Eukaryotic Gene Expression: Basics & Benefits

P N RANGARAJAN

Lecture 39

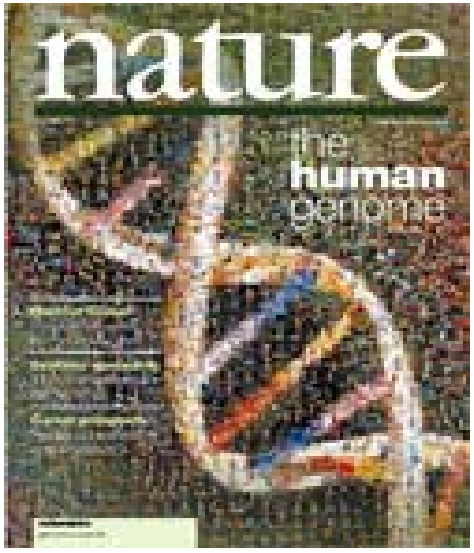
Genomics & Proteomics

UNDERSTANDING GLOBAL CHANGES IN GENE EXPRESSION:

NEW TECHNOLOGIES & NEW CHALLENGES

Lander ES, Linton LM, Birren B, *et al.* (February 2001). "Initial sequencing and analysis of the human genome". *Nature* **409** (6822): 860–921.

Venter JC, Adams MD, Myers EW, *et al.* (February 2001). "The sequence of the human genome". *Science* **291** (5507): 1304–51.



Post-Sequencing Era

Sequencing genomes of several organisms has provided a wealth of information, leading to the creation of several new disciplines (microarray, bioinformatics, proteomics and pharmocogenetics etc.)

40-60 percent of all identified genes across species are of unknown function.

GenBank:

- Doubles ~18 months
- > 190 billion bases

- Genomes:
  - Eukaryotes: ~200
  - Prokaryotes: ~600

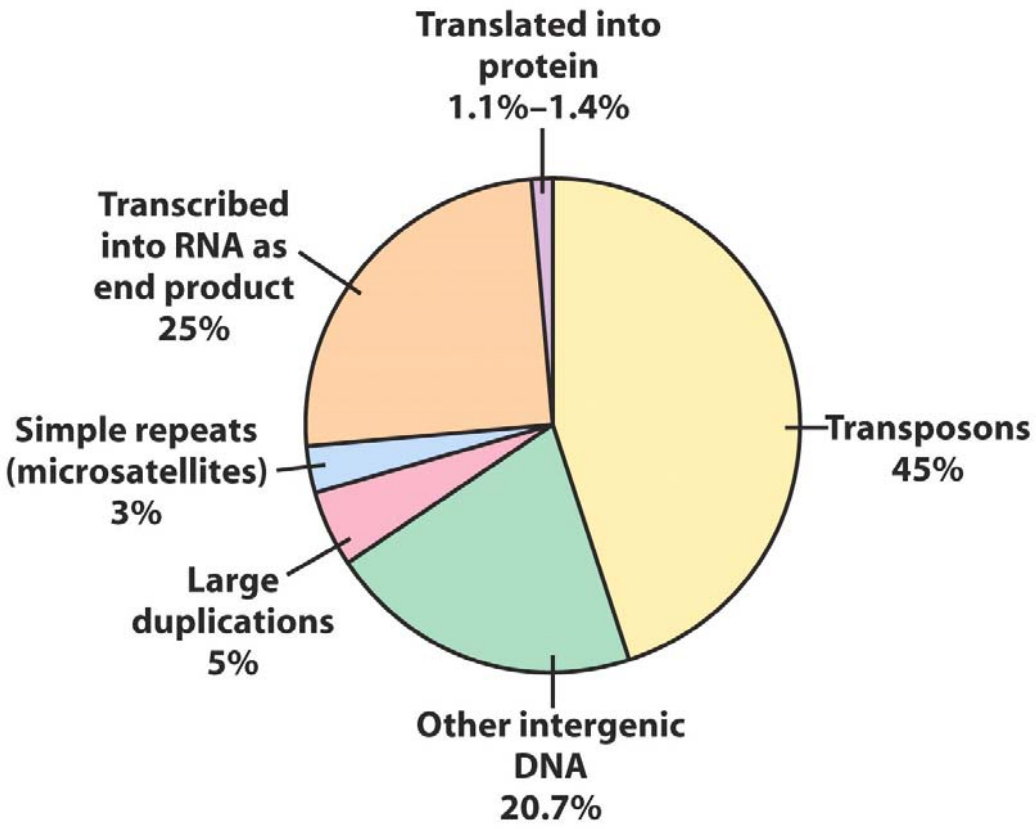
To handle this huge volume of data, a new area of

computational biology known as

BIOINFORMATICS

came into prominence.

Human genome



- Only 1.1% to 1.4% of human genome DNA actually encodes proteins (~ 30,000 GENES).
- More than 50% of genome consists of short, repeated sequences.
- 45% of genome consists of transposons (short movable DNA sequences).

Although there are ~30,000 protein coding genes, all of them are not expressed in any given cell type.

A conservative estimate is that ~ 18,000 transcripts may be present which are translated into 18,000 proteins.

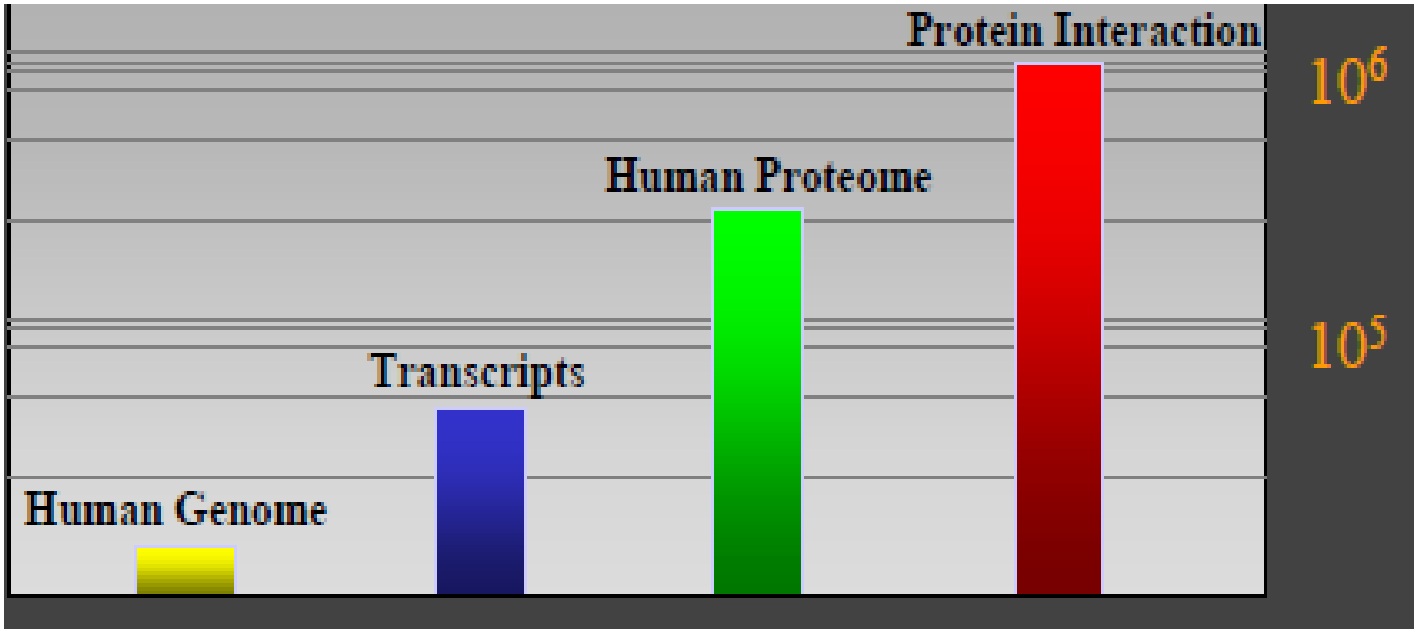
Among these ~ 2500 may be common to all cell types, 2000 may be secreted.

18,000 – 4500 = 13,500

Assuming that there are about 300 different cell types:

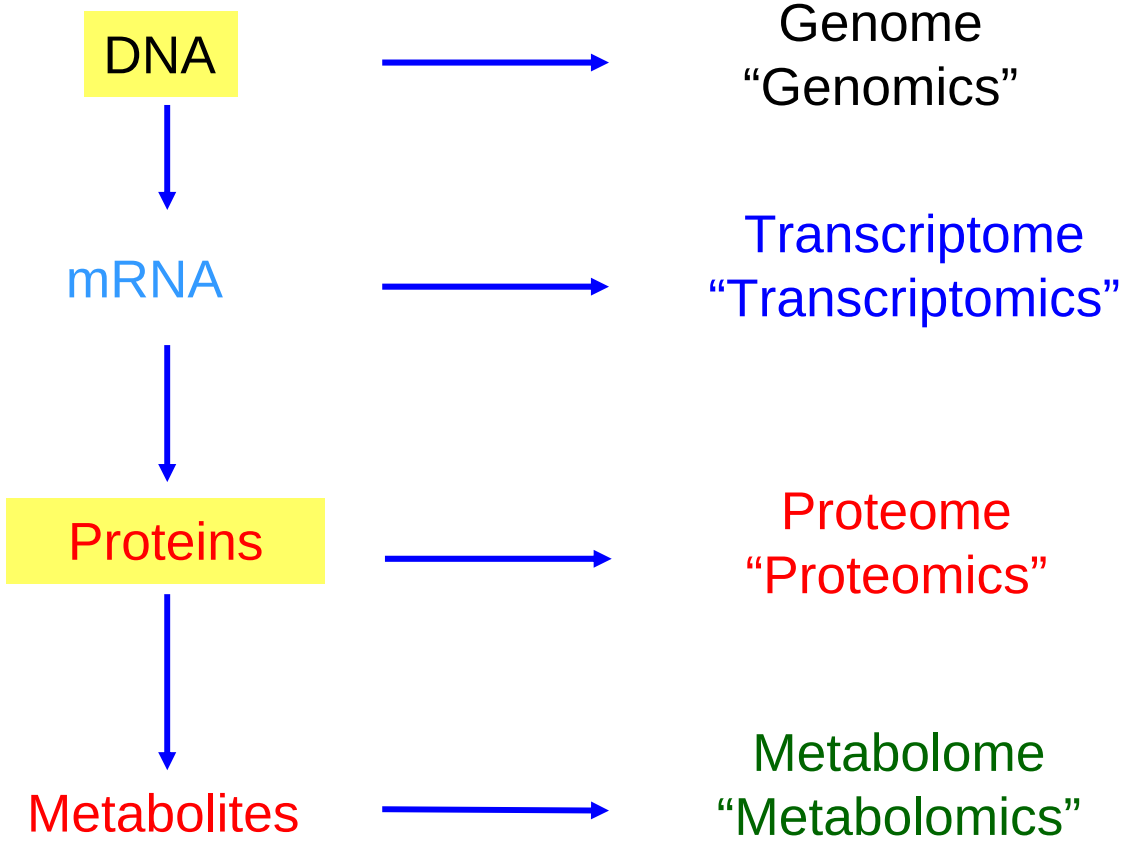
13,500/300 = 45 proteins which cell type specific

Genome:	30,000 genes
Transcriptome:	40,000-100,000 mRNAs
Proteome:	100,000-400,000 proteins
Interactome:	>1,000,000 interactions



http://www.uta.edu/biology/michalak/classnotes/genomics/lect35.ppt

‘Omics’ era


DNA
mRNA
Proteins
Metabolites
Genome
“Genomics”
Transcriptome
“Transcriptomics”
Proteome
“Proteomics”
Metabolome
“Metabolomics”

Genomics

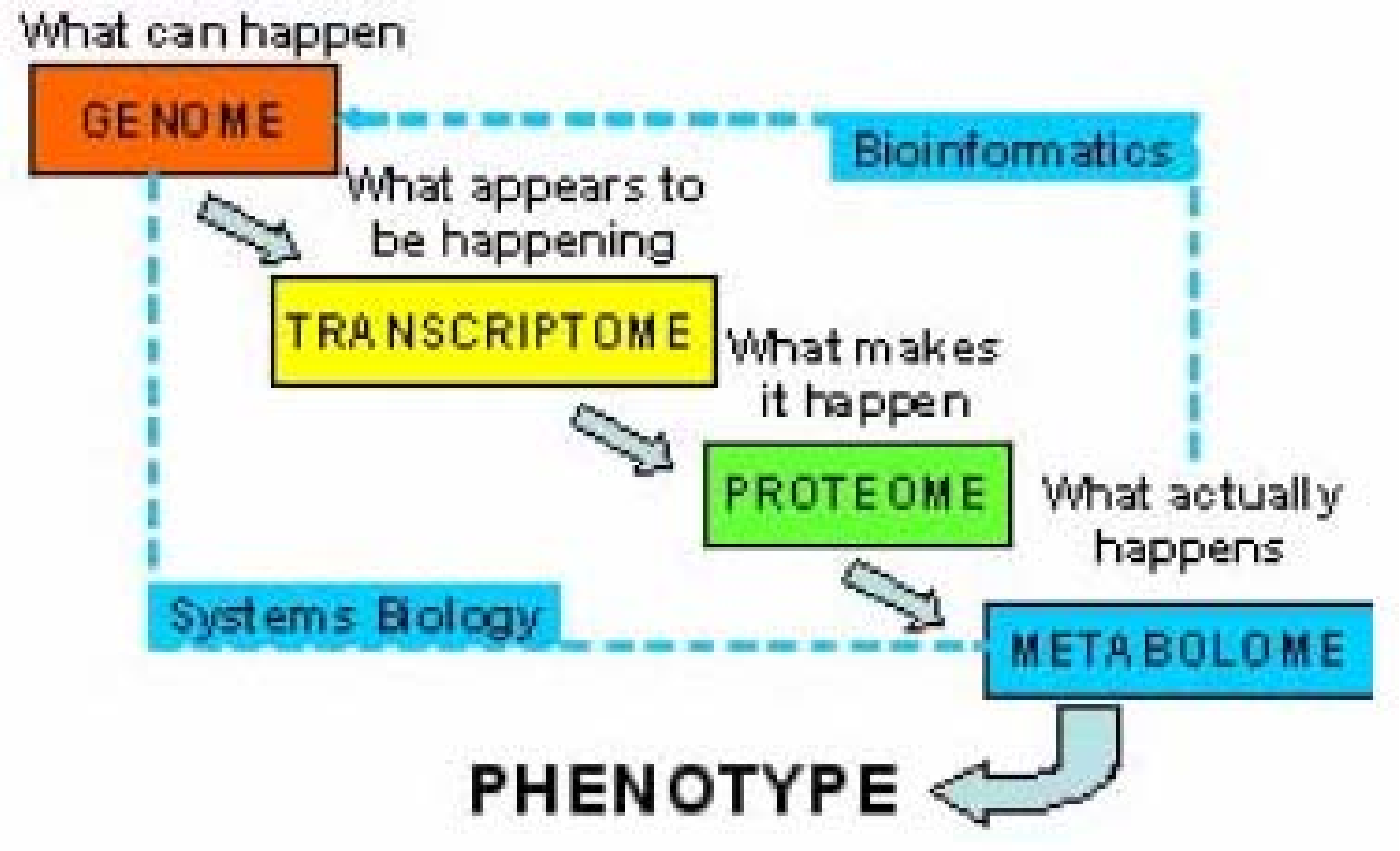
Functional genomics - Describes the way in which genes and their products, proteins, interact together in complex networks in living cells. Disruption of these interactions can lead to diseases.

Structural genomics - Architectural features of genes and chromosomes.

Comparative genomics - the evolutionary relationships between the genes and proteins of different species.

Epigenomics (epigenetics) - genetic effects not caused by changing DNA sequences (usually involving DNA methylation and histone modifications).

Pharmacogenomics - finding new biological targets and new ways to design drugs and vaccines.



http://www.criticalcarecanada.com/presentations/the-omics-revolution.pdf

Genomics

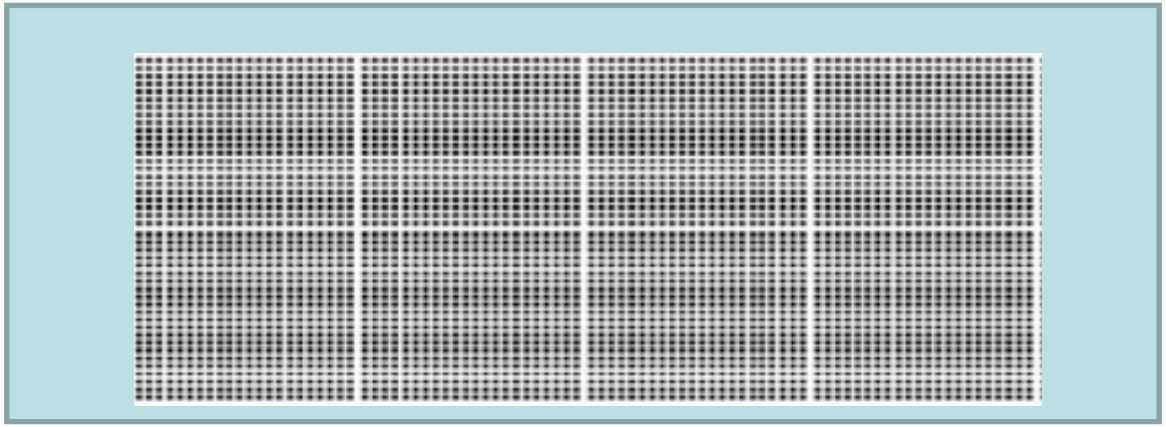
GENOMICS

Rapid identification of all the genes expressed in a cell or tissue

DNA MICROARRAY

- A small 1 square centimeter chip that’s divided into thousands of squares.
- Each square contains many copies of a single gene.
- Initially developed by Patrick Brown at the Stanford University School of Medicine to determine which genes are involved in yeast cell sporulation.

DNA Microarrays allow you look at expression of all the potential mRNAs in a cell at the same time



- Microarrays are composed of short DNA oligomers attached to an inert substrate – glass slide
- Typically contain a grid of 105-106 spots, each with a different DNA molecule
- Fluorescently-labeled DNA or RNA hybridize to complementary probes
- Hybridized array is scanned with a laser to produce a signal for each spot

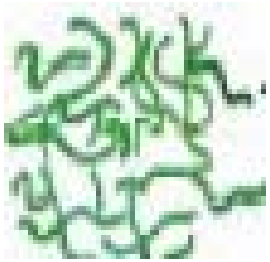
DNA microarray methodology: Animation

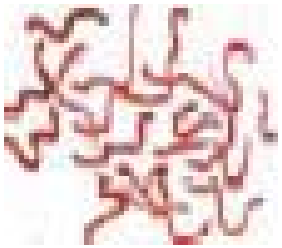
http://www.bio.davidson.edu/Courses/genomics/chip/chip.html

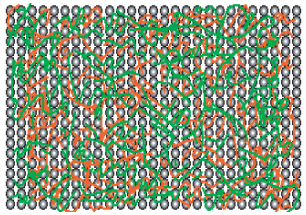
http://media.pearsoncmg.com/bc/bc_campbell_genomics_2/medialib/method/chip/chip.html

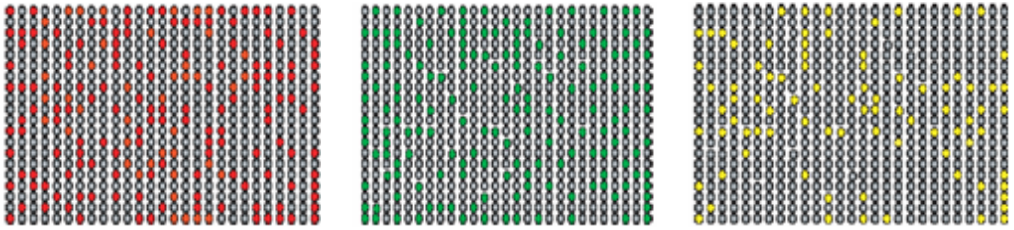
Condition A
(normal/control)

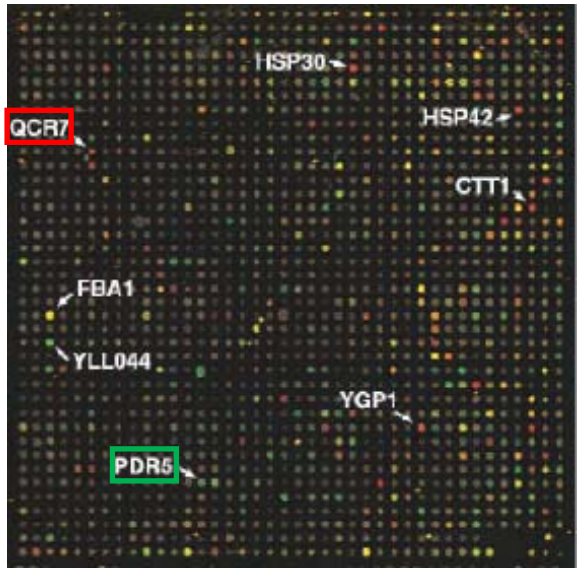
Condition B
(experimental)







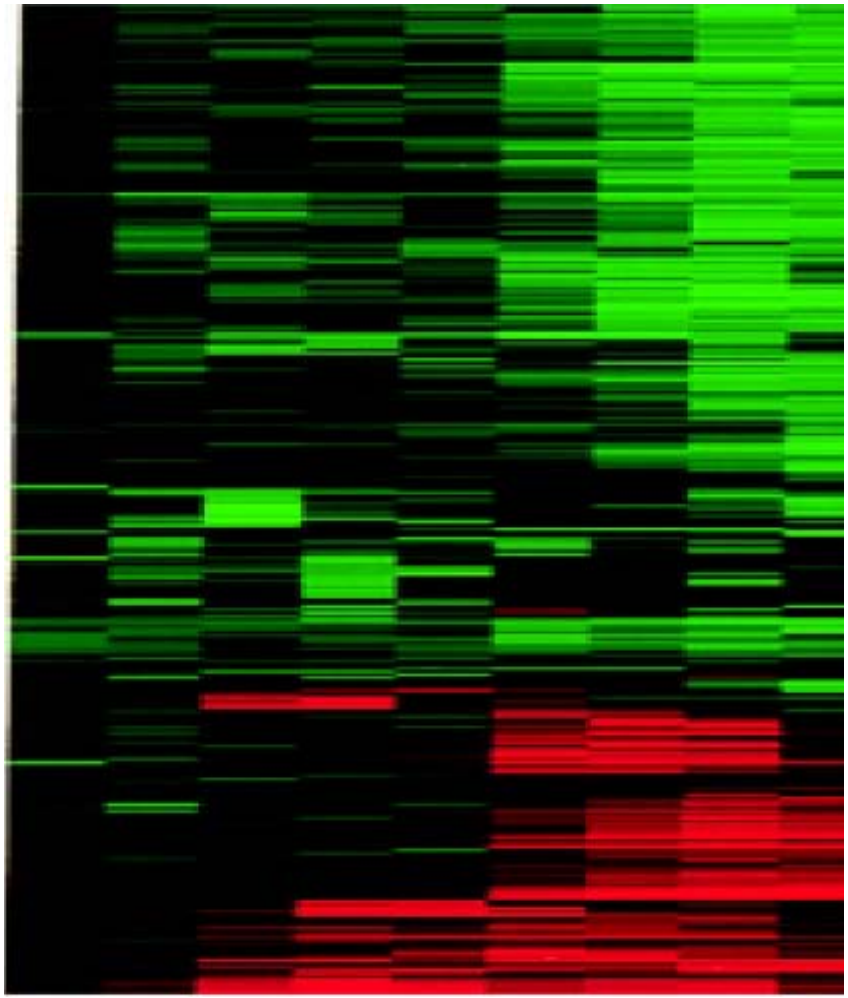




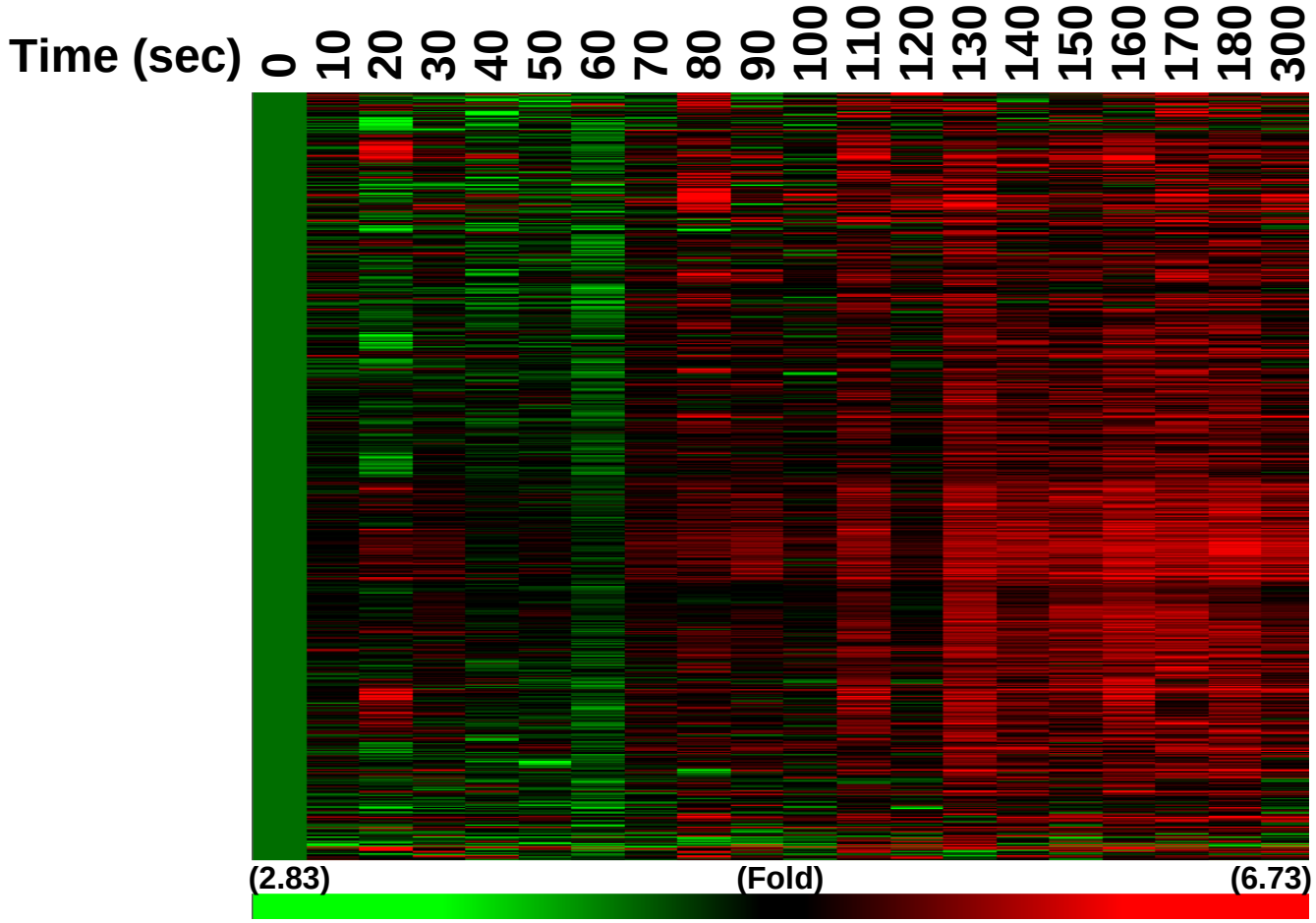
QCR7

PDR5

Time

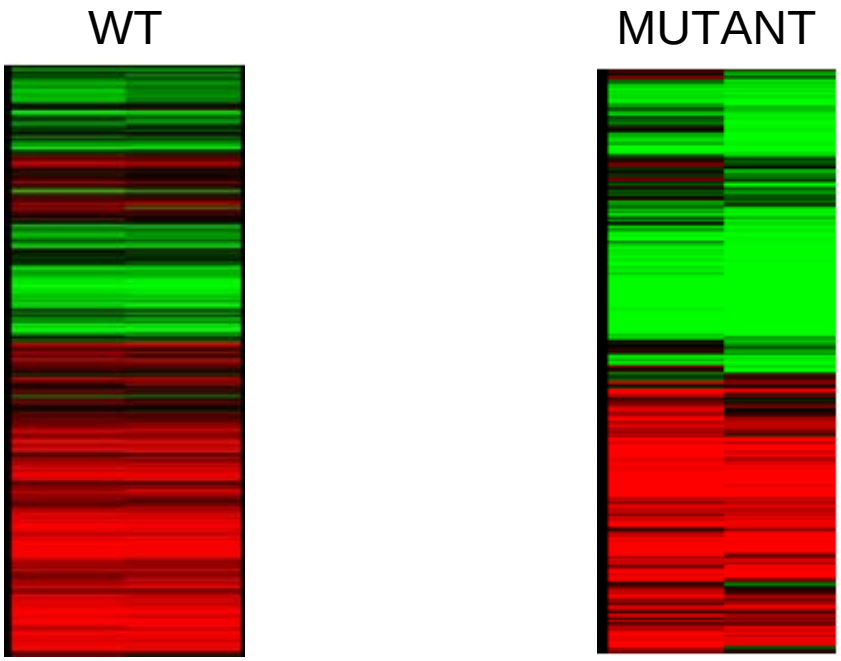


Time course of serum stimulation of mouse fibroblasts



Gene expression profile of cells subjected to oxidative stress at 10s intervals

Microarray analysis of a RNA Polymerase II mutant



Understanding global changes in gene expression

Expression profiling

MICROARRAYS

SAGE

MPSS

TOGA

Alternatives to microarrays

SAGE (Serial Analysis of Gene Expression)

Velculescu, V.E., Zhang, L., Vogelstein, B., Kinzler, K.W., 1995. Serial analysis of gene expression. Science 270, 484–487.

Massively Parallel Signature Sequencing (MPSS)

Brenner, S et al., 2000. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nat. Biotechnol. 18, 630–634.

TOtal Gene expression Analysis (TOGA)

Sutcliffe, J.G et al., 2000. TOGA: an automated parsing technology for analyzing expression of nearly all genes. Proc. Natl. Acad. Sci. U.S.A. 97, 1976–1981.

Proteomics

PROTEOMICS

Rapid identification of all the proteins synthesized in a cell or tissue

Proteomics is the study of proteome, which is the protein complement of the genome

Proteomics is the study of protein expression, regulation, modification, and function in living systems for understanding how living systems use proteins.

Using a variety of techniques, proteomics can be used to study how proteins interact within a system, or how proteins change under different conditions, including post translational modifications.

Proteomics

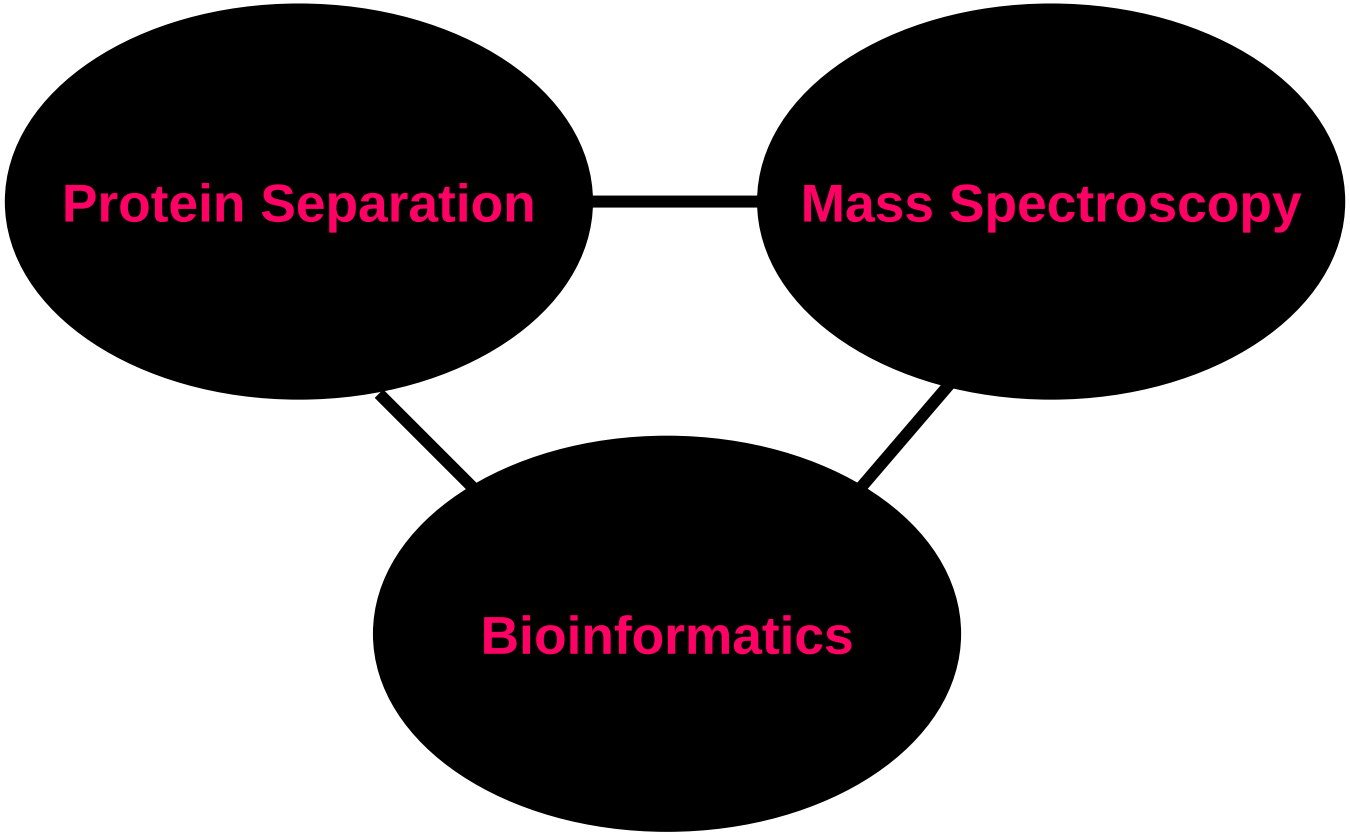
- First coined in 1995
- Defined as the large-scale characterization of the entire protein complement of a cell line, tissue, or organism.
- Goal:

-To obtain a more global and integrated view of biology by studying all the proteins of a cell rather than each one individually.

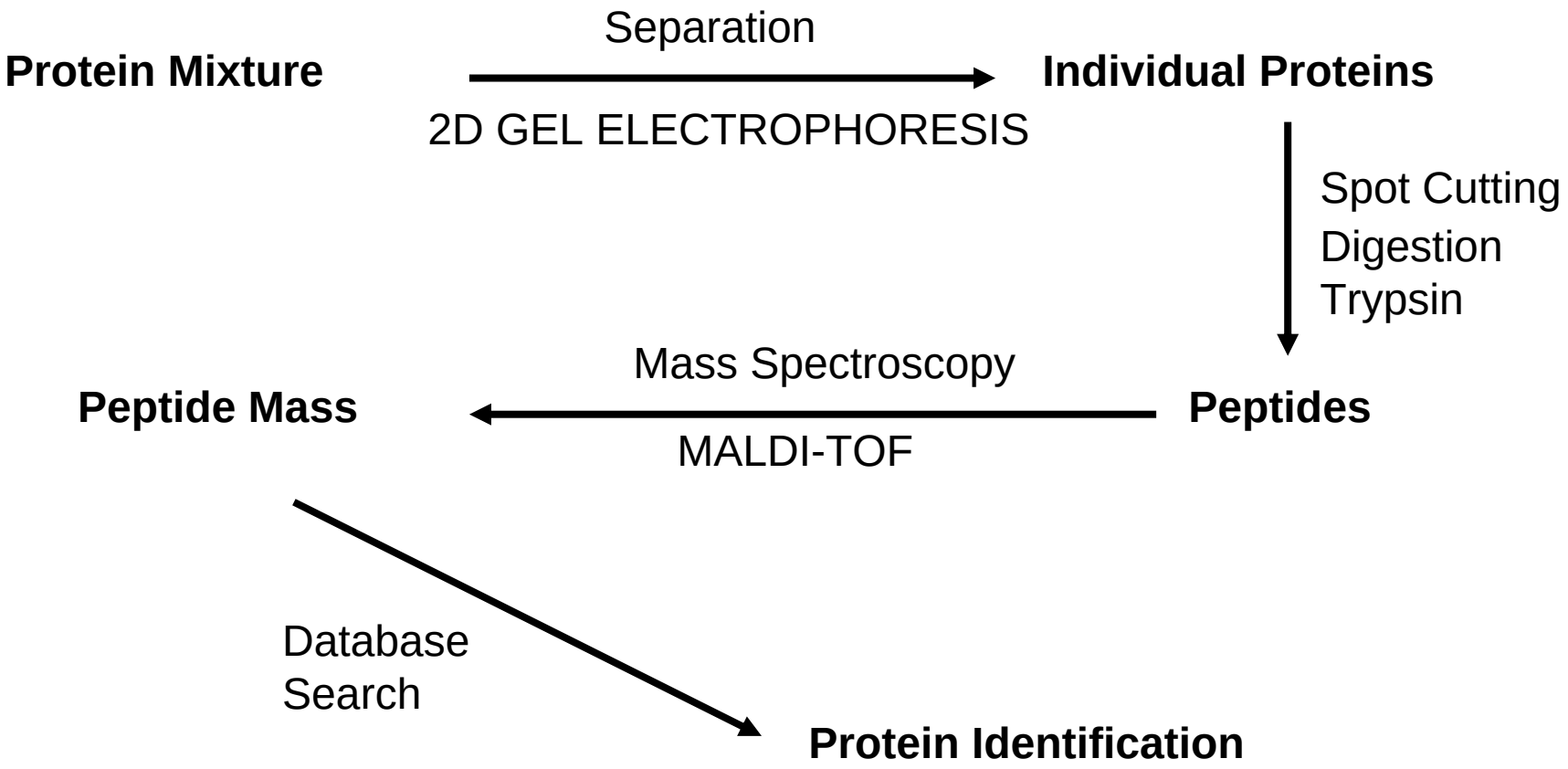
Why is Proteomics necessary?

- Having complete sequences of genome is not sufficient to elucidate biological function.
- A cell is normally dependent upon several metabolic and regulatory pathways for its survival
- Modifications of proteins can be determined only by proteomic methodologies
- It is necessary to determine the protein expression levels, post tranlsational modifications, protein localization as well as protein-protein interactions.
- Proteins are direct drug targets.

Components of Proteomics

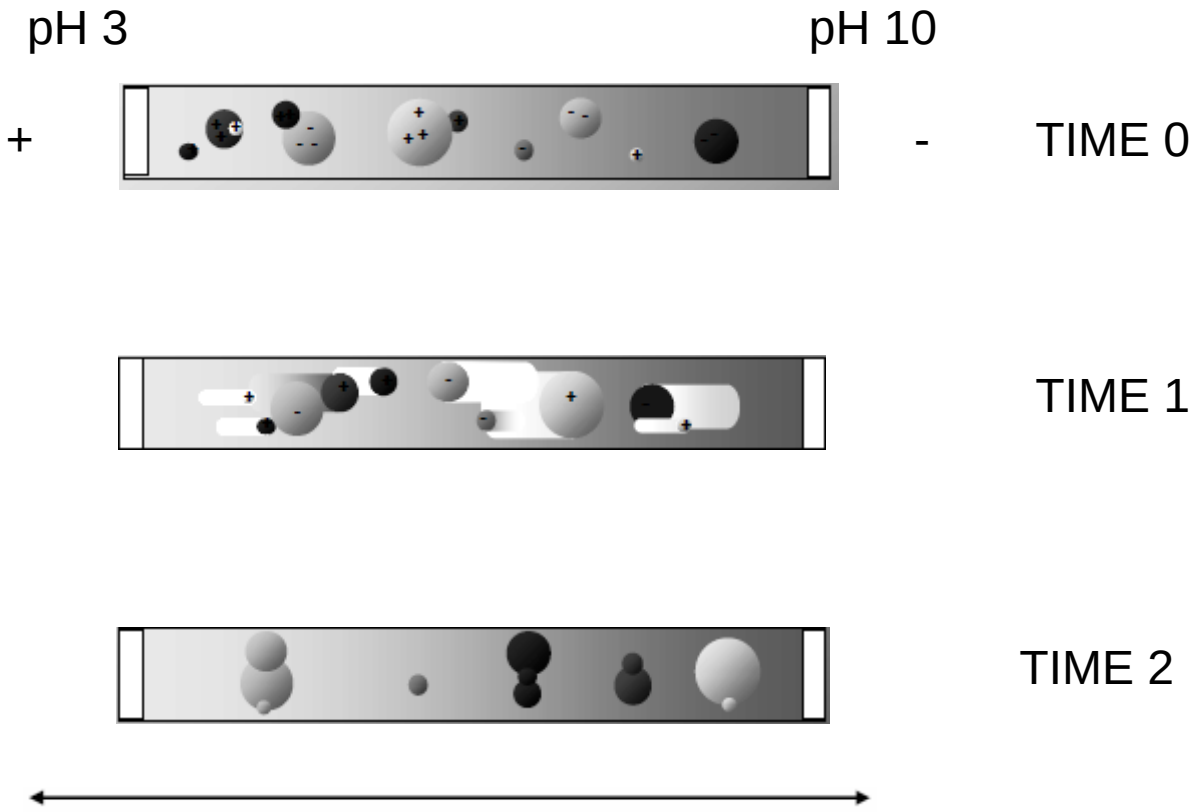


Basic Proteomic Analysis Scheme

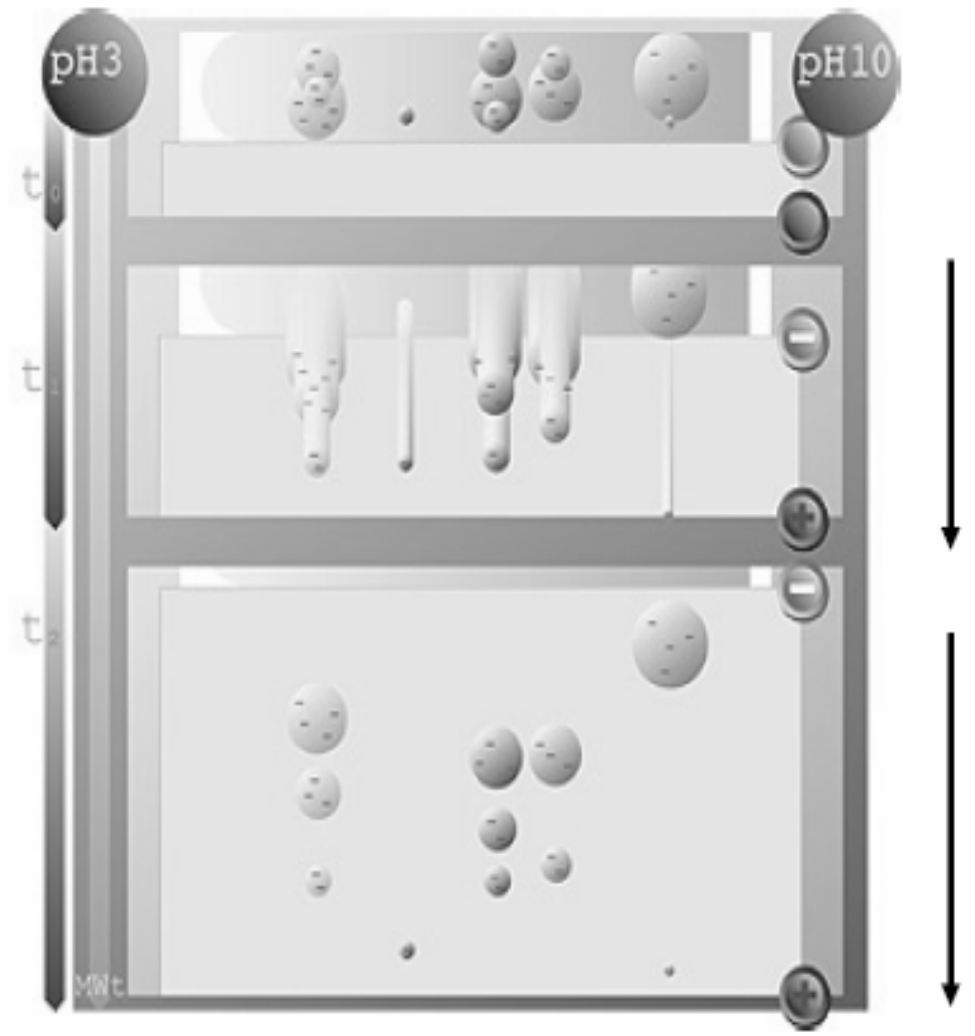


2DGE: ISOELECTRIC FOCUSSING, SDS-PAGE

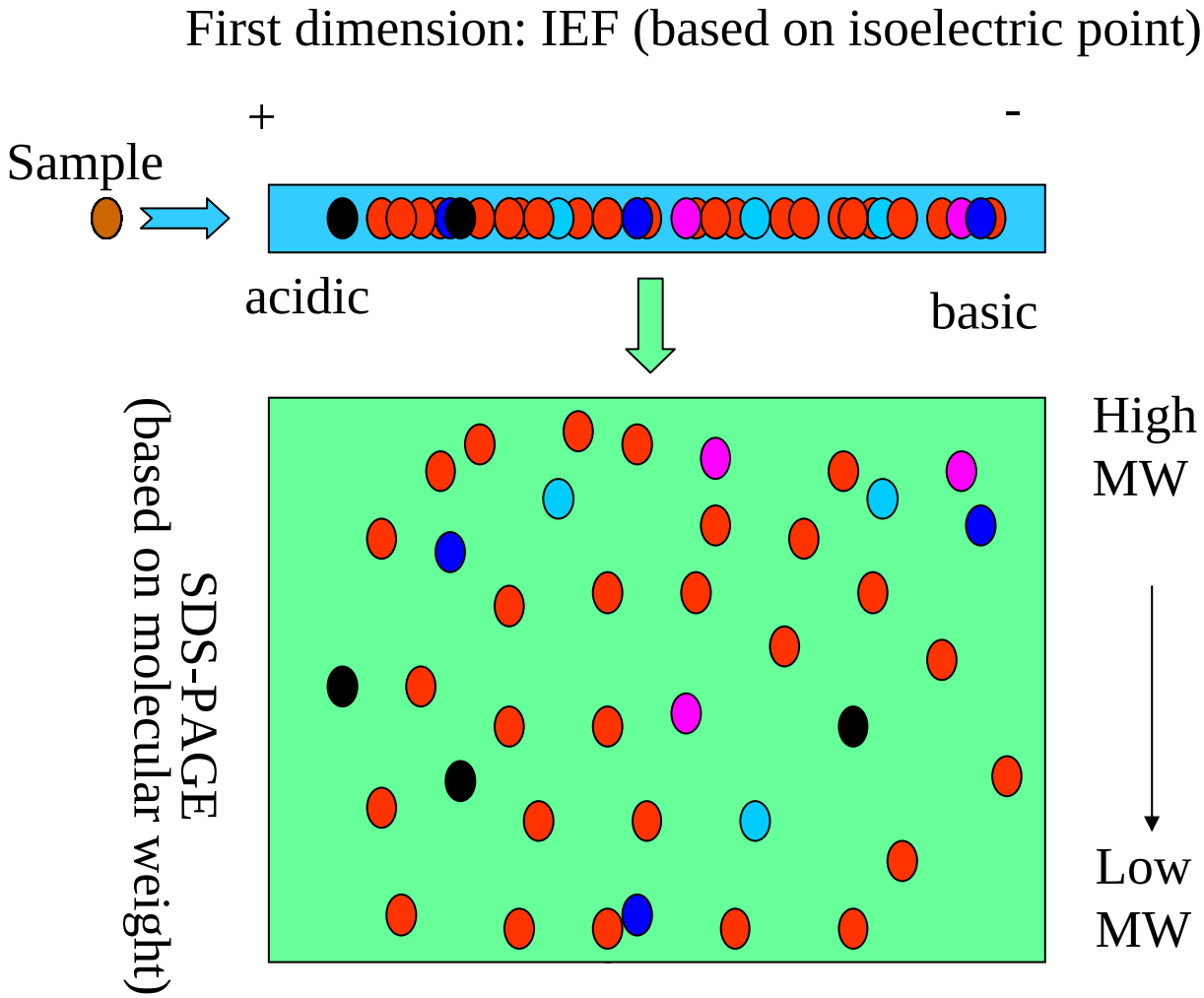
ISOELECTRIC FOCUSSING



SDS-PAGE

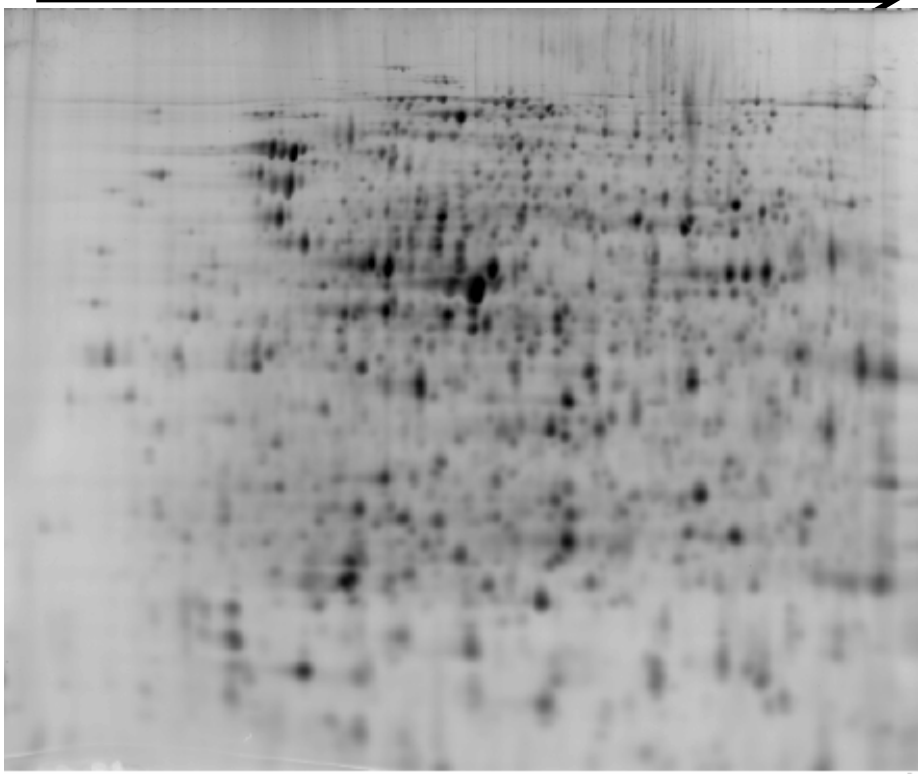


Two-dimensional Gel Electrophoresis (2DGE)



ISOELECTRIC FOCUSSING

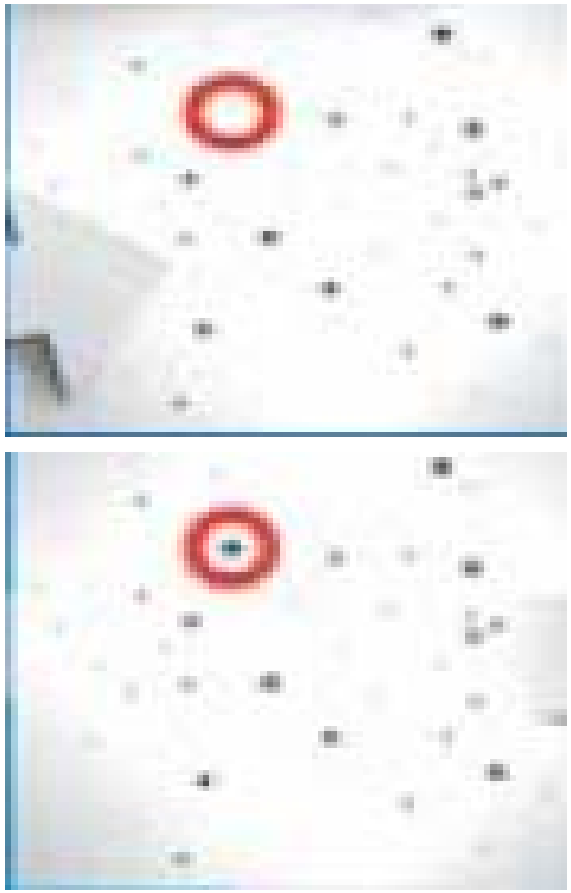
SDS PAGE



Analysis of differential Protein Expression by 2DGE

CONTROL

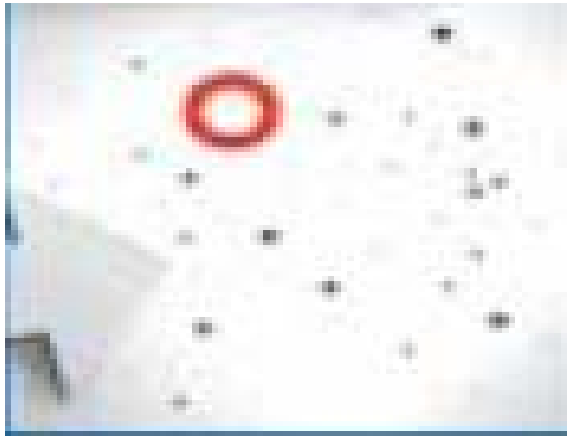
EXPERIMENTAL



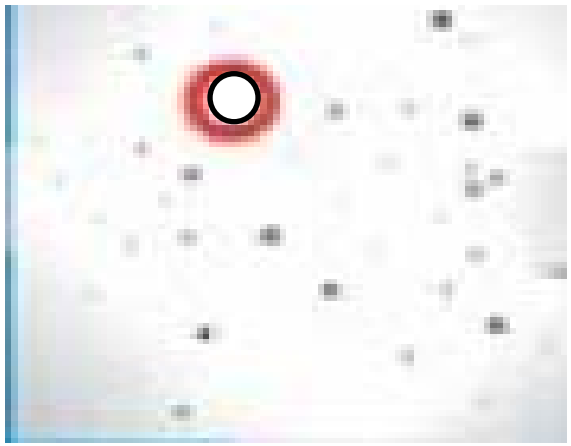
Once proteins are seperated by 2DGE,

how are the proteins identified?

CONTROL

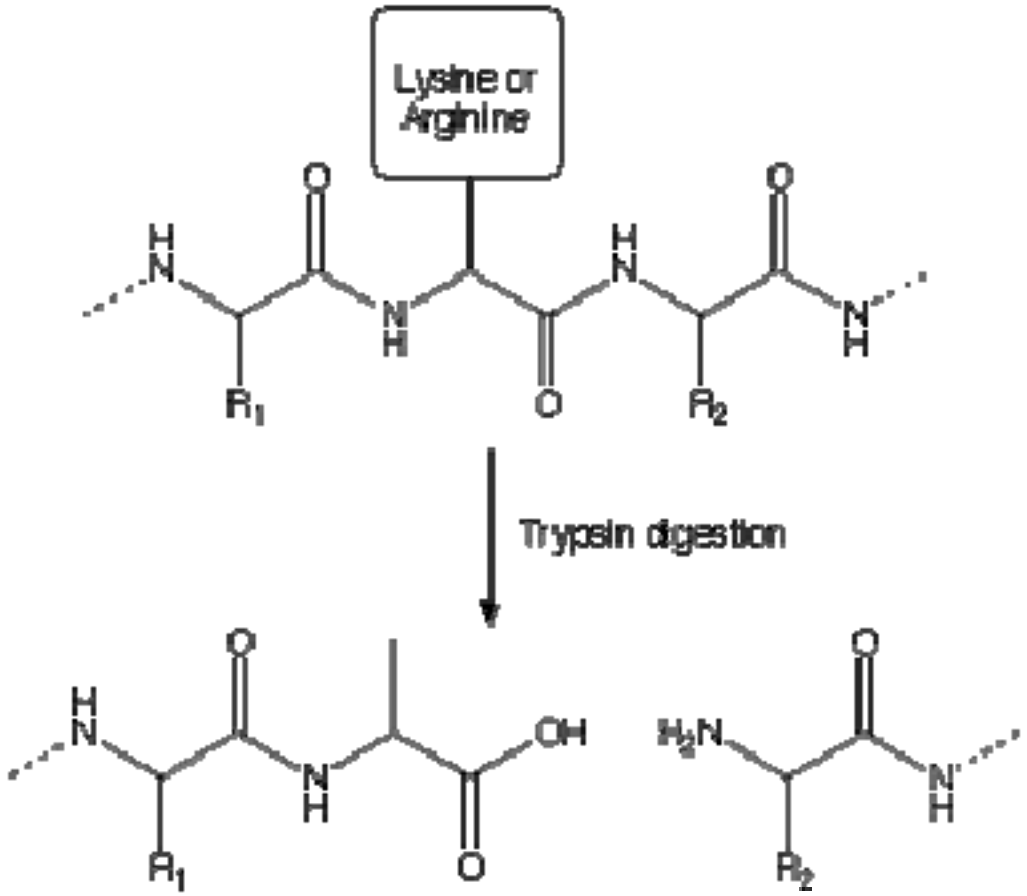


EXPERIMENTAL



MASS SPECTROMETRY

In-gel trypsin digestion



Peptides are introduced into the mass spectrometer and identified by peptide fingerprinting or tandem mass spectrometry (MSMS).

This approach is called "bottom-up" proteomics wherein proteins are identified at the peptide level.

Peptide mass fingerprinting uses the masses of proteolytic peptides as input to a search of a database of predicted masses that would arise from digestion of a list of known proteins.

If a protein sequence in the reference list gives rise to a significant number of predicted masses that match the experimental values, there is some evidence that this protein was present in the original sample.

In tandem mass spectrometry (MSMS), sample proteins are broken up into short peptides using an enzyme like trypsin and separated in time using liquid chromatography.

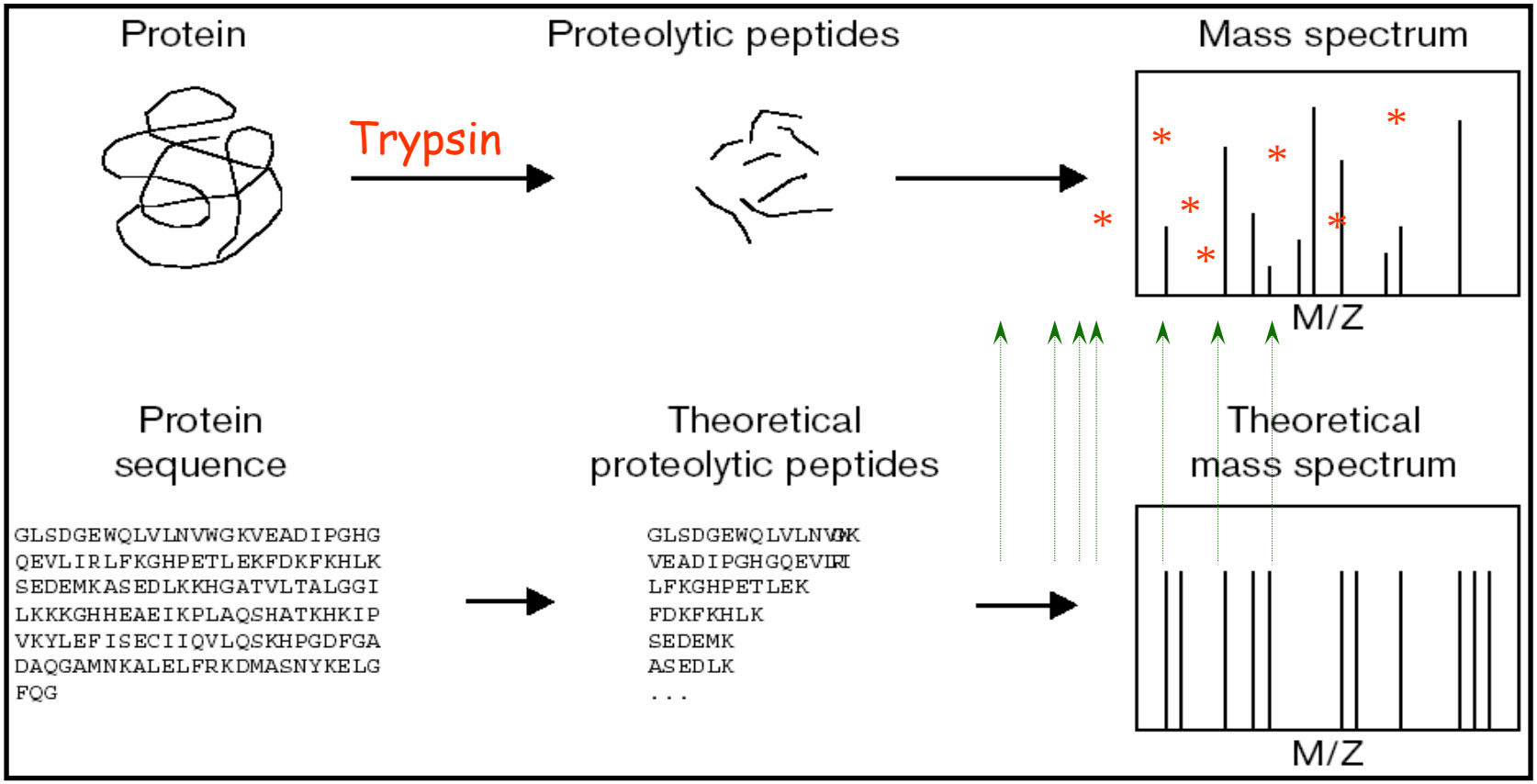
They are then sent through one mass spectrometer to separate them by mass.

Peptides having a specific mass are then typically fragmented using collision-induced dissociation and sent through a second mass spectrometer, which will generate a set of fragment peaks from which the amino acid sequence of the peptide may often be inferred.

Peptide identification software is used to try to reliably make these inferences

Mass Spectrometry

Peptide mass fingerprinting



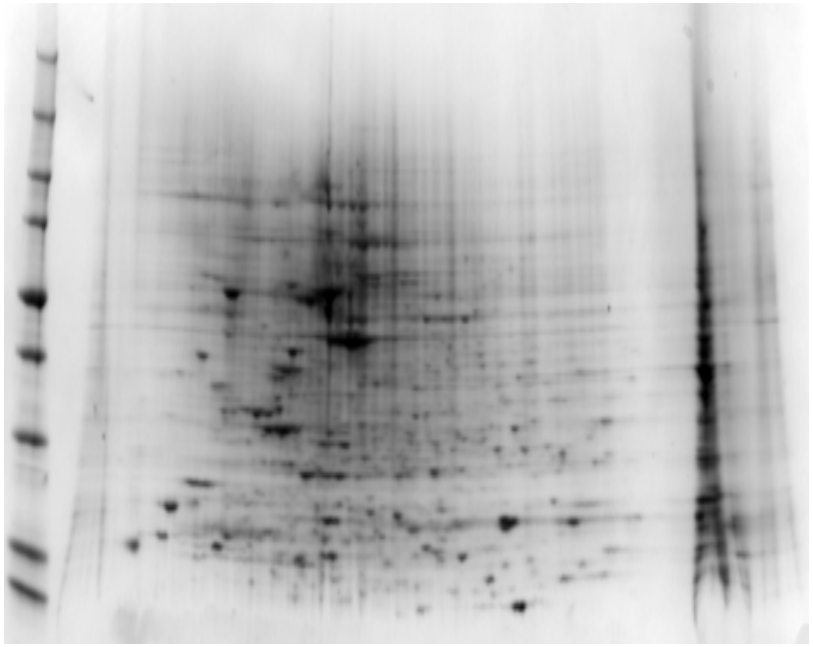
Identification of proteins by mass spectrometry

John B. Fenn

Koichi Tanaka

The Nobel Prize in Chemistry 2002 was awarded *"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"*

http://nobelprize.org/nobel_prizes/chemistry/laureates/2002/

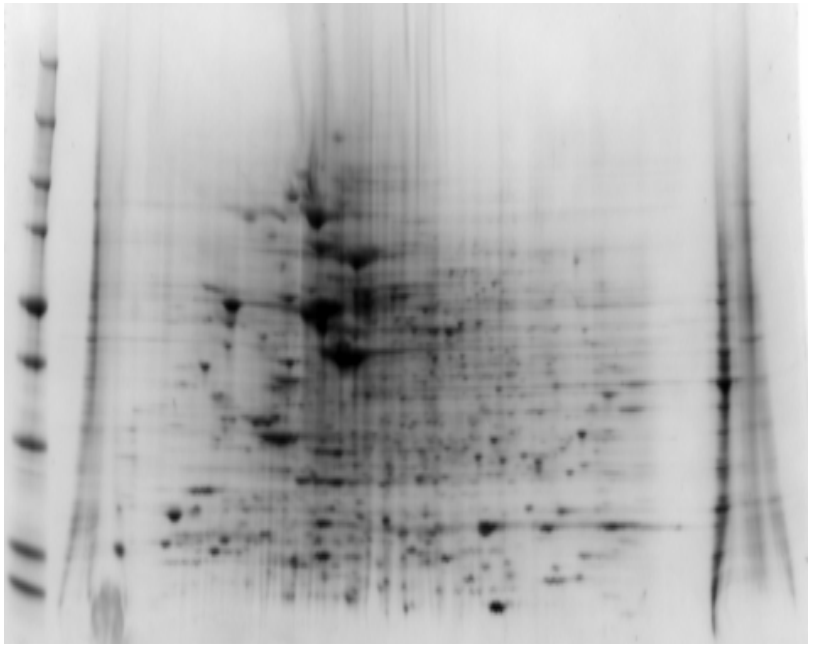


Undifferentiated cells

Uninduced

Normal

-hormone



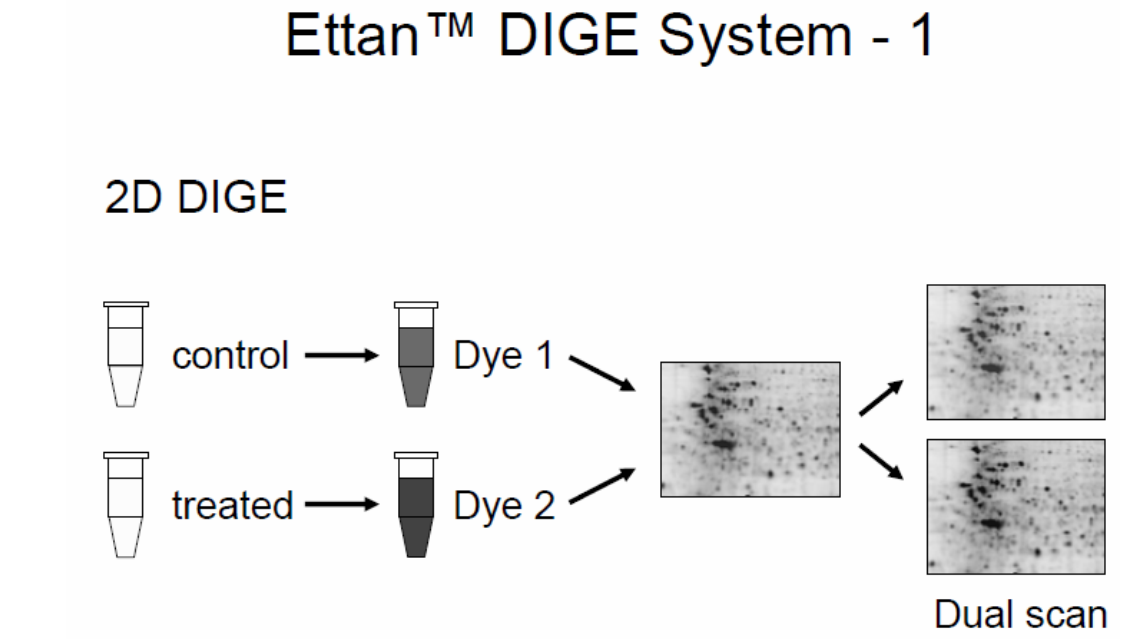
Differentiated cells

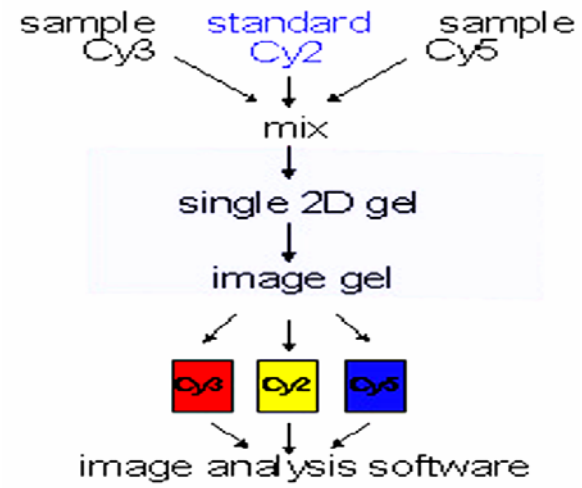
Induced

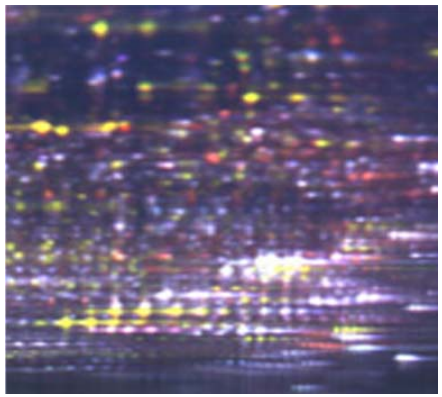
Mutant

+ hormone

Two Dimensional Differential Gel Electrophoresis (2D DIGE)







http://www.gelifesciences.com/aptrix/upp00919.nsf/Content/B9A1E643A0DEB1D1C1257628001D1C81/$file/11001340AD.pdf

Liquid chromatography mass spectrometry (LC-MS)

LC-MS combines the physical separation capabilities of liquid chromatography (or HPLC) with the mass analysis capabilities of mass spectrometry

Electrospray ionization Mass spectropmetry (ESI-MS)

ESI-MS: Intact proteins are ionized and then introduced to a mass analyzer. This approach is referred to as "top-down" strategy of protein analysis.

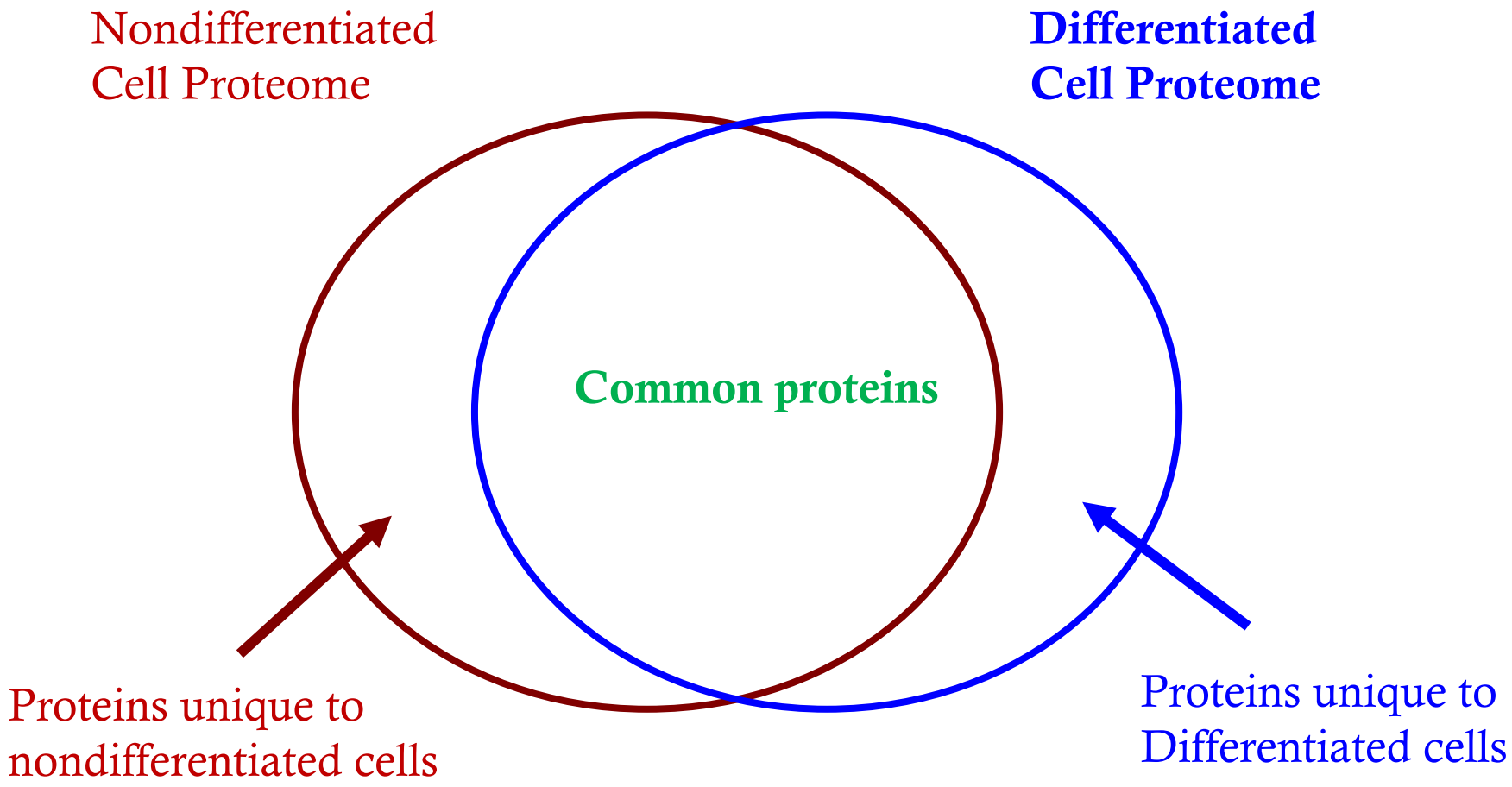
MS-MS databases

- PepSea (disabled)
  - http://195.41.108.38/PA_SequenceOnlyForm.html
- ProteinProspector
    - http://prospector.ucsf.edu/
- PeptideSearch (limited)
    - http://www.narrador.embl-heidelberg.de/GroupPages/Homepage.html
- Mascot (probably the best)
    - www.matrixscience.com

Applications of Proteomics

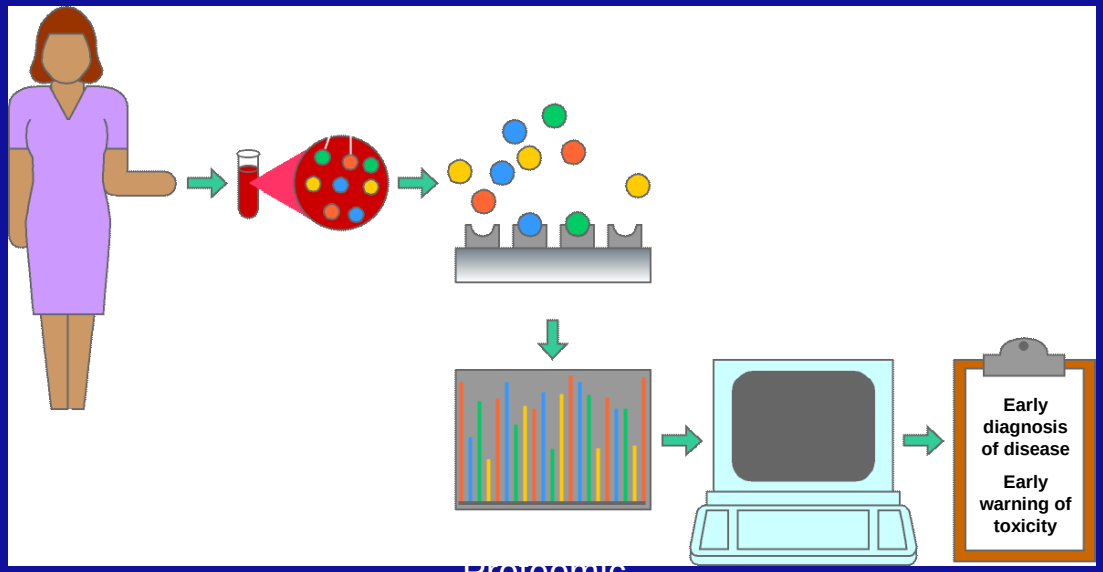
- Protein Mining – catalog all the proteins present in a tissue, cell, organelle, etc.
- Differential Expression Profiling – Identification of proteins in a sample as a function of a particular state: differentiation, stage of development, disease state, response to drug or stimulus.
- Network Mapping – Identification of proteins in functional networks: biosynthetic pathways, signal transduction pathways, multiprotein complexes
- Mapping Protein Modifications – Characterization of posttranslational modifications: phosphorylation, glycosylation, acetylation etc.

Differential protein expression

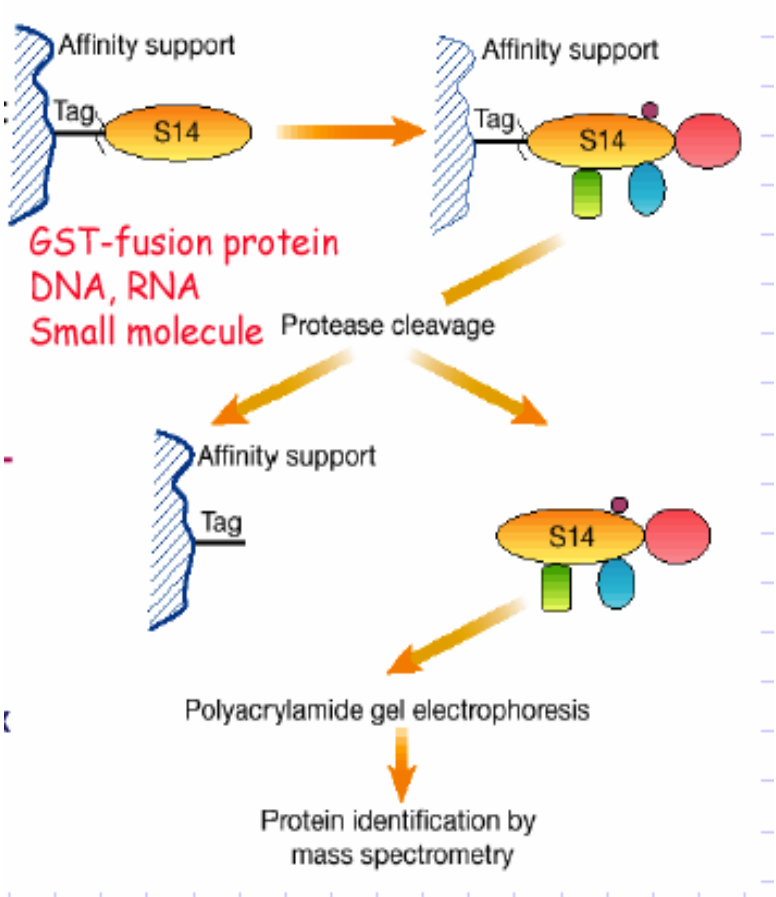


CLINICAL PROTEOMICS

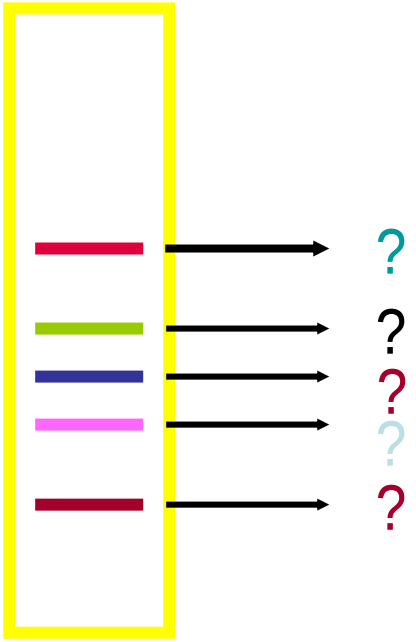
Serum Protein Pattern Diagnostics



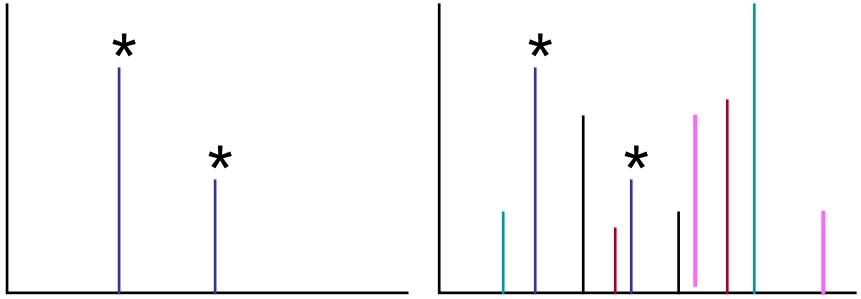
The study of protein-protein interaction by Mass Spectrometry



SDS- PAGE



MASS

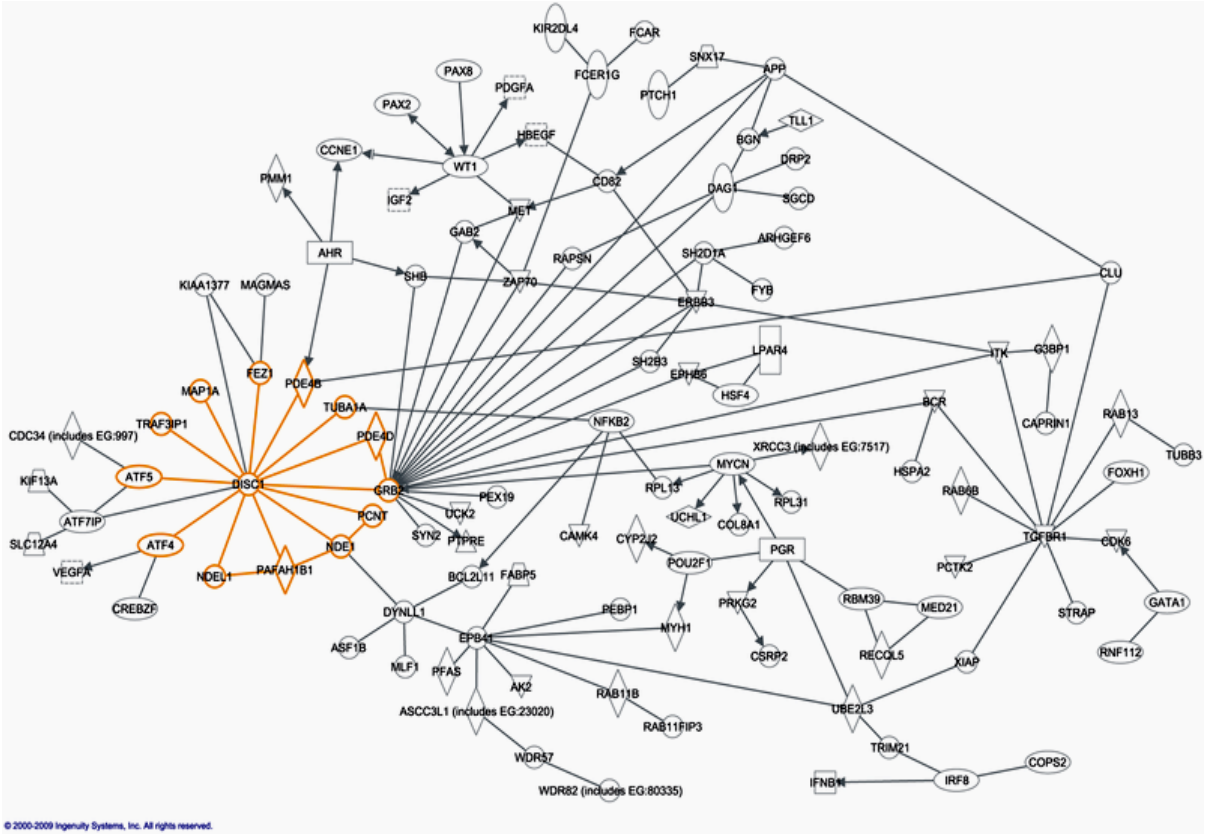


Interactome

whole set of molecular interactions in cells

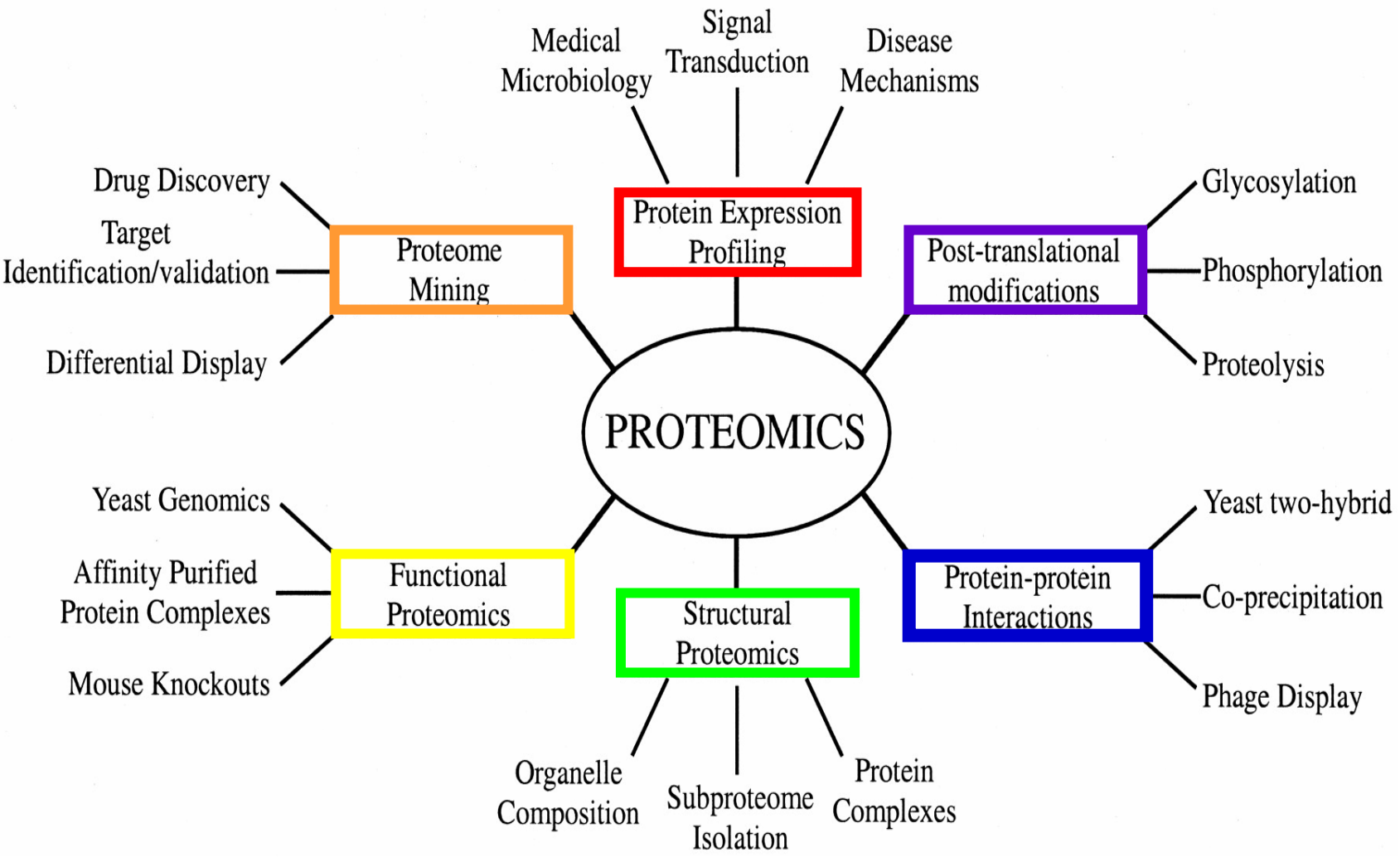
Hennah, Porteous, The DISC1 Pathway Modulates Expression of Neurodevelopmental, Synaptogenic and Sensory Perception Genes

http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0004906]

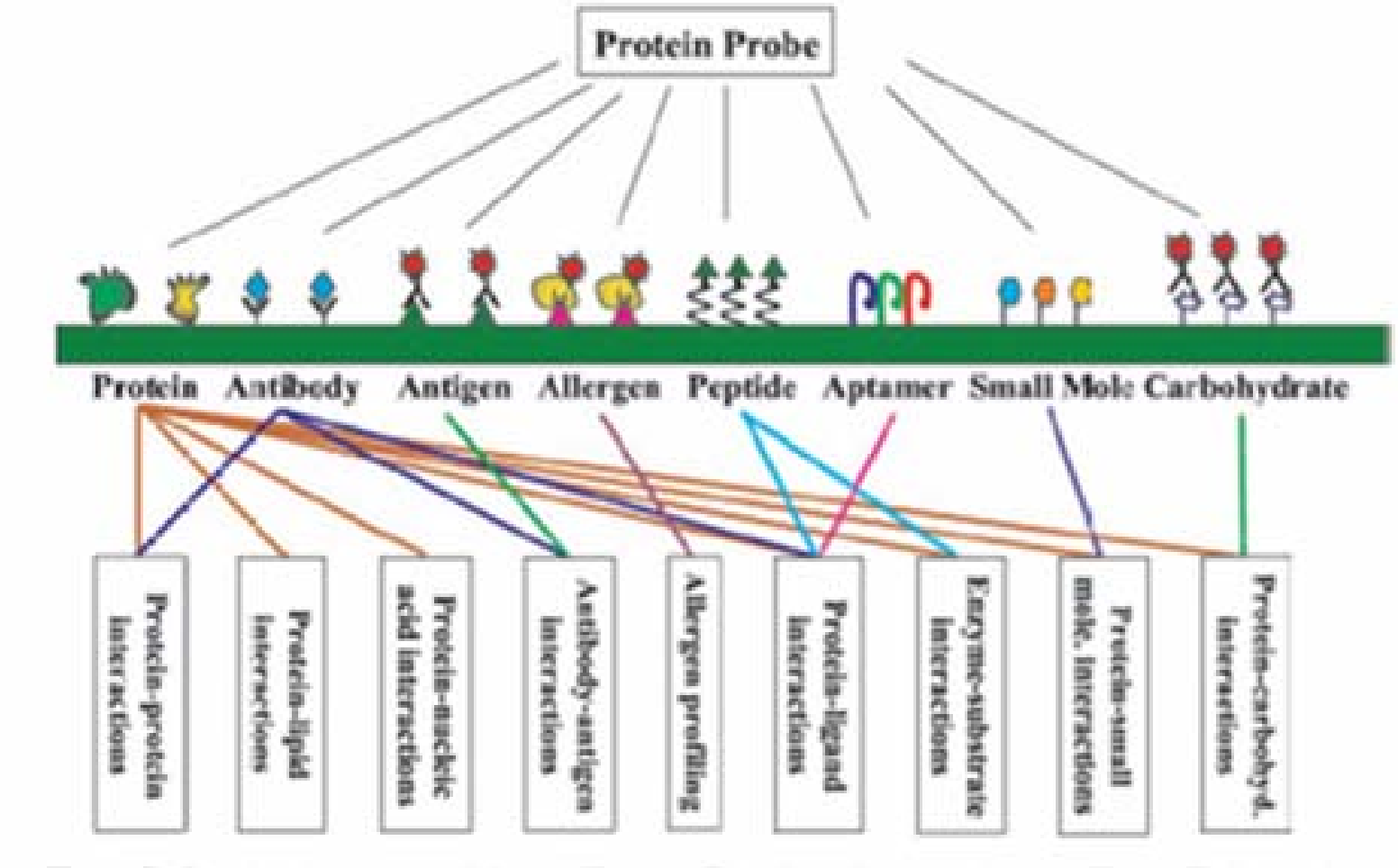


http://en.wikipedia.org/wiki/File:Network_of_how_100_of_the_528_genes_identified_with_significant_differential_expression_relate_to_DISC1_and_its_core_interactors.png

Types of Proteomics and Their Applications to Biology



Protein microarrays



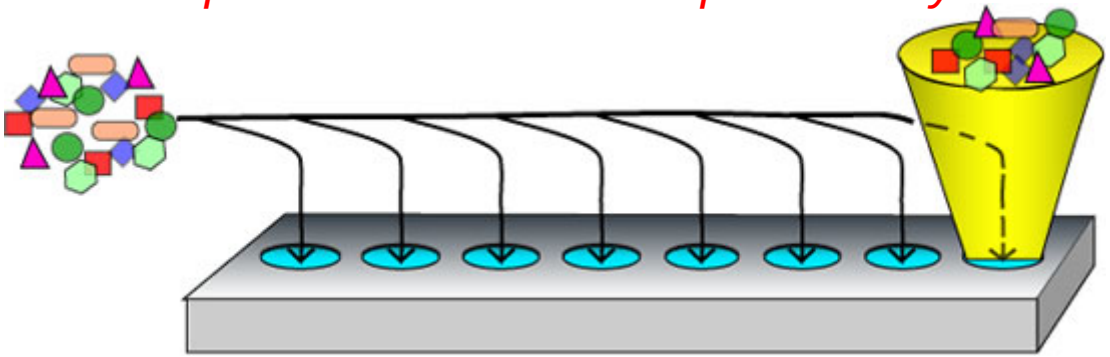
Zhu H, **Snyder M**. Protein arrays and microarrays.
Curr Opin Chem Biol. 2001;5: 40-45.

Fasolo, J. & **Snyder, M.** Protein microarrays.
Methods Mol Biol 2009 548, 209-222.

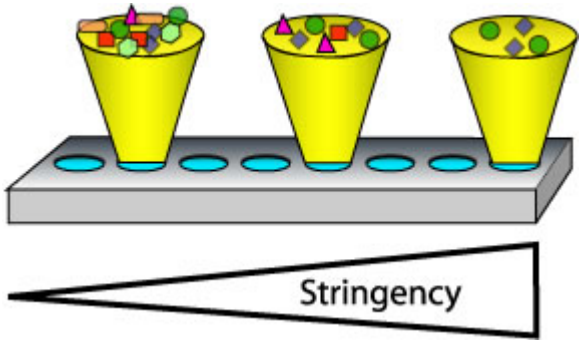
Protein Analysis by SELDI-MS

Surface Enhanced Laser Desorption/Ionization Mass Spectrometry

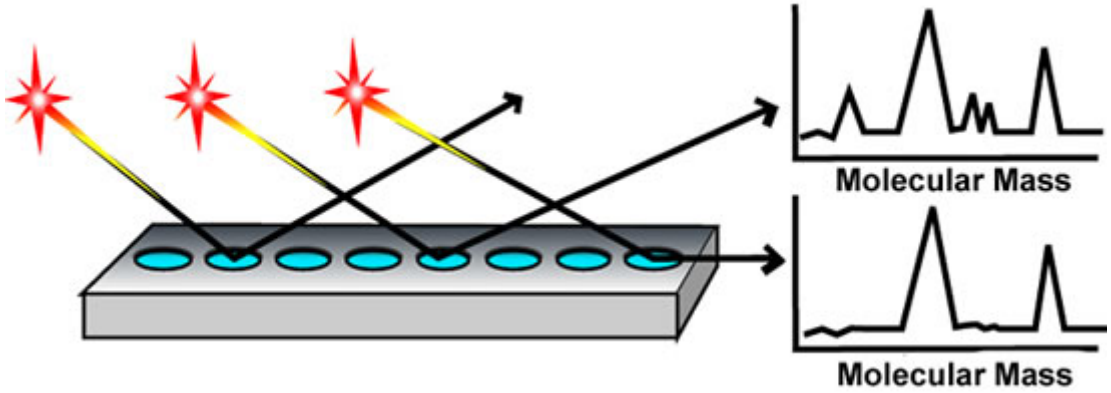
1) Apply sample (serum, tissue extract, etc.) to Protein Chip array.



2) Wash sample with increasing stringency to remove non-specific proteins.



3) Energy absorbing molecules are added to retained proteins. Following laser desorption and ionization of proteins, Time-of Flight (TOF) mass spectrometry accurately determines their masses



Source:http://dir.niehs.nih.gov/proteomics/emerg3.htm

J. Biomed. & Biotechnol., 2003 Vol 4 237-241

DNA microarray

Affymetrix website: www.affymetrix.com

Stanford University: genome-www.stanford.edu

Nature Genetics, vol. 21 supplement, “The Chipping Forecast”

www.microarray.org

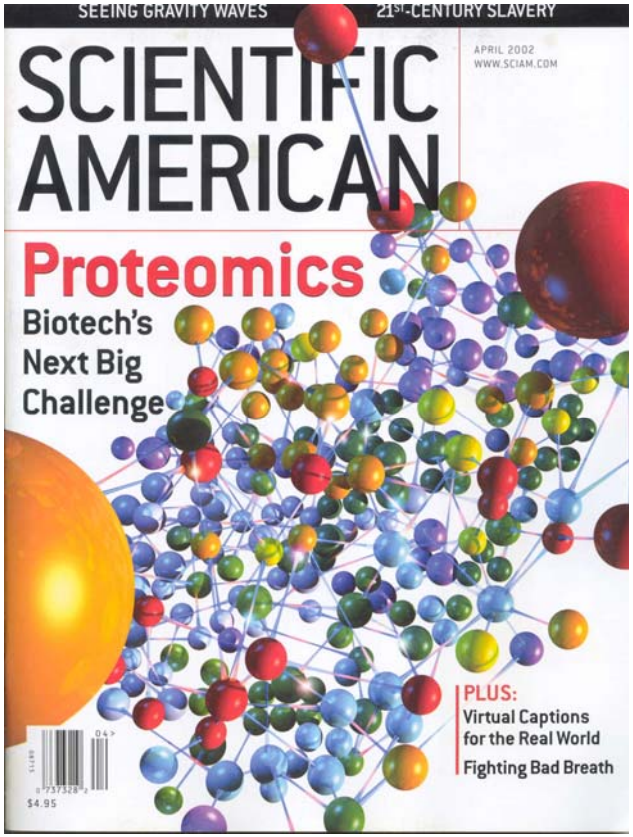
www.gene-chips.com/

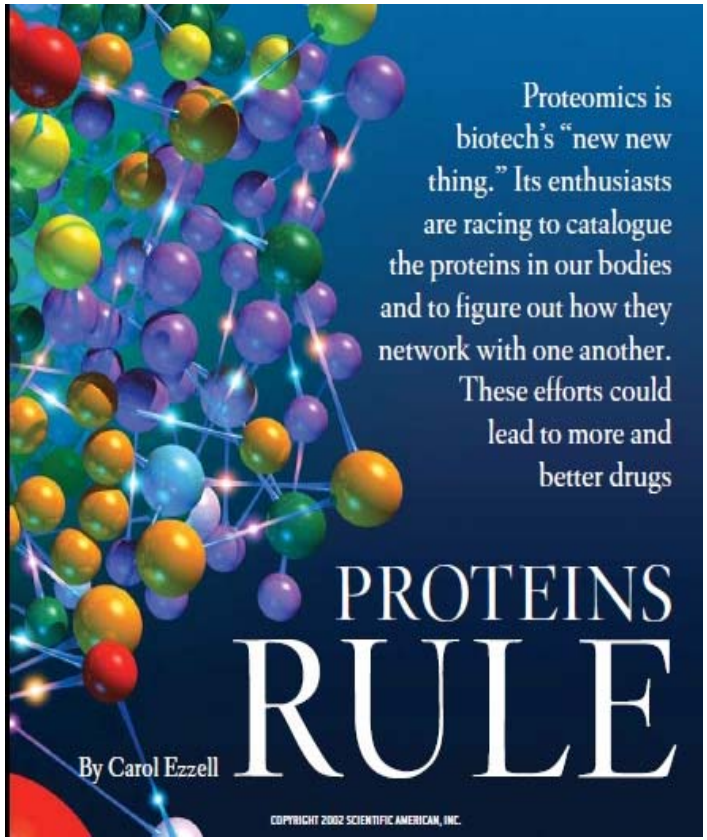
ihome.cuhk.edu.hk/~b400559/array.html

www.stat.wisc.edu/~yandell/statgen/reference/array.html

HUMAN GENOME ORGANIZATION

http://www.hugo-international.org/





http://facstaff.uwa.edu/dsalter/biochemistry%20sp02/biochem%20lit%20first%20one/1483392.pdf

High-Speed Biologists Search for Gold in Proteins.
*Science (2001) 294:*2074–2077

Proteomics' new order
Nature (2005) 437:169-170

HUMAN PROTEOME ORGANIZATION
http://www.hupo.org

New England Journal of Medicine - Getting to the Heart of Proteomics – January 29, 2009

Science - Proteomics Ponders Prime Time –
September 26, 2008

Nature - Biologists initiate plan to map human proteome
– Nature Vol 452 24 April 2008

JPR - HUPO's Human Proteome Project: the next big thing?
– May 6, 2008

ProteoMonitor - HUPO Plans Ambitious 10-Year, $1B Project to Map Entire Human Proteome – May 1, 2008

Naturenews - Biologists initiate plan to map human proteome
– April 28, 2008

Nature - The big ome It's time to make the case for proteins.
– April 24, 2008

John B. Fenn **Koichi Tanaka**

Fenn, J.B., 2003. Electrospray wings for molecular elephants (Nobel lecture). Angew. Chem. Int. Ed Engl. 42, 3871–3894.

Tanaka, K., 2003. The origin of macromolecule ionization by laser irradiation (Nobel lecture). Angew. Chem. Int. Ed. Engl. 42, 3860–3870.

KEGG – Kyoto Encyclopedia of Genes and Genomes

A grand challenge in the post-genomic era is a complete computer representation of the cell, the organism, the ecosystem, and the biosphere, which will enable computational prediction of higher-level complexity of cellular processes and organism behaviors from genomic and molecular information.

240 organisms; 20,000 organism-specific pathways; 782,135 genes

A database for post-genome analysis
Minoru Kanehisa. Trends in Genetics 13: 375-376

KEGG for representation and analysis of molecular networks
involving diseases and drugs.
Minoru Kanehisa etal.,
Nucleic Acids Research, 2010, Vol. 38, D355–D360

http://www.genome.jp/kegg/