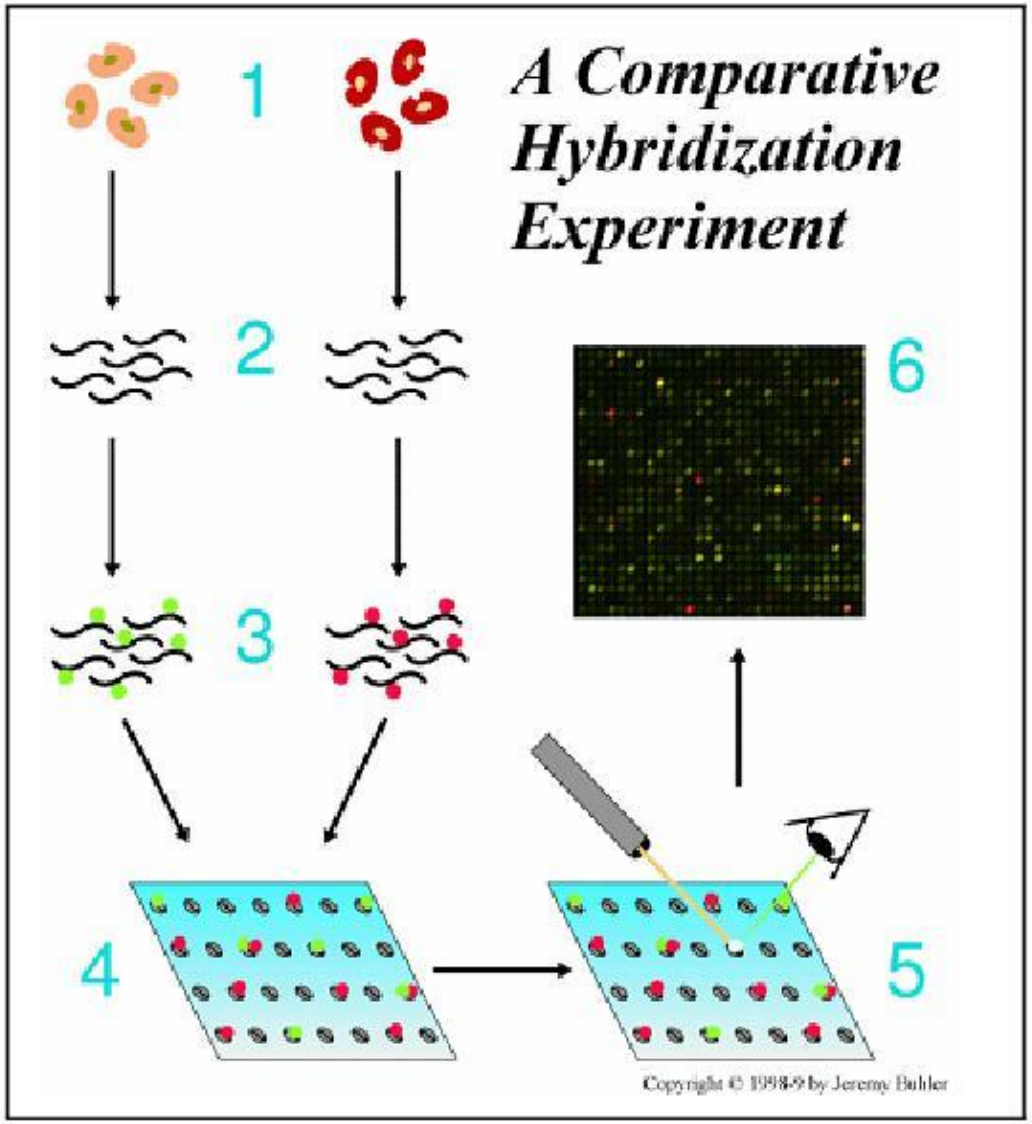
# CISC 636 Computational Biology & Bioinformatics (Fall 2016)

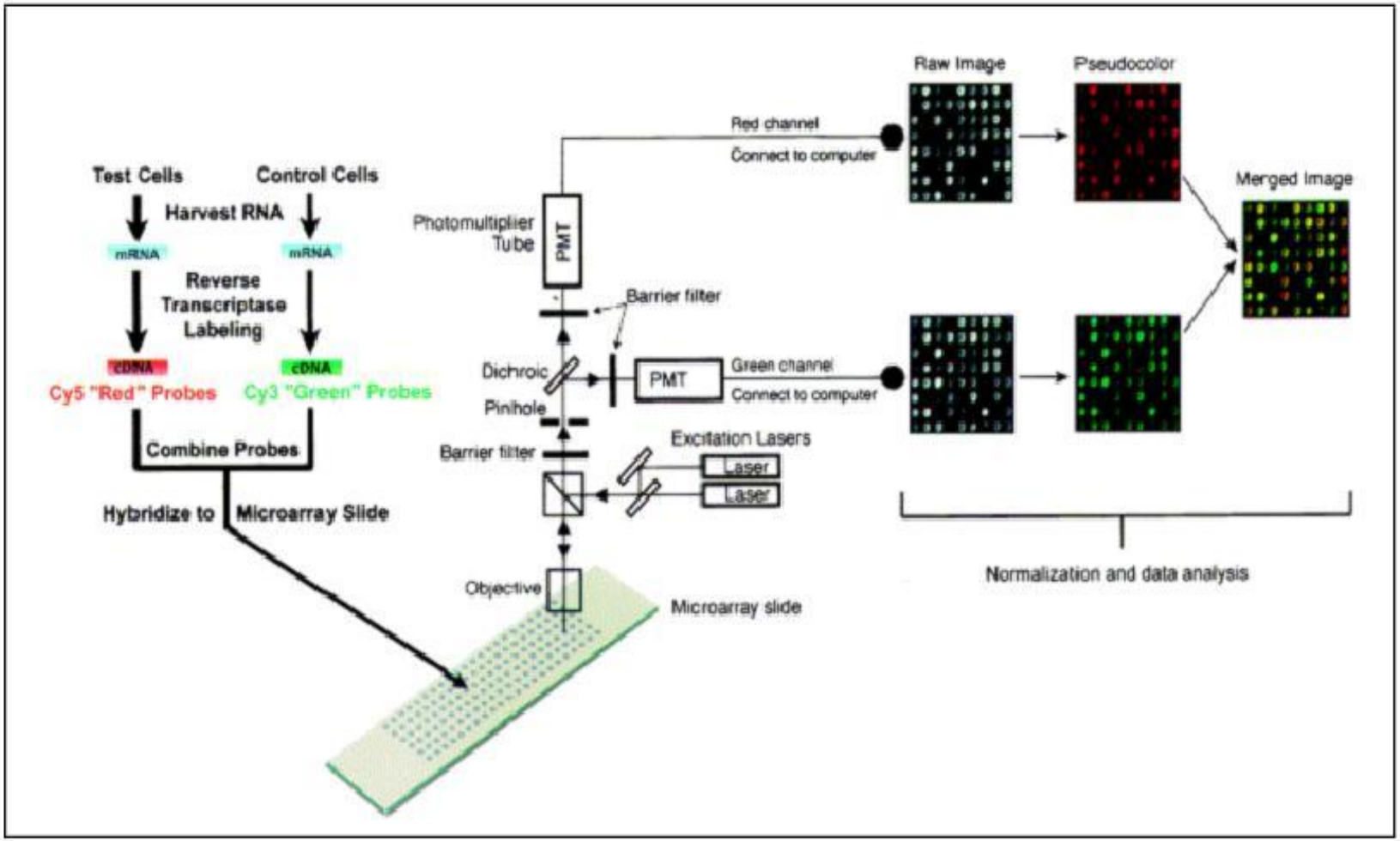
## DNA Microarray, 2d gel, MSMS, yeast 2-hybrid.

# Gene expression

- How many copies of a gene (its product) is present in the cell?
- For experimental reasons, gene expressions are measured by numbers of mRNAs, not directly by proteins. (See Proteomics)
- Various cell types are due to different genes expressed.
- The difference between diseased (e.g., cancerous) and non-diseased
- Diseased cells are often resulted from the abnormal levels of expression of key genes.


A Comparative Hybridization Experiment
1
2
3
4
5
6
Copyright © 1998-9 by Jeremy Buhler

- Microarray
  - Oligonucleotide (Affymetrix) array
    - Oligo (~ 25 bases long)
    - High density (1cm$^2$ contain 100k oligos)
  - cDNA array
    - cDNA (RT-PCR), much longer (> 1000 bases)
    - Varied density of cDNA on each spot, hybridization depends on length
    - Less possibility for false positives
  - Image processing
  - Background subtraction
  - Normalization

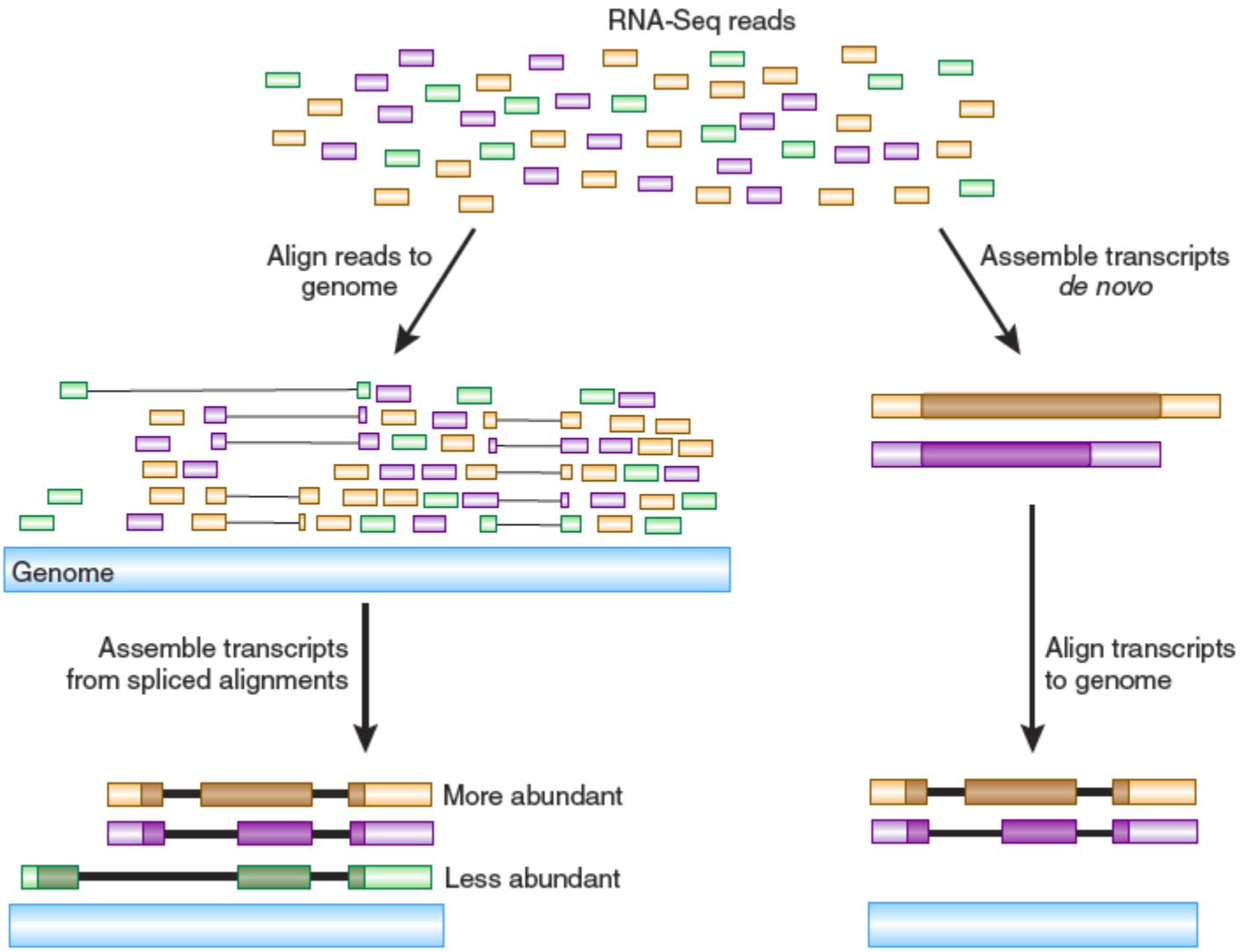

Test Cells
Control Cells
Harvest RNA
mRNA
mRNA
Reverse Transcriptase Labeling
cDNA
cDNA
Cy5 "Red" Probes
Cy3 "Green" Probes
Combine Probes
Hybridize to Microarray Slide
Photomultiplier Tube
PMT
Barrier filter
Dichroic
PMT
Green channel
Connect to computer
Pinhole
Barrier filter
Excitation Lasers
Laser
Laser
Objective
Microarray slide
Red channel
Connect to computer
Raw Image
Pseudocolor
Merged Image
Normalization and data analysis


JOURNAL OF BACTERIOLOGY
AUGUST 2000, VOLUME 182, NUMBER 16, PAGES 4375–4676
Journal of Bacteriology
AMERICAN SOCIETY FOR MICROBIOLOGY
AUGUST 2000, VOLUME 182, NUMBER 16
Published Twice Monthly

# Applications

- Inferring transcription regulatory networks
- Understanding correlation between genotype and phenotype
- predicting genotype <=> phenotype
- Phenotypes:
  - drug/therapy response
  - drug-drug interactions for expression
  - drug mechanism
  - interacting pathways of metabolism

# Transcriptome by RNA-seq



What is proteomics?

- Like genomics is the study of all genes in a genome, proteomics is the study of all proteins of a cell at a given time.
- Three aspects
  - Biological process (why is this being done? e.g. movement of cell)
  - Molecular function (what kind of molecule is this? e.g., ATPase)
  - Cellular component (where is this located? e.g., ribosome)

Why is it difficult?

- Moving target
  - Cell-to-cell variations
  - Cell behavior changes with time
- Lack of high throughput technology
  - Protein chips? Protein sequences do not have hybridization that DNA sequences have.

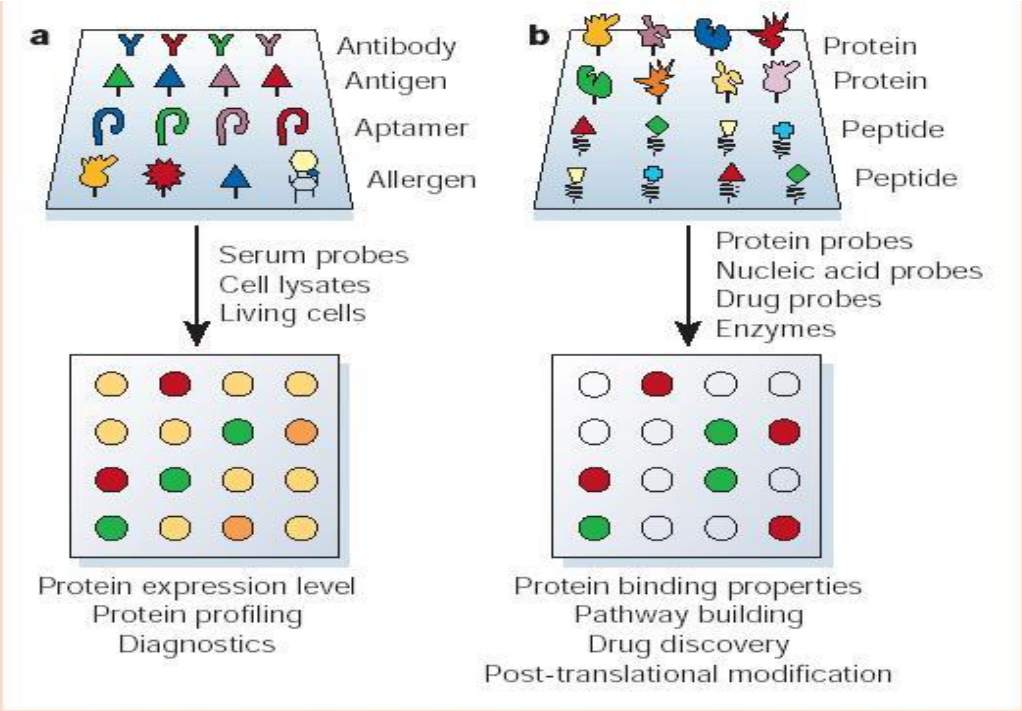



**Figure 1** Analytical versus functional protein microarrays. **a**, Analytical protein microarray. Different types of ligands, including antibodies, antigens, DNA or RNA aptamers, carbohydrates or small molecules, with high affinity and specificity, are spotted down onto a derivatized surface. These chips can be used for monitoring protein expression level, protein profiling and clinical diagnostics. Similar to the procedure in DNA microarray experiments, protein samples from two biological states to be compared are separately labelled with red or green fluorescent dyes, mixed, and incubated with the chips. Spots in red or green colour identify an excess of proteins from one state over the other. **b**, Functional protein microarray. Native proteins or peptides are individually purified or synthesized using high-throughput approaches and arrayed onto a suitable surface to form the functional protein microarrays. These chips are used to analyse protein activities, binding properties and post-translational modifications. With the proper detection method, functional protein microarrays can be used to identify the substrates of enzymes of interest.

# 2D gel electrophoresis

- Isoelectric points (first dimension)
- Molecular weights (second dimension)

Both pI and MW are functions of amino acid sequence of a protein.

Some proteins do not resolve well by 2D gels.

Issues:

- Detection of spots (image processing)
- Quantification of each spot
- Identification of each spot (Mass Spectrometry)

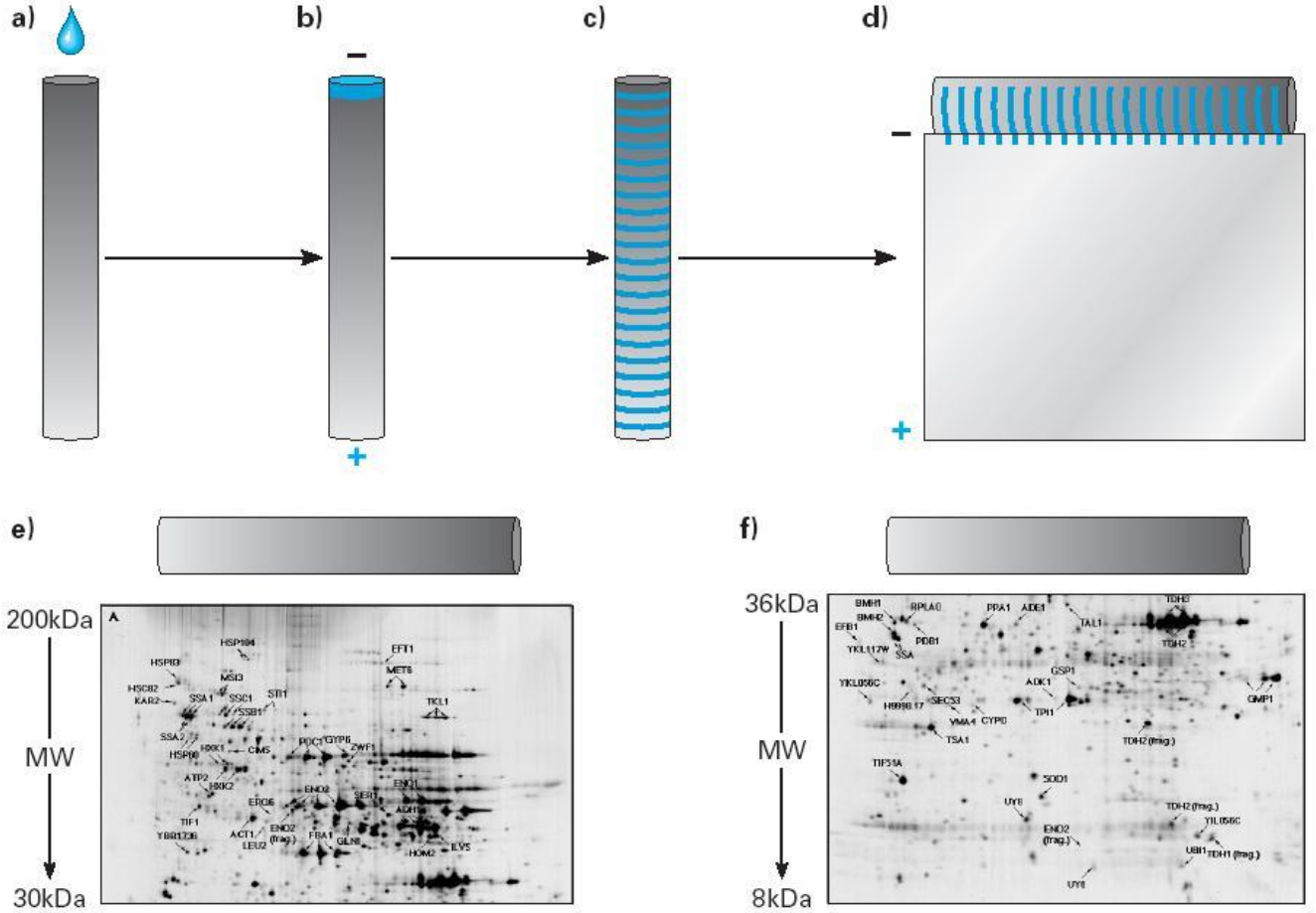
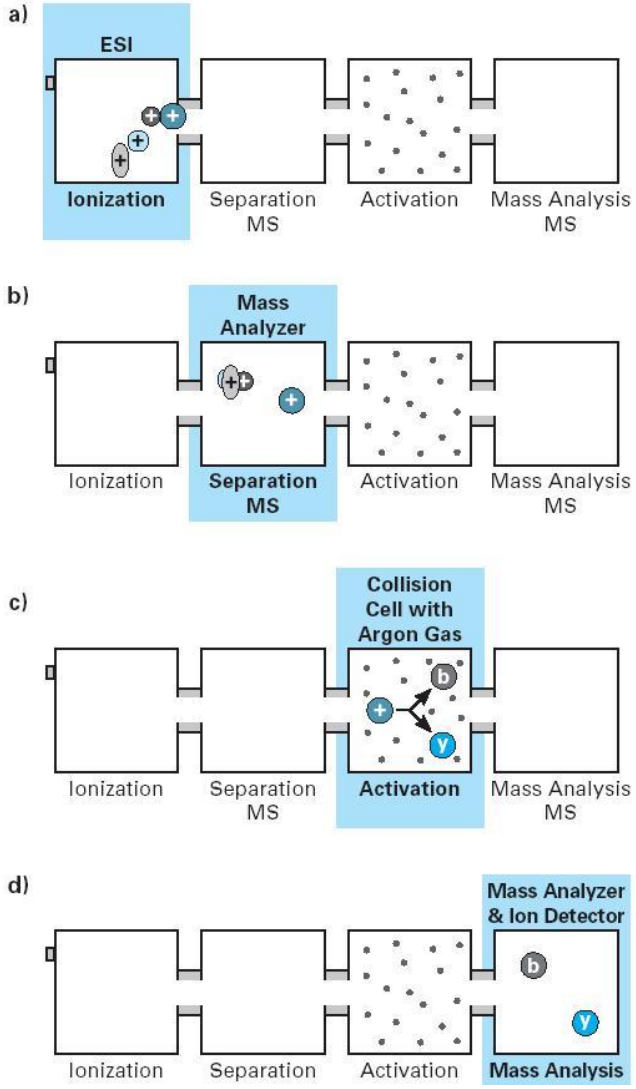



**FIGURE 6.18 • Two-dimensional (2D) gel electrophoresis.** Each column (a–c) with a gradation of gray shading represents the isoelectric focusing gel with a pH gradient. **a)** A mixture of proteins (blue drop) is applied to the isoelectric focusing gel and **b)** exposed to an electrical current. **c)** Proteins migrate to their isoelectric points (pI) and stop moving. **d)** This tubular gel is placed on top of a slab polyacrylamide gel that contains SDS and is subjected to electrophoresis (SDS-PAGE). Proteins migrate into the slab gel according to their molecular weights. Yeast cells were grown in rich media and subjected to 2D gel analysis. Using duplicate isoelectric focusing gels, large **e)** and small **f)** proteins were analyzed on separate gels. The same spots appear at the bottom of e) and the top of f). Molecular weights are resolved on the Y-axis and pIs on the X-axis. Panels e) and f) are from the Swiss 2D database at ExPASy.


a)
ESI
Ionization
Separation MS
Activation
Mass Analysis MS
b)
Mass Analyzer
Ionization
Separation MS
Activation
Mass Analysis MS
c)
Collision Cell with Argon Gas
b
y
Ionization
Separation MS
Activation
Mass Analysis MS
d)
Mass Analyzer & Ion Detector
b
y
Ionization
Separation
Activation
Mass Analysis

a)

**S–P–A–F–D–S–I–M–A–E–T–L–K**
(protonated mass 1410.6)

| Mass$^+$ | b-ions | y-ions | Mass$^+$ |
|---|---|---|---|
| 81.1 | S | PAFDSIMAETLK | 1323.6 |
| 185.2 | SP | AFDSIMAETLK | 1226.4 |
| 256.3 | SPA | FDSIMAETLK | 1155.4 |
| 403.5 | SPAF | DSIMAETLK | 1008.2 |
| 518.5 | SPAFD | SIMAETLK | 893.1 |
| 605.6 | SPAFDS | IMAETLK | 806.0 |
| 718.8 | SPAFDSI | MAETLK | 692.3 |
| 850.0 | SPAFDSIM | AETLK | 561.7 |
| 921.1 | SPAFDSIMA | ETLK | 490.6 |
| 1050.2 | SPAFDSIMAE | TLK | 361.5 |
| 1151.3 | SPAFDSIMAET | LK | 260.4 |
| 1264.4 | SPAFDSIMAETL | K | 147.2 |

b)

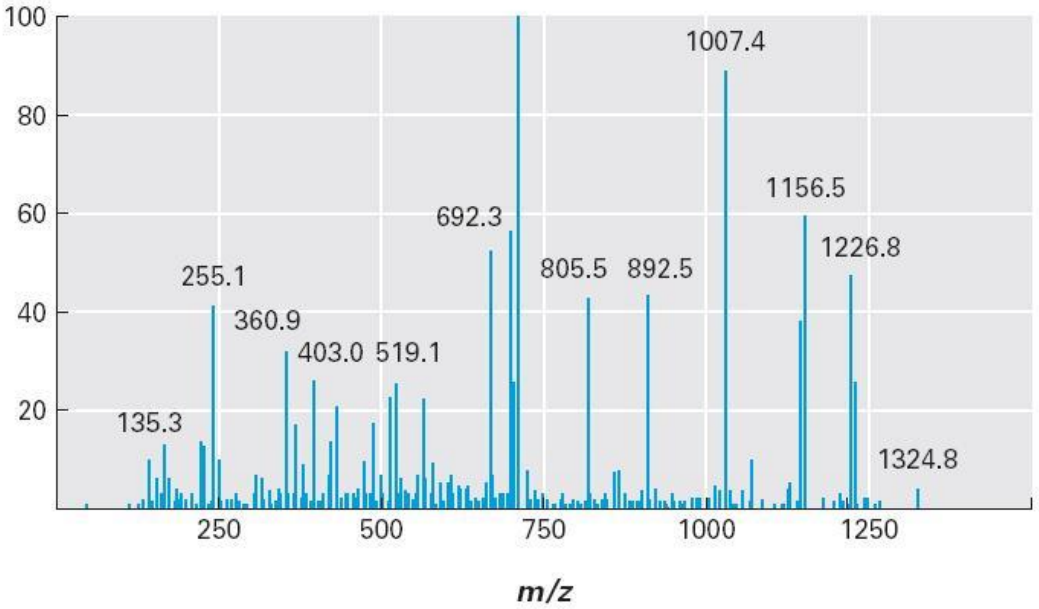



**FIGURE 6.20 • Protein identification through peptide fragment formation and separation.** When a group of identical proteins is broken into its peptide pieces, more than one pair of b and y peptides will be formed. **a)** One protein sequence and its calculated mass on top, with the b peptides/masses (gray) and the y peptides/masses (blue) below. **b)** An experimentally determined mass/charge spectrum from the peptide in a). Notice that some peaks are higher than others, which means that some b/y peptide pieces were more abundant than others. The spectrum is used to determine the peptide's amino acid sequence and protein identity.

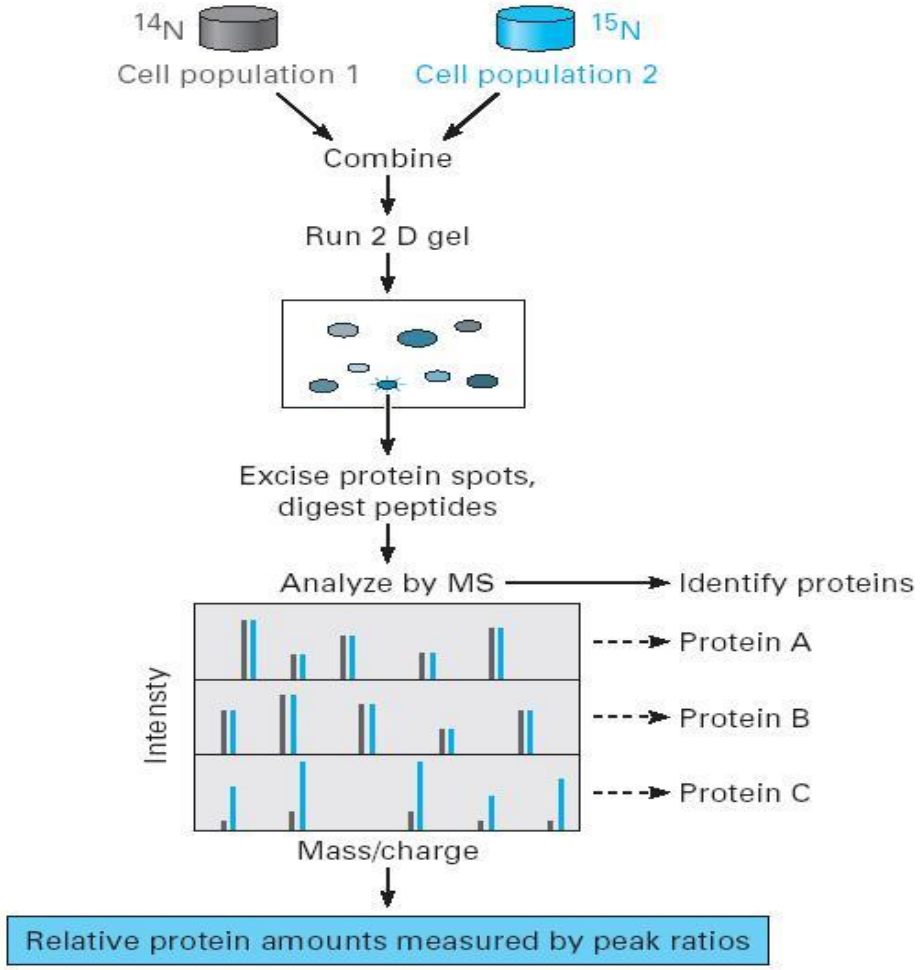



**FIGURE 6.24 • Quantifying differences in proteomes.** Two cell pools are grown in the presence (blue = cell population 2) or absence (gray = cell population 1) of heavy nitrogen $^{15}N$. The proteins are extracted, pooled, and subjected to 2D gel analysis; spots excised; and proteins identified and quantified by MS/MS. The relative areas under the pairs of heavy and light peptide peaks indicate relative abundance of each protein pair.

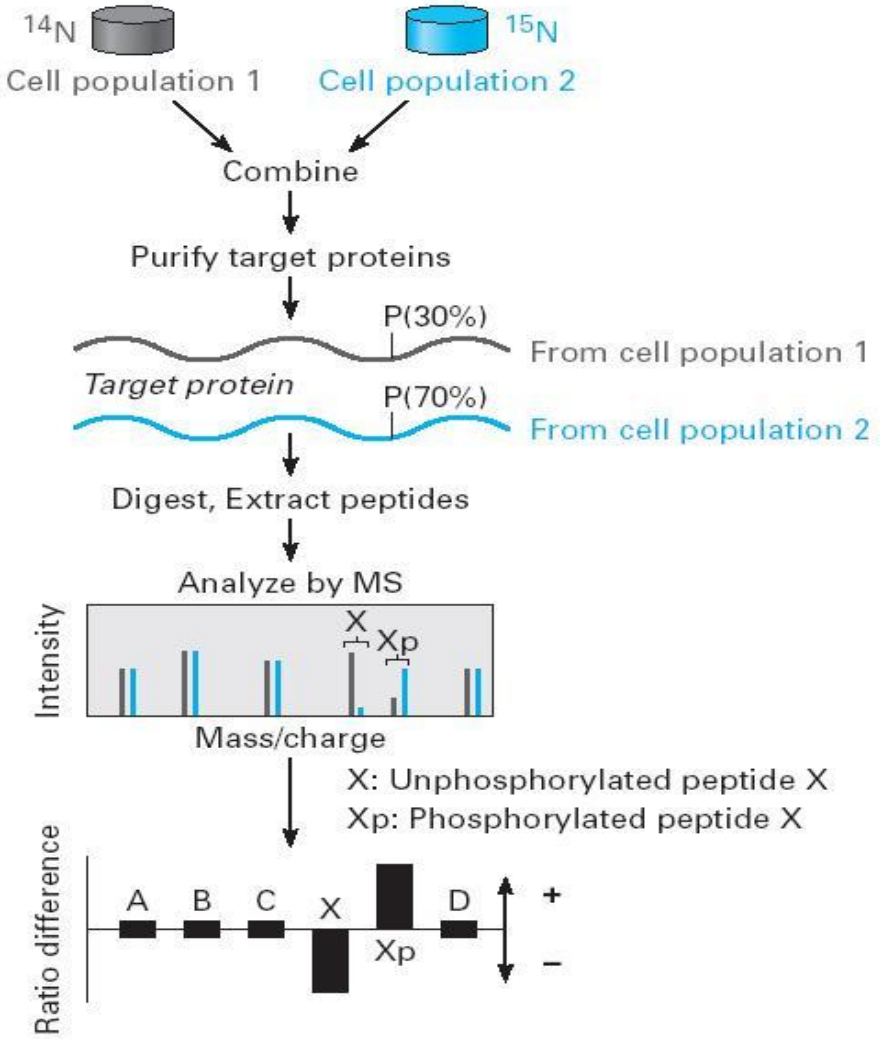



**FIGURE 6.25 • Quantifying relative levels of phosphorylation.** Bar graphs A, B, C, and D indicate no difference for proteins A–D in the two populations. Compared to population 1, unphosphorylated protein X was reduced in population 2. Phosphorylated protein X (Xp) was increased in population 2.

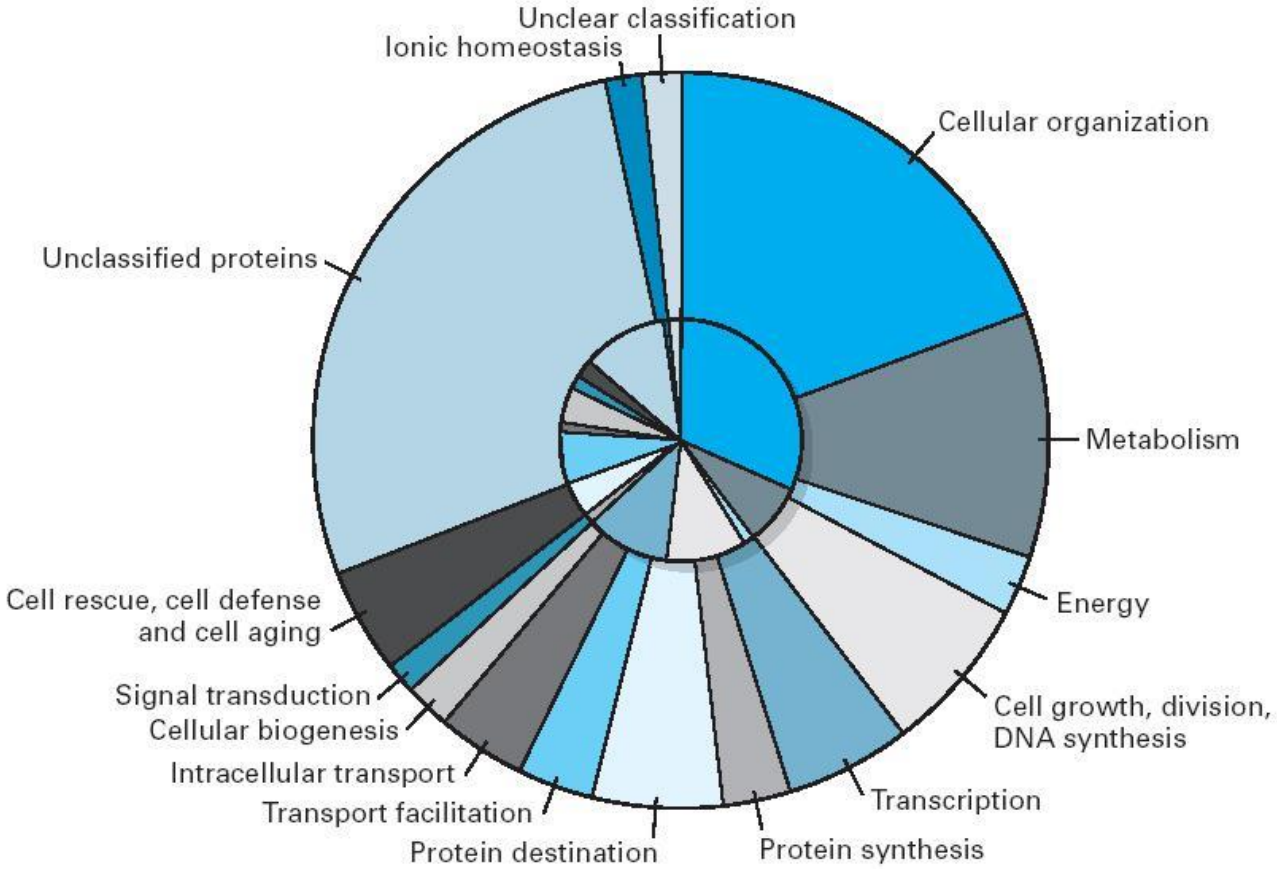



**FIGURE 6.9 • Bar code analysis of biological processes.** Distribution of functional classes of essential (inner circle) and nonessential (outer circle) genes using criteria from the Munich Information Center for Protein Sequences (MIPS).

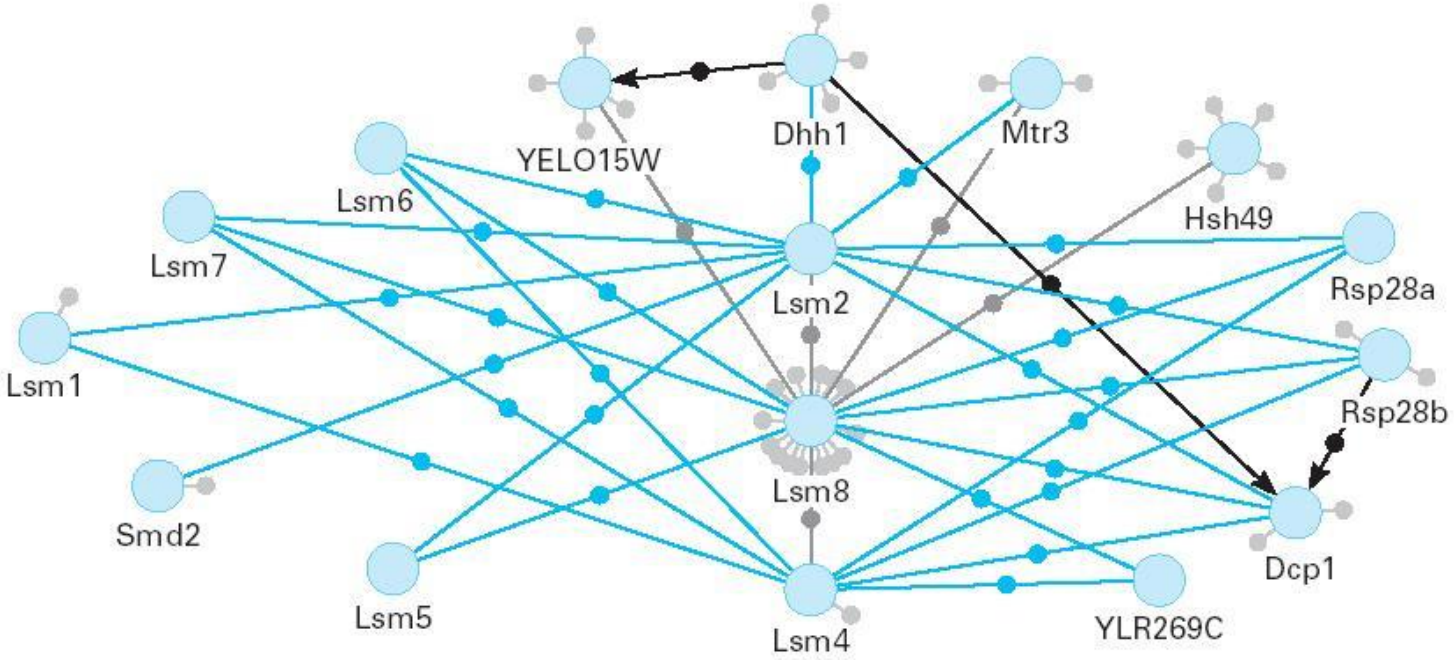



**FIGURE 6.15 • Proteomics circuit showing the interactions of RNA splicing proteins.** The proteins are indicated by large blue nodes and their interactions with lines. The dots on the lines help you follow each line. Blue lines show interactions that were detected by traditional Y2H array screens, black from multiple high-throughput screens, and gray from literature and array screens. The black arrows point away from the protein used as bait in the screens. Small gray nodes indicate other protein-protein interactions not highlighted here.

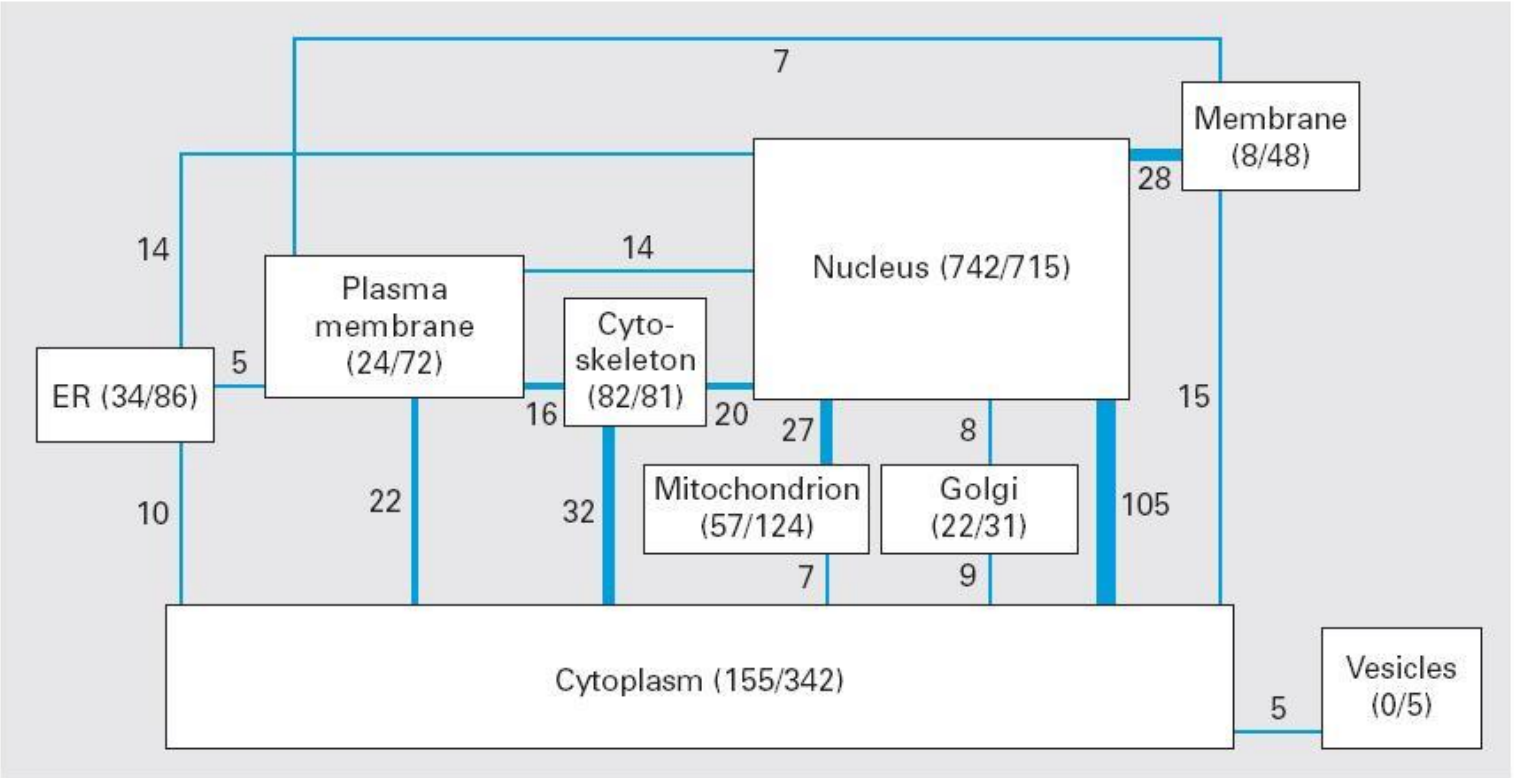



**FIGURE 6.17 • Protein interactions grouped by cellular compartments.** Numbers in parentheses indicate the number of interactions in the circuit diagram among proteins of this compartment/the total number of proteins in the same compartment. Lines connecting compartments indicate the number of protein interactions by the thickness of each line (numbers of connections are near each line). For example, there are 7 interactions between the 48 membrane proteins and the 72 plasma membrane proteins in "Benno Figure 1."

# Yeast two hybrid System 1

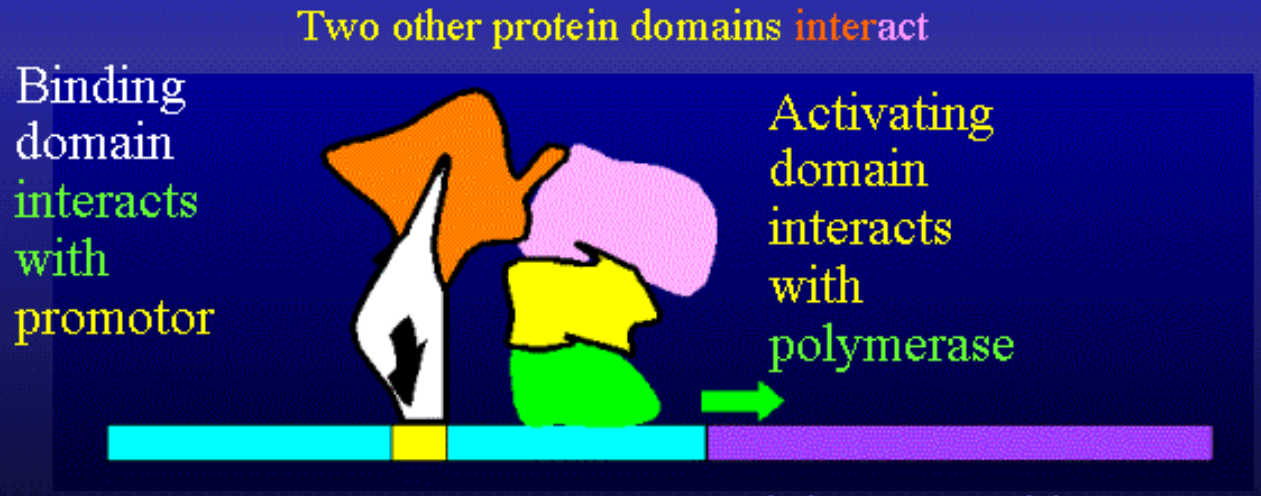
Gal4 protein: comprises DNA binding and activating domains


Measure reporter enzyme activity as e.g. blue colonies.

David B. Collinge KVL

# Yeast two hybrid System 2

- Gal4 protein: the two domains of the protein do not need to be transcribed in a single protein
- Just as long as they come to interact




Measure reporter enzyme activity as e.g. blue colonies.

David B. Collinge KVL

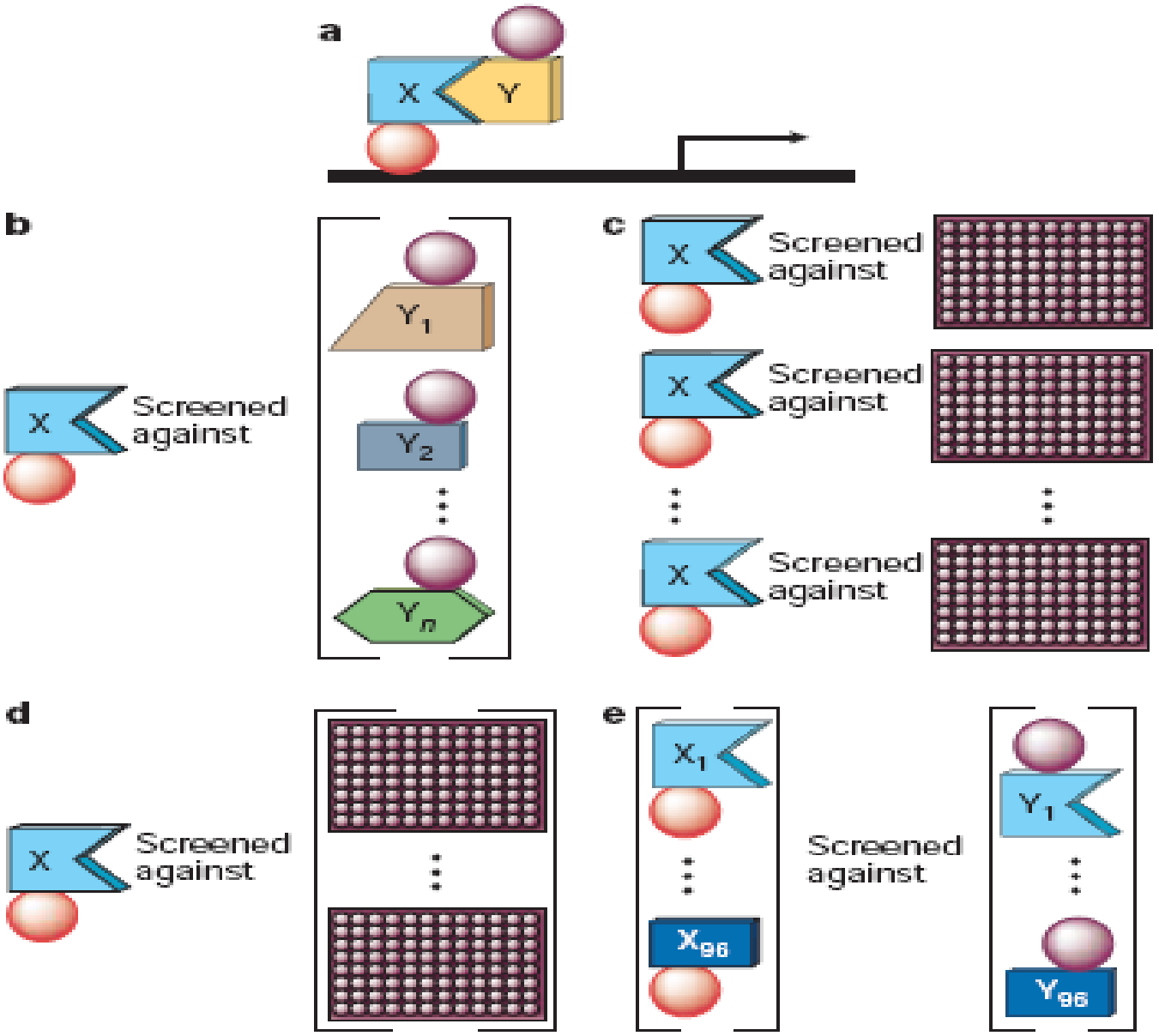
# Yeast two hybrid System (3)

- This is achieved using gene fusions:
- Plasmids carrying different constructs can be expressed in yeast.

Binding domain as a translational fusion with the gene encoding another protein in one plasmid

Activating domain as a translational fusion with the gene encoding a third protein in a second plasmid

David B. Collinge KVL


a
X
Y
b
X
Screened
against
$Y_1$
$Y_2$
$Y_n$
c
X
Screened
against
X
Screened
against
X
Screened
against
d
X
Screened
against
e
$X_1$
$X_{96}$
Screened
against
$Y_1$
$Y_{96}$