Chapter 14:
Functional Genomics

Jonathan Pevsner, Ph.D.
pevsner@kennedykrieger.org
Bioinformatics and Functional Genomics
(Wiley-Liss, 3rd edition, 2015)

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

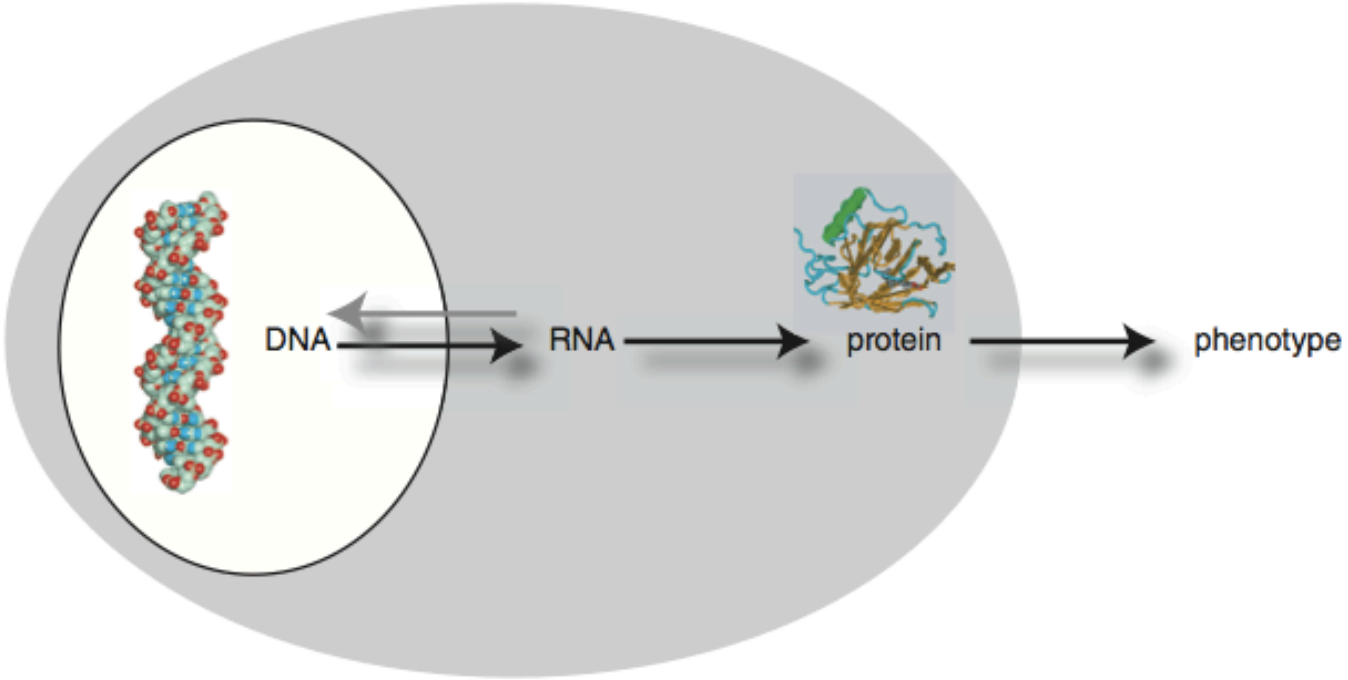
Introduction: Functional genomics

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

B&FG 3e
Fig. 14.1
Page 636

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)(Chapter 17)
- 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

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

- The budding yeast *S. cerevisiae* is the best-characterized organism among the eukaroytes
- Single-celled fungus (see Chapter 18 on fungi)
- 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; see Chapter 12).

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.

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).
- ~7000 genes have been deleted

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

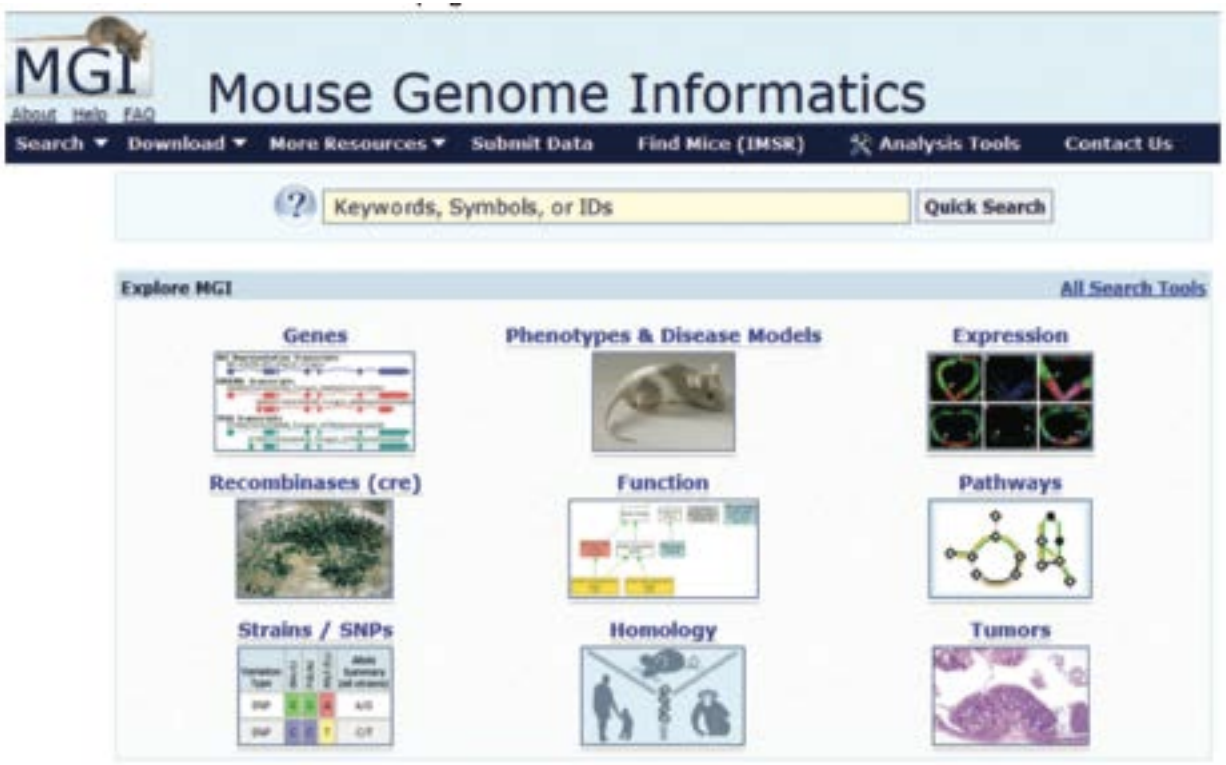
8 model organisms: (6) zebrafish *Danio rerio*

- 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.

B&FG 3e
Fig. 14.6
Page 646

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.

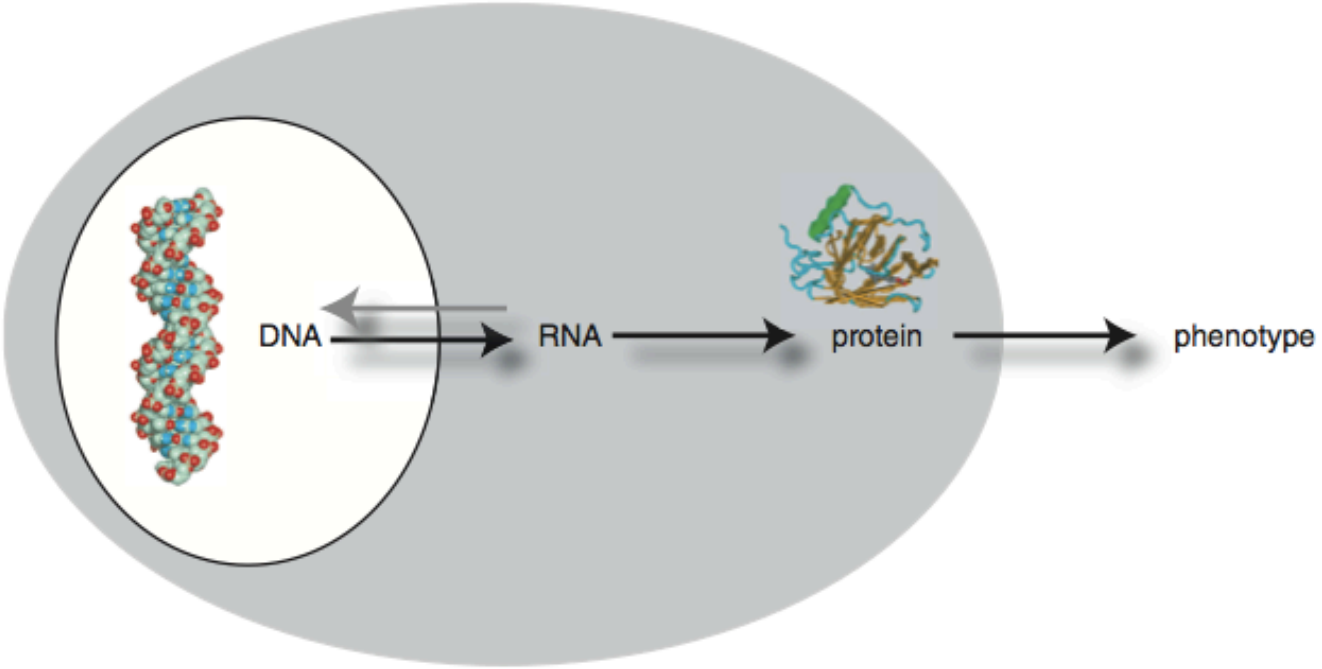
Functional genomics using reverse and forward genetics

- 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

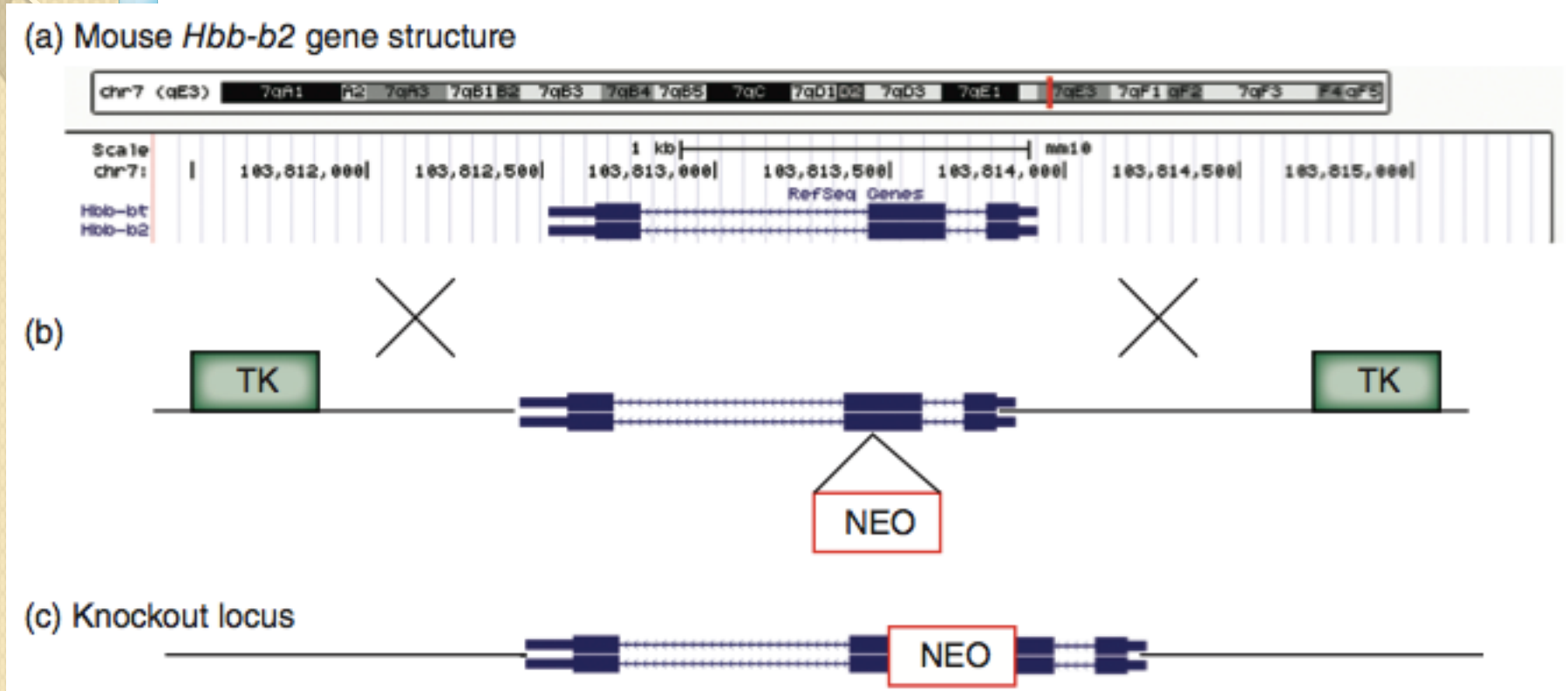
Forward genetics ("phenotype-driven" screen)

B&FG 3e
Fig. 14.7
Page 649

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 survive. Confirm by PCR.

Method of gene knockout by homologous recombination



B&FG 3e
Fig. 14.9
Page 652

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

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

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

B&FG 3e
Table 14.1
Page 658

Knockout

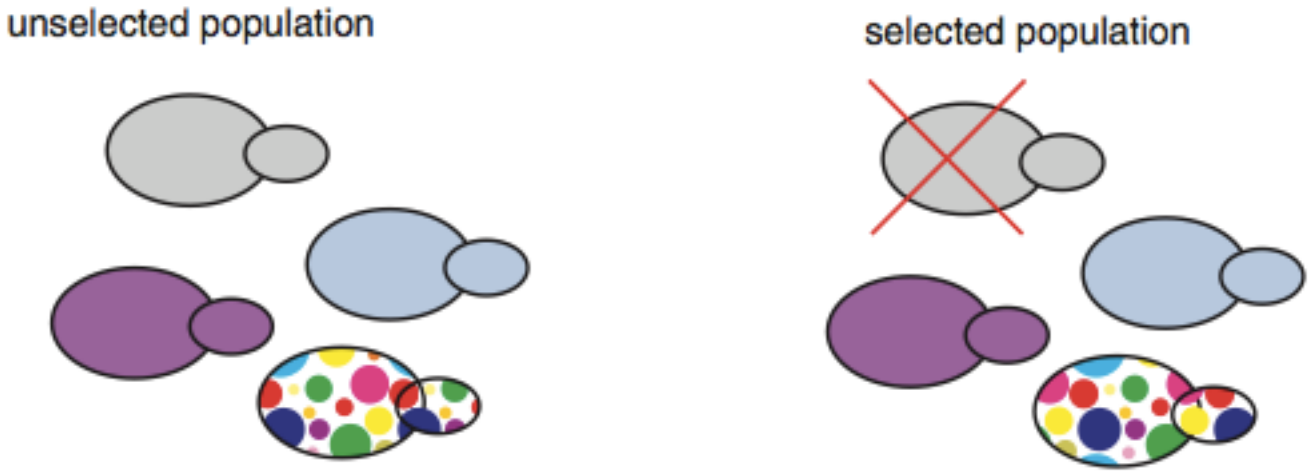
Reverse genetics: insertional mutagenesis in yeast

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.

B&FG 3e
Fig. 14.15
Page 661

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