Chapter 14:
Functional Genomics

Learning objectives

Upon reading this chapter, you should be able to:

- define functional genomics;
- describe the key features of eight model organisms;
- explain techniques of forward and reverse genetics;
- discuss the relation between the central dogma and functional genomics; and
- describe proteomics-based approaches to functional genomics.

Outline : Functional genomics

Introduction

Relation between genotype and phenotype

Eight model organisms

E. coli; yeast; *Arabidopsis*; *C. elegans*; *Drosophila*; zebrafish; mouse; human

Functional genomics using reverse and forward genetics

Reverse genetics: mouse knockouts; yeast; gene trapping; insertional mutatgenesis; gene silencing

Forward genetics: chemical mutagenesis

Functional genomics and the central dogma

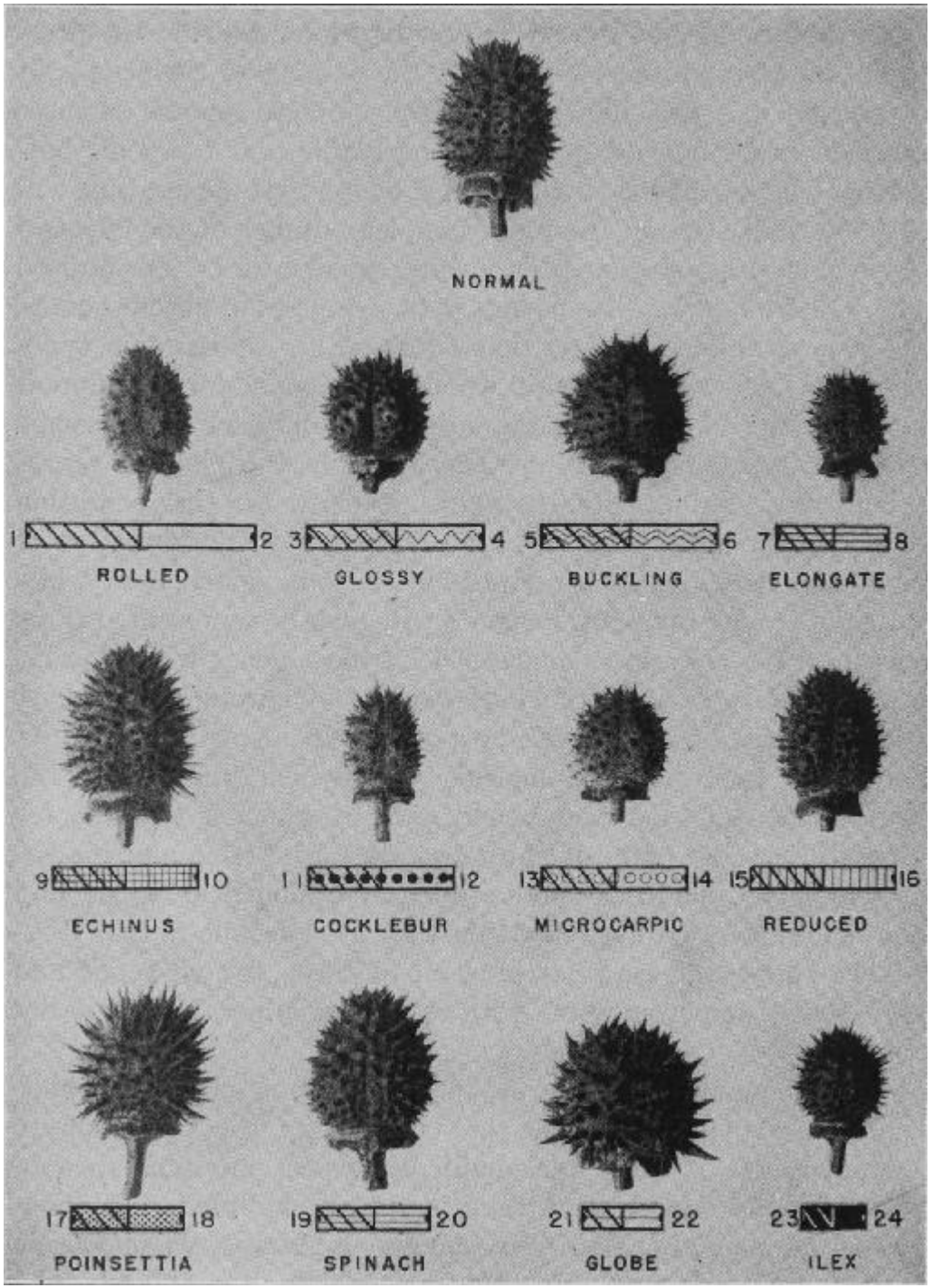
Approaches to function; Functional genomics and DNA; …and RNA; …and protein

Proteomic approaches to functional genomics

CASP; protein-protein interactions; protein networks

Perspective

Albert Blakeslee (1874–1954) studied the effect of altered chromosome numbers on the phenotype of the jimson-weed *Datura stramonium*, a flowering plant.



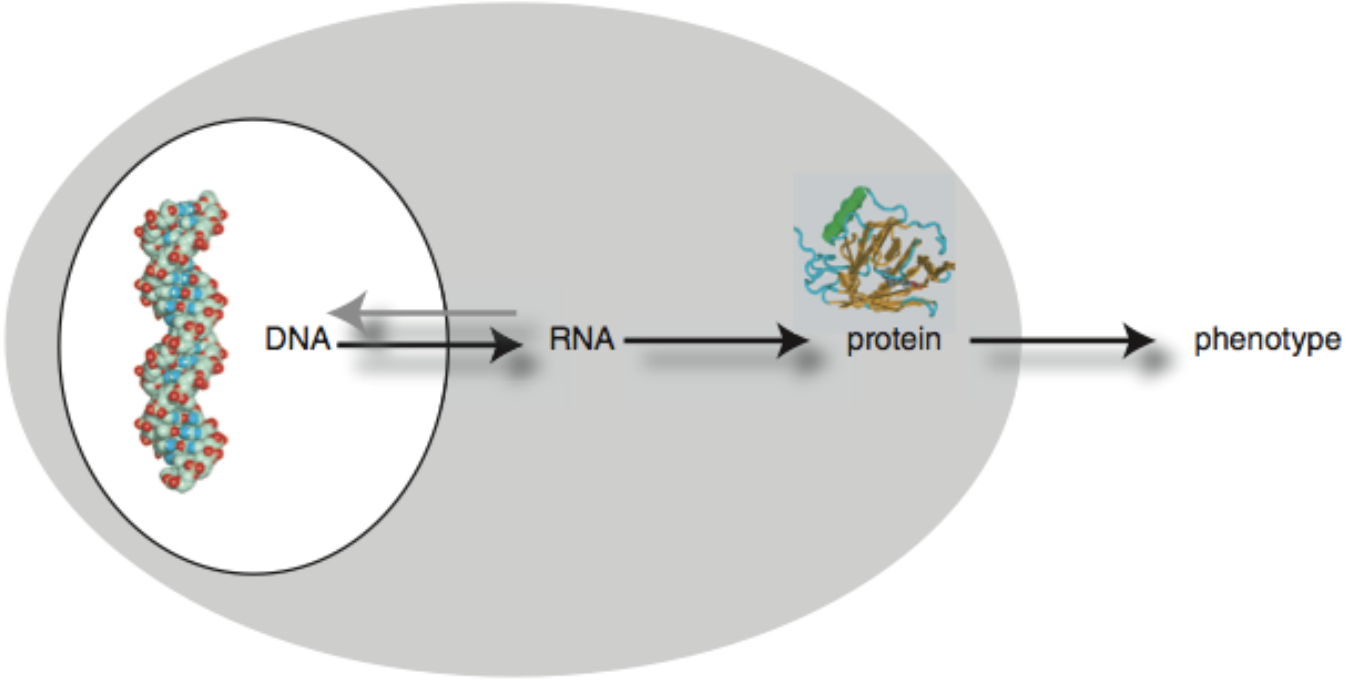
Introduction: Functional genomics

Functional genomics is the genome-wide study of the function of DNA (including both genes and non-genic regions), as well as RNA and proteins encoded by DNA.

The term "functional genomics" may apply to

- the genome, transcriptome, or proteome
- the use of high-throughput screens
- the perturbation of gene function
- the complex relationship of genotype and phenotype

Functional genomics approaches to high throughput analyses



	DNA	RNA	protein
Natural variation --across development --across body regions --across species, strains	SNPs; epigenomics	transcriptome profiling (RNA-seq)	protein localization; protein-protein interactions; pathways
Functional disruptions			
--experimental	knockout collections transgenic animals	RNAi; siRNA	chemical modification
--in nature	Williams syndrome Down syndrome cancer chromosomal changes	nonsense-mediated RNA decay	myasthenia gravis

Relationship between genotype and phenotype

The genotype of an individual consists of the DNA that comprises the organism. The phenotype is the outward manifestation in terms of properties such as size, shape, movement, and physiology. We can consider the phenotype of a cell (e.g., a precursor cell may develop into a brain cell or liver cell) or the phenotype of an organism (e.g., a person may have a disease phenotype such as sickle-cell anemia).

A great challenge of biology is to understand the relationship between genotype and phenotype. We can gather information about either one alone, but how they are connected very often remains obscure.

Outline : Functional genomics

Introduction

Relation between genotype and phenotype

Eight model organisms

E. coli; yeast; *Arabidopsis*; *C. elegans*; *Drosophila*; zebrafish; mouse; human

Functional genomics using reverse and forward genetics

Reverse genetics: mouse knockouts; yeast; gene trapping; insertional mutatgenesis; gene silencing

Forward genetics: chemical mutagenesis

Functional genomics and the central dogma

Approaches to function; Functional genomics and DNA; …and RNA; …and protein

Proteomic approaches to functional genomics

CASP; protein-protein interactions; protein networks

Perspective

Functional genomics: 8 model organisms

We introduce 8 model organisms that have particularly important roles in functional genomics. The list is not comprehensive, but highlights important principles as well as advantages (and disadvantages) of studying various model systems.

Eight model organisms for functional genomics

Bacterium *Escherichia coli*

Yeast *Saccharomyces cerevisiae*

Plant *Arabidopsis thaliana*

Nematode *Caenorhabditis elegans*

Fruitfly *Drosophila melanogaster*

Zebrafish *Danio rerio*

Mouse *Mus musculus*

Homo sapiens: variation in humans

8 model organisms: (1) Bacterium *Escherichia coli*

The bacterium *Escherichia coli* serves as the best-characterized bacterial organism, if not the best-characterized living organism. For decades it served as a leading model organism for bacterial genetics and molecular biology studies.

- 4.6 megabase (million base pairs) genome was sequenced by Blattner *et al.* (1997)Principal website is **EcoCyc**, the Encyclopedia of *Escherichia coli* K-12 Genes and Metabolism
- EcoCyc assigns a function to >75% of the 4501 annotated genes

https://ecocyc.org/

8 model organisms: (2) Yeast *S. cerevisiae*

- The budding yeast *S. cerevisiae* is the best-characterized organism among the eukaroytes
- Single-celled fungus. First eukaryote to have its genome sequenced
- 13 megabase genome encodes 6000 proteins
- *Saccharomyces* Genome Database (SGD) is principal database and community resource
- ~6600 annotated open reading frames (ORFs, corresponding to genes), including ~5000 that are verified, 750 that are uncharacterized
- ~4200 gene products have been annotated to the root gene ontology terms (molecular function, biological process, cellular component;

SGD (*Saccharomyces* Genome Database) entry for *SEC1* (www.yeastgenome.org)

SEC1/YDR164C Summary

Help

Summary | Locus History | Literature | Gene Ontology | Phenotype | Interactions | Expression | Protein

Alternative single page format

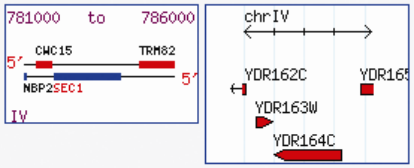
SEC1 BASIC INFORMATION

Standard Name	*SEC1* [1]
Systematic Name	YDR164C
Feature Type	ORF, Verified
Description	Sm-like protein involved in docking and fusion of exocytic vesicles through binding to assembled SNAI complexes at the membrane; localization to sites of secretion (bud neck and bud tip) is dependent on SNARE function (2)
Name Description	SECretory [3]
GO Annotations	*All **SEC1** GO evidence and references* *View Computational GO annotations for **SEC1***
Molecular Function Manually curated	• SNARE binding (IDA)
Biological Process Manually curated	• exocytosis (IMP, IPI) • vesicle docking during exocytosis (IPI) • vesicle fusion (IDA, IMP)
Cellular Component Manually curated	• cellular bud neck (IDA) • cellular bud tip (IDA) • plasma membrane (IDA)
Mutant Phenotype Systematic deletion Free text	***SEC1** Phenotype details and references* Order mutant strains used in systematic deletion project • inviable • accumulates secretory vesicles

SEC1 RESOURCES

Click on map for expanded view

SGD ORF map | GBrowse



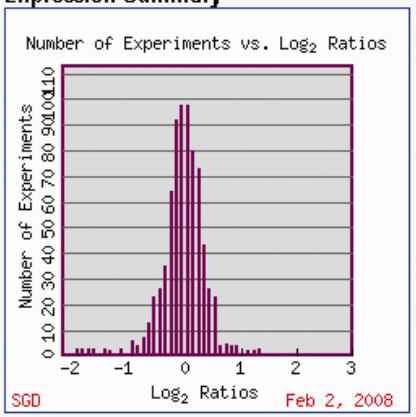
- **Literature** — Literature Guide — View
- **Retrieve Sequences** — Genomic DNA — View
- **Sequence Analysis Tools** — BLASTP — View
- **Protein Info & Structure** — Protein Info — View
- **Localization Resources** — GFP DB at UCSF — View
- **Interactions** — BioGRID (Toronto) — View
- **Phenotype Resources** — PROPHECY — View

SGD (*Saccharomyces* Genome Database) entry for *SEC1*

Interactions	***SEC1*** *All interactions details and references*
Physical Interactions	***SEC1*** *Physical Interactions details and references*
Affinity Capture-MS	There are **7 total** Affinity Capture-MS interactions
Affinity Capture-RNA	There is **1 total** Affinity Capture-RNA interactions
Affinity Capture-Western	There are **13 total** Affinity Capture-Western interactions
Reconstituted Complex	There are **7 total** Reconstituted Complex interaction
Two-hybrid	There are **3 total** Two-hybrid interactions
Genetic Interactions	***SEC1*** *Genetic Interactions details and references*
Dosage Lethality	There are **5 total** Dosage Lethality interactions resulting in the following phenotype: **inviable**
Dosage Rescue	There are **13 total** Dosage Rescue interactions resulting in the following phenotype: **wildtype**
Phenotypic Suppression	There are **5 total** Phenotypic Suppression interactio resulting in the following phenotype: **Not available**
Synthetic Growth Defect	There is **1 total** Synthetic Growth Defect interactions resulting in the following phenotype: **slow growth**
Synthetic Lethality	There are **38 total** Synthetic Lethality interactions resulting in the following phenotype: **inviable**

- **Maps & Displays**
  Chromosomal Features Map [View]
- **Comparison Resources**
  PSI-BLAST Results [View]
- **Functional Analysis**
  Expression Connection Summary [View]

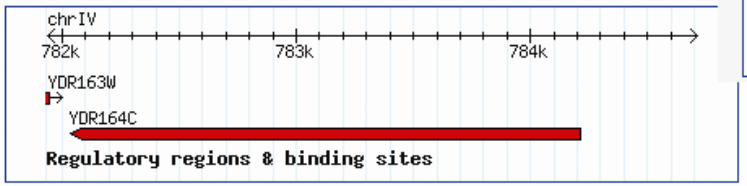
Click on histogram for expression summary
Expression Summary



Sequence Information

ChrIV:784212 to 782038 | ORF Map | GBrowse

Note: *this feature is encoded on the Crick strand.*



Genetic position: 94.77 cM

Last Update: Coordinates: 2006-04-13 | Sequence: 1996-07-31

Subfeature details

	Relative Coordinates	Chromosomal Coordinates	Most Recent Updates Coordinates	Most Recent Updates Sequence
CDS	1..2175	784212..782038	*2006-04-13*	*1996-07-31*

ORF Genomic DNA [Get Sequence]

External Links	All Associated Seq | Entrez Gene | Entrez RefSeq Protein | MIPS | UniProt/Swiss-Prot
Primary SGDID	S000002571

SSO1/YPL232W Summary

Help

Summary | Locus History | Literature | Gene Ontology | Phenotype | Interactions | Expression | Protein

Alternative single page format

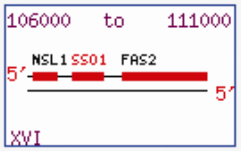
SSO1 BASIC INFORMATION

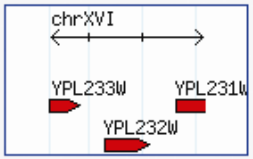
Standard Name	*SSO1*
Systematic Name	YPL232W
Feature Type	ORF, Verified
Description	Plasma membrane t-SNARE involved in fusion of secretory vesicles at the plasma membrane and in vesicle fusion during sporulation; forms a complex with Sec9p that binds v-SNARE Snc2p; syntaxin homolog; functionally redundant with Sso2p (1, 2, 3, 4)
GO Annotations	*All **SSO1** GO evidence and references* *View Computational GO annotations for **SSO1***
Molecular Function Manually curated	• SNAP receptor activity (IDA, IPI)
Biological Process Manually curated	• Golgi to plasma membrane transport (TAS) • membrane fusion (IDA, IMP) • prospore formation (IMP) • sporulation (sensu Fungi) (IMP)
Cellular Component Manually curated	• plasma membrane (IDA) • prospore membrane (IDA) • SNARE complex (IDA)
Mutant Phenotype Systematic deletion Free text	***SSO1** Phenotype details and references* Order mutant strains used in systematic deletion project • viable • SSO1, SSO2 double null mutant is inviable; high copy number of either SSO1 or SSO2 suppresses mutations in late-acting sec genes (sec1,3,5,9,15)

SSO1 RESOURCES

Click on map for expanded view

SGD ORF map | GBrowse





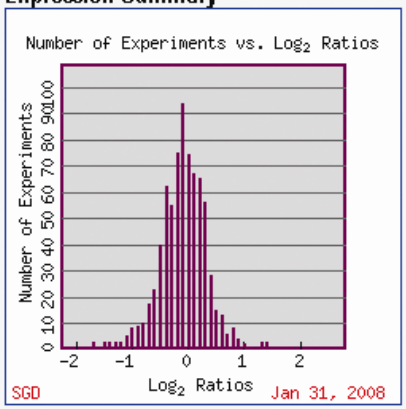
- **Literature**
  Literature Guide [View]
- **Retrieve Sequences**
  Genomic DNA [View]
- **Sequence Analysis Tools**
  BLASTP [View]
- **Protein Info & Structure**
  Protein Info [View]
- **Localization Resources**
  GFP DB at UCSF [View]
- **Interactions**
  BioGRID (Toronto) [View]
- **Phenotype Resources**
  PROPHECY [View]

Interactions	***SSO1*** *All interactions details and references*
Physical Interactions	***SSO1*** *Physical Interactions details and references*
Affinity Capture-Western	There are **12 total** Affinity Capture-Western interactions
Biochemical Activity	There are **2 total** Biochemical Activity interactions
Co-crystal Structure	There are **2 total** Co-crystal Structure interactions
Reconstituted Complex	There are **7 total** Reconstituted Complex interactions
Two-hybrid	There is **1 total** Two-hybrid interactions
Genetic Interactions	***SSO1*** *Genetic Interactions details and references*
Dosage Lethality	There are **3 total** Dosage Lethality interactions resulting in the following phenotype: **inviable**
Dosage Rescue	There are **6 total** Dosage Rescue interactions resulting in the following phenotype: **wildtype**
Phenotypic Enhancement	There is **1 total** Phenotypic Enhancement interactions resulting in the following phenotype: **Not available**
Phenotypic Suppression	There are **2 total** Phenotypic Suppression interactions resulting in the following phenotype: **Not available**
Synthetic Lethality	There is **1 total** Synthetic Lethality interactions resulting in the following phenotype: **inviable**
Synthetic Rescue	There is **1 total** Synthetic Rescue interactions resulting in the following phenotype: **wildtype**

- **Maps & Displays**
  Chromosomal Features Map [View]
- **Comparison Resources**
  PSI-BLAST Results [View]
- **Functional Analysis**
  Expression Connection Summary [View]

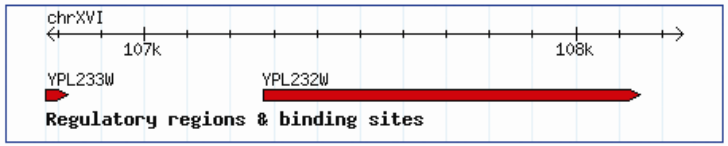
Click on histogram for expression summary

Expression Summary



Sequence Information

ChrXVI:107275 to 108147 | ORF Map | GBrowse



Last Update: Coordinates: 1996-07-31 | Sequence: 1996-07-31

Subfeature details

	Relative Coordinates	Chromosomal Coordinates	Most Recent Updates Coordinates	Most Recent Updates Sequence
CDS	1..873	107275..108147	*1996-07-31*	*1996-07-31*

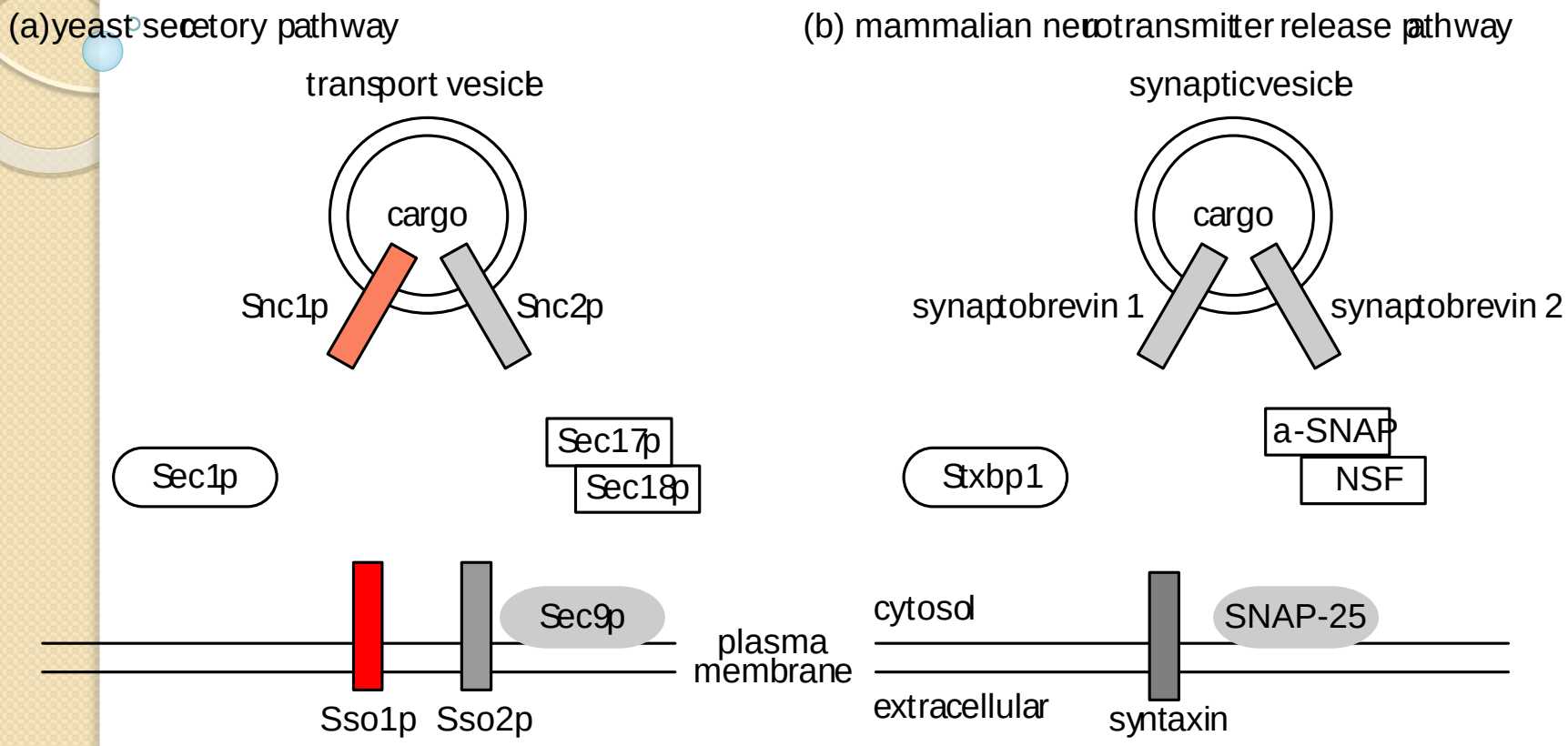
ORF Genomic DNA [Get Sequence]

External Links	All Associated Seq | Entrez Gene | Entrez RefSeq Protein | MIPS | UniProt/Swiss-Prot
Primary SGDID	S000006153

ADDITIONAL INFORMATION for *SSO1*

Community wiki	Domains/Motifs	Expression Connection	Function Junction
Gene/Sequence Resources	Global Gene Hunter	Locus History	PDB Homologs
Protein Info	Researchers		

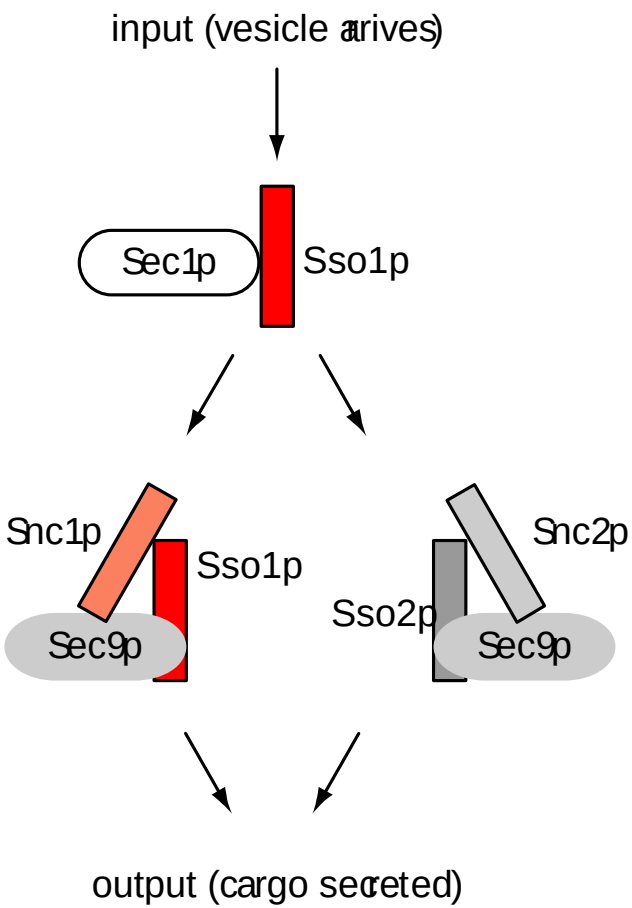
Diagram of *S. cerevisiae* and mammalian proteins involved in secretion to illustrate functional genomics principles and approaches



A constitutive trafficking pathway exists in yeast (left) with a set of proteins having orthologs in a regulated trafficking pathway in mammals (right).

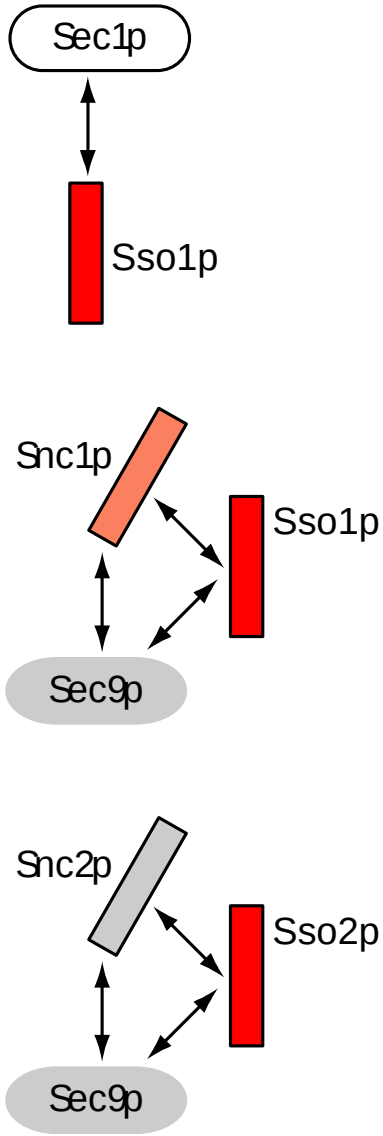
Diagram of *S. cerevisiae* and mammalian proteins involved in secretion to illustrate functional genomics principles and approaches

(c) pathway

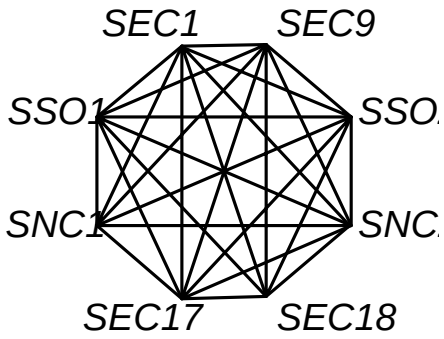


Pathway diagram: parallel pathways

(d) protein interctions



(e) genetic interactions



Biochemical protein interactions (left) and genetic interactions (above) are complementary methods to elucidate pathways

8 model organisms: (3) Plant *Arabidopsis thaliana*

- The thale cress *Arabidopsis thaliana* was the first plant to have its genome sequenced (and the third finished eukaryotic genome sequence).
- Model for eukaryotic functional genomics projects
- Principal web site is The *Arabidopsis* Information Resource (TAIR)
- Appealing features as a model plant: short generation time, prolific seed production, compact genome size, and opportunities for genetic manipulation.

The *Arabidopsis* Information Resource (TAIR): principal genome database for *Arabidopsis*

tair

Home Help Contact About Us Login/Register

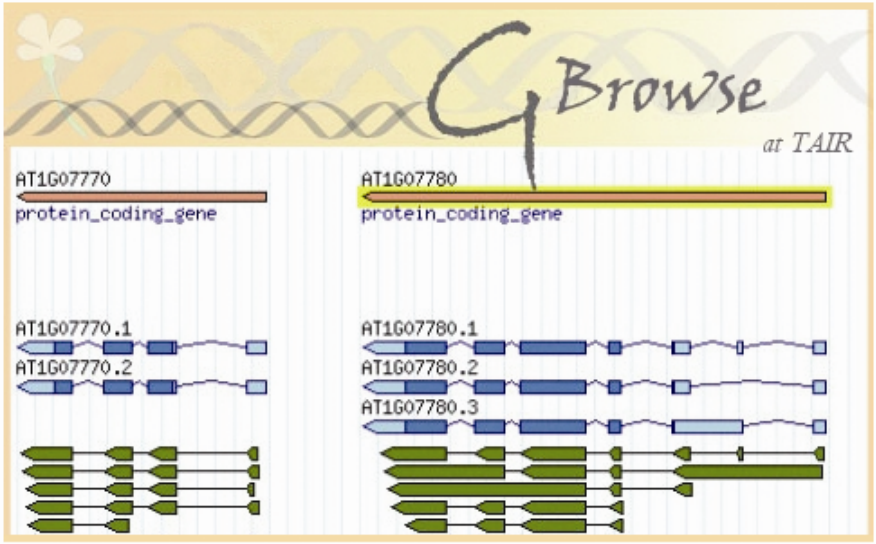
Gene Search

Search | Browse | Tools | Stocks | Portals | Download | Submit | News

The Arabidopsis Information Resource

The Arabidopsis Information Resource (TAIR) maintains a database of genetic and molecular biology data for the model higher plant *Arabidopsis thaliana*. Data available from TAIR includes the complete genome sequence along with gene structure, gene product information, metabolism, gene expression, DNA and seed stocks, genome maps, genetic and physical markers, publications, and information about the Arabidopsis research community. Gene product function data is updated every two weeks from the latest published research literature and community data submissions. Gene structures are updated 1-2 times per year using computational and manual methods as well as community submissions of new and updated genes. TAIR also provides extensive linkouts from our data pages to other Arabidopsis resources.

The Arabidopsis Biological Resource Center at The Ohio State University collects, reproduces, preserves and distributes seed and DNA resources of *Arabidopsis thaliana* and related species. Stock information and ordering for the ABRC are fully integrated into TAIR.



Breaking News

Change to seed ordering process for European users.
NASC can no longer accept orders placed through TAIR, because of changes to their database and ordering system.

AraCyc 4.1 release
23 pathways were significantly updated in the last release in October. More **details.**

New GO bar charts
Try our new bar charts to visualize GO annotation categories for your gene set or the whole genome. (**see details**)

GBrowse now at TAIR
View TAIR genome map data using the GMOD generic genome browser, or upload your own genome data track (**see details**)

Perlegen SNPs now available
249,052 high-quality SNPs from Perlegen resequencing arrays now available from **TAIR polymorphism search** and **SeqViewer**. Over 1 million

Search

- Search Overview
- DNA/Clones
- Ecotypes
- Genes
- GO Annotations
- Keywords
- Locus History
- Markers
- Microarray Element
- Microarray Experiment
- Microarray Expression
- People/Labs
- Polymorphisms/Alleles
- Proteins
- Protocols
- Publication
- Seed/Germplasm
- Sequences

Tools

- Tools Overview
- GBrowse
- Seqviewer
- Mapviewer
- AraCyc Metabolic Pathways
- BLAST
- WU-BLAST
- FASTA
- Patmatch
- Motif Analysis
- VxInsight
- Java Tree View
- Bulk Data Retrieval
- Chromosome Map Tool
- Gene Hunter
- Restriction Analysis
- Gene Symbol Registry

Stocks

- Stocks Overview
- ABRC Home
- Browse ABRC Catalog
- Supplement to ABRC Catalog
- Search ABRC DNA/Clone Stocks
- Search ABRC Seed/Germplasm Stocks
- ABRC Stock Order History
- ABRC Fee Structure
- Place ABRC Order
- Search My ABRC Orders
- Search ABRC Invoices
- How to Make Payments to ABRC
- ABRC Stock Donation

Download

- Download Overview
- Genes
- GO and PO Annotations
- Maps
- Pathways
- Proteins
- Protocols
- Microarray Data
- Sequences
- Software
- User Requests

Browse

- Browse Overview
- ABRC Catalog
- 2010 Projects
- Monsanto SNP and Ler Collections
- Gene Families
- Gene Class Symbols
- Ontologies/Keywords
- Archived e-Journals

Portals

- Portals Overview
- Clones/DNA Resources
- Education and Outreach
- Gene Expression Resources
- Genome Annotation
- MASC/Functional Genomics
- Mutant and Mapping Resources
- Nomenclature
- Proteome Resources

The *Arabidopsis* Information Resource (TAIR)

8 model organisms: (4) Nematode *C. elegans*

- First multicellular animal to have its genome sequenced
- Capable of complex behaviors
- Body is simple and all the 959 somatic cells in its body have been mapped including their lineages throughout development

- **Wormbase** is the main database/resource
- Genome encodes ~20,400 protein-coding genes (same number as in humans).

8 model organisms: (5) Fruitfly *Drosophila*

- Metazoan (animal) invertebrate
- Early studies of *Drosophila* resulted in the descriptions of the nature of the gene as well as linkage and recombination, producing gene maps a century ago
- Sequencing of many *Drosophila* genomes (and inbred lines) providing unprecedented insight into mechanisms of genome evolution
- Genomic changes can be induced with extreme precision, from single-nucleotide changes to introducing large-scale chromosomal deletions, duplications, inversions, or other modifications

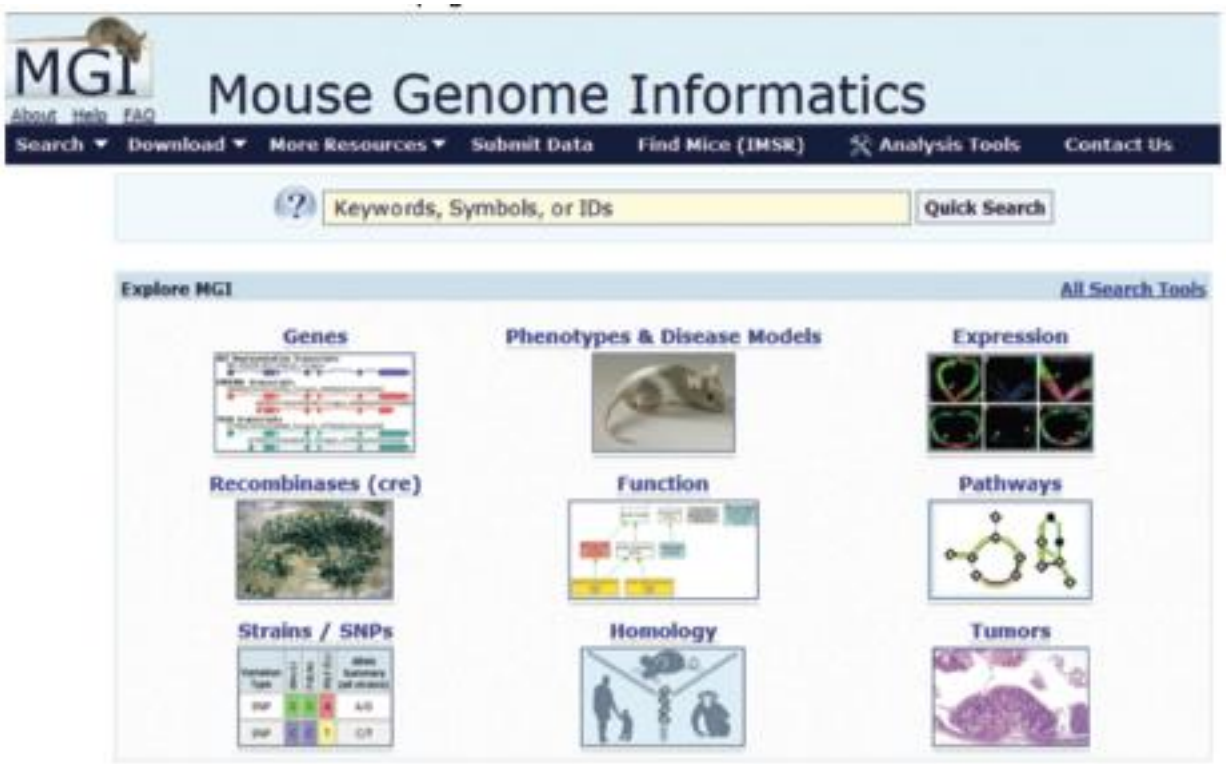
8 model organisms: (6) zebrafish *Danio rerio*

- Lineages leading to modern fish and humans diverged approximately 450 million years ago
- Freshwater fish having a genome size of 1.8 billion base pairs (Gb) organized into 25 chromosomes
- >26,000 protein-coding genes
- Mutations in large numbers of human disease gene orthologs have been generated and characterized, using both forward and reverse genetic screens
- Short generation time
- Large numbers of progeny
- Developing embryo is transparent (transgenes can be visualized)

8 model organisms: (7) Mouse *Mus musculus*

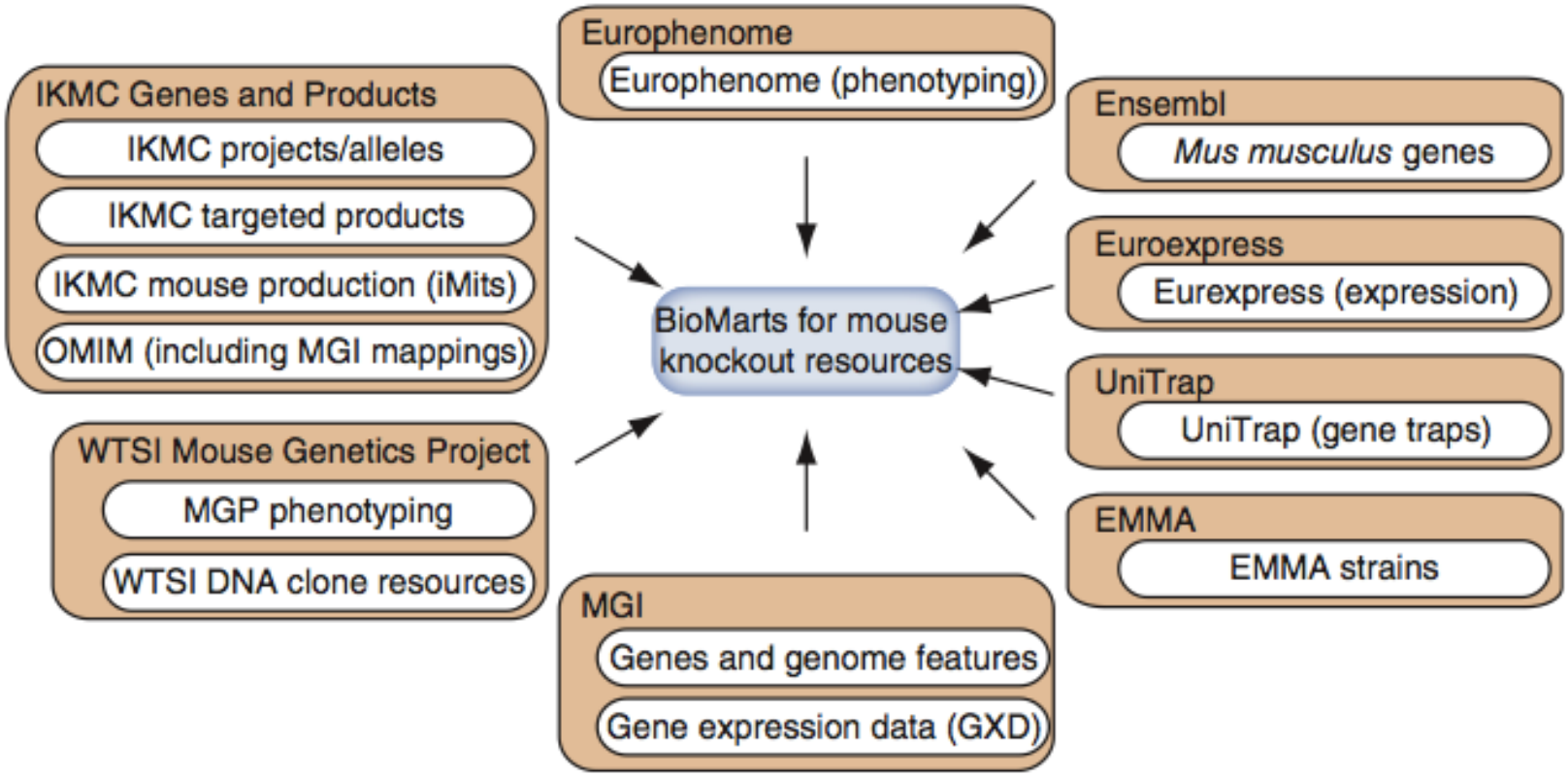
- Shared common ancestor with humans ~90 million years ago
- Close structural and functional relationship between mouse and human genomes
- Relatively short generational span
- Powerful tools developed to manipulate its genome
- Main mouse genome website is the **Mouse Genome Informatics (MGI)**
- ~10,000 genes **knocked out**
- Collaborative Cross: 1000 recombinant inbred strains of mouse are being bred, producing large numbers of genetically related mice that have nonlethal phenotypic diversity

Mouse genome informatics (MGI) database



MGI database is the principal website for mouse genomics information. The home page provides a portal to a vast number of resources.

Mouse genome informatics (MGI) database



MGI offers customized Biomarts for mouse functional genomics projects.

International Knockout Mouse Consortium (IKMC)

8 model organisms: (8) humans

We do not think of humans as model organisms per se. But nature performs functional genomics experiments on us constantly.

Motivation for studying humans: to understand the causes of disease in order to search for more effective diagnoses, preventions, treatments, and ultimately cures.

Outline : Functional genomics

Introduction
- Relation between genotype and phenotype

Eight model organisms
- *E. coli*; yeast; *Arabidopsis*; *C. elegans*; *Drosophila*; zebrafish; mouse; human

Functional genomics using reverse and forward genetics
- Reverse genetics: mouse knockouts; yeast; gene trapping; insertional mutatgenesis; gene silencing
- Forward genetics: chemical mutagenesis

Functional genomics and the central dogma
- Approaches to function; Functional genomics and DNA; …and RNA; …and protein

Proteomic approaches to functional genomics
- CASP; protein-protein interactions; protein networks

Perspective

Functional genomics using reverse and forward genetics

- Reverse genetic screens: a large number of genes (or gene products) is systematically inhibited one by one. This can be accomplished in many ways, for example by deleting genes using homologous recombination, gene trapping, or by selectively reducing messenger RNA abundance. One or more phenotypes of interest are then measured.

  Main challenges of this approach:
  - for some organisms it difficult to disrupt large numbers of genes (such as tens of thousands) in a systematic fashion.
  - It can also be challenging to discern the phenotypic consequences for a gene that is disrupted.

Functional genomics using reverse and forward genetics

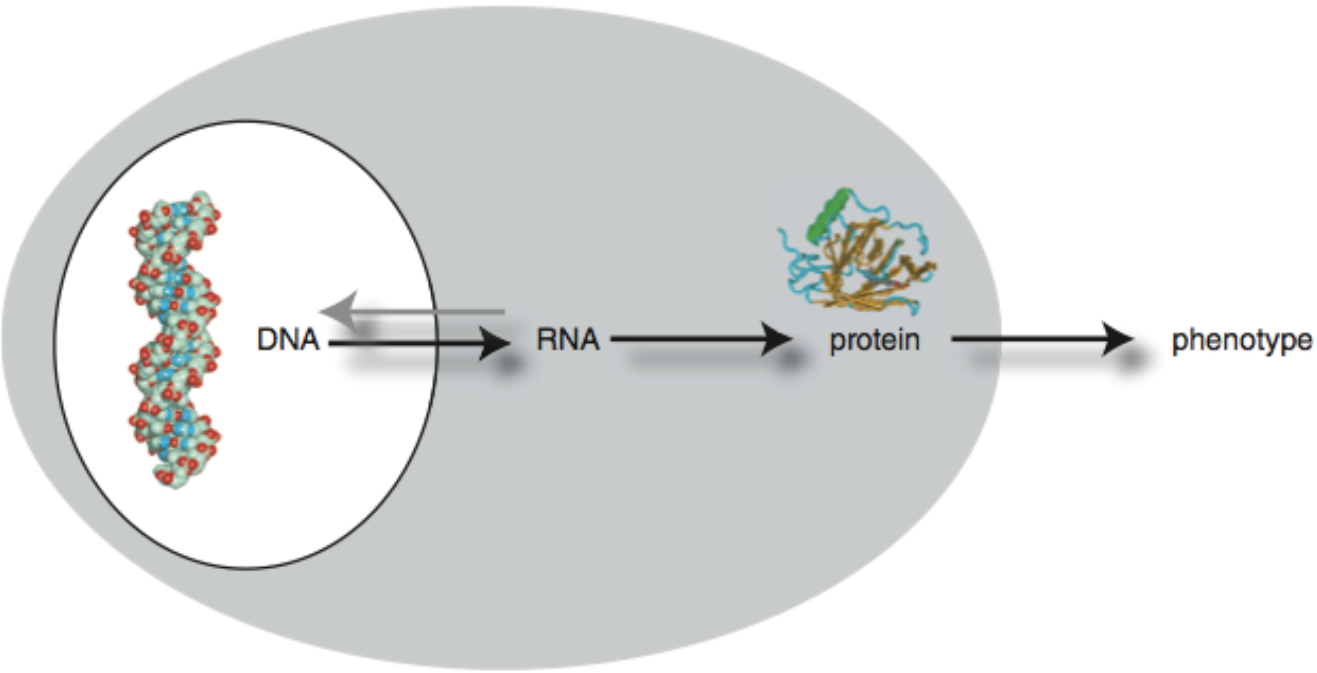
- Forward genetic screens:
  - the starting point is a defined phenotype of interest, such as the ability of plants to grow in the presence of a drug, neurons to extend axons to appropriate targets in the mammalian nervous system, or an eukaryotic cell to transport cargo
  - An experimental intervention is made, such as administering a chemical mutagen or radiation to cells (or to an organism). This results in the creation of mutants.
  - The phenotype of interest is observed in rare representatives among a large collection of mutants.

Reverse genetics (mutate genes then examine phenotypes)

Strategy:

Systematically inhibit the function of every gene in a genome
Approach 1: gene targeting by homologous recombination
Approach 2: gene trap mutagenesis
Approach 3: inhibit gene expression using RNA interference
Measure the effect of gene disruption on a phenotype

Reverse genetics



Strategy:

Identify a phenotype (e.g. growth in the presence of a drug)
Mutate genomic DNA (e.g. by chemical mutagenesis)
Identify individuals having an altered phenotype
Identify the gene(s) that were mutated
Confirm those genes have causal roles in influencing the genotype

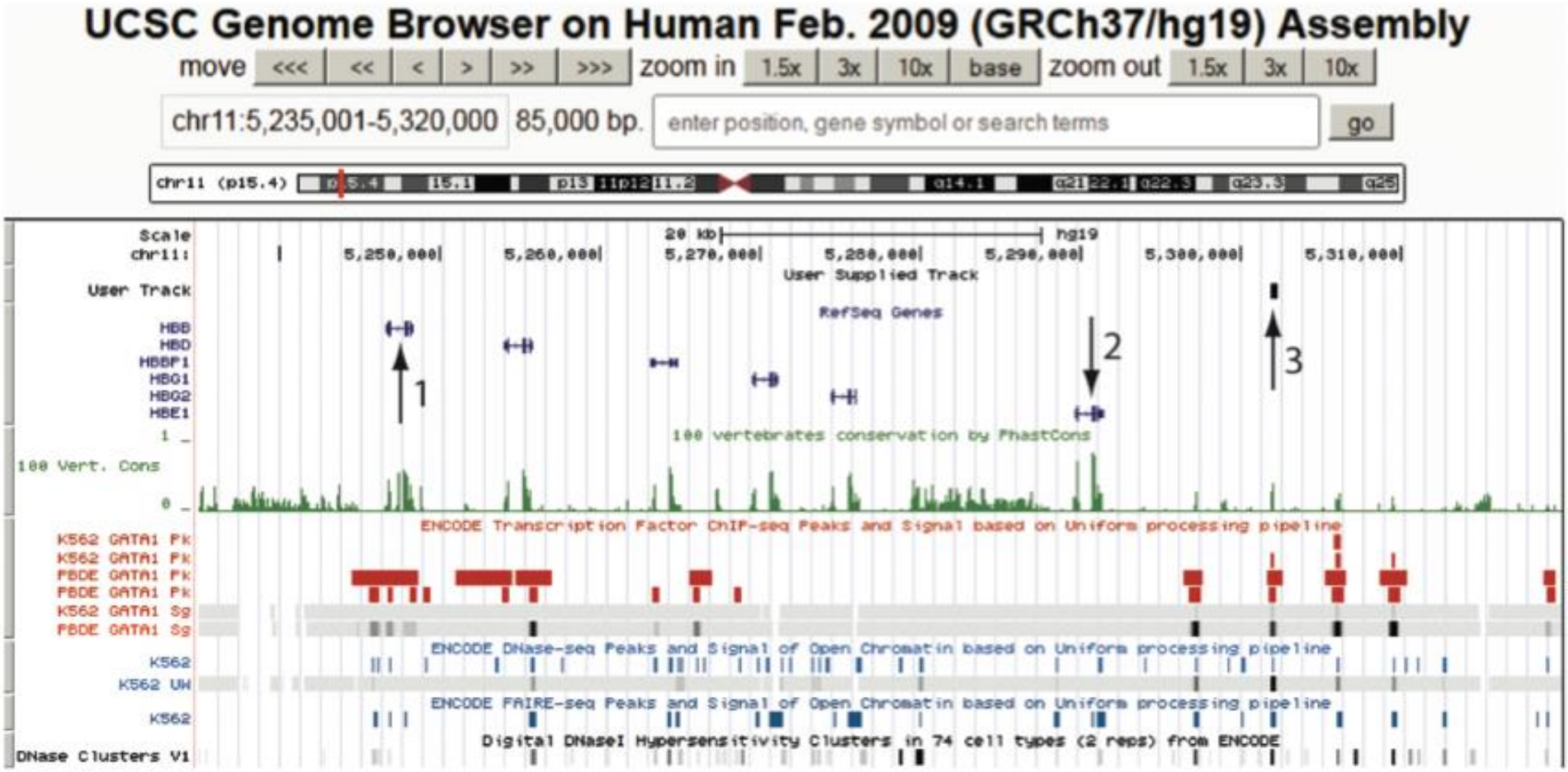
Forward genetics

Forward genetics ("phenotype-driven" screen)

Reverse genetics: mouse knockouts and the β-globin gene

- Knocking out a gene: create an animal model in which a homozygous deletion is created, that is, there are zero copies (denoted (−/−) and referred to as a null allele) instead of the wildtype situation of two copies in a diploid organism (+/+).
- In a hemizygous deletion, one copy is deleted and one copy remains (+/−).
- Use a targeting vector that includes the β-globin gene having a portion modified by insertion of the *neo* gene into exon 2.
- This targeting vector is introduced into embryonic stem cells by **electroporation**. When the cells are cultured in the presence of the drug G418, wildtype cells die whereas cells having the *neo* cassette (gene casette)survive. Confirm by PCR.

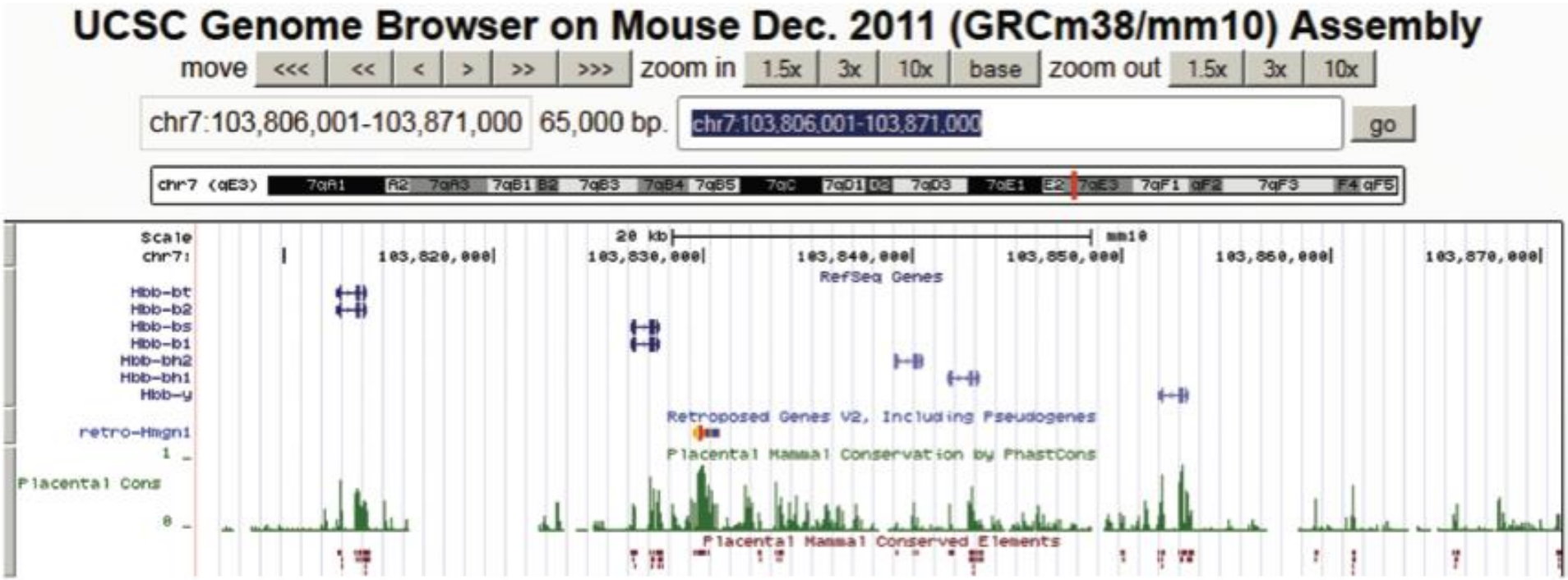
The β globin locus



B&FG 3e
Fig. 14.8
Page 651

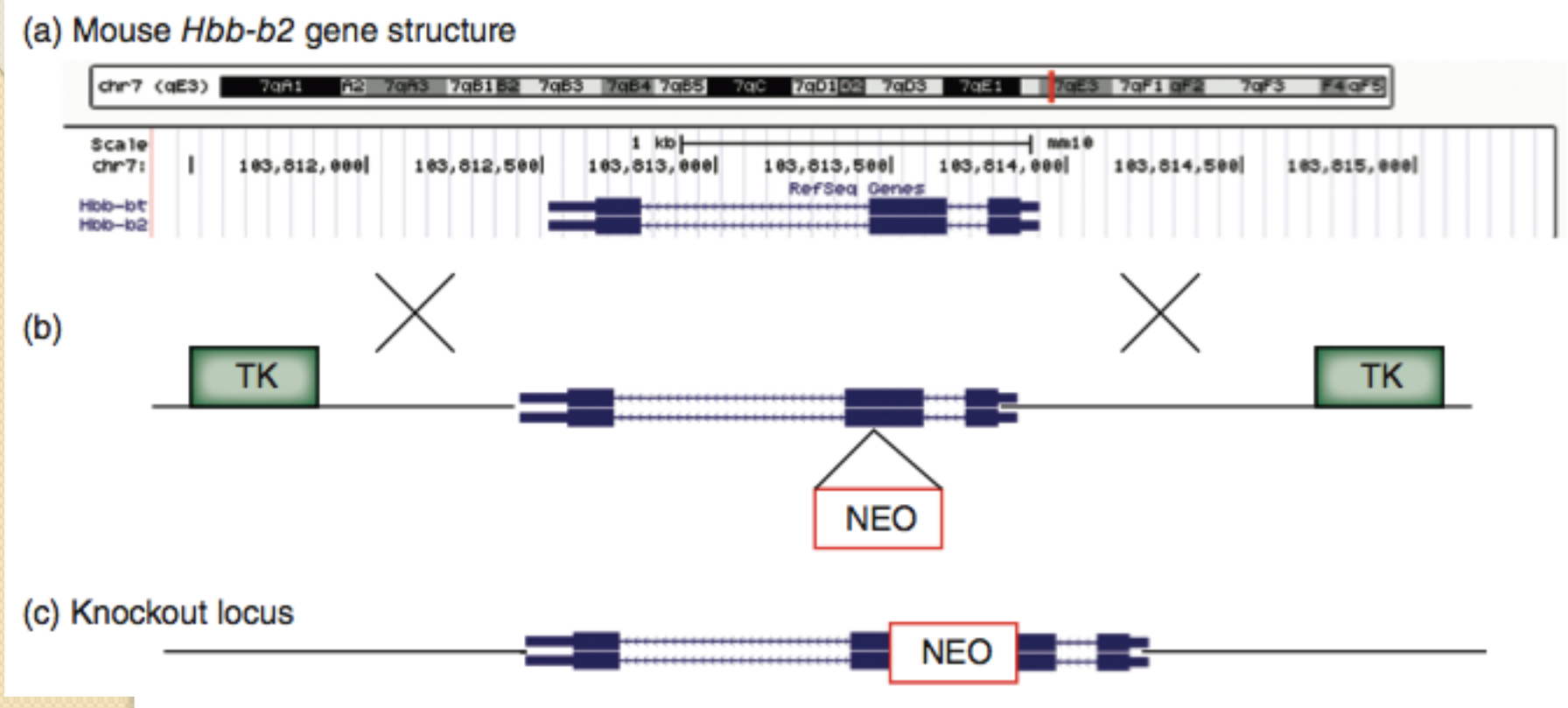
Human (85 kilobases on chr11:5,235,001–5,320,000, GRCh37/hg19 assembly)

The β globin locus



Mouse (65 kilobases on chr7:103,806,001–103,871,000, GRCm38/mm10 assembly)

Method of gene knockout by homologous recombination



The successfully targeted locus includes a β globin gene that is interrupted by the *neo* gene.

Mouse Genome Informatics (MGI) website entry for the major beta globin gene (*Hbb-b1*)

MGI

About Help FAQ

Keywords, Symbols, or IDs | Quick Search

Home | Genes | Phenotypes | Expression | Recombinases | Function | Pathways | Strains / SNPs | Homology | Tumors

Search | Download | More Resources | Submit Data | Find Mice (IMSR) | Analysis Tools | Contact Us

Hbb-b1
Gene Detail

Your Input Welcome

Symbol	**Hbb-b1**
Name	**hemoglobin, beta adult major chain**
ID	MGI:96021
Synonyms	beta1, beta maj, beta major globin, MommeD7
Feature Type	protein coding gene
Genetic Map	Chromosome 7 54.85 cM Detailed Genetic Map ± 1 cM Mapping data(6)
Vertebrate homology	HomoloGene:68066 Vertebrate Homology Class 1 human; 2 mouse; 4 rat; 1 chimpanzee; 1 rhesus macaque; 2 cattle; 3 dog Protein SuperFamily: globin
Human homologs	Human Homolog HBB, hemoglobin, beta NCBI Gene ID 3043 neXtProt AC NX_P68871 Human Synonyms beta-globin, CD113t-C Human Chr (Location) 11p15.5; chr11:5246696-5248301 (-) *GRCh37.p10* Disease Associations (6) Diseases Associated with Human HBB
Alleles and phenotypes	All alleles(7) : Targeted(5) Chemically induced(2) Homozygotes for a deletion of both adult hemoglobin-beta genes resemble human Cooley anemia and die perinatally. Heterozygotes are anemic with hematological abnormalities and iron accumulation. Homozygotes for a missense mutation have hemolytic anemia. Human Diseases Modeled Using Mouse Hbb-b1 (2) Alleles Annotated to Human Diseases(4)
Gene Ontology (GO) classifications	All GO classifications: (17 annotations) Process erythrocyte development, hemopoiesis, ... Component haptoglobin-hemoglobin complex, hemoglobin complex Function haptoglobin binding, hemoglobin alpha binding, ... External Resources: FuncBase

Strain-Specific Marker

Prkcg | Gpi1 | Klk1b3 | Myod1 | Gca2 | Fes | Tyr | Hbb-b1 | Itga1 | Fgf3

0 10 20 30 40 50 60 70 80 90

The entry summarizes molecular data on that gene and includes a phenotype category, indicating that seven mutant alleles are indexed (five targeted and two chemically induced).

MGI description of beta globin mutants

MGI

About Help FAQ

Keywords, Symbols, or IDs | Quick Search

Home | Genes | Phenotypes | Expression | Recombinases | Function | Pathways | Strains / SNPs | Homology | Tumors

Search ▾ Download ▾ More Resources ▾ Submit Data Find Mice (IMSR) Analysis Tools Contact Us

Phenotypic Alleles

Query Results -- Summary

Symbol	**Hbb-b1**
Name	**hemoglobin, beta adult major chain**
ID	MGI:96021

7 matching Alleles (1 Gene/Marker represented)

Allele Symbol Gene; Allele Name	Chr	Synonyms	Category	Abnormal Phenotypes Reported in these Systems	Human Disease Models
Hbb-b1^{MommeD7} hemoglobin, beta adult major chain; modifier of murine metastable epialleles, D7	7	RBC14	Chemically induced (ENU)	hematopoietic, immune, integument, liver/biliary, mortality/aging	Beta-Thalassemia 613985
Hbb-b1^{Rbc13} hemoglobin, beta adult major chain; red blood cell mutant 13	7	RBC13	Chemically induced (ENU)	hematopoietic, immune, integument, liver/biliary, mortality/aging	Beta-Thalassemia 613985
Hbb-b1^{tm1Ley} hemoglobin, beta adult major chain; targeted mutation 1, Timothy J Ley	7		Targeted (knock-in)	no abnormal phenotype observed	
Hbb-b1^{tm1Shs} hemoglobin, beta adult major chain; targeted mutation 1, Takuji Shirasawa	7	Hbb^{Pres}	Targeted (knock-in)	cellular, hematopoietic, homeostasis, immune, muscle, respiratory	Hemoglobin--Beta Locus; HBB 141900
Hbb-b1^{tm1Unc} hemoglobin, beta adult major chain; targeted mutation 1, University of North Carolina	7	Hbb^{th-3}, Hbb^{th3}	Targeted (knock-out)		
Hbb-b1$^{tm1(KOMP)Mbp}$ hemoglobin, beta adult major chain; targeted mutation 1, Mouse Biology Program, UCDavis	7		Targeted (knock-out) ***(Cell Line)***		
Hbb-b1$^{tm1(KOMP)Wtsi}$ hemoglobin, beta adult major chain; targeted mutation 1, Wellcome Trust Sanger Institute	7		Targeted (knock-out) ***(Cell Line)***		

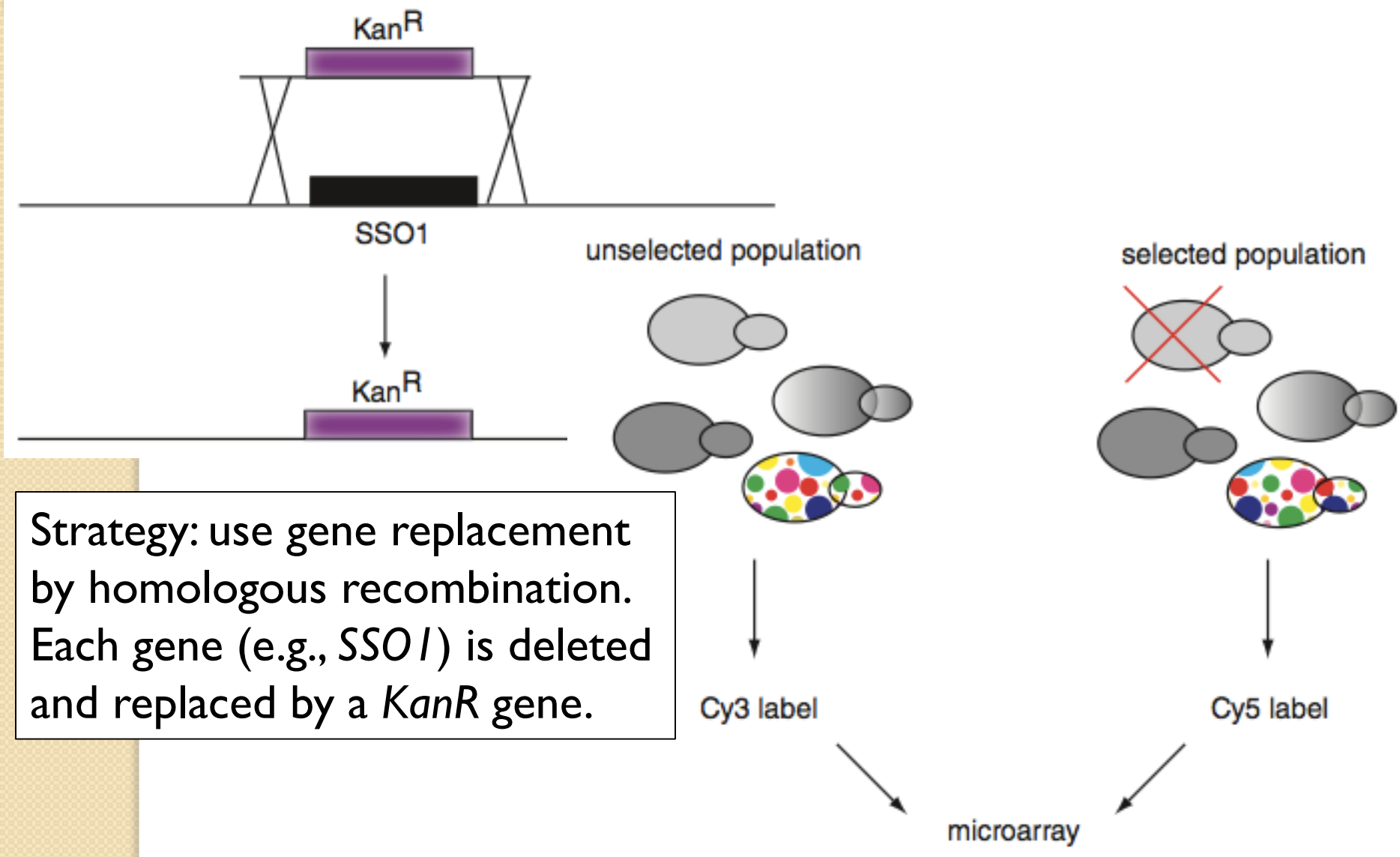
The entry includes phenotypic data such as type of mutation, human disease relevance, genetic background.

Reverse genetics: knocking out genes in yeast using molecular barcodes

Knockout studies in the yeast *S. cerevisiae* are far more straightforward and also much more sophisticated than in the mouse :

- The yeast genome is extremely compact, having very short noncoding regions and introns in fewer than 7% of its ~6000 genes.
- Homologous recombination can be performed with high efficiency

Targeted deletion of virtually all *S. cerevisiae* genes



Strategy: use gene replacement by homologous recombination. Each gene (e.g., *SSO1*) is deleted and replaced by a *KanR* gene.

Reverse genetics: random insertional mutagenesis (gene trapping)

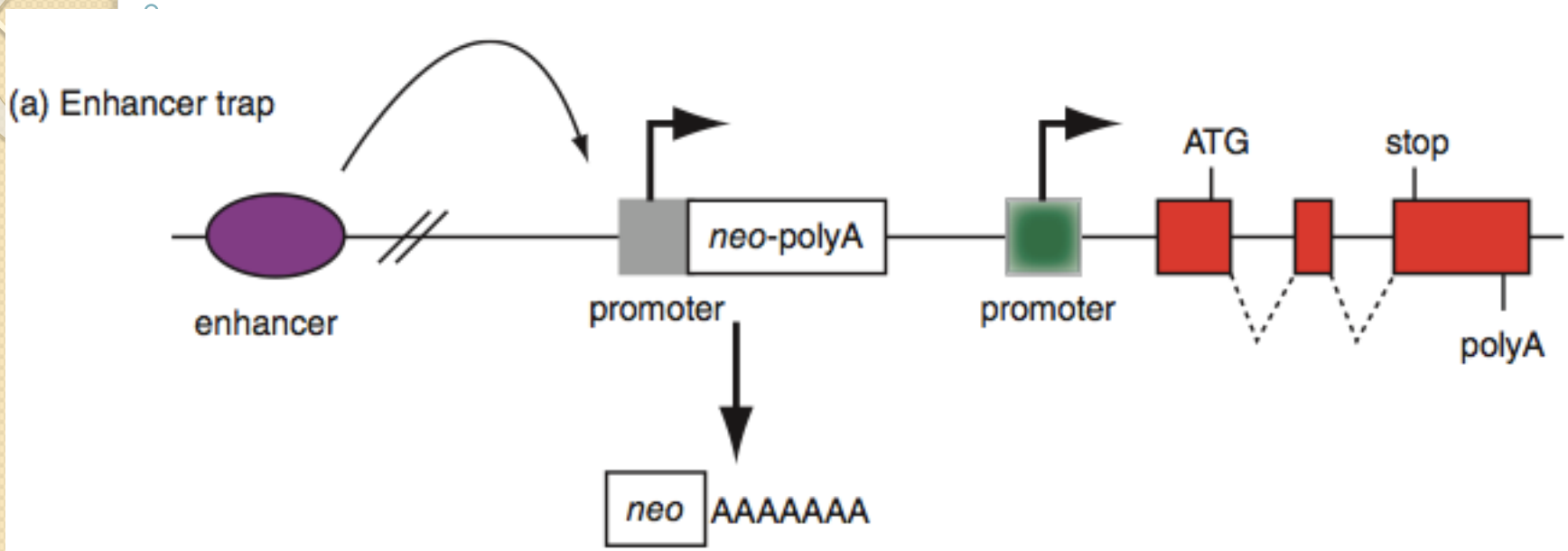
- Insertional mutations are introduced across the genome in embryonic stem cells.
- Vectors insert into genomic DNA leaving sequence tags that often include a reporter gene.
- In this way, mutagenesis of a gene can be accomplished and the gene expression pattern of the mutated gene can be visualized.

Reverse genetics techniques

Method	Advantages	Disadvantages
Homologous recombination (e.g., gene knockouts)	A targeted gene can be replaced, deleted, or modified precisely; stable mutations are produced; specific (no off-target effects)	Low throughput; low efficiency
Gene silencing (e.g., RNAi)	Can be high-throughput; can be used to generate an allelic series; can restrict application to specific tissues or developmental stages	Unpredictable degree of gene silencing; phenotypes not stable; off-target effects are possible
Insertional mutagenesis	High-throughput; used for loss-of-function and gain-of-function studies; results in stable mutations	Random or transposon-mediated insertions target only a subset of the genome; limited effectiveness on tandemly repeated genes; limited usefulness for essential genes
Ectopic expression	Similar to gene silencing	Similar to gene silencing

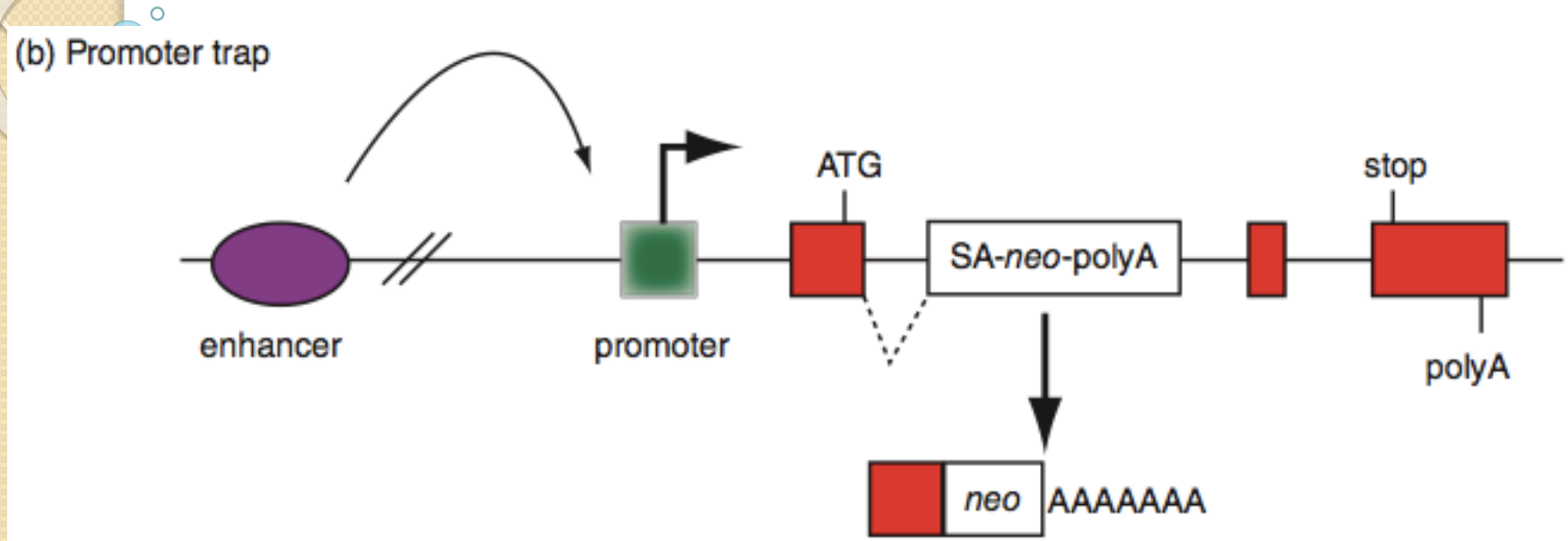
Knockout

Strategies for gene trap mutagenesis



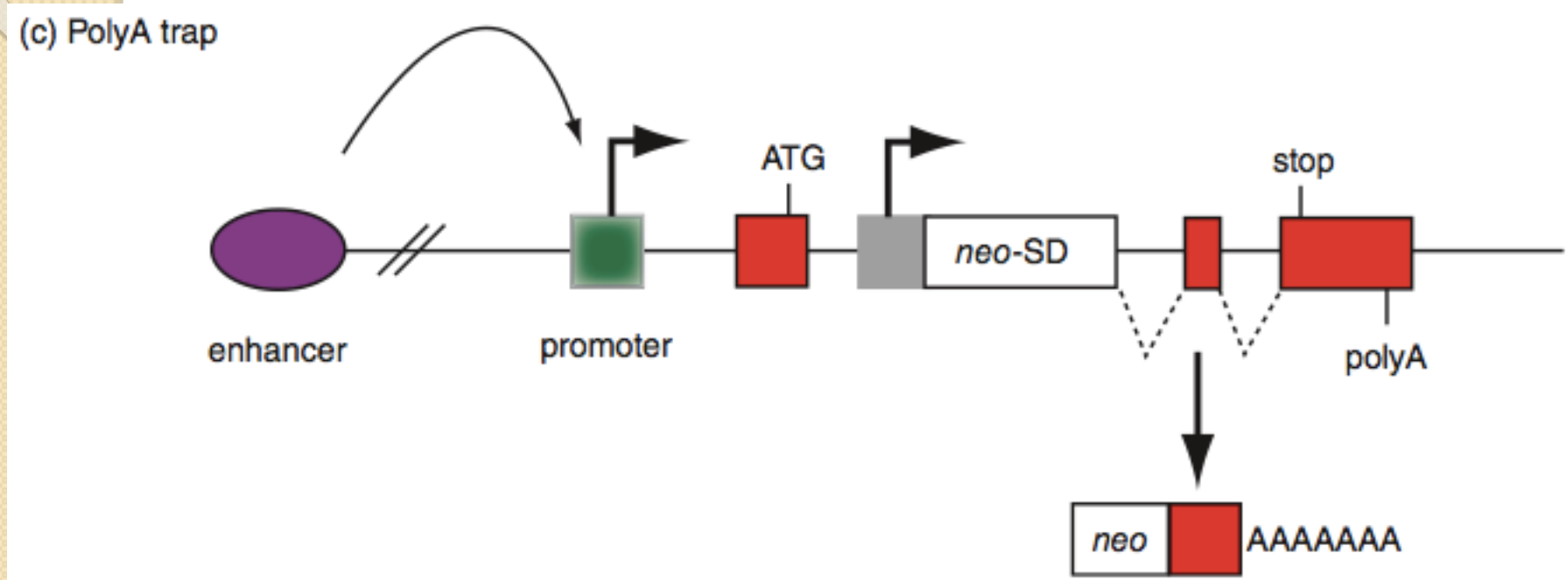
An enhancer trap consists of a vector containing a promoter, a *neo* gene that confers **antibiotic resistance** (and therefore allows for selection of successfully integrated sequences), and a polyadenylation signal (polyA). This construct is activated by an endogenous enhancer, and disrupts the function of the endogenous gene.

Strategies for gene trap mutagenesis



A promoter trap lacks an exogenous promoter and instead depends on an endogenous enhancer and promoter. It includes a splice acceptor (SA), *neo* cassette, and polyadenylation site. Integration of this vector disrupts the expression of an endogenous gene.

Strategies for gene trap mutagenesis



A poly(A) trap vector includes its own promoter and *neo* cassette but depends on an endogenous polyadenylation signal for successful expression.

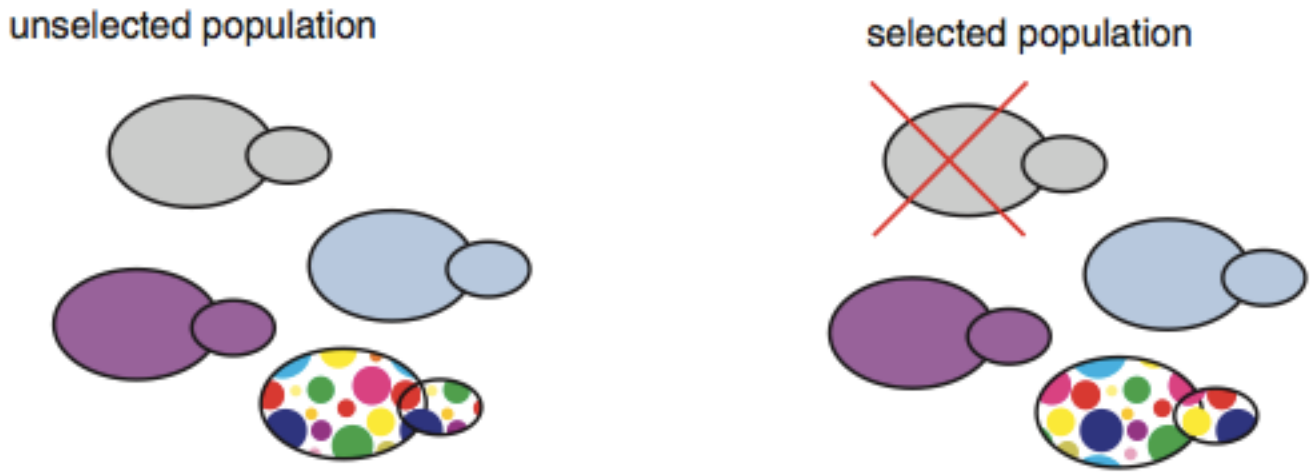
Reverse genetics: insertional mutagenesis in yeast

Two powerful approaches to gene disruption in yeast (in addition to homologous recombination) are:

(1) genetic footprinting using transposons; and

(2) harnessing exogenous transposons.

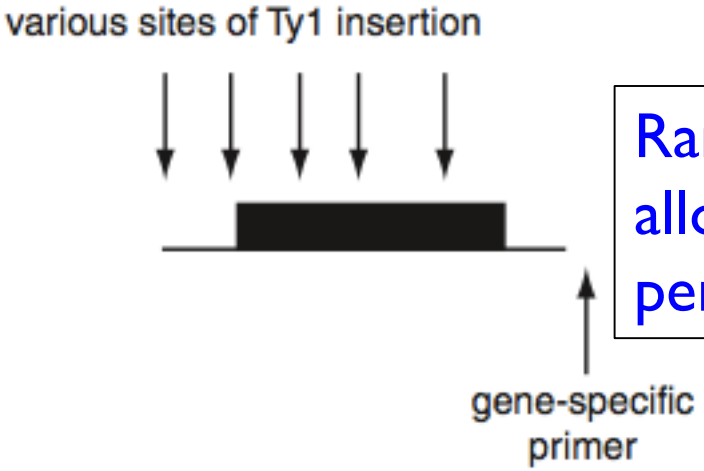
Genetic footprinting



A population of yeast is selected (e.g., by changing the medium or adding a drug); some genes will be unaffected by the selection process.

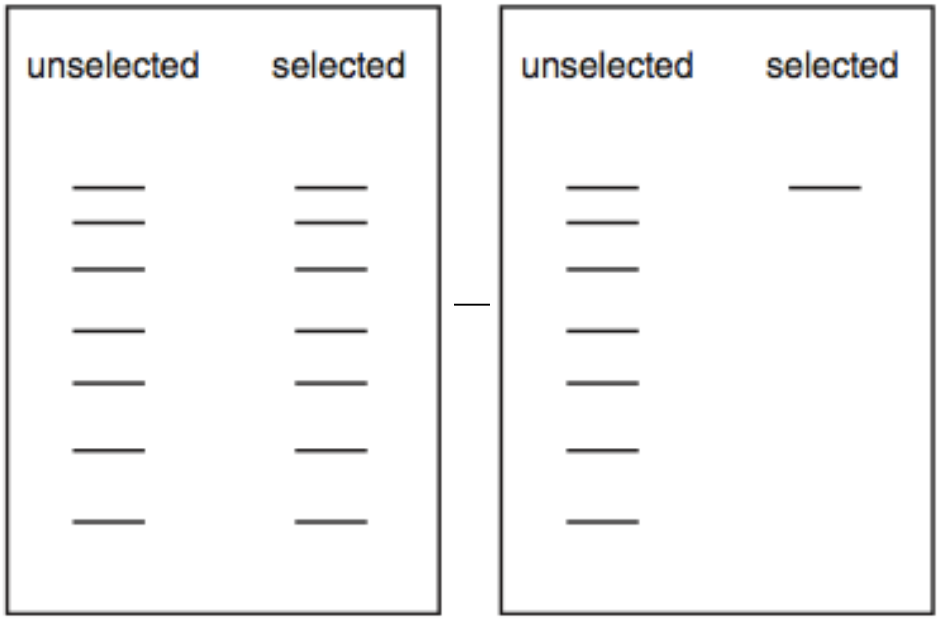
Genetic footprinting

The yeast transposable element **Ty1** is present in about 35 copies per genome;



Random insertion of a transposon allows gene-specific PCR to be performed.

Visualization of DNA products electrophoresed on a gel. Some genes will be unaffected by the selection process (panel at left). Other genes, tagged by the transposition, will be associated with a reduction in fitness. Less PCR product will be observed, therefore identifying this gene as necessary for survival of yeast in that selection condition.



Reverse genetics: gene silencing by disrupting RNA

- Another approach to identifying gene function is to **disrupt the messenger RNA** rather than the genomic DNA. RNA interference (RNAi) is a powerful, versatile technique that allows **genes to be silenced** by double-stranded RNA.

Forward genetics: chemical mutagenesis

- Forward genetics approaches are sometimes called **phenotype-driven screens.**
- ***N*-ethyl-*N*-nitrosurea (ENU)** is a powerful chemical **mutagen** used to alter the male germline to induce **point mutations** (applied to mouse, *Arabidopsis*, other organisms).
- After ENU is given a phenotype of interest is observed. Recombinant animals are created by inbreeding and the phenotype can then be demonstrated to be heritable.
- The mutagenized gene is mapped by **positional cloning** and identified by sequencing the genes in the mapped interval.

Comparison of reverse and forward genetics

- Reverse genetics asks "**What is the phenotype of this mutant?**" Forward genetics asks "**What mutants have this particular phenotype?"**
- Reverse genetics approaches attempt to generate null alleles as a primary strategy (and conditional alleles in many cases).
- Forward genetics strategies such as chemical mutagenesis are "blind" in that **multiple mutant alleles** are generated that affect a phenotype.

Outline : Functional genomics

Introduction
- Relation between genotype and phenotype

Eight model organisms
- *E. coli*; yeast; *Arabidopsis*; *C. elegans*; *Drosophila*; zebrafish; mouse; human

Functional genomics using reverse and forward genetics
- Reverse genetics: mouse knockouts; yeast; gene trapping; insertional mutatgenesis; gene silencing
- Forward genetics: chemical mutagenesis

Functional genomics and the central dogma
- Approaches to function; Functional genomics and DNA; …and RNA; …and protein

Proteomic approaches to functional genomics
- CASP; protein-protein interactions; protein networks

Perspective

Approaches to function and definitions of function

The ENCODE project claimed that >80% of genomic DNA is functional.

We now consider three different *definitions* of function:

- evolutionary selected effect
- causal role
- inferred selected effect

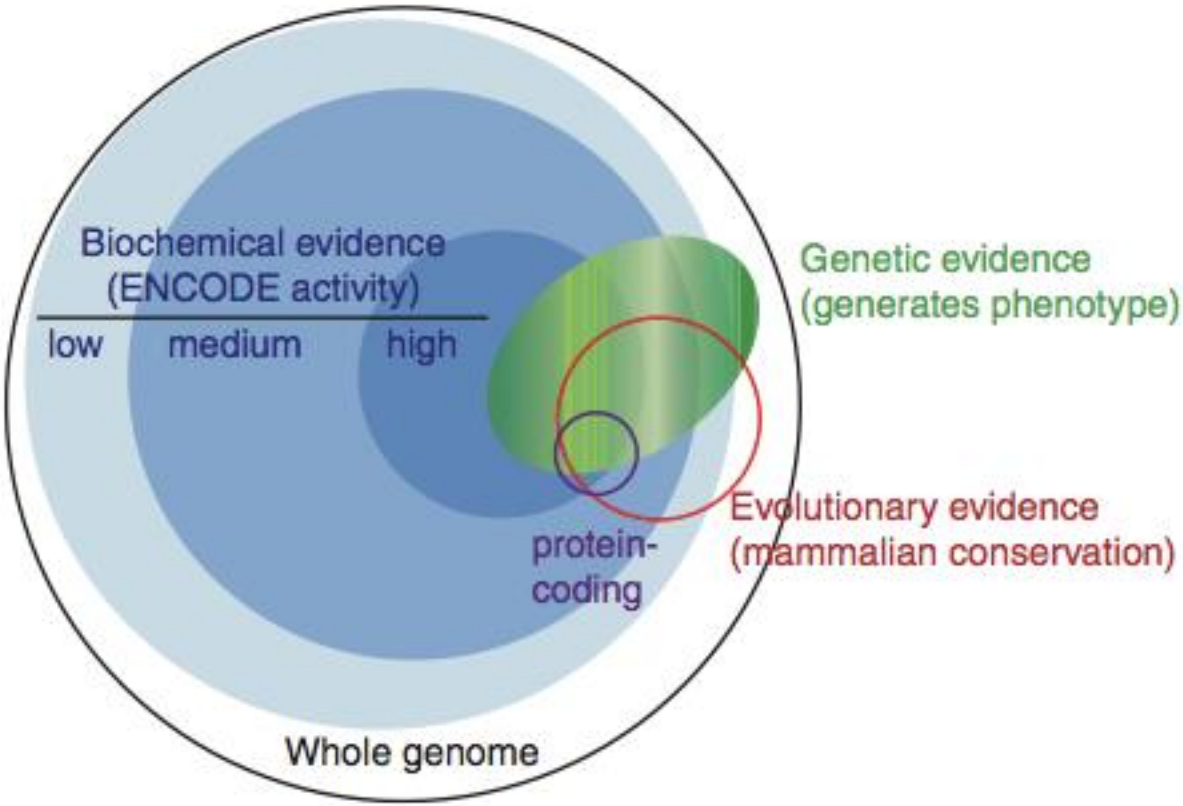
And consider three *approaches* to studying function:

- genetic
- evolutionary
- biochemical

Distinguishing different approaches to function (columns) from definitions of function (rows)

	Approach to function		
	Genetic	Evolutionary	Biochemical
Definition of function	Establish consequence of sequence alterations	Comparative genomics: align DNA, proteins	Measure an activity in a given cell type
Evolutionary selected effect	• Naturally occurring or targeted mutations can be a "gold standard" • Possible to infer function based on selection	• <15% of genome under constraint • Noncoding regions often hard to align	
Causal role	• Example: knockout generates a phenotype • Caveat: some phenotypes depend on a particular condition to be identified	• Many conserved loci functionally important • Caveat: some ultra-conserved loci dispensible • Caveat: some poorly conserved loci are functionally equivalent	• There are increasing numbers of examples of mutations in enhancer regions that cause disease
Inferred selected effect	• Question inspired by ENCODE biochemical map: do most biochemical signatures correspond to functional sites that impact fitness?	• Creation of ENCODE biochemical map may inspire new discoveries of sequence conservation in biochemically functional noncoding regions	• Majority of genome functional • An uncertain % drift, noise • ENCODE biochemical map will facilitate hypothesis testing

Distinguishing different approaches to function from definitions of function



Three circles corresponding to the magnitude of functional findings in ENCODE

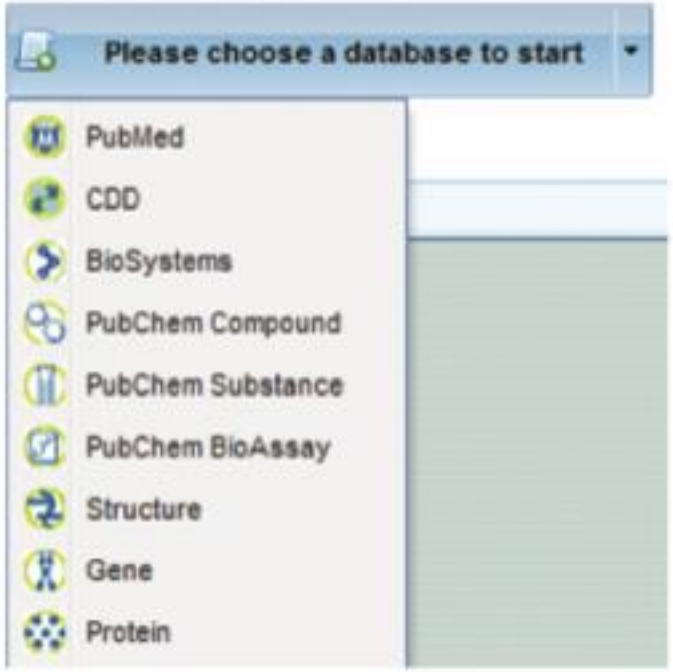
Functional genomics and DNA: integrating information

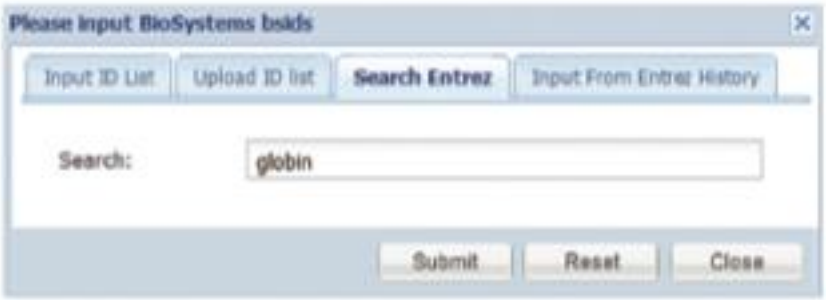
A goal of functional genomics is to provide integrated views of DNA, RNA, protein, and pathways. Many resources (such as those at Ensembl, EBI, and NCBI) offer this integrated view.

An example is the Frequency weighted links (FLink) tool at NCBI. Input a list of genes (or proteins or small molecules) and obtain a ranked list of biosystems.

NCBI FLink: identify connections between an input list of proteins, genes, or other molecules and associated database entries

FLink: select database





FLink: input identifiers or search terms

NCBI FLink: identify connections between an input list of proteins, genes, or other molecules and associated database entries

FLink: table of globin results

BioSystems

Clear Selections | Show | LinkTo | Download CSV | Summary

BSID	Source	Name	Type	Organism
437	KEGG	Two-component system	conserved biosystem	
438	KEGG	Bacterial chemotaxis	conserved biosystem	
451	KEGG	Base excision repair	conserved biosystem	
83043	KEGG	Base excision repair	organism-specific biosystem	Homo sapiens
83240	KEGG	Base excision repair	organism-specific biosystem	Mus musculus
105837	REACTOME	DNA Repair	organism-specific biosystem	Homo sapiens
105838	REACTOME	Base Excision Repair	organism-specific biosystem	Homo sapiens
105839	REACTOME	Base-Excision Repair, AP Site Formation	organism-specific biosystem	Homo sapiens
105840	REACTOME	Depurination	organism-specific biosystem	Homo sapiens
105841	REACTOME	Recognition and association of DNA glycosylase with site containing an affected purine	organism-specific biosystem	Homo sapiens

Functional genomics and RNA

Surveys of RNA transcript levels across different regions (for multicellular organisms) and times of development provide fundamental information about an organism's program of gene expression.

As an example, the *Saccharomyces* Genome Database (SGD) offers many resources to describe gene expression in yeast. For each gene, an expression summary plots the log2 ratio of gene expression (*x* axis) versus the number of experiments. That plot is clickable, so experiments in which SEC1 RNA is dramatically up- or down- regulated can be quickly identified.

Outline : Functional genomics

Introduction
- Relation between genotype and phenotype

Eight model organisms
- *E. coli*; yeast; *Arabidopsis*; *C. elegans*; *Drosophila*; zebrafish; mouse; human

Functional genomics using reverse and forward genetics
- Reverse genetics: mouse knockouts; yeast; gene trapping; insertional mutatgenesis; gene silencing
- Forward genetics: chemical mutagenesis

Functional genomics and the central dogma
- Approaches to function; Functional genomics and DNA; …and RNA; …and protein

Proteomic approaches to functional genomics
- CASP; protein-protein interactions; protein networks

Perspective

Proteomic approaches to functional genomics

Basic features of proteins include their sequence, structure, homology relationships, post-translational modifications, localization, and function. In addition to the study of individual proteins, high throughput analyses of thousands of proteins are possible. We describe three approaches:

- identifying pairwise interactions between protein using the yeast two-hybrid system;
- identifying protein complexes involving two or more proteins using affinity chromatography with mass spectrometry; and
- analyzing protein pathways.

While protein studies have been studied in depth in a variety of model organisms, studies in *S. cerevisieae* are particularly advanced.

Proteomic approaches to functional genomics

We usually think of forward and reverse approaches in terms of genetics, but these terms can apply to proteomics.

Forward proteomics:

- Select experimental system (e.g. normal versus diseased tissue).
- Proteins are extracted and may be labeled with fluorescent dyes or other tags
- Proteins are separated and analyzed by techniques such as mass spectrometry.
- Spectra are analyzed and differentially regulated proteins are identified.
- These regulated proteins may reflect functional differences in the comparison of the original samples.

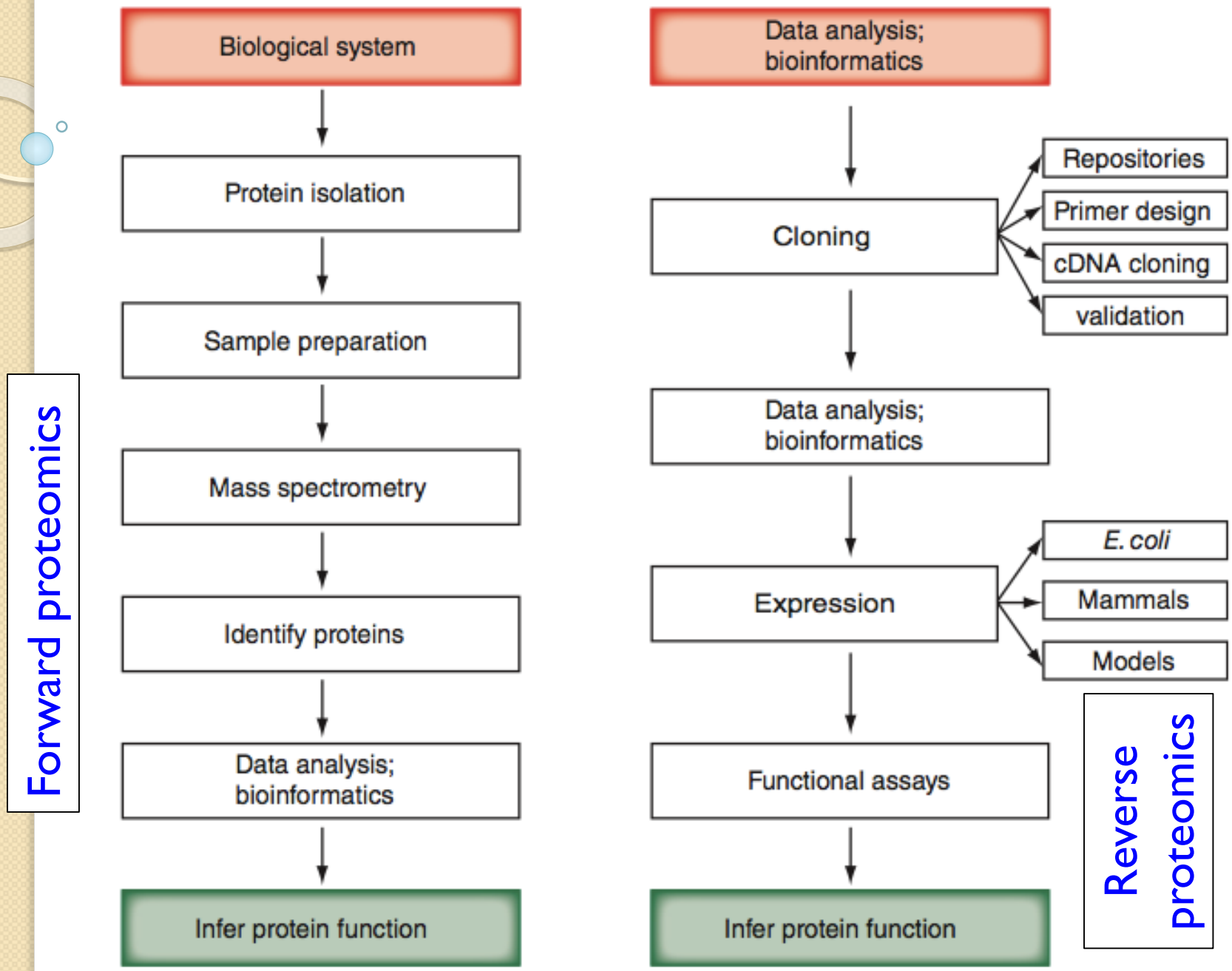
Proteomic approaches to functional genomics

We usually think of forward and reverse approaches in terms of genetics, but these terms can apply to proteomics.

Reverse proteomics:

- A genome sequence of interest is analyzed and genes, transcripts, and proteins are predicted.
- Complementary DNAs (cDNAs) are cloned based on information about open reading frames.
- cDNAs are validated by sequence analysis and expressed in systems such as *E. coli* (for the production of recombinant proteins), mammalian cells, or other model organism systems.
- Functional assays are performed; assays include the yeast two-hybrid system or other protein interaction assays.

Forward and reverse proteomics



Critical assessment of protein function annotation

CAGI involved many challenges inherent in the nature of protein function:

- Protein function is defined at multiple levels, involving the role of a protein on its own and in pathways, cells, tissues, and organisms.
- Protein function is context dependent (e.g., many proteins change function in the presence of a signal such as calcium or a binding partner).
- Proteins are often multifunctional.
- Functional annotations are often incomplete and may be incorrect.
- Curation efforts map protein function to gene names, but multiple isoforms of a gene may have different functions.

Protein-protein interactions

Most proteins perform their functions in networks associated with other proteins and other biomolecules. As a basic approach to discerning protein function, pairwise interactions between proteins can be characterized.

Proteins often interact with partners with high affinity. (The two main parameters of **any binding interaction are the affinity,** measured by the **dissociation constant K_D**, and the maximal number of **binding sites B_{max}**.)

Protein-protein interactions

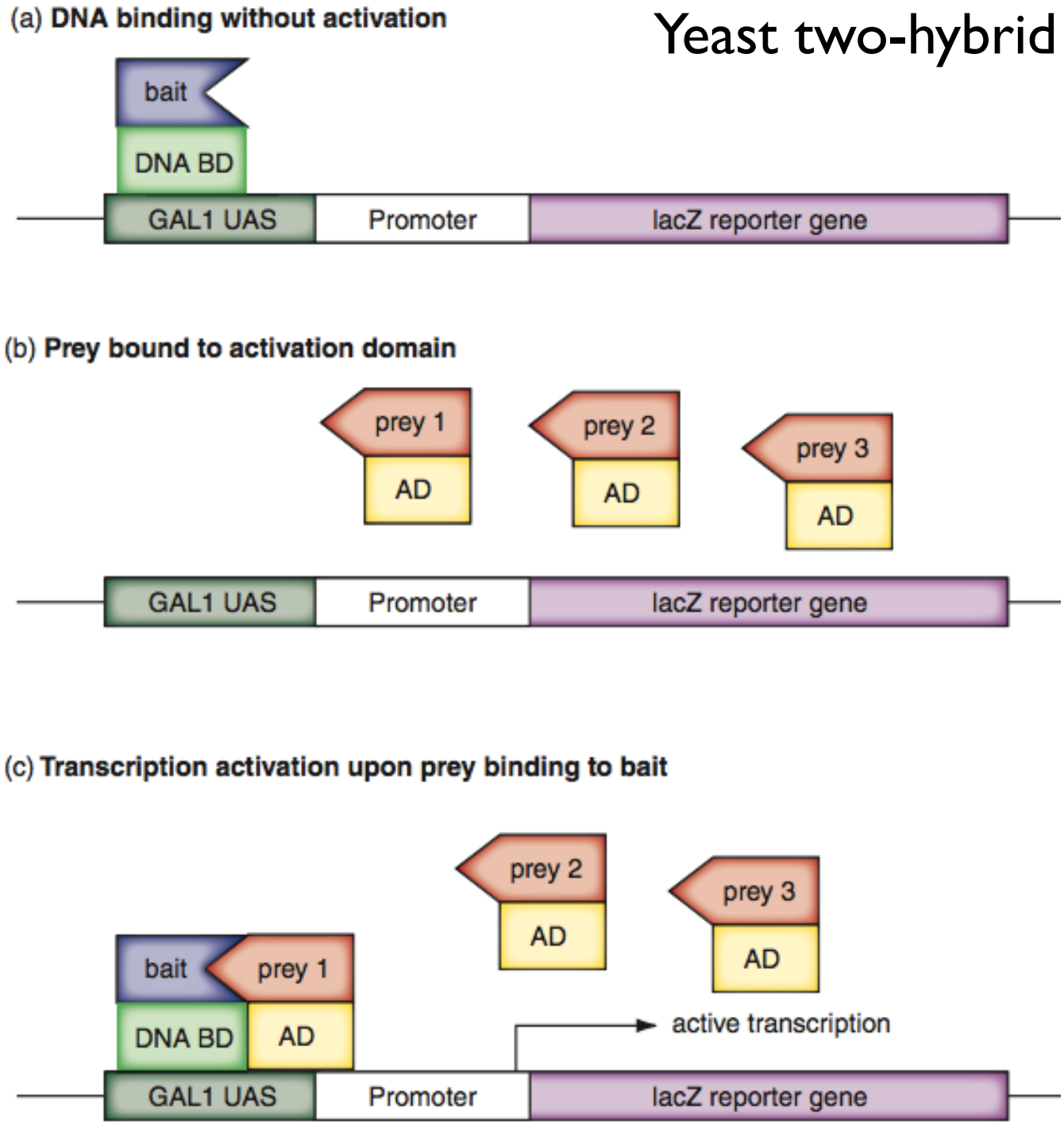
The interactions of two purified proteins can be **measured** with dozens of techniques such as the following:

- *Co-immunoprecipitation*: specific **antibodies** directed against a protein are used to precipitate the protein along with any associated binding partners.
- *Affinity chromatography*: a cDNA construct encodes a protein of interest in frame with **glutathione *S*-transferase (GST)** or some other tag. A resin to which glutathione is covalently attached is incubated with a **GST fusion protein**, and it binds to the resin along with any binding partners. Irrelevant proteins are eluted and then the specific binding complex is eluted and its protein content is identified.

Protein-protein interactions

- *Cross-linking with* ***chemicals*** *or* ***ultraviolet radiation***: a protein is allowed to bind to its partners and then **cross-linking** is applied and the interactors are identified.
- *Surface plasmon resonance* (with the BIAcore technology of GE Healthcare): a **protein is immobilized to a surface and kinetic binding properties of interacting proteins are measured.**
- *Equilibrium dialysis and filter binding assays*, in which bound & free ligands are separated and quantitated.
- *Fluorescent resonance energy transfer* (FRET): two labeled proteins yield a characteristic change in **resonance energy** upon sharing a close physical interaction.

Yeast two-hybrid system

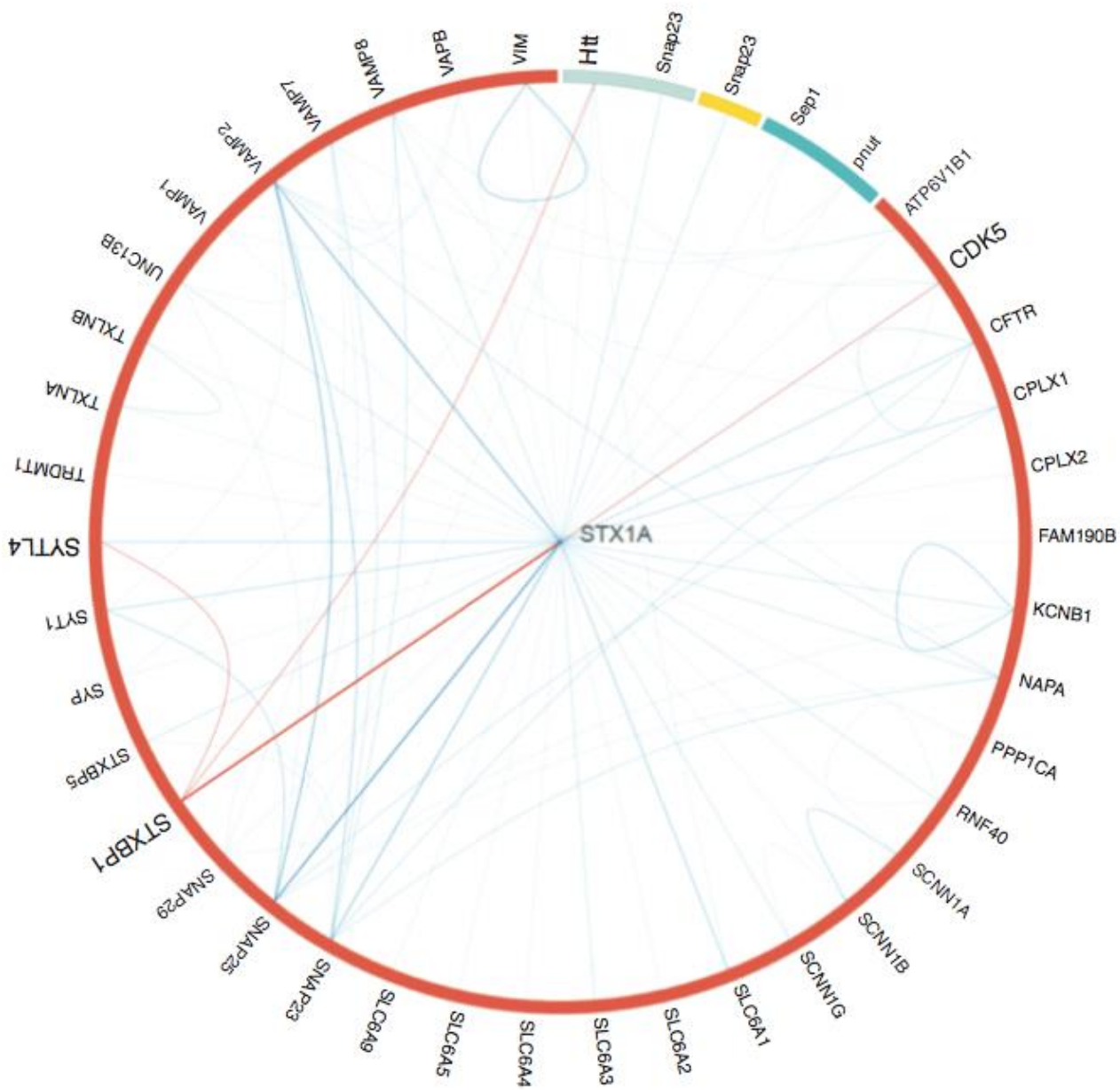


Protein–protein interaction databases

Database	Comment	URL
BioGrid	Repository for interaction datasets	http://www.thebiogrid.org/
Biomolecular Object Network Databank (BOND)	Requires log-in; formerly BIND	http://bond.unleashedinformatics.com/
Comprehensive Yeast Genome Database (CYGD)	From the Munich Information Center for Protein Sequences (MIPS)	http://mips.helmholtz-muenchen.de/genre/proj/yeast/
Database of Interacting Proteins (DIP)	From UCLA	http://dip.doe-mbi.ucla.edu/
Human Protein Reference Database (HPRD)	From Akhilesh Pandey's group at Johns Hopkins	http://www.hprd.org/
IntAct	At the European Bioinformatics Institute	http://www.ebi.ac.uk/intact/
Molecular Interactions (MINT) Database	Rome	http://mint.bio.uniroma2.it/mint/
PDZBase	Database of PDZ domains	http://abc.med.cornell.edu/pdzbase
Reactome	Curated resource of core human pathways and reactions	http://reactome.org/
Search Tool for the Retrieveal of Interacting Genes/Proteins (STRING)	Database of known and predicted protein–protein interactions	http://string.embl.de/

There are many prominent protein–protein interaction databases

Example of a protein-protein interaction database entry: BioGrid network map for syntaxin and its binding partners



From pairwise interactions to protein networks

A typical mammalian genome has ~20,000 to 25,000 protein-coding genes, a subset of which (perhaps 10,000 to 15,000) are expressed in any given cell type. These **proteins are localized to particular compartments** (or are secreted) where many of them interact as part of their function.

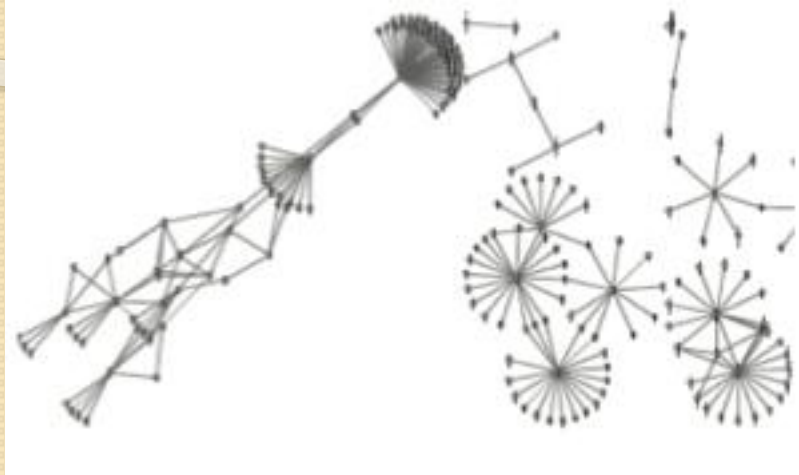
Many databases show protein network data. We next show PSICQUIC and Cytoscape as examples.

Protein interaction networks

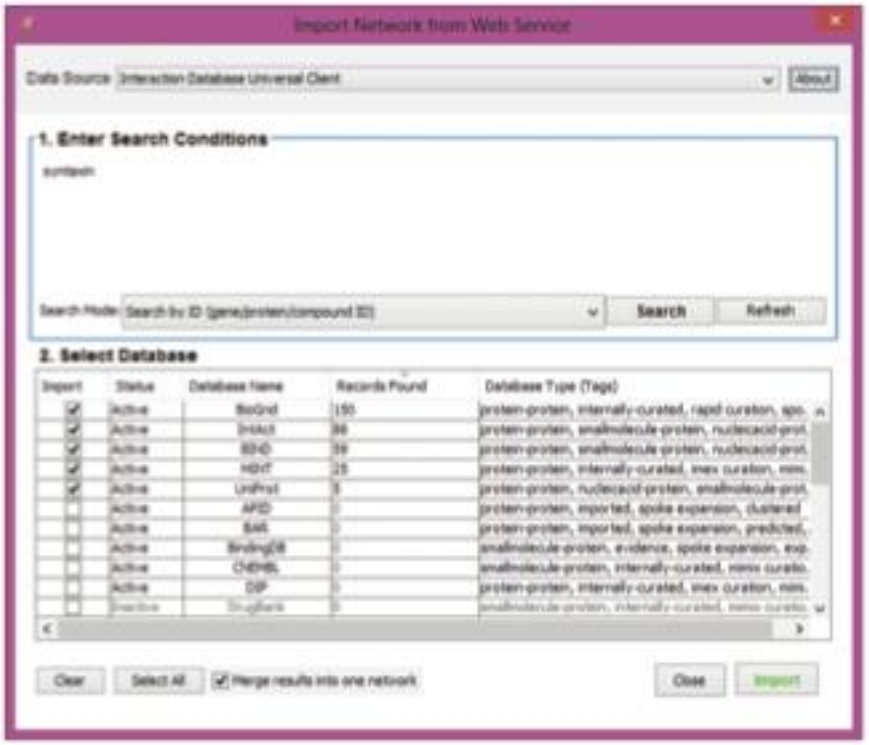
PSICQUIC databases of protein interactions.

B&FG 3e
Fig. 14.21
Page 679

Protein interaction networks

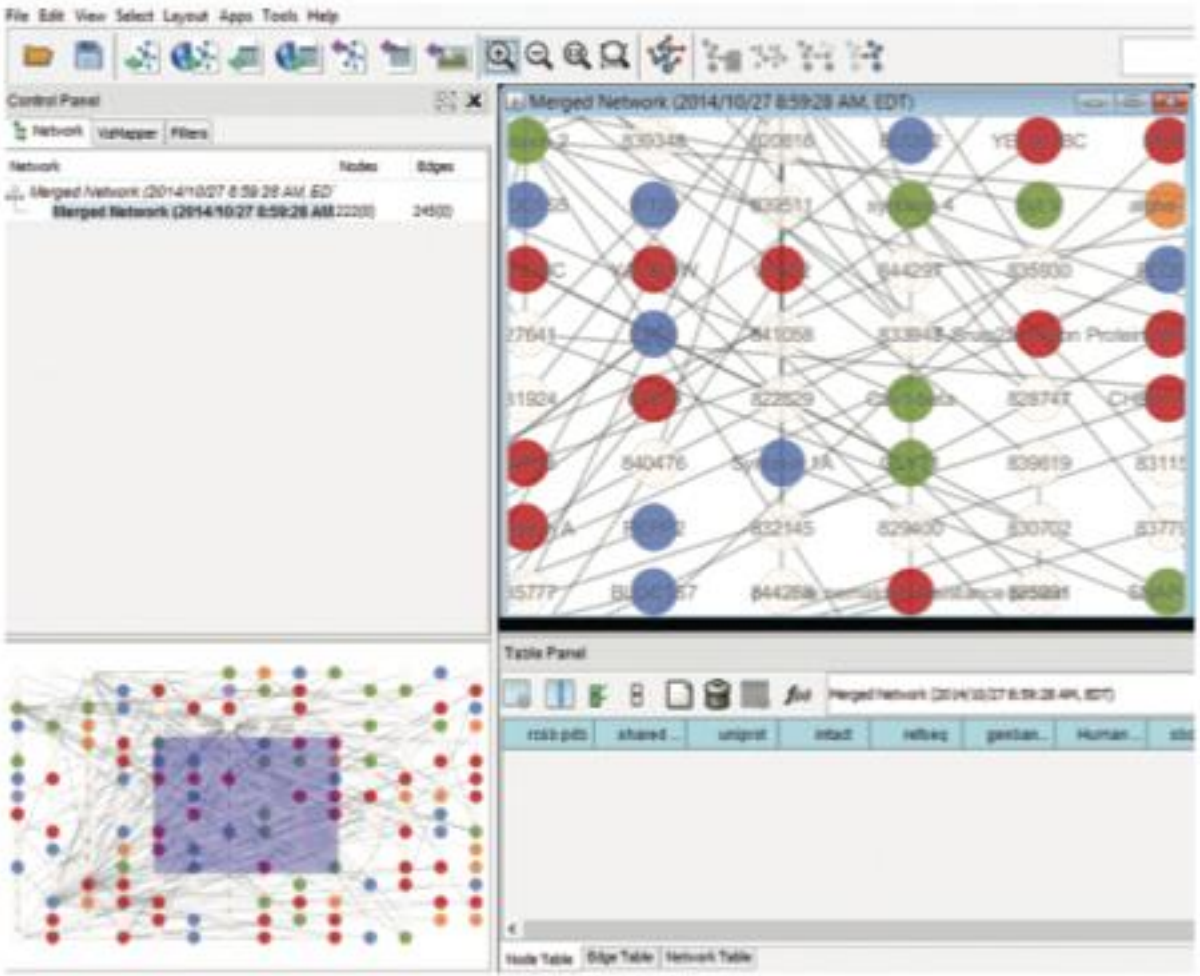


PSICQUIC display of Cytoscape network for syntaxin



Cytoscape data import

Protein interaction networks



Zoom of Cytoscape diagram showing syntaxin binding partners

From pairwise interactions to protein networks

There are many issues regarding protein interaction networks.

- Assessment of accuracy. How likely is it that a false positive or false negative error has occurred? Benchmark ("gold standard") datasets are required that consist of trustworthy pathways.
- Choice of data. Many researchers integrate data from genomic sequences, expression of RNA transcripts, and protein measurements. But RNA and protein levels may be poorly correlated.
- Experimental organism. Function may be better conserved between paralogs than between orthologs!

From pairwise interactions to protein networks

There are many issues regarding protein interaction networks.

- Variation in Pathways. Some pathways (e.g. Krebs cycle) are characterized in great detail; many not. Some are transient, others stable.
- Categories of maps. Maps may be of metabolic pathways, physical and/or genetic interaction data, summaries of the scientific literature, or signalling pathways. **Maps may be based on experimental data or inferred relationships**.

Pathways, networks, and integration: bioinformatics resources

There are many database resources.

- PathGuide lists >500 biological pathway resources.
- BioGRID database provides manual curation of ~32,000 publications describing physical and genetic interactions.
- MetaCyc is a database of metabolic pathways. https://metacyc.org/
- Kyoto Encyclopedia of Genes and Genomes (KEGG) contains a detailed map of metabolism based on 120 metabolic pathways, with links to various organisms.
- KEGG pathways are a collection of manually drawn maps in six areas: metabolism; genetic information processing; environmental information processing; cellular processes; human diseases; and drug development.

KEGG database

KEGG includes pathway maps, data for a broad range of organisms, and a variety of analysis tools.

KEGG: Kyoto Encyclopedia of Genes and Genomes

KEGG is a database resource for understanding high-level functions and utilities of the biological system, such as the cell, the organism and the ecosystem, from molecular-level information, especially large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies (See Release notes for new and updated features).

Main entry point to the KEGG web service

KEGG2	KEGG Table of Contents	Update notes

Data-oriented entry points

KEGG PATHWAY	KEGG pathway maps	[Pathway list]
KEGG BRITE	BRITE functional hierarchies	[Brite list]
KEGG MODULE	KEGG modules	[Module list]
KEGG DISEASE	Human diseases	[Cancer | Infectious disease]
KEGG DRUG	Drugs	[ATC drug classification]
KEGG ORTHOLOGY	Ortholog groups	[KO system]
KEGG GENOME	Genomes	[KEGG organisms]
KEGG GENES	Genes and proteins	Release history
KEGG COMPOUND	Small molecules	[Compound classification]
KEGG REACTION	Biochemical reactions	[Reaction modules]

Entry point for wider society

KEGG MEDICUS	Health-related information resource

Organism-specific entry points

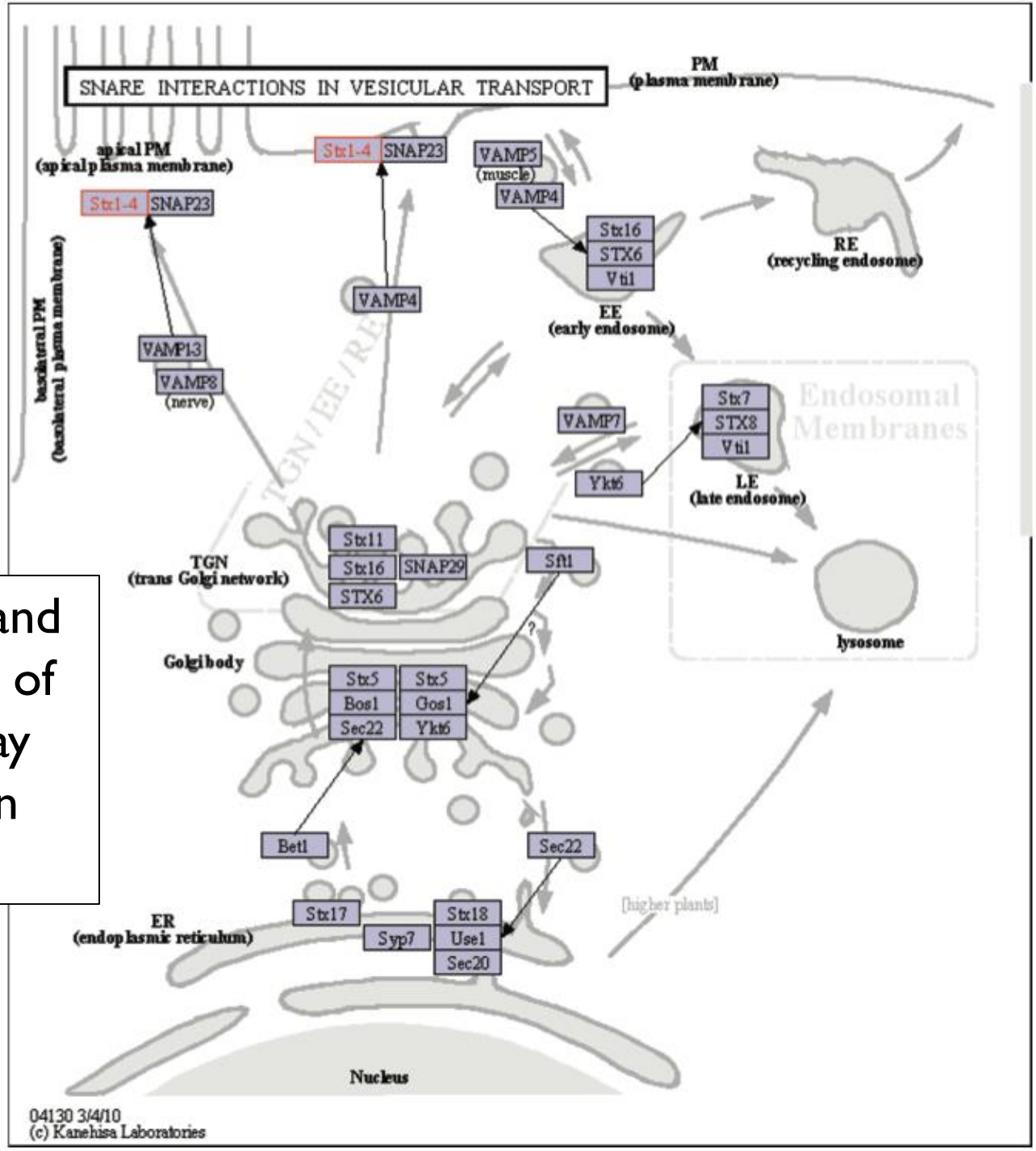
KEGG Organisms Enter org code(s) [ ] Go hsa hsa eco

Analysis tools

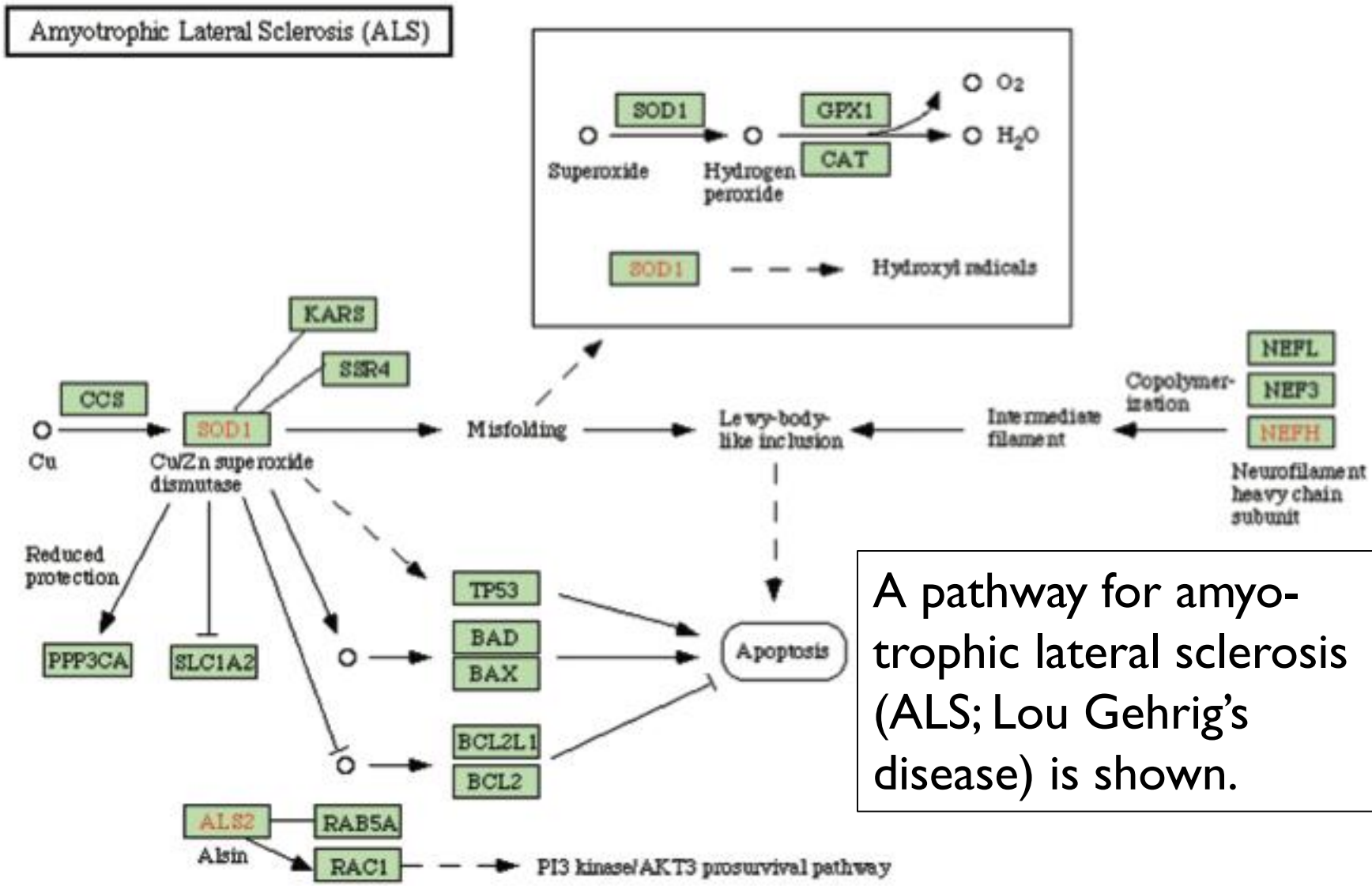
KEGG Mapper	KEGG PATHWAY/BRITE/MODULE mapping tools
KEGG Atlas	Navigation tool to explore KEGG global maps
KAAS	KEGG automatic annotation server
BLAST/FASTA	Sequence similarity search

KEGG database

KEGG includes maps and data for a broad range of organisms. This pathway shows SNARE function including syntaxin



KEGG database: disease pathways



A pathway for amyotrophic lateral sclerosis (ALS; Lou Gehrig's disease) is shown.

Outline : Functional genomics

Introduction
- Relation between genotype and phenotype

Eight model organisms
- *E. coli*; yeast; *Arabidopsis*; *C. elegans*; *Drosophila*; zebrafish; mouse; human

Functional genomics using reverse and forward genetics
- Reverse genetics: mouse knockouts; yeast; gene trapping; insertional mutatgenesis; gene silencing
- Forward genetics: chemical mutagenesis

Functional genomics and the central dogma
- Approaches to function; Functional genomics and DNA; …and RNA; …and protein

Proteomic approaches to functional genomics
- CASP; protein-protein interactions; protein networks

Perspective

Perspective

The field of functional genomics is broad, and can be considered using many different categories.

- What type of organism do we wish to study? We highlighted eight model organisms, although many other models are commonly used.
- What type of questions do we want to address: natural variation or experimental manipulations used to elucidate gene function?
- What type of experimental approach do we wish to apply (e.g., forward versus reverse genetics)?
- What type of molecules do we wish to study (i.e., from genomic DNA to RNA to protein or metabolites)?
- What types of biological questions are we trying to address?

Perspective

- We are beginning to confront a problem that is perhaps even harder than identifying genes: identifying their function. Function has many definitions.