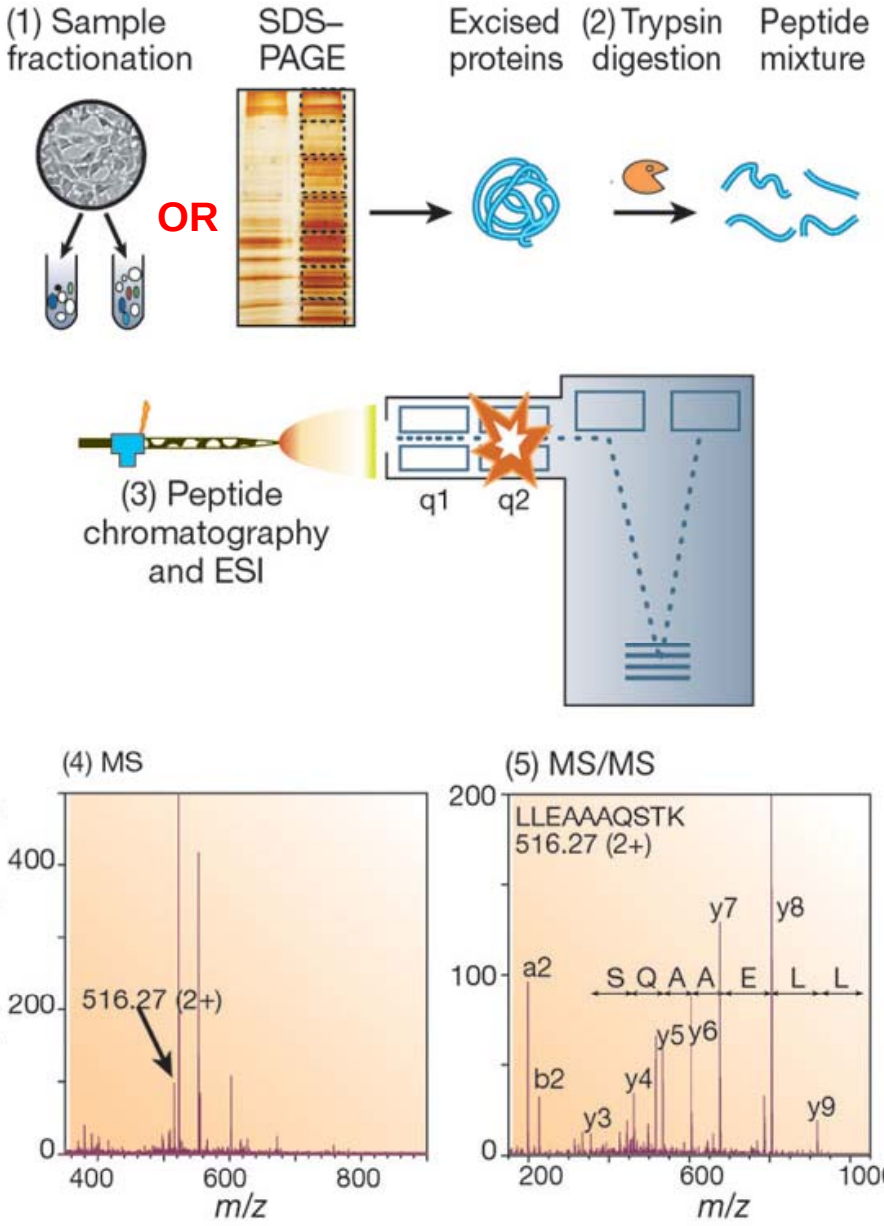


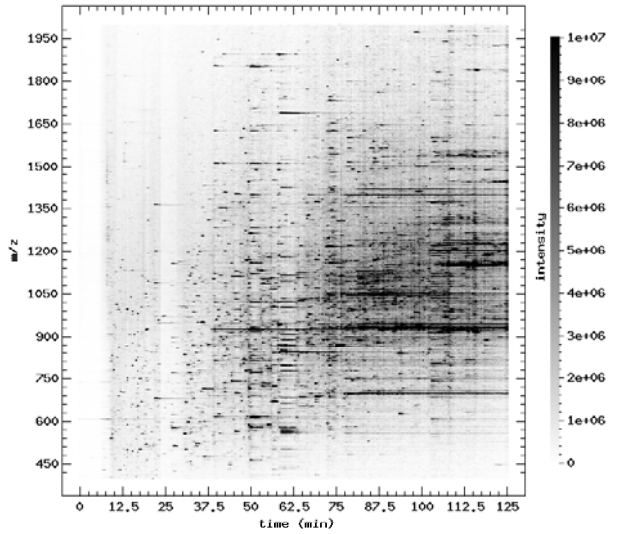
STEP 4 {MS/MS/MS} [post translation modification]

Shotgun Proteomics

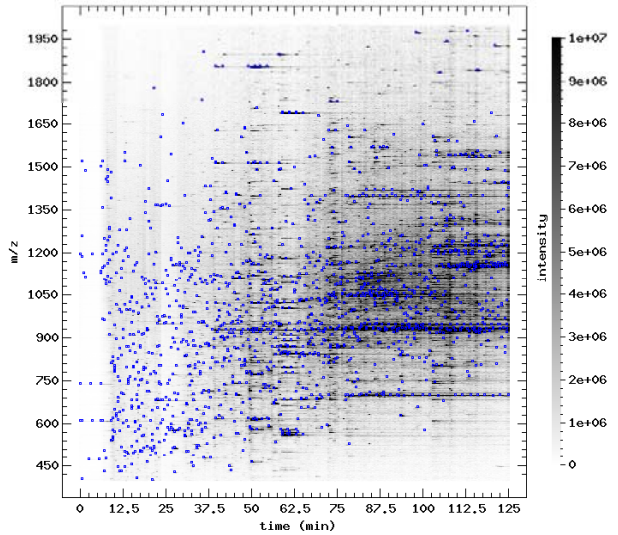
- Separation of peptides
  - Extensive chromatographic separation (one or multiple dimensional separations)
- Data acquisition
  - Data-dependent acquisition (Automated acquisition of MS/MS spectra from as many precursor ions as possible)
- Data analysis
  - Automated interpretation of the MS/MS spectra (Database Search)



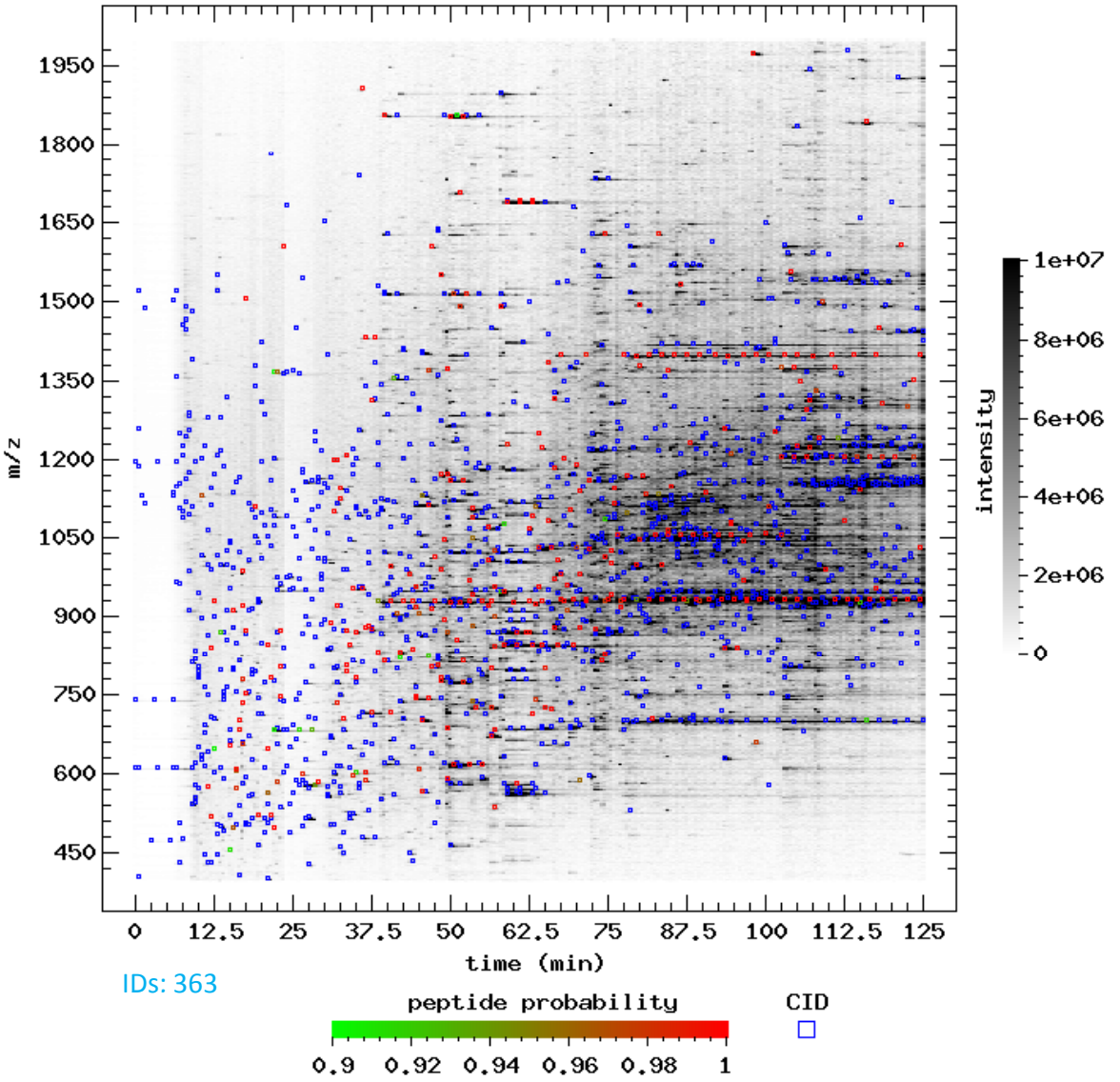




Features: 2720



CIDs: 1633

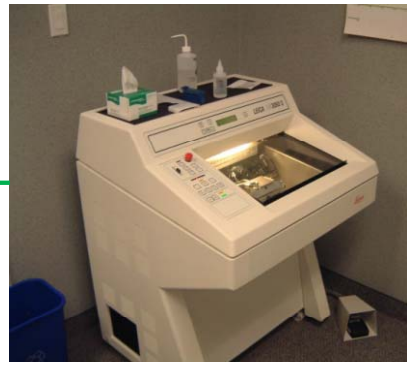


IDs: 363

Finally: ID/CID: 22%, ID/feature: 13%

MALDI Imaging Mass Spectrometry (IMS)

1. The analysis of a sample surface for its molecular content
   - Sample consists of a thinly sliced section of tissue/organ/whole animal
2. A 2 dimensional array of MALDI spectra are obtained over the surface of the sample
   - each spectra has location component
3. An ion intensity map can then be produced for any mass that is detected over the scanned area.

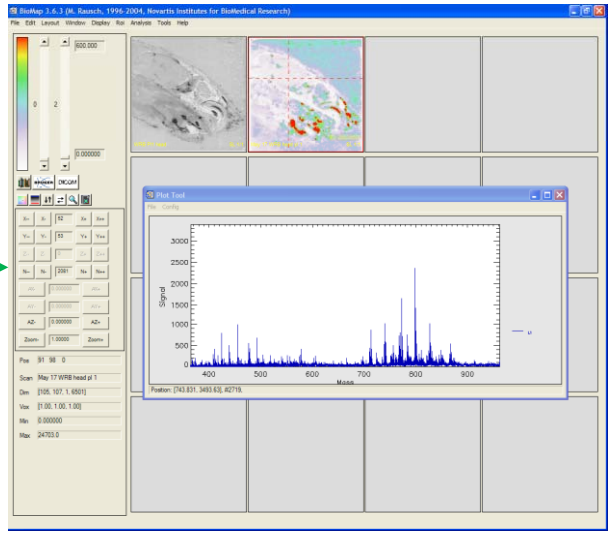


Slice Tissue Sections

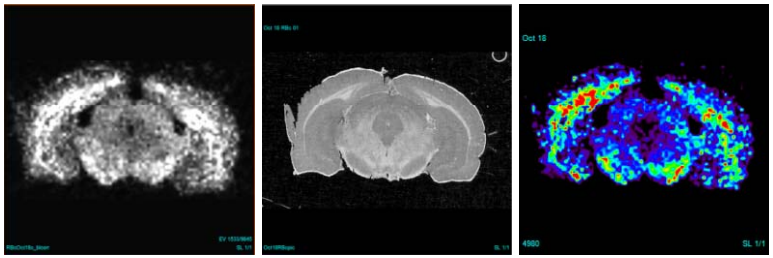
Mount Sample

Spray Coat with Matrix

MALDI/TOF/TOF



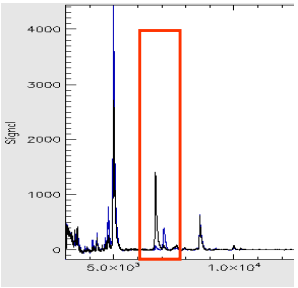
Data processing and analysis

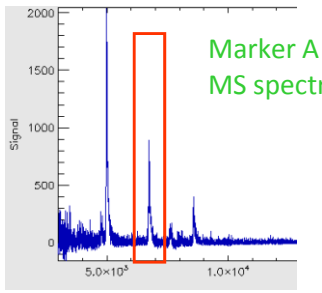


MS image Optical image merged image

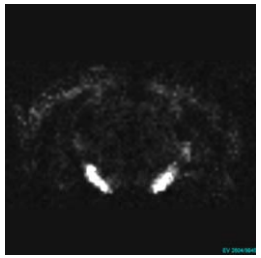
House keep protein

compare spectra from multiple ROIs



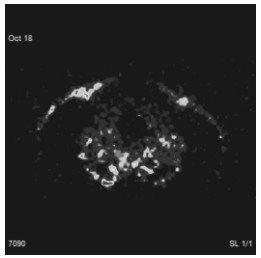


Marker A
MS spectrum

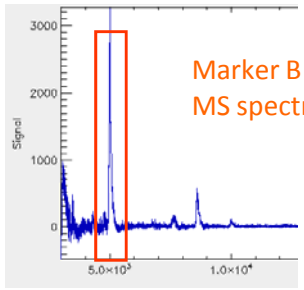


Marker A
MS image

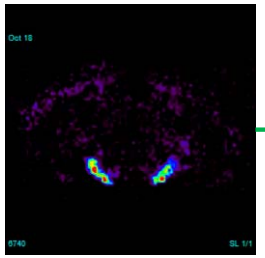
relative intensity

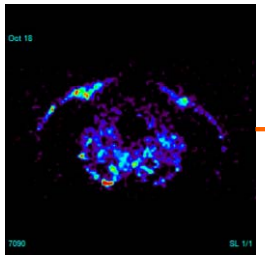


Marker B
MS image



Marker B
MS spectrum





Optical image

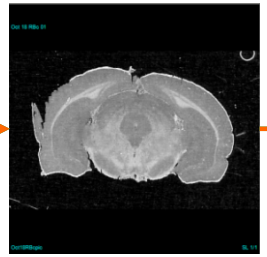
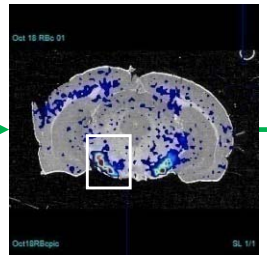
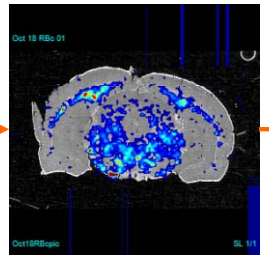
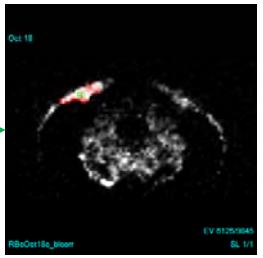


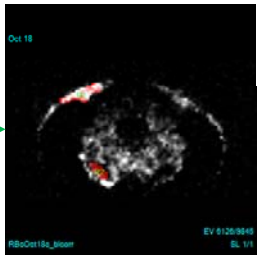
image overlay



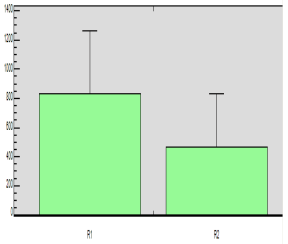


mark of ROIs
(regions of interest)



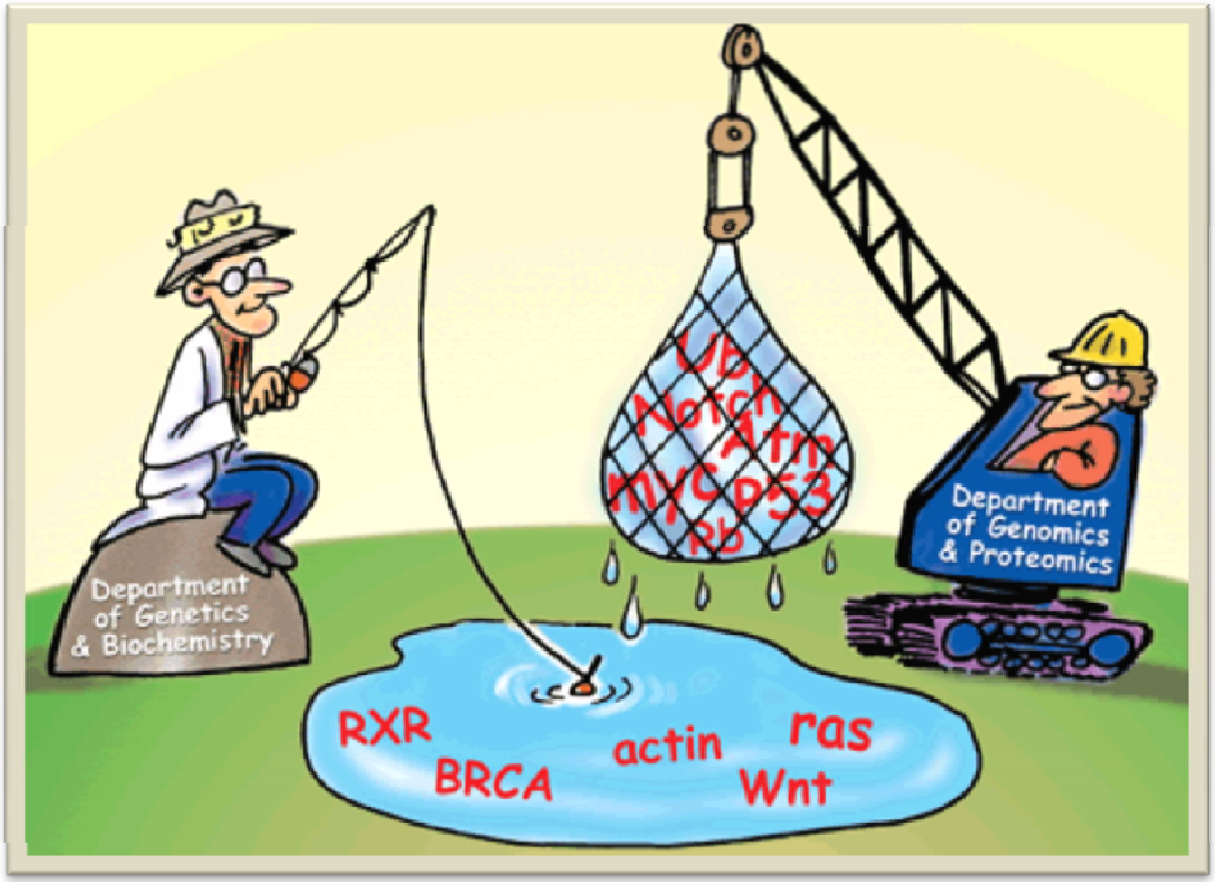


compare intensity of detection protein in specific regions



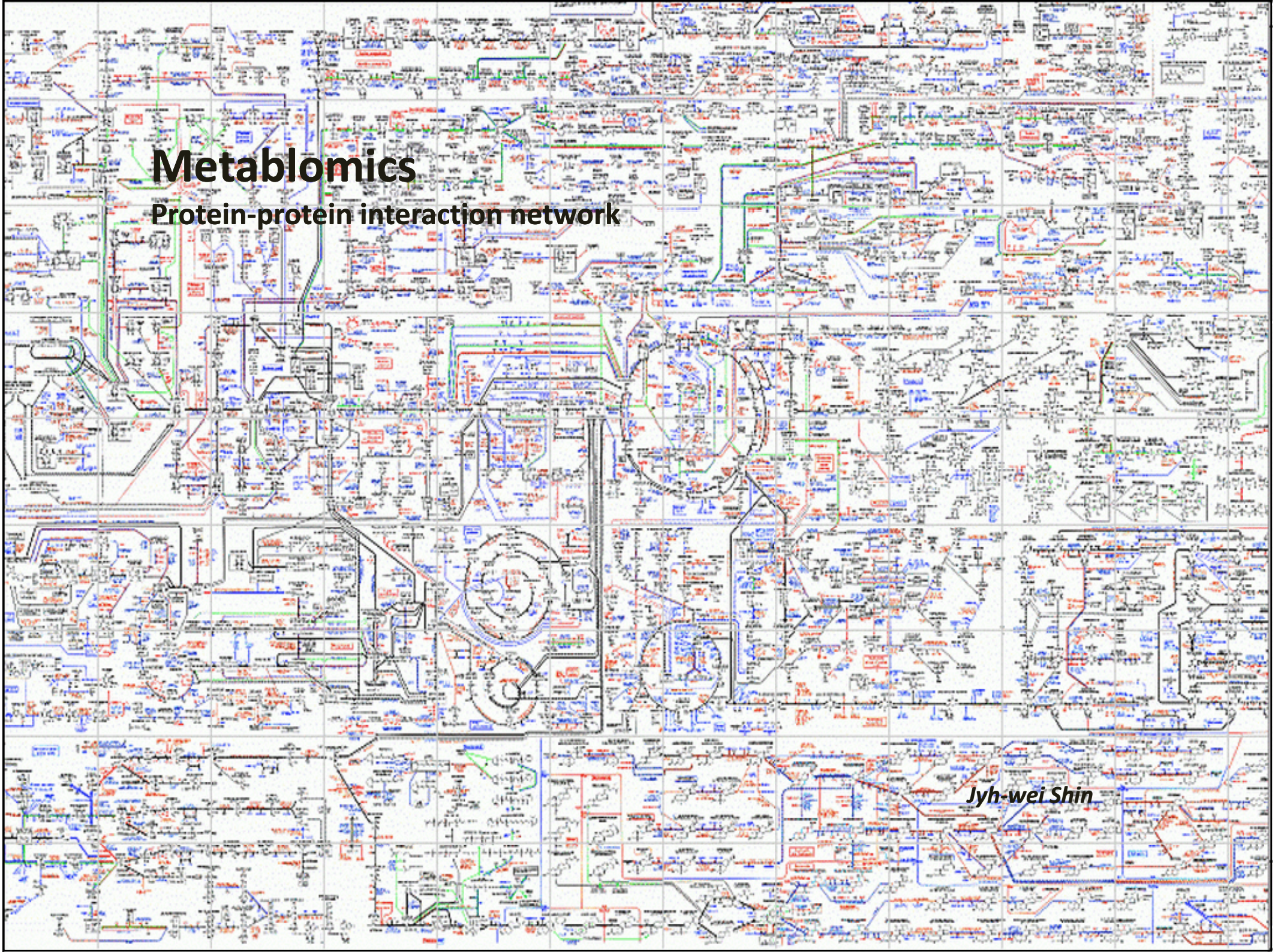
compare average intensity of a detected protein in specific regions

Proteomics includes not only the identification and quantification of proteins, but also the determination of their **localization, modifications, interactions, activities, and, ultimately, their function**. Initially encompassing just two-dimensional (2D) gel electrophoresis for protein separation and identification, proteomics now refers to any procedure that characterizes large sets of proteins.



Science 16 February 2001: Vol. 291. no. 5507, pp. 1221 – 1224
PROTEOMICS: Proteomics in Genomeland
Stanley Fields

Metabolomics

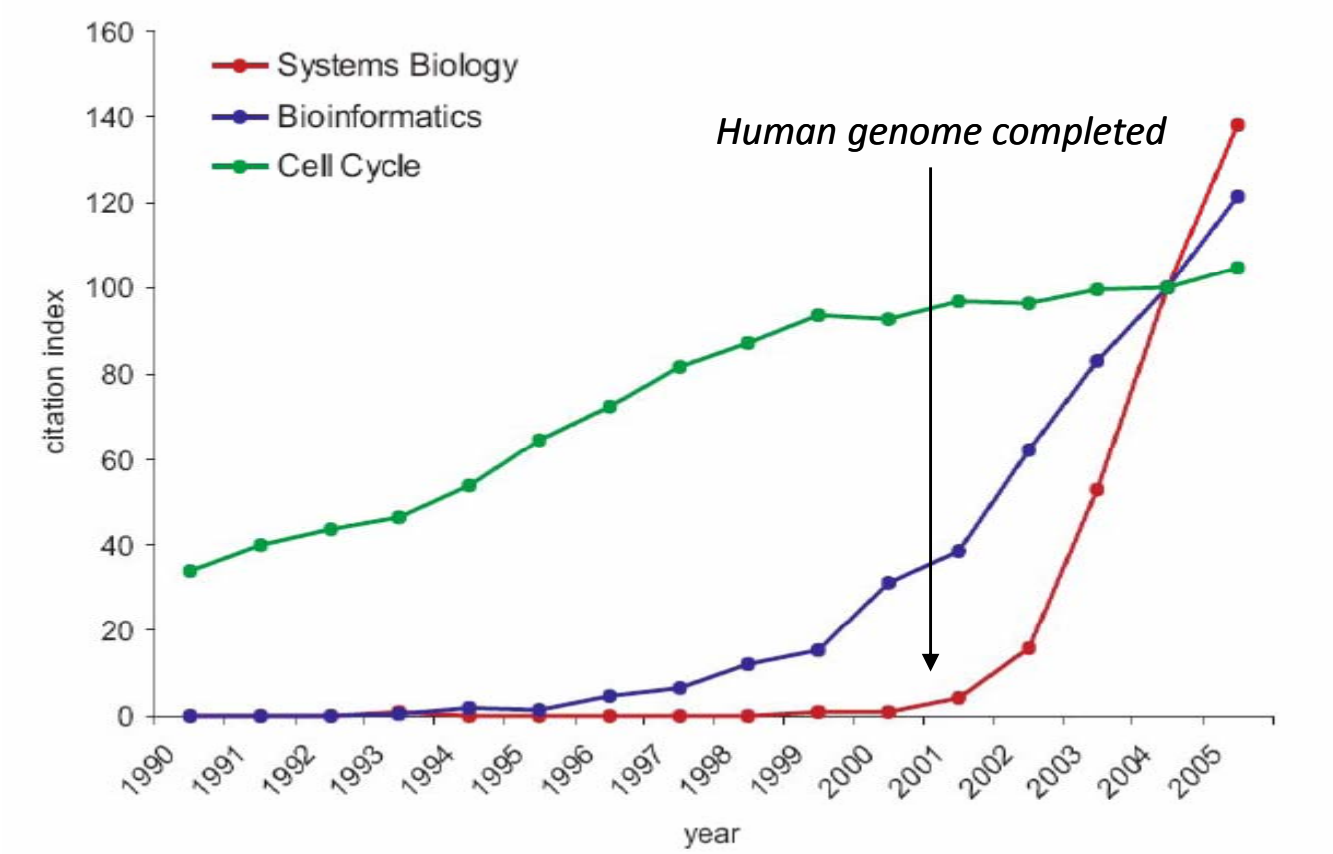

Metablomics
Protein-protein interaction network
Jyh-wei Shin

Potential benefits of genomics, proteomics and metabolomics for the patients

Patient		Genomics	Proteomics	Metabolomics
Susceptibility	Predisposition	●●●	●	●
	Environmental Factors	●	●●●	●●●
	Risk	●●	●●	●●
	Prevention	●	●●	●●
Diagnosis	Early Diagnosis	●	●●●	●●●
	Differential D Stratification	●●	●●●	●●●
Treatment	Choice	●	●●●	●●●
	Response	●	●●●	●●●
	Alternatives	●	●●	●●
Aftercare	Side Effects	●	●●	●●
	Monitoring	●	●●●	●●●
	Prevention	●	●●	●●

Interest in Systems Biology?

1. Understand the *structure* of the system (Regulatory and biochemical networks)
2. Understand the *dynamics* of the system (Construct model with predictive capabilities)
3. Understand the *control* methods



A Broad Definition of Bioinformatics

- **Informatics**
- Its carrier is a set *of digital codes and a language*.
- In its manifestation in the space-time continuum, it has utility (e.g. to decrease entropy of an open system).

- **Bioinformatics**
- *The essence of life is information* (i.e. from digital code to emerging properties of biosystems.)
- Bioinformatics is the study of information content of life

Framework for Systems Biology

- **The components**. *Discover all* of the genes in the genome and the subset of genes, proteins, and other small molecules constituting the pathway of interest. If possible, define an initial *model* of the molecular interactions governing pathway function (how?).

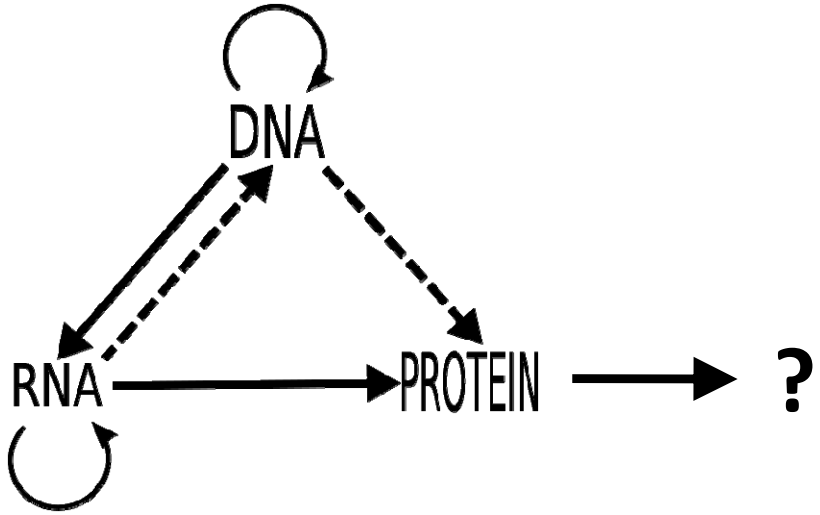
- **Pathway perturbation**. Perturb each pathway component through a series of *genetic* or *environmental* manipulations. Detect and quantify the corresponding *global* cellular response to each perturbation.

- **Model Reconciliation.** *Integrate* the observed mRNA and protein responses with the current, pathway-specific model and with the global network of protein-protein, protein-DNA, and other known physical interactions.

- **Model verification/expansion.** Formulate new *hypotheses* to explain observations not predicted by the model. Design additional perturbation experiments to test these and iteratively repeat steps (2), (3), and (4).

Crick's 1958 central dogma

The central dogma of molecular biology deals with the detailed residue-by-residue transfer of sequential information. It states that information cannot be transferred back from protein to either protein or nucleic acid.

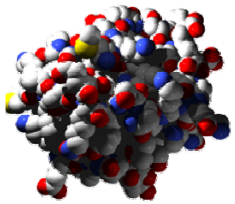
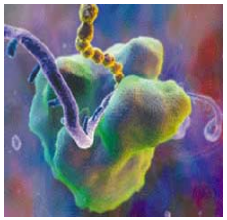
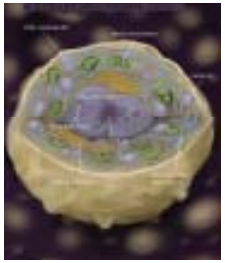
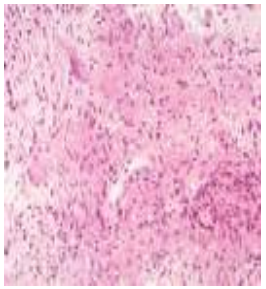


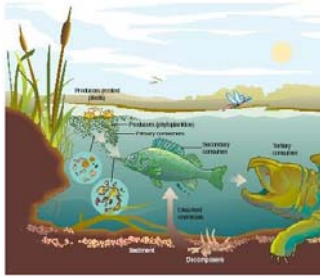
Genotype, Phenotype, Environment

- ***Genotype*** *is the identity of the genes (genetic constitution) in an individual*
  - *Specifies "form" (**allele**) of genes*
- ***Phenotype*** *is the observed traits*
- ***Environment*** *is outside world in which the organism lives*

- *What is the relationship between genotype, phenotype, and environment?*
- *Most traits are complex and depend on multiple genetic and environmental factors (**multifactorial**)*
- *Can genetics alone unravel multifactorial traits?*
- *Can systems biology help us understand multifactorial traits?*

System heterogeneity in size and timescale

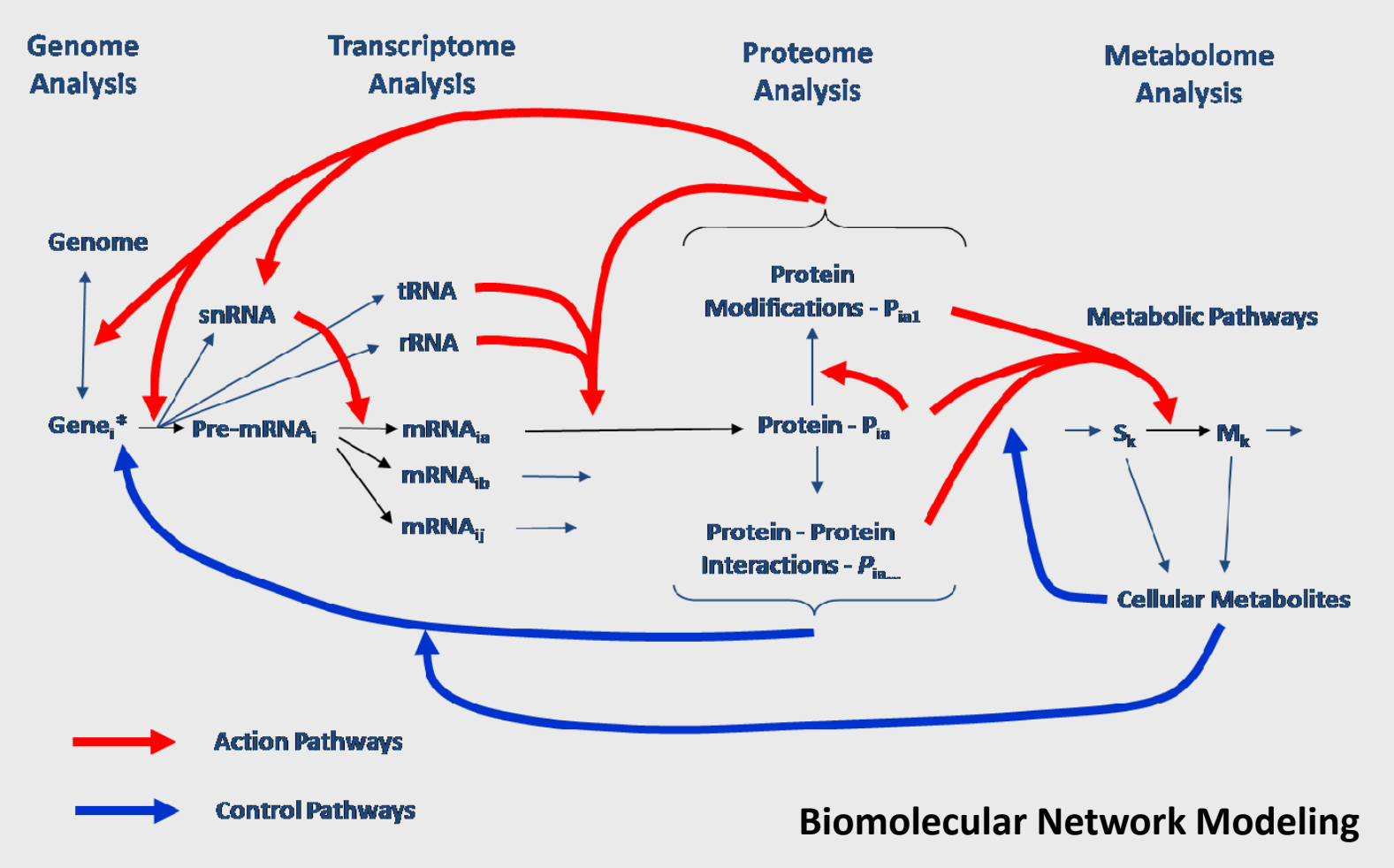


Scale	Atomic	Molecular	Cellular	Tissue	Organism	Ecosystem
Size	0.1 - 1.0 nm	1.0 - 10 nm	10 - 100 nm	0.01 - 1.0 m	0.01 – 4.0 m	1 – 1,000 km
Data format	Coordinate data Dynamic data	Interaction data K_{on}, K_{off}, K_d	Concentrations Diffusion rates	Metabolic input Metabolic output	Behaviors Habitats	Environmental impact Nutrient flow
Rx time	0.1 - 10 ns	10 ns - 10 ms	10 ms - 1000 s	1 s – 1 hr	1 hr – 100 yrs	1 yr – 1000 yrs
Rx pattern	Molecular dynamics	Interactions	Fluid dynamics	Process flow	Mechanics	Network Dynamics

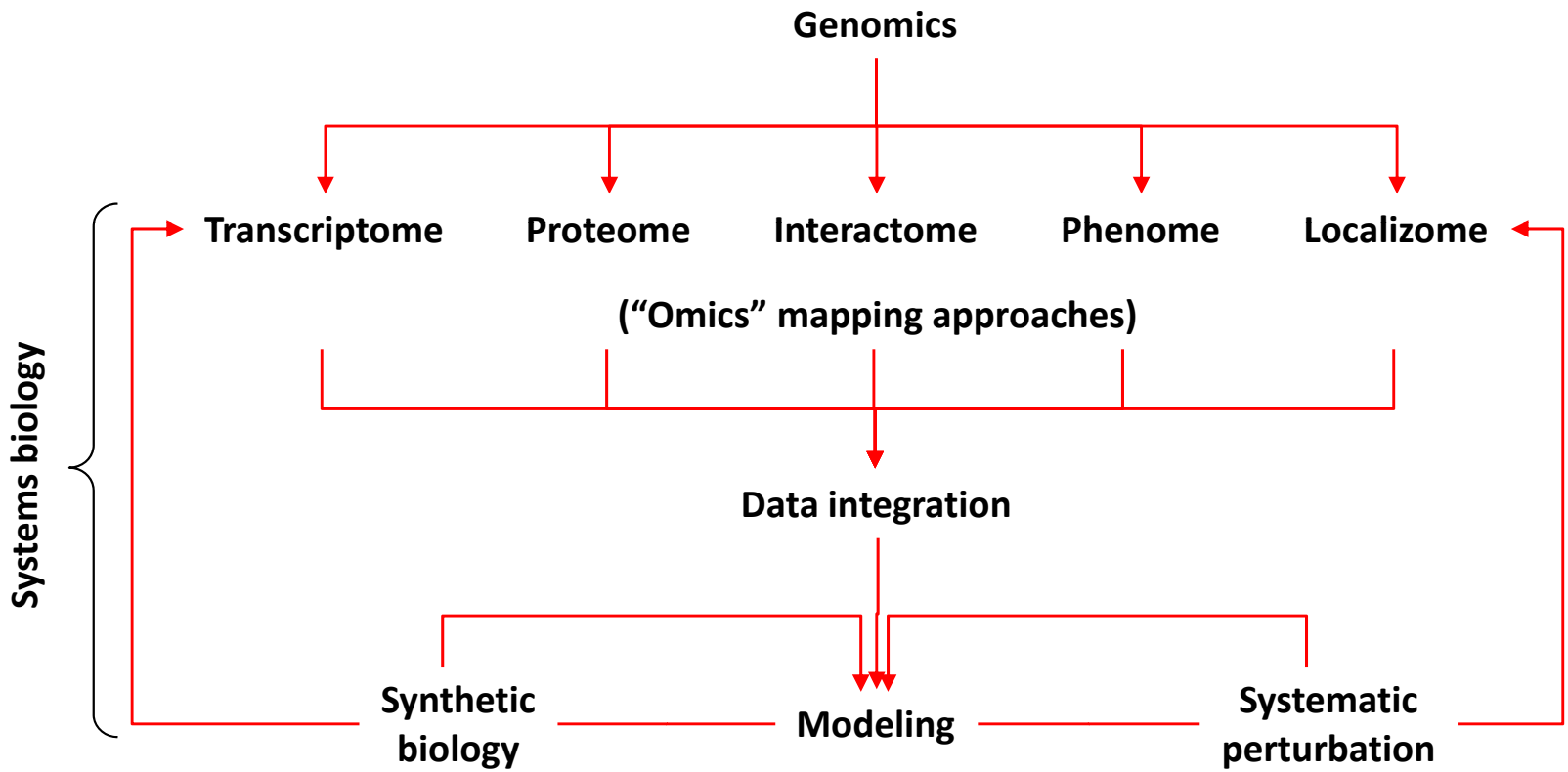
- If one scale (e.g., protein-protein interactions) behaves deterministically and with isolated components, then we can use plug-n-play approaches
- If it behaves chaotically or stochastically, then we cannot
- *Most biological systems lie between this deterministic order and chaos: Complex systems*

What we talked about this section

- *Protein-protein interactions*, chemical-protein interactions
- *Biological networks* – structures, properties & behavior



So how can we meaningfully integrate the data?



Increasing Level of Complexity

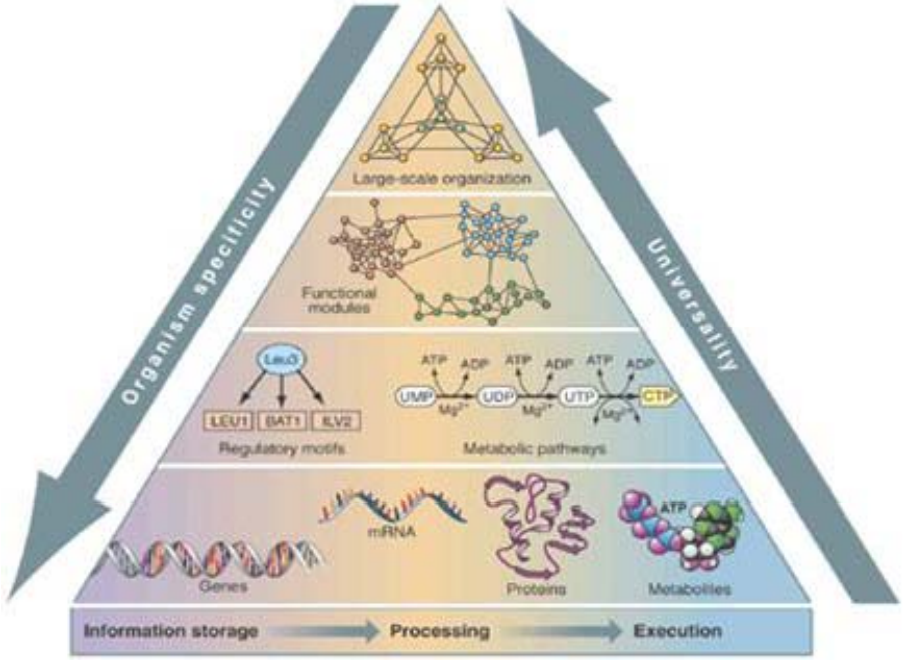
- **The genome**
  - **4** *bases*
  - 3 billion bp total
  - 3 billion bp/cell, identical

- **The proteome**
  - **20** *amino acids*
  - ~60K genes, ~200K proteins
  - ~10K proteins/cell; different cells/conditions, different expressions

- **The metablome**
  - ~200K *reactions*
  - ~20K pathways
  - ~1K pathways/cell; different cells/conditions, different expressions

A complex problem

- 35,000 genes either on or off (huge simplification!) would have $2^{35,000}$ solutions
- Things can be simplified by grouping and finding key genes which regulate many other genes and genes which may only interact with one other gene
- In reality there are lots of subtle interactions and non-binary states.

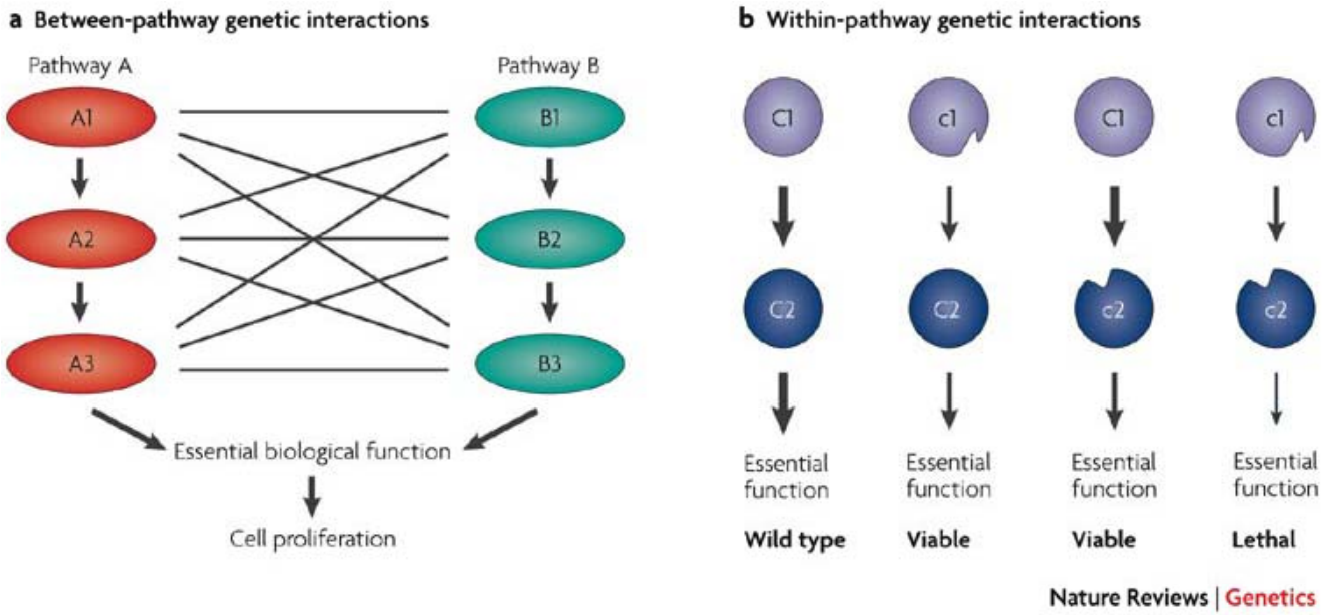


Interactions → Networks → Pathways

- **Interaction** : record stores a description of the binding event between two objects, A and B, which are generally molecules.
- **Network** : encod into a *complex intracellular web* of molecular interactions or molecular complex.
- **Pathway** : define as a group of molecules that are generally free from each other, but form a network of interactions usually to mediate some cellular function.
- There are **3** type of interactions that can be mapped to pathways:
  1) *protein (enzyme) - metabolite (ligand)* : metabolic pathways
  2) *protein - protein* : cell signaling pathways, complexes for cell processes
  3) *protein - gene* : genetic networks

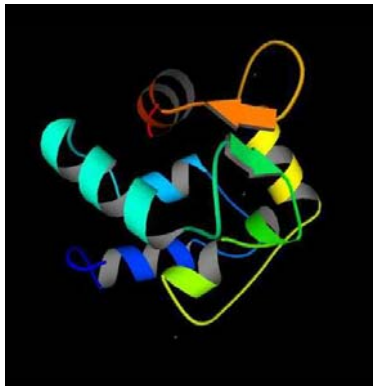
Interactions → Networks → Pathways

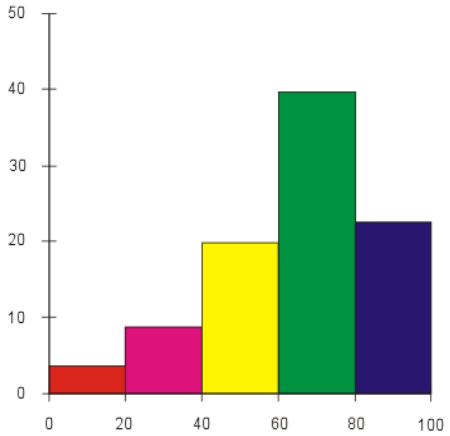
- A collection of interactions defines a network
- Pathways are a subset of networks
  - ✓ All pathways are networks of interactions, however *not all networks are pathways!*
  - ✓ Difference in the level of annotation/understanding
- We can define a pathway as a biological network that relates to a **specific** physiological process or phenotype.

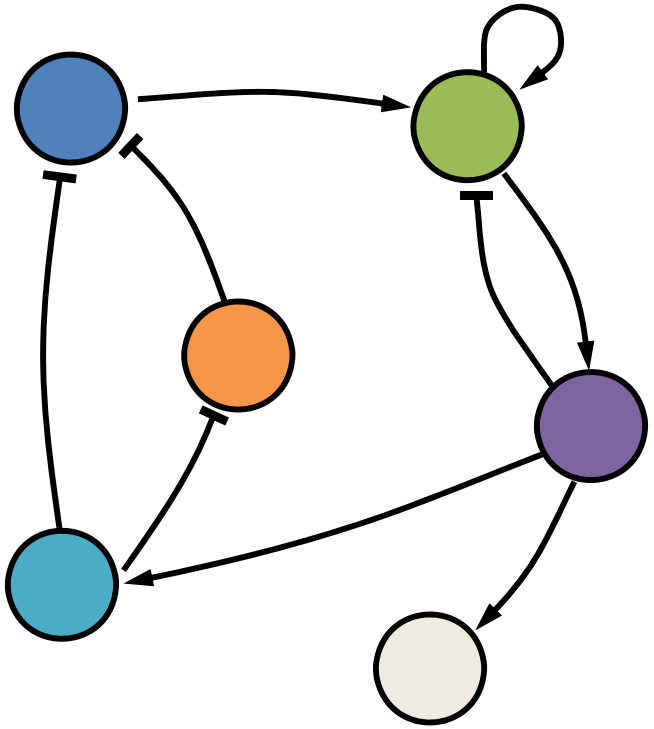


Comparing Genetic Interaction to Protein - Protein Interaction and Gene Regulatory Networks

- PPI (protein-protein interaction) = binding
- GRN (gene regulatory network) = expression
- SGA (synthetic gene array) = function
- Network of functional interactions
- SGA complements PPIs and GRNs

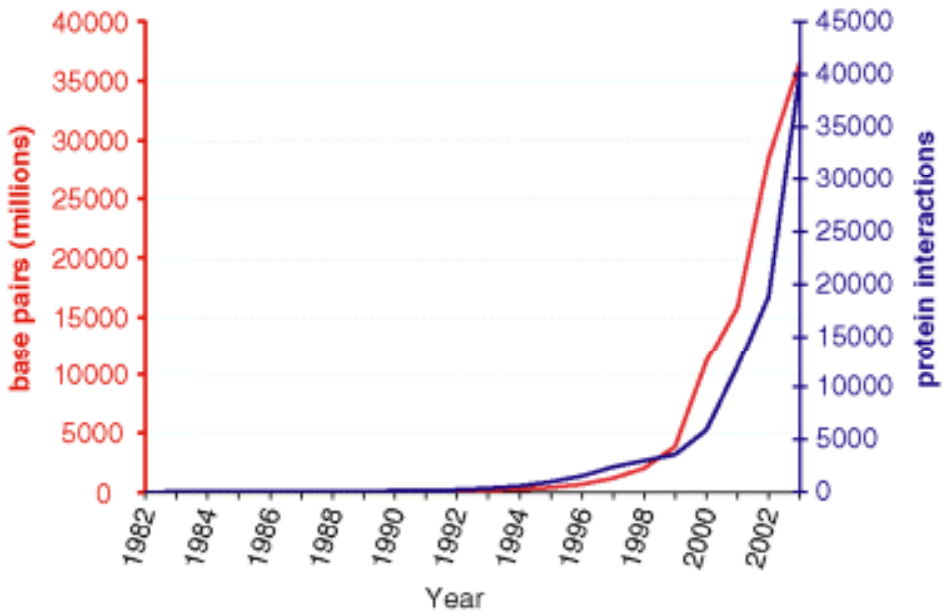






Importance of protein-protein interactions

- Many cellular processes are regulated by multiprotein complexes
- Distortions of protein interactions can cause diseases
- Protein function can be predicted by knowing functions of interacting partners ("guilt by association")



A comparison of sequence (GenBank) and protein-protein interaction data (DIP database)

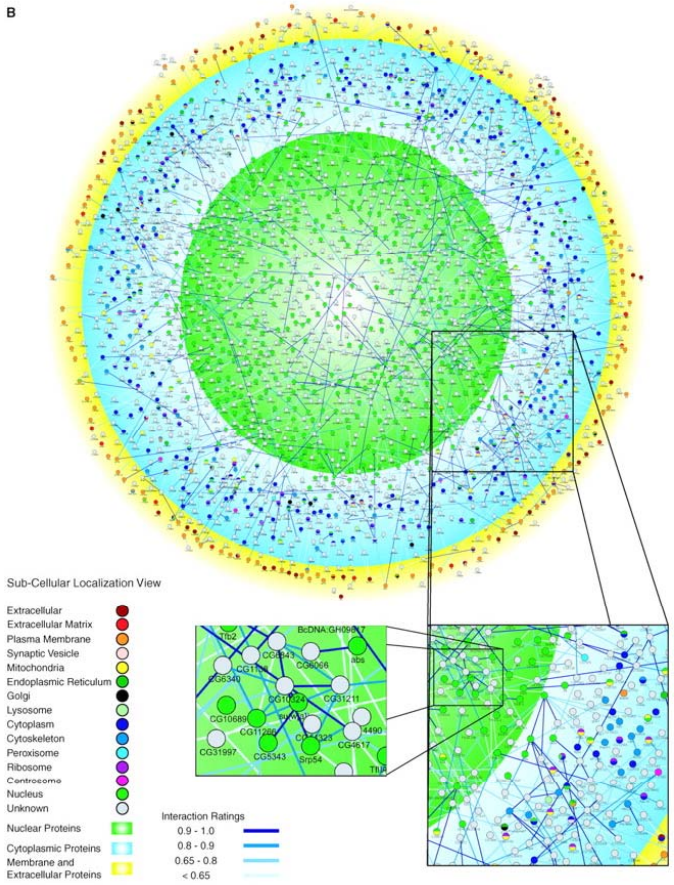
Methods of identifying PPIs

Experimental

- Protein-protein arrays
- Yeast Two Hybrid assay
- Tandem Affinity Purification assay

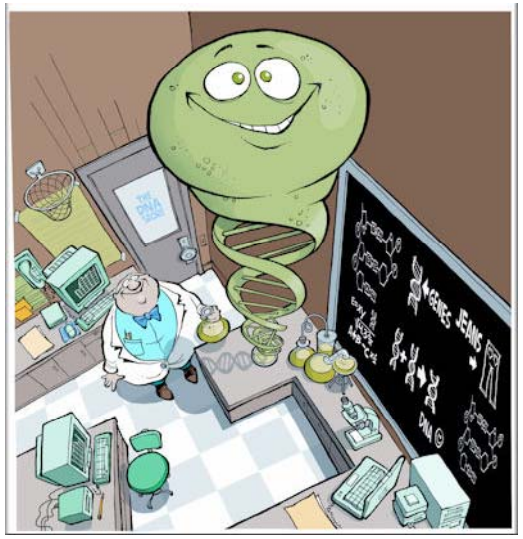
Computational/Inferential

- Interolog analysis
- Co-localization, co-expression
- Correlated mutations
- Text-mining



Two ways of looking a problem

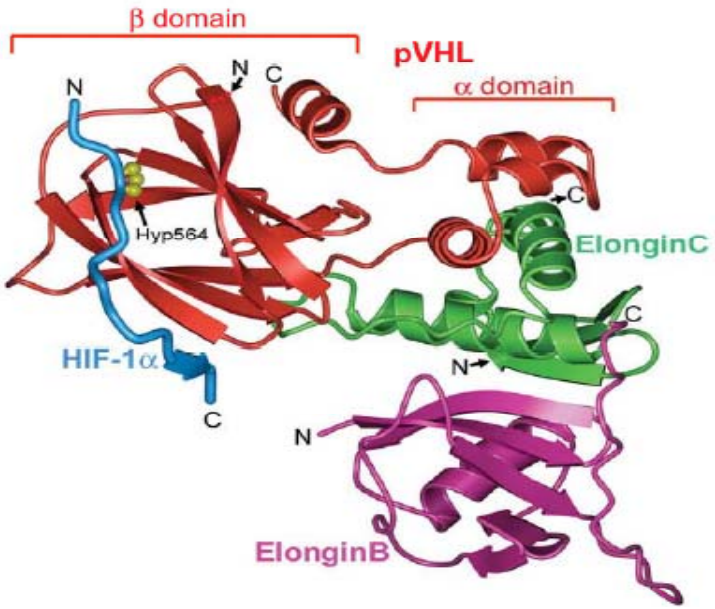
- *Top down* or *bottom up*
  - Either look at the whole organism and abstract large portions of it
  - Or try to understand each small piece and then after understanding every small piece assemble into the whole
  - Both are used, valid and complement each other



Interactions → Networks → Pathways

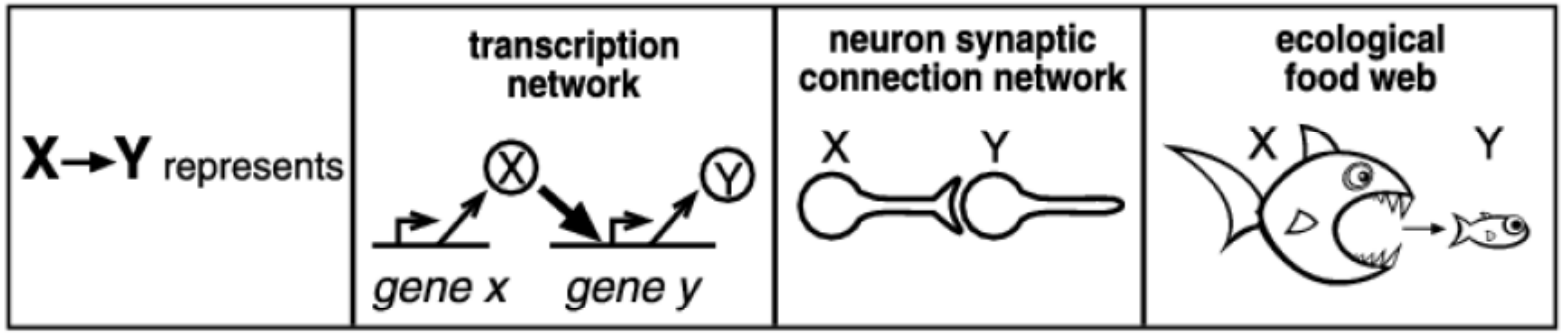
Why do we need protein interaction information

- Learning protein functions:
  - If two proteins interact, there is a very high possibility that their *functions* are related as well
- Cellular operations are largely endured by interactions among proteins
- From protein pathways to understanding cells, tissues, … to life and evolution



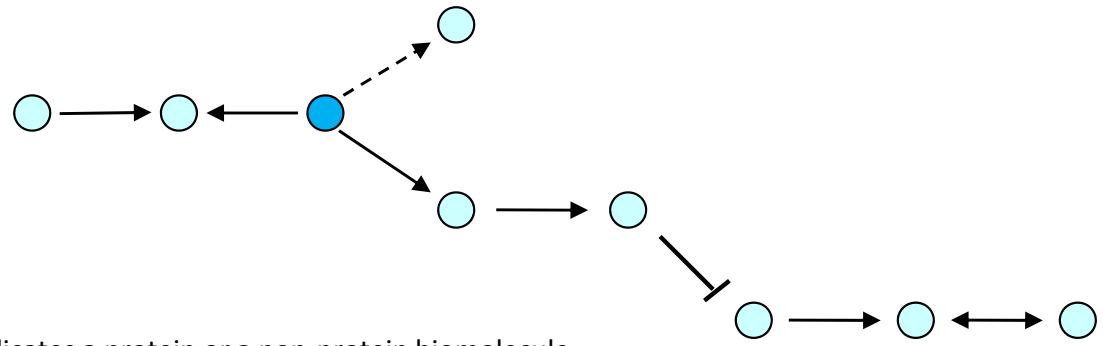
Interactions → **Networks** → Pathways

- A network is the *backbone* of a complex system.
- A group or system of (electric) components and *connecting circuitry* designed to function in a specific manner.
- Such motifs are found in networks from:
  - *Engineering* : Electronic circuits, World Wide Web *et al*.,
  - *Biochemistry* : Transcriptional regulation networks
  - *Neurobiology* : Neuron connectivity
  - *Ecology* : Food webs



Interactions → Networks → **Pathways**

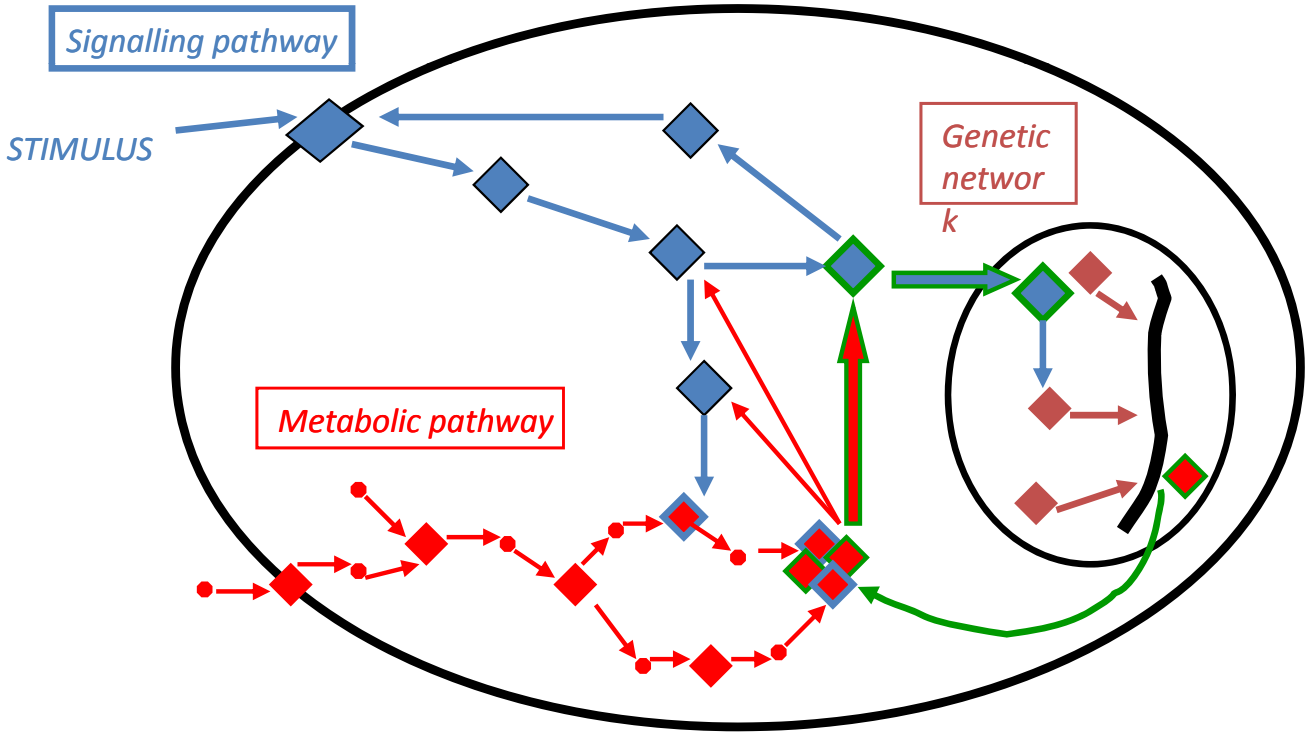
- It can be defined ad a modular unit of interacting molecules to fulfill a cellular function.
- It is usually represented by a *2-D diagram with characteristic symbols* linking the protein and non-protein entities.



A circle indicates a protein or a non-protein biomolecule.
An symbol in between indicates the nature of molecule-molecule interaction.

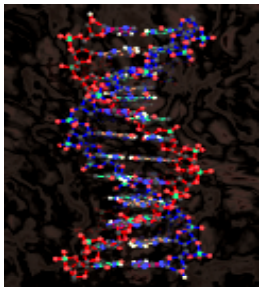
Biological Pathways

- There are 3 type of interactions that can be mapped to pathways:
  1) *enzyme – ligand* : metabolic pathways
  2) *protein – protein* : cell signaling and complexes for cell processes
  3) *gene regulatory elements* – gene products : genetic networks



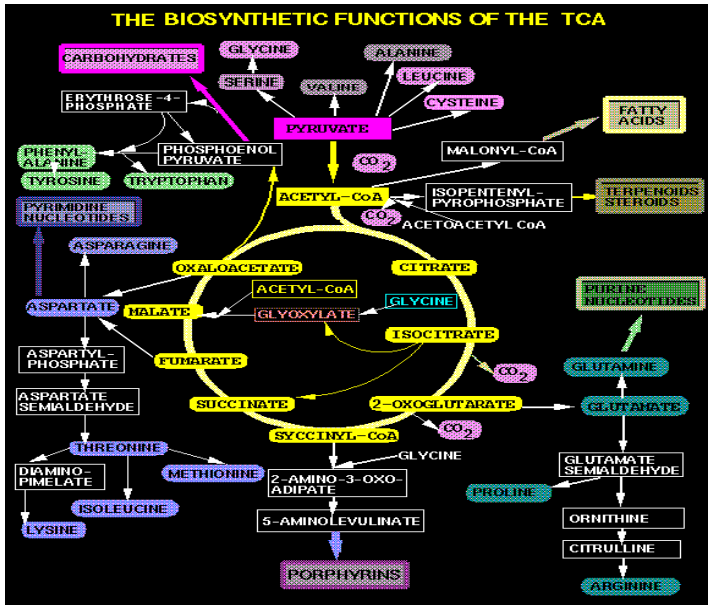
Characterization of Biological Systems

Sequenced Genome



Step 1. Identification of components of the system

Identify potential ORFs and assign functions to the genes



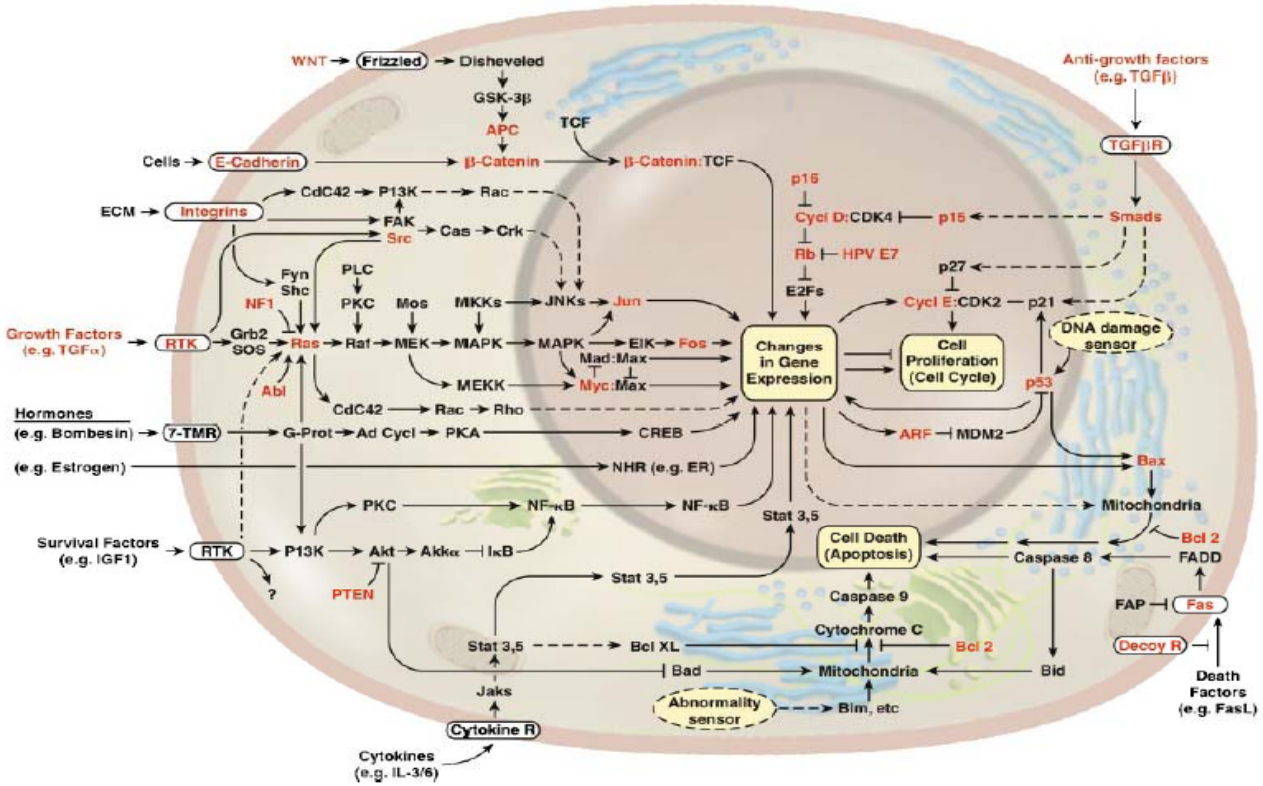
Step 2. The development of a static model.

Reconstruct Metabolic and Functional networks using sequence data, biochemical, expression and other experimental data

Step 3. The Development of Dynamic Models

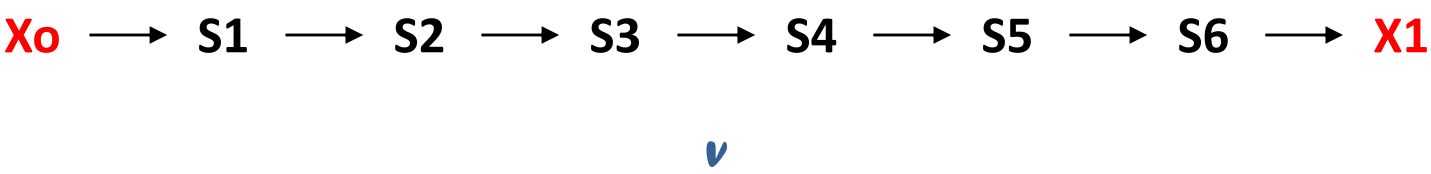
Verification and refinement of models using Experimental data

Biological Complex Systems



- *Why is it so complicated?*
- *Can we make sense of this complexity?*
- *Can we convey our understanding of this complexity?*

Are Biologists Ready?

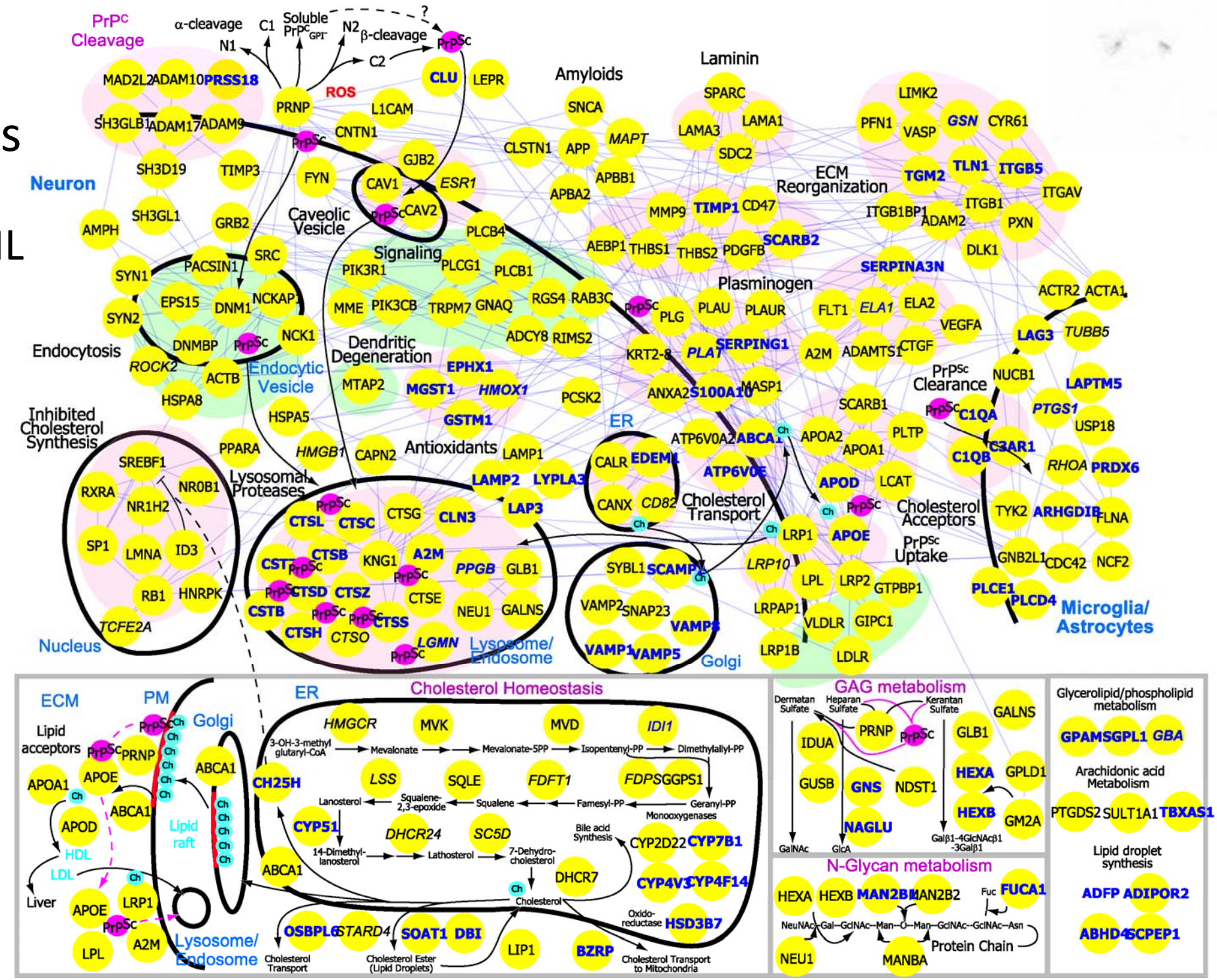


Xo and X1 fixed,
all reactions reversible, assume stable *steady* state.

6 weeks

B6-RML

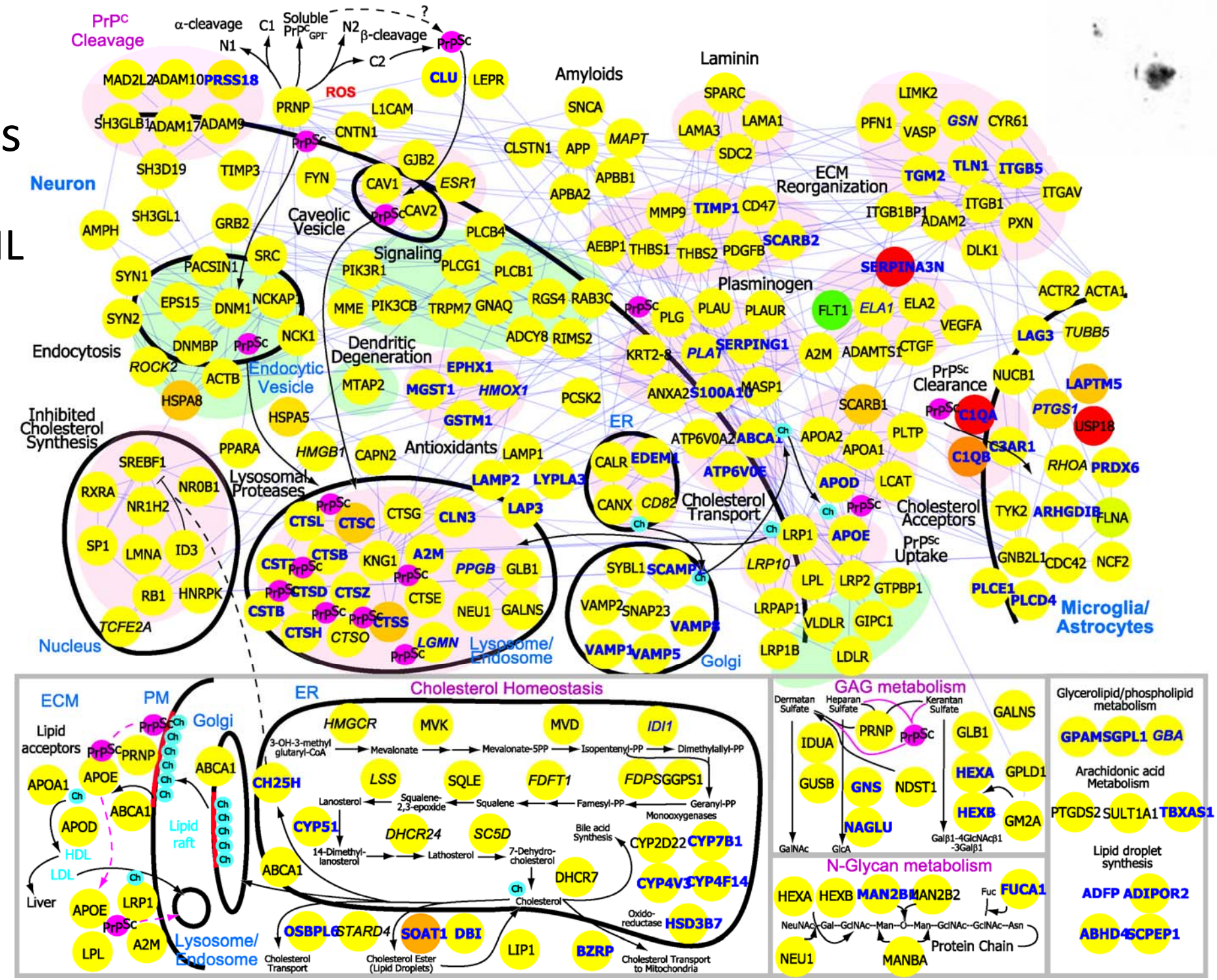
No signs



10 weeks

B6-RML

No signs



14 weeks

B6-RML

No signs

16 weeks

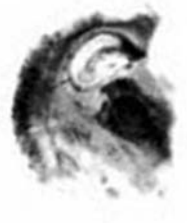
B6-RML

No signs

18 weeks

B6-RML

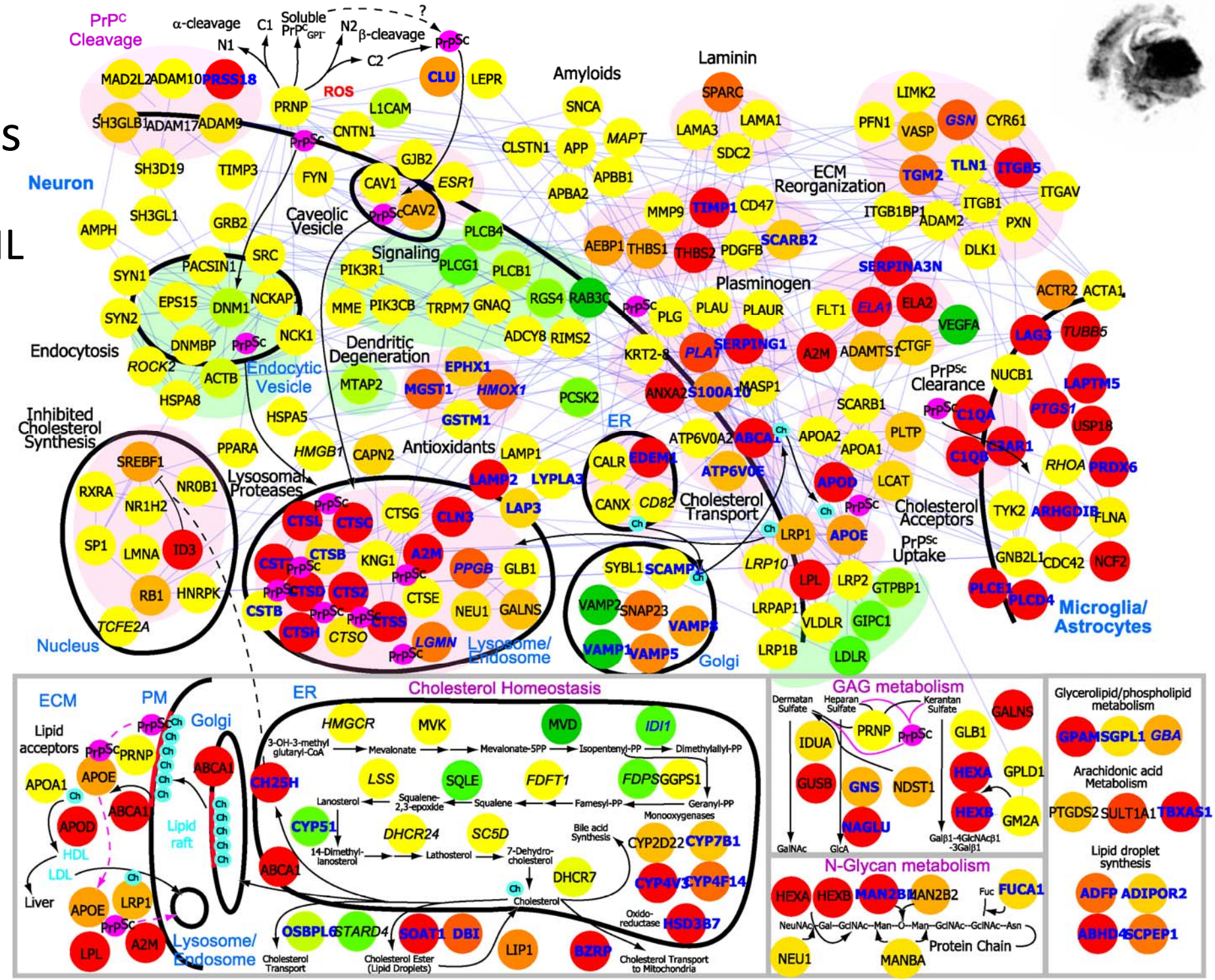
No signs



20 weeks

B6-RML

Slow, clumsy



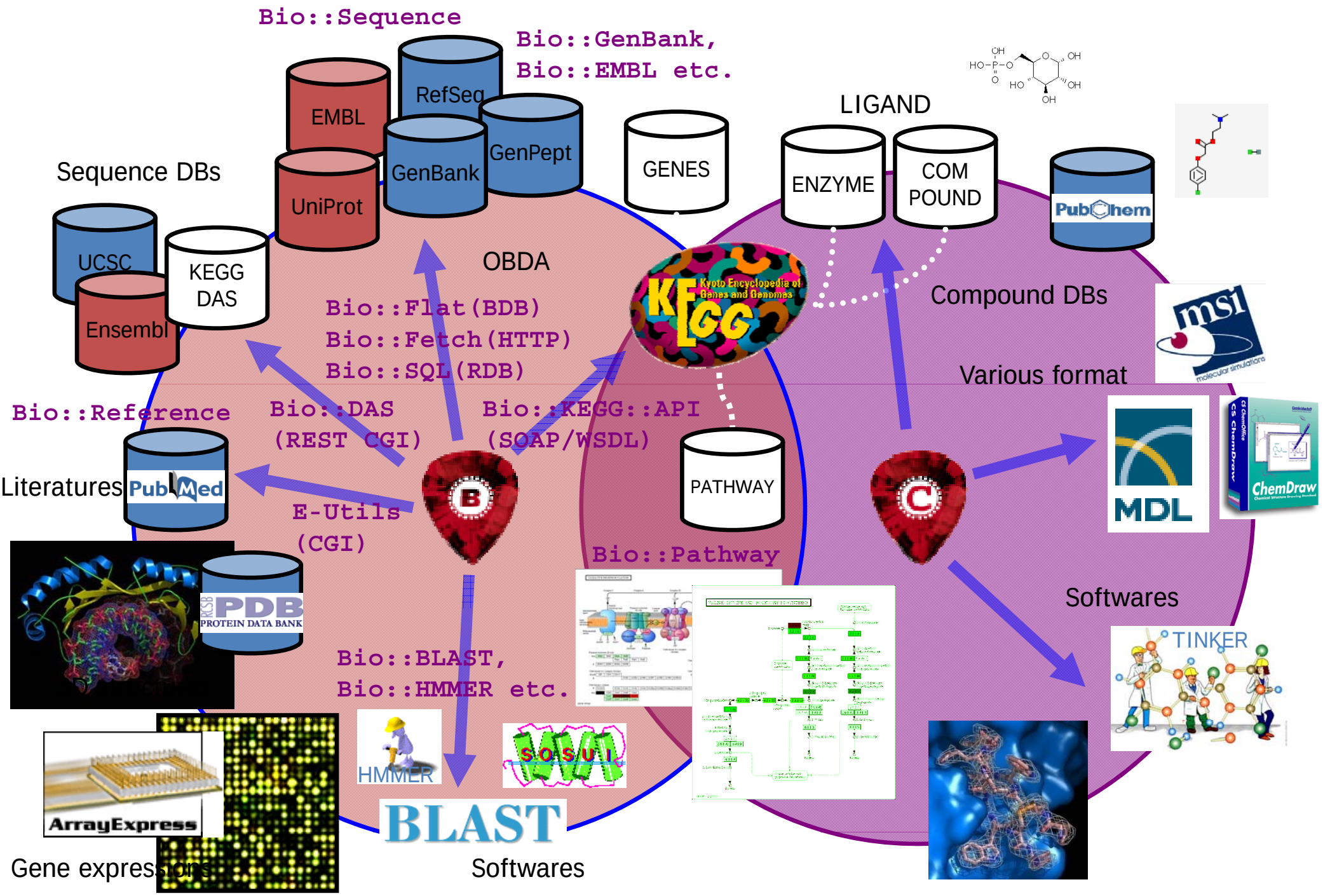
22 weeks

B6-RML

Head tilt, kyphosis, ataxia

Human Metabolome Project

- $7.5 million Genome Canada Project launched in Jan. 2005
- Mandate to quantify (normal and abnormal ranges) and identify all metabolites in urine, CSF, plasma and WBC's
- Make all data freely and electronically accessible (HMDB)
- Make all cmpds publicly available (HML)


Bio::Sequence
Bio::GenBank,
Bio::EMBL etc.
EMBL
RefSeq
GenBank
GenPept
UniProt
Sequence DBs
UCSC
Ensembl
KEGG
DAS
GENES
LIGAND
ENZYME
COM
POUND
PubChem
OBDA
Bio::Flat(BDB)
Bio::Fetch(HTTP)
Bio::SQL(RDB)
Kyoto Encyclopedia of Genes and Genomes
KEGG
Compound DBs
msi
molecular simulations
Various format
Bio::Reference
Bio::DAS
(REST CGI)
Bio::KEGG::API
(SOAP/WSDL)
Literatures
PubMed
E-Utils
(CGI)
B
C
PATHWAY
MDL
ChemDraw
Bio::Pathway
PDB
PROTEIN DATA BANK
Softwares
TINKER
Bio::BLAST,
Bio::HMMER etc.
HMMER
SOSUI
BLAST
ArrayExpress
Gene expressions
Softwares