Forward and Reverse Genetics

1. Background

What is the function of a particular gene? The standard genetic approach to answer this question is to look at mutant phenotypes or gene "knockouts". The assumption is that if we know what goes wrong with the organism when a particular gene is mutated, we can infer what the gene does in its wild-type state.

2. Forward (classical) genetics – mutant phenotype -> gene

Typically, mutant phenotypes are known before their corresponding genes have been identified. These can be phenotypes in model organisms, such as white eyes in *Drosophila*, or heritable human diseases such as cystic fibrosis or Huntington's disease. In most cases, much work went into mapping and cloning the mutant genes using pedigree or association studies and "clone walking". Many genes were named after their mutant phenotype before they were cloned. For example the *Drosophila rosy* gene is named for the rose eye color of mutant flies. It encodes the enzyme xanthine dehydrogenase. The *Drosophila white* gene is named for the white eye color of mutant flies. Thus, the wild-type *white* gene is responsible for red eyes.

Forward saturation genetics – treat organism (bacteria, *C. elegans*, *Drosophila*, *Arabidopsis*, *etc.*), with a mutagen, then screen offspring for particular phenotypes of interest. Examples: inability of bacteria to grow on certain sugars, problems in fly embryonic development, plants lacking a response to light. The goal is to find all of the genes involved in a trait. This approach is known as a "genetic screen".

Mutagens:

a) X-rays – cause breaks in double-stranded DNA, resulting in large deletions of pieces of chromosome or chromosomal re-arrangements. These mutations are typically easy to map by cytological examination of chromosomes, but are often not limited to single genes. Not good for fine-scale mutagenesis.
b) Chemical – for example, the chemical ethylmethanesulfonate (EMS) causes point mutations, which are changes at a single nucleotide position. Mutations may be non-sense (introduce a premature stop codon) or mis-sense (cause an amino acid replacement). They may also be in non-coding sequence, affecting splicing signals or regulatory elements that control gene expression. This approach allows for many different mutations within gene regions, but these are difficult to map.
c) Insertional (transposon) mutagenesis – Transposable elements (TEs) containing a marker gene(s) are mobilized in the genome. The TE can insert within a coding region and disrupt the amino acid sequence or it may insert into neighboring non-coding DNA and affect intron splicing or gene expression. The major advantage is that the TE insertion can easily be mapped and the region of genome cloned.

After mutants are identified, they can be separated into complementation groups, mapped to general chromosomal location by linkage with known markers, and eventually cloned and sequenced.

3. Reverse genetics – gene -> mutant phenotype

In the post-genomic era, the classical problem has been reversed. We now know all of the genes in an organism, but we do not know the function of many of them (usually >50% of the predicted genes in eukaryotic genomes) and we do not know what phenotypes are caused by mutations in an even larger fraction of the predicted genes.

a) Large-scale random mutagenesis and screening – use forward mutagenesis as above (for example EMS), except instead of screening for a particular phenotype, screen your gene of interest for nucleotide changes. This typically requires that you screen 1000's or 10,000's of individuals. This is done by performing PCR on your gene of interest and looking for slight differences in the migration of the PCR product on a gel or column. In theory, you could sequence the DNA of each individual and look for changes, but there are more efficient methods of detection. Some examples:

DHPLC – Denaturing High Performance Liquid Chromatography
DGGE – Denaturing Gradient Gel Electrophoresis
SSCP – Single-Stranded Conformation Polymorphism

These methods can be automated for large-scale screening and can also be used to identify naturally-occurring variants by comparing a large number of individuals from within a population or species. Not all changes will knockout the gene. Some changes will be "silent" or at non-essential nucleotide positions. This method is good for fine-scale mutagenesis.

b) Homologous recombination (HR) – works well in bacteria, yeast, mice (and some other mammals). It does not work well in *Drosophila*, although a complex experimental approach has been developed for this. HR has been used to knockout every predicted gene in yeast. It is possible to buy a set of about 6,000 yeast strains, each with a different gene knocked-out. Many mouse genes have been knocked out by this method. It has also been used to knockout a pig enzyme that links sugars into a form recognized as an antigen by the human body, with the long-term goal of engineering pig organs to be used for human transplants.

c) Transposable element excision – especially useful in *Drosophila*, where the Berkeley Drosophila Genome Project (BDGP) has a large collection of fly lines, each with a marked TE inserted at a unique chromosomal location. When a source of transposase is introduced, the TE will excise with some low frequency, resulting in a loss of the marker gene. Often the TE excision will also result in a deletion of the flanking DNA. Thus, if you have a TE insertion near your gene of interest, you may try to knockout your gene by excising the nearby TE.

The "standard" P-element method – uses a single TE with a single marker gene.

The wHy method – A *Hobo* transposable element is placed inside a *P*-element and is flanked by *white* and *yellow* genes on either side. When the Hobo TE is excised, deletions occurring to either side are easily identified by loss of the phenotypic marker. A large collection of these inserts has been developed with the goal of deleting almost all genes in the *Drosophila* genome.

d) RNA interference (RNAi) – double stranded RNA (dsRNA) can lead to specific post-transcriptional gene silencing (PTGS). This mechanism is part of a natural response of the host that most likely evolved to control virus or TE replication. RNAi works in worms, insects, plants, mammalian cell cultures, *etc*. Many short RNAs (known as micro- or miRNAs) are encoded by genomes and play a role in the regulation of gene expression. Sometimes RNAi does not completely eliminate expression of the target gene, but only reduces it. For this reason, it is often referred to as "knock down" instead of "knock out".

Methods of dsRNA delivery:
injection – works well in *C. elegans*, where injection into the body knocks-out gene expression in the injected worm. The knock-out even persists into the next generation. Injection also works well in *Drosophila* (and other insect) eggs to knockout genes involved in early development. However, it is much less efficient for genes expressed in adult tissues.

feeding/soaking – dsRNA can be added directly to *C. elegans* food or the worms can be fed bacteria that produce dsRNA. This method is efficient and can be used on a large scale with little effort. A similar approach works with cultured cells (*Drosophila*, mammal, etc.), where dsRNA is added to the cell culture medium.

transgenic methods – a source of dsRNA is introduced permanently into the host genome. Thus, it is heritable and can be used to create stable knock-down lines. In addition, specific promoters can be chosen to limit RNAi to particular tissues or developmental stages. Thus, it is much more specific than injection and can be used to knockout genes expressed in adult tissues.

Two basic approaches for transgenic constructs:
two inverted promoters – one on either side of the "source" gene
hairpin RNA constructs – fold into dsRNA after transcription

e) Genome editing - in recent years, several methods have been developed to target mutations to a specific location in the genome. These can be used to knock-out target genes, make specific point mutations in a target gene, or even insert new genes or DNA sequences at a specific target site in the genome. Examples include: zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and clustered regularly interspaced short palindromic repeat (CRISPR) and CRISPR associated (Cas) nucleases. The CRISPR-Cas system has become the most common method for genome engineering.

4. Site-directed mutagenesis and transgenics
Can make a specific change at an exact nucleotide(s) in gene of interest *in vitro*, then put the mutated gene back into the genome.
Requires: Gene of interest cloned into a plasmid. Ideally, a host with null background (knockout) of the target gene.

Mutagenesis is performed *in vitro* on the cloned gene in a plasmid replicated in bacteria, then the mutated gene is inserted into the host genome in a transposable element vector.

QuikChange Mutagenesis Method (Stratagene)
- dsDNA denatured by heating
- cooled in presence of 2 complementary "mismatch" primers
- plasmid replicated *in vitro* with non-strand-displacing polymerase (*Pfu*)
- above steps repeated12-18 times to get a pool of new and parental plasmids
- digest methylated parental DNA with *Dpn*I enzyme, leaving *in vitro* synthesized DNA
- transform remaining plasmids into bacteria and replicate
- sequence DNA to ensure proper nucleotide change(s)

Applications
a) Drosophila Alcohol dehydrogenase (*Adh*) 3' untranslated region (UTR)
Specific deletions were made in a highly-conserved sequence in the *Adh* 3' UTR.
High conservation implies functional importance.
The mutant *Adh* genes were introduced into *Adh*-null flies.
Flies with a mutant *Adh* 3' UTR showed a twofold increase in alcohol dehydrogenease enzymatic activity, thus the wild-type sequence must be a negative regulatory element; when it is deleted, expression increases.

b) Heterologous reporter genes
Use a gene from a different organism that is easily detected. This can be used to test regulatory sequences
Examples: β-galactosidase (enzyme encoded by the *lacZ* gene of bacteria), GFP (Green fluorescent protein, naturally found in jellyfish)

A 150-bp sequence from the 5' flanking region of the *Drosophila ocnus* gene was cloned in front of the *E. coli lacZ* gene, then inserted into the *D. melanogaster* genome using a P-element vector. The *ocnus* gene encodes a sperm protein and is expressed specifically in testis. The transgenic *lacZ* gene was also expressed specifically in testis. This indicates that the 150-bp flanking sequence is sufficient to confer testis-specific expression.