Outline: Chapter 24, part II

1. Genes and genomes (Ch 24.1 and Ch 9.2)

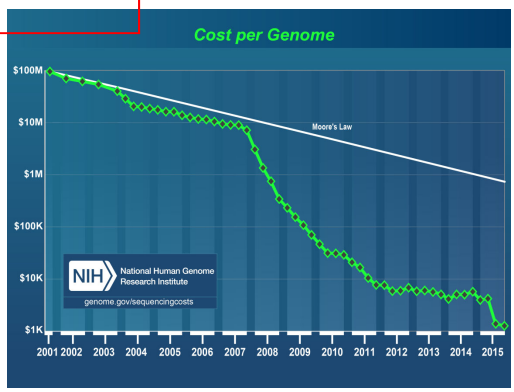
2. Biochemical approaches to understanding genes and genomes (Ch 9.3)

Explosion in genome sequence data

Organism	Year	bases
Phage Φ174	1977	5386
Human mito DNA	1981	16,589
H. influenzae	1995	1,830,137
S. cerevisiae	1996	12 million
C. elegans	1998	97 million
Human	2003 (2000)	~3 billion

GOLD database:
Jan 2019, ~150,000 projects
- ~130,000 Bacteria
- ~4,000 Eukarya

(gold.jgi.doe.gov)



Genomes are very different

TABLE 24–2 DNA, Gene, and Chromosome Content in Some Genomes

	Total DNA (bp)	Number of chromosomes*	Approximate number of genes
Escherichia coli K12 (bacterium)	4,639,675	1	4,435
Saccharomyces cerevisiae (yeast)	12,080,000	16†	5,860
Caenorhabditis elegans (nematode)	90,269,800	12‡	23,000
Arabidopsis thaliana (plant)	119,186,200	10	33,000
Drosophila melanogaster (fruit fly)	120,367,260	18	20,000
Oryza sativa (rice)	480,000,000	24	57,000
Mus musculus (mouse)	2,634,266,500	40	27,000
Homo sapiens (human)	3,070,128,600	46	29,000

Note: This information is constantly being refined. For the most current information, consult the websites for the individual genome projects.

*The diploid chromosome number is given for all eukaryotes except yeast.

†Haploid chromosome number. Wild yeast strains generally have eight (octoploid) or more sets of these chromosomes.

‡Number for females, with two X chromosomes. Males have an X but no Y, thus 11 chromosomes in all.

Table 24-2
Lehninger Principles of Biochemistry, Fifth Edition

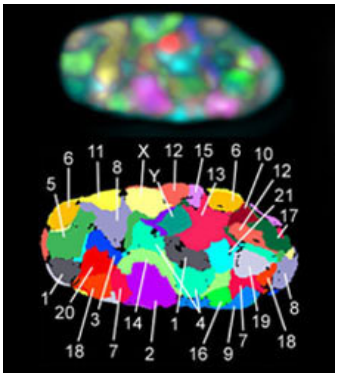
Understanding the genome: health and disease

Explore the human genome:

ncbi.nlm.nih.gov/genome
omim.org

Era of Genomic Medicine:

- **disease-causing genes**
- **pharmacogenomics**
- **cancer genomics**
- **epigenetics/microbiome**



unlockinglifescode.org/the-genome-ball

Understanding the genome: biochemical function

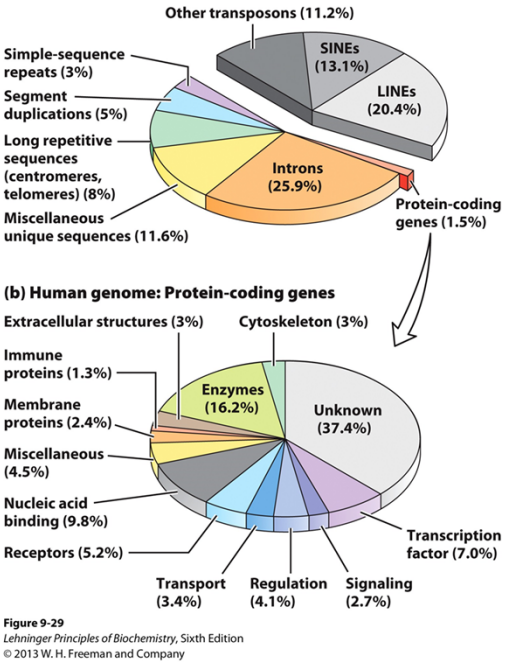


Figure 9-29
Lehninger Principles of Biochemistry, Sixth Edition

Understanding the genome: many "omes"

Genome: Complete set of DNA sequences in chromosome(s)

Transcriptome: Complete set of RNA (coding and non-coding) made from those DNA sequences

Proteome: Complete set of protein molecules made from all mRNAs

Metabolome: Complete set of small molecule metabolites

Interactome: Complete set of protein complexes, or protein-protein interactions

Microbiome: Complete set of microorganisms living in part of body

Understanding the genome: biochemical approaches

What is a genomic approach?

Use methods designed to yield information about many gene products simultaneously

Learn about roles for RNA or proteins produced in cells

Requires "high-throughput" technologies

Goals: to define molecular function of all expressed macromolecules in an organism

Outline: Chapter 24, part II

1. Genes and genomes (Ch 24.1 and Ch 9.2)

2. Biochemical approaches to understanding genes and genomes (Ch 9.3)

What do biochemists want to know?
What experiments address these questions?

Tools for genome-wide analysis

1. Comparative genomics
- identify similarities to known proteins

2. Genetic analysis
- deletion ("knockout") collections

3. Cellular expression patterns
- where and when is it expressed?

4. Determine interacting partners
- who does it interact with?

Tools for genome-wide analysis

1. Comparative genomics

- Assign gene function based on comparison to known genes

- Look for homologous sequences:
  Homolog- similar proteins due to shared ancestry

- Significant advance:
  Bioinformatics- how to identify and QUANTIFY similarities

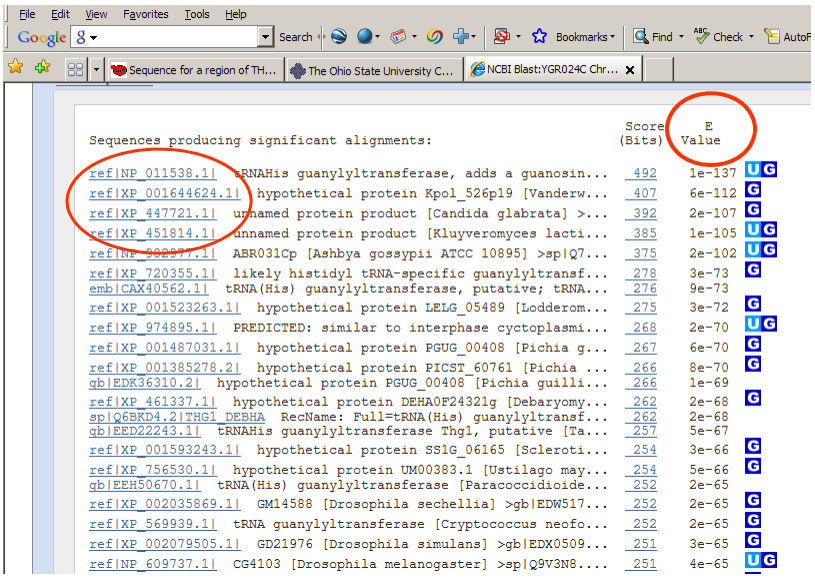
BLAST: Basic Local Alignment Search Tool

Input: A known protein sequence (can use nucleotide also)

BLAST will use this sequence as the query to search for similarity to all other proteins in any biological database

Uses statistical methods to determine how much confidence in the match: **E-value**
E = 0.05 is 1/20 chance of similarity occurring by chance alone, lower number is better

BLAST at NCBI (www.ncbi.nlm.nih.gov)



Altschul et al. (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402.

Tools for genome-wide analysis

2. Genetic analysis

- make a collection of strains; each strain contains a single, different deletion of one of known genes

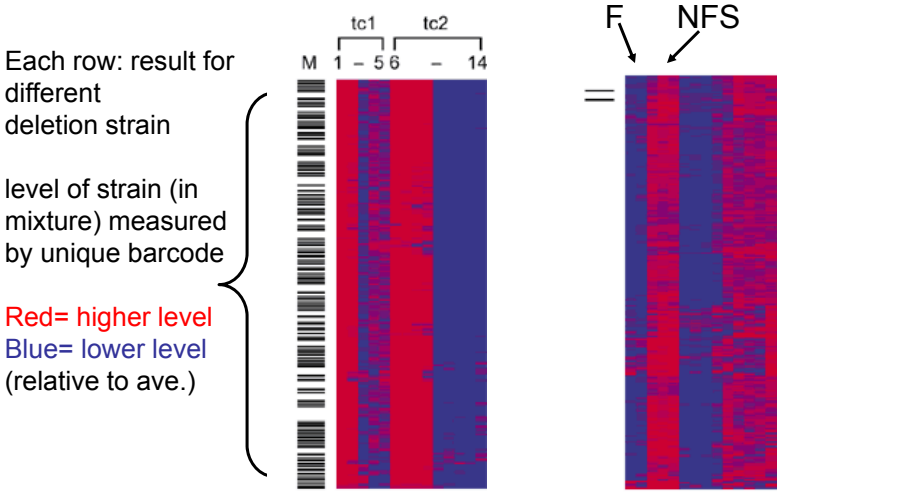
- fundamental advance: yeast deletion collection (2002); other organisms at various stages

- requirements:
  genes are non-essential for growth
  know phenotype to look for

Systematic screen for human disease genes in yeast

Lars M. Steinmetz[1,3]*, Curt Scharfe[2,3]*, Adam M. Deutschbauer[1], Dejana Mokranjac[4], Zelek S. Herman[3], Ted Jones[3], Angela M. Chu[2], Guri Giaever[3], Holger Prokisch[4], Peter J. Oefner[2,3] & Ronald W. Davis[1–3]

Nature Genetics
vol 31,p.400
July 22, 2002



Each row: result for different deletion strain

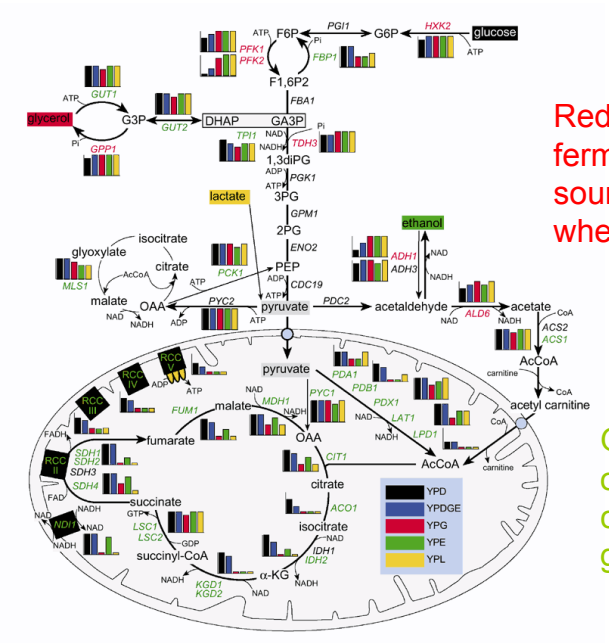
level of strain (in mixture) measured by unique barcode

Red= higher level
Blue= lower level
(relative to ave.)

Respiration defects- mitochondrial function?

Fermentation defects (do not rely on respiration)

Growth fitness is consistent with known roles of genes



Red = lower fitness on fermentable carbon source (ie glucose) when deleted

Green = lower fitness on non-fermentable carbon source (i.e. glycerol) when deleted

From group of previously unknown mitochondria-associated genes (259 new ones identified this study), identify ones for which human homolog is found in region associated with known mito disease- 11 new candidates for mito-related disease genes

Table 1 • Candidate genes associated with putative human mitochondrial-related diseases

Putative mitochondrial-related disorder	OMIM	Cytogenic location	Genetic markers flanking disease interval	Interval location (cM)	New human candidates in interval; marked (*) if yeast deletion phenotype	Previously known human candidates in interval; marked (*) if yeast deletion phenotype
Spastic paraplegia 5A	270800	8p12–q13	*PLAT–D8S279* (ref. 23)	64.6–91.5	*CGI-11*[3] (*) *LOC85479*[a] (*) *PDE7A* (*)	*MRPL15* (*)
Friedreich ataxia, FRDA2	601992	9p23–p11	*D9S285–D9S1874* (ref. 24)	27.9–59.9	*ACO1* (*) *DNAJA1* (*) *MGC14836*[a] *SR-BP1*[a] (*)	*NDUFB6* *ALDH1B1*
Optic atrophy, OPA4	605293	18q12.2–q12.3	*D18S34–D18S479* (ref. 25)	62.3–71.3	*DKFZP667C165*[a] (*)	*ATP5A1*(*) *ACAA2*
Optic atrophy, OPA2	311050	Xp11.4–p11.21	*DXS993–DXS991* (ref. 26)	66.1–86.9	*APEXL2*[a] (*) *PFKFB1* (*)	*TIMM17B* *ALAS2*
Neuropathy, motor-sensory type II, with deafness	310490	Xq24–q26	*DXS425–HPRT* (ref. 27)	126.3–152.5	*PLS3* (*)	*NDUFA1* *SLC25A14* *PDCD8*
Ptosis, hereditary congenital 2	300245	Xq24–q27.1	*DXS1047–DXS984* (ref. 28)	150.3–159.5	*MGC14797*[a] (*)	*SLC25A14* *PDCD8*
Mental retardation with optic atrophy, deafness	309555	Xq26	*DXS424–DXS297* (ref. 29)	116.8–167.3	*PLS3* (*) *MGC14797*[a] (*)	*SLC25A5* *NDUFA1* *SLC25A14* *PDCD8* *SLC9A6* (*)

[a]UniGene identifier for putative genes not listed in HUGO nomenclature database.

Tools for genome-wide analysis

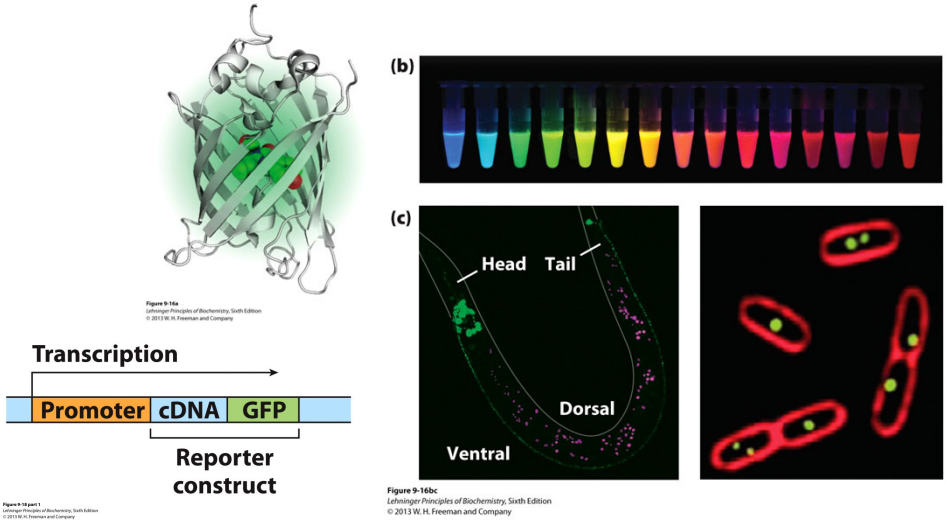
3. Cellular expression pattern

- When or where is a protein or RNA expressed?
  - specific tissues, times in development?
  - change in level in response to environment/conditions?

1) Protein expression: **fluorescent-labeled proteins** to determine location in cell

2) RNA expression: **DNA microarray** to detect and quantify RNA abundance
   - Note: RNA-Seq is a more common alternative

GFP-fusion proteins to detect cellular location

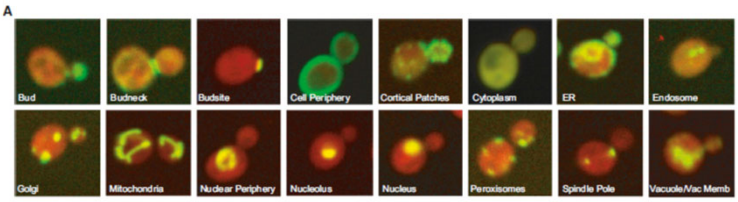


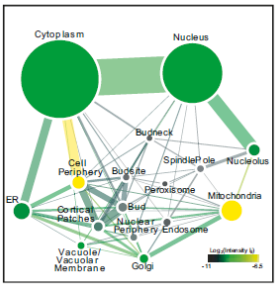
Yeast Proteome Dynamics from Single Cell Imaging and Automated Analysis

Yolanda T. Chong,[1,4,6] Judice L.Y. Koh,[1,4,7] Helena Friesen,[1] Supipi Kaluarachchi Duffy,[1,2] Michael J. Cox,[1,2] Alan Moses,[3] Jason Moffat,[1,2,5] Charles Boone,[1,2,5] and Brenda J. Andrews[1,2,5,*]

Cell *161*:1413-1424 2015

~4100 visible fusion GFP-fusion proteins (RFP as cell boundary marker)

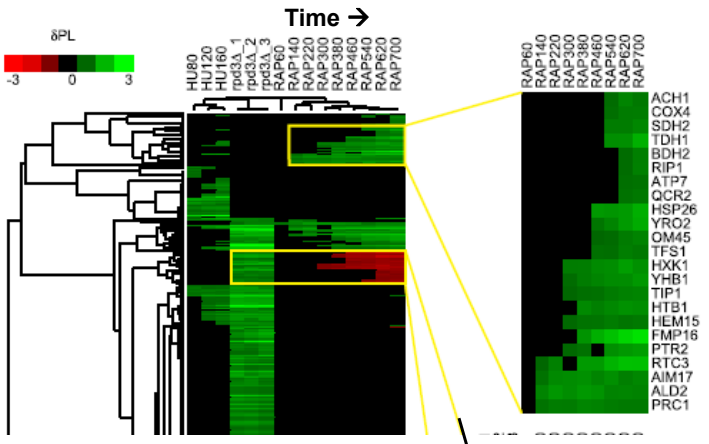




Abundance Localization Map (ALM)
- summarize all data to identify trends

Proteome abundance changes in response to perturbations

- treat collection with rapamycin (growth inhibitor) and visualize



- Vaculoar protein degradation
- Energy regeneration

Ribosomes and biogenesis factors

Differential gene expression by microarray

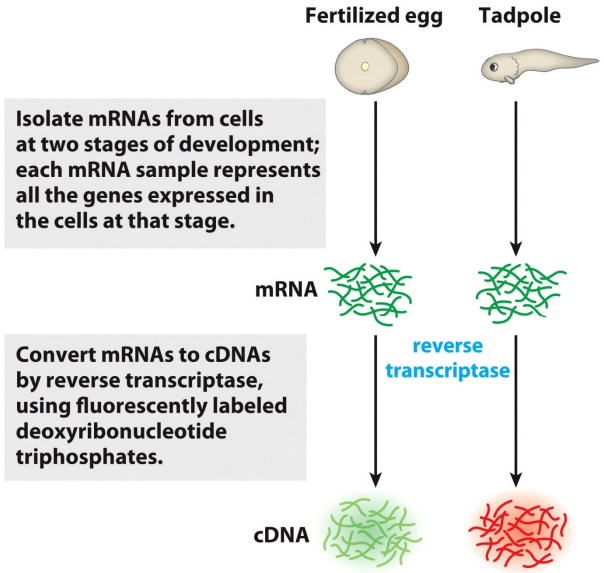


Figure 9-23 part 1
Lehninger Principles of Biochemistry, Sixth Edition
© 2013 W. H. Freeman and Company

Differential gene expression by microarray

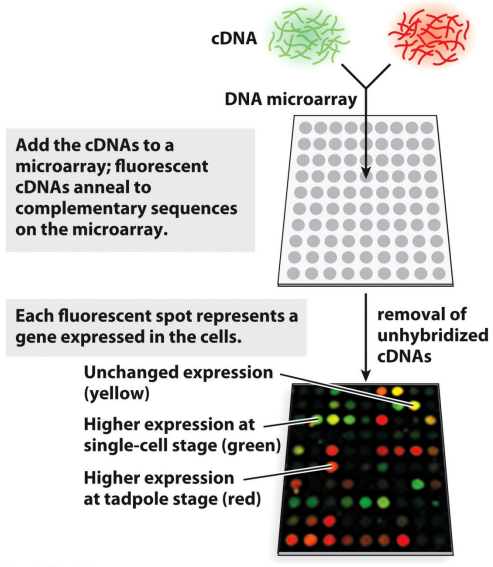


Figure 9-23 part 2
Lehninger Principles of Biochemistry, Sixth Edition
© 2013 W. H. Freeman and Company

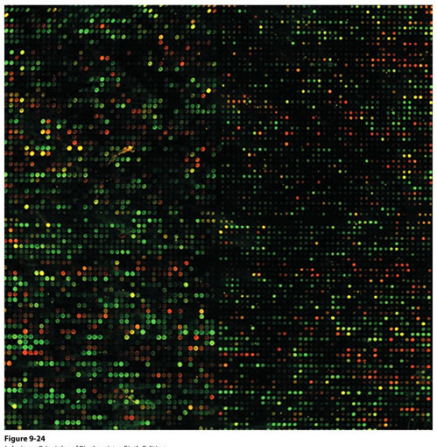


Figure 9-24
Lehninger Principles of Biochemistry, Sixth Edition
© 2013 W. H. Freeman and Company

Tools for genome-wide analysis

4. Determine interacting partners

- Known protein interacts with unknown: same process?

- Purification of protein complexes
  Immunoprecipitation/tandem affinity purification of tagged proteins

-**Yeast two-hybrid analysis**
  express fusion proteins using 2 domains of Gal4 activator
  "bait" fused with DNA binding domain (BD)
  "prey" fused with activation domain (AD)
  If they interact: Gal4 comes together and activates a transcriptional response that causes a phenotype

Gal4 transcription factor is modular and separable

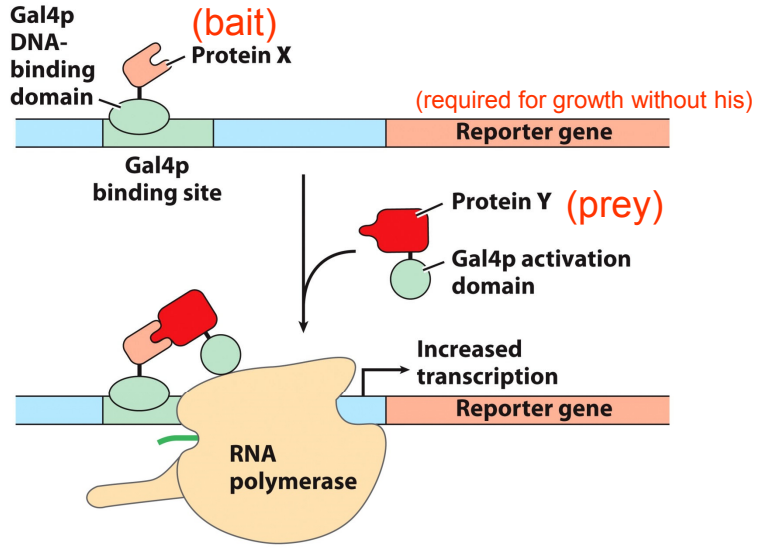


Figure 9-21a
Lehninger Principles of Biochemistry, Sixth Edition
© 2013 W. H. Freeman and Company

Gal4 transcription factor is modular and separable

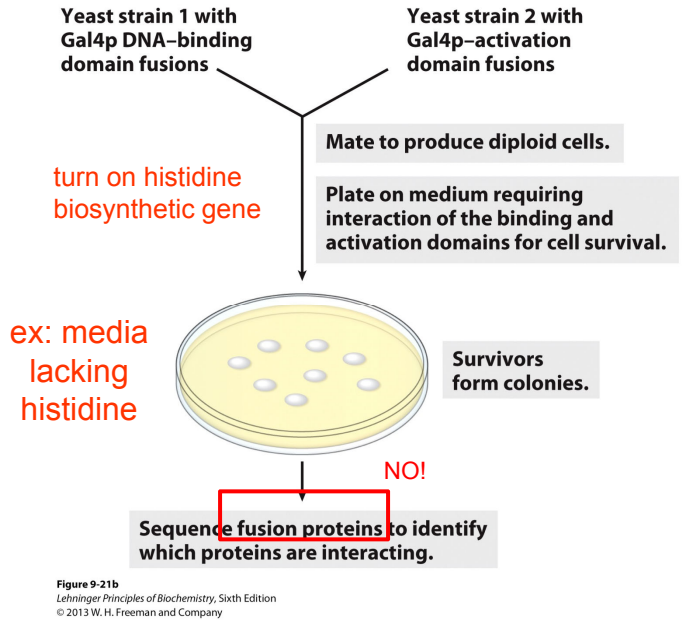


Figure 9-21b
Lehninger Principles of Biochemistry, Sixth Edition

Tools for genome-wide analysis

1. Comparative genomics
- identify similarities to known proteins

2. Genetic analysis
- deletion ("knockout") collections

3. Cellular expression patterns
- where and when is it expressed?

4. Determine interacting partners
- who does it interact with?

Summary

Genes and Genomes

- Genes are defined by a DNA sequence that direct production of a macromolecular product
- Genomes are very different, not predictable
- The human genome: many insights, still many to come

Biochemical approaches to understanding genomes

- Understand the function of all gene products encoded by an organism in a given environment
- Advances/new techniques: high-throughput

Nature:
01 April 2010