Reverse genetics - Knockouts

Paper to read for this section :

Melton, D.W. (1994) Gene targeting in the mouse. *Bioessays* **16**:633-8

Up until now, we've concentrated on ways to get a cloned gene. We've seen how we can go from a mutant with an observable phenotype to a cloned gene, and how to obtain genes based on proteins or mRNAs. Once you have a cloned gene, what can you do with it? The first thing most people do once they've identified the part of a DNA clone that comprises their favorite gene is sequence the DNA. The sequence allows you to determine the primary structure of the gene product, and to identify regulatory elements that might be involved in the expression of the gene. Having the gene enables a lot more than sequencing. In particular, the cloned gene can be the starting point for identifying new mutations by targeted mutagenesis methods.

The most basic form of mutation that can be evaluated is a null mutation; the consequences of removing a gene product completely should provide information about its normal function. Knocking out the expression or activity of a gene is one of the first things one does when a candidate gene is identified by reverse genetics or sequence gazing. Even where mutations have already been identified by classical mutant hunts, clean knockouts can be useful.

In this chapter, we will examine some of the methods available to construct knockout mutations in specific genes. As you might expect based on our discussion of other genetic methods, the available approaches are dependent on the genetic system being used. As usual, there are advantages in using easily transformed microorganisms like *E. coli* and yeast.

Engineering knockouts by homologous recombination in microorganisms

Homologous recombination can be used to construct precise deletions in schemes that are very similar to the methods used in marker rescue. However, instead of using a library for recombination, one engineers a specific construct that can replace the wild-type gene.

Figure 8-1 shows a 2-step method for replacing a chromosomal gene; the markers shown are

designed for gene replacement in yeast. A plasmid clone is constructed with a deletion in the gene of interest. The plasmid can be constructed in *E. coli*, with an origin of replication that can function in *E. coli*. The clone also carries a selectable *URA3* marker in the vector part, and lacks an origin of replication that can function in yeast.

Diploid *ura3* cells are transformed with this plasmid and plated on media lacking uracil, selecting for integration of the plasmid into a chromosome so that the *URA3*[+] gene will be linked to an origin of replication. The major pathway for this to happen in yeast is for a homologous recombination event to occur between the chromosomal locus and the shared sequences in the plasmid. A single crossover generates a duplication wiith the plasmid sequences in between the two copies of the gene. One of the two copies contains the engineered deletion.
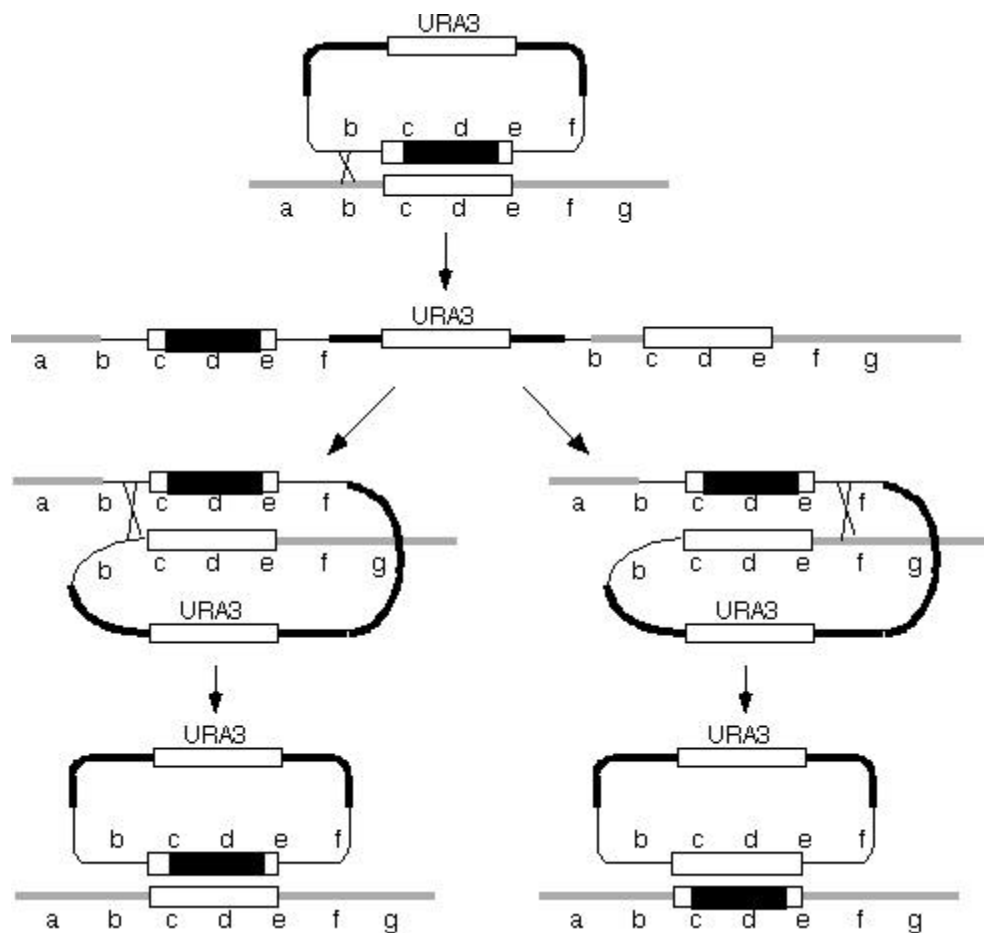

Figure 8-1 Two step method for gene replacement

The strain containing the inserted plasmid is grown up in the absence of selection and then plated on 5-FOA + uracil. This selects for cells that have lost the duplication by a recombination event between the two copies. Depending on whether the recombination events that integrate and excise the plasmids occur on the same side or on different sides of the gene, the excision can have two possible results. If the two recombination events occur on the same side, the original wild-type gene is restored. If the recombination that removes the plasmid occurs on the other side of the gene relative to the integration, then the wild-type gene is lost with the plasmid, and the deleted gene is left behind in the chromosome.

The presence of the deletion in one of the chromosomal copies of the gene can be assayed by Southern blotting or by PCR analysis. The construction of the deletion is done in a diploid cell in case the function of the gene is essential. When the diploid is sporulated, the deleted copy of the gene and the wild-type copy will segregate 2:2; if the gene is essential for viability, only 2 out of four spores in each tetrad will grow into a colony. If all four spores grow, the ones containing the deletion are identified and characterized to determine the effects of losing the gene.

Similar methods can be used in *E. coli*. Since *E. coli* is a haploid organism, strains where construction of the deletion will be done should be made partially diploid, or **merodiploid**, for the gene of interest. Since construction of the plasmid with the deletion requires that it contain an *E. coli* origin of replication, we need a way to prevent replication in order to select for integration events. This can be done either by using plasmids with conditional origins, such as *ts* mutants, or by using switching to strains that don't support plasmid replication. For example, many *E.coli* plasmids are unable to replicate in strains with mutations in *polA*, the gene that encodes DNA polymerase I.

An alternative approach to select for *E.coli* chromosomes that have undergone recombination with a plasmid is based on the ability of certain strains to mobilize the *E. coli* genome for transfer into other strains by conjugation (Figure 8-2). *E. coli* strains that contain integrated copies of the F factor are called Hfr strains. In an Hfr donor, replication starting from an origin of transfer on the F plasmid transfers DNA into a recipient strain that lacks the F plasmid. If the engineered plasmid is introduced

into the Hfr strain, it can integrate into the chromosome of the donor and get transfered into the recipient. In the recipient strain, the duplication can recombine into the chromosome and then resolve in the same way that the duplications resolve in Figure 8-1.

The recipient strain is genetically marked so that the donor can't grow after the mating occurs. Selection for a marker on the plasmid selects for only those cells that have received the plasmid; this can only occur if the plasmid integrated into the donor DNA.
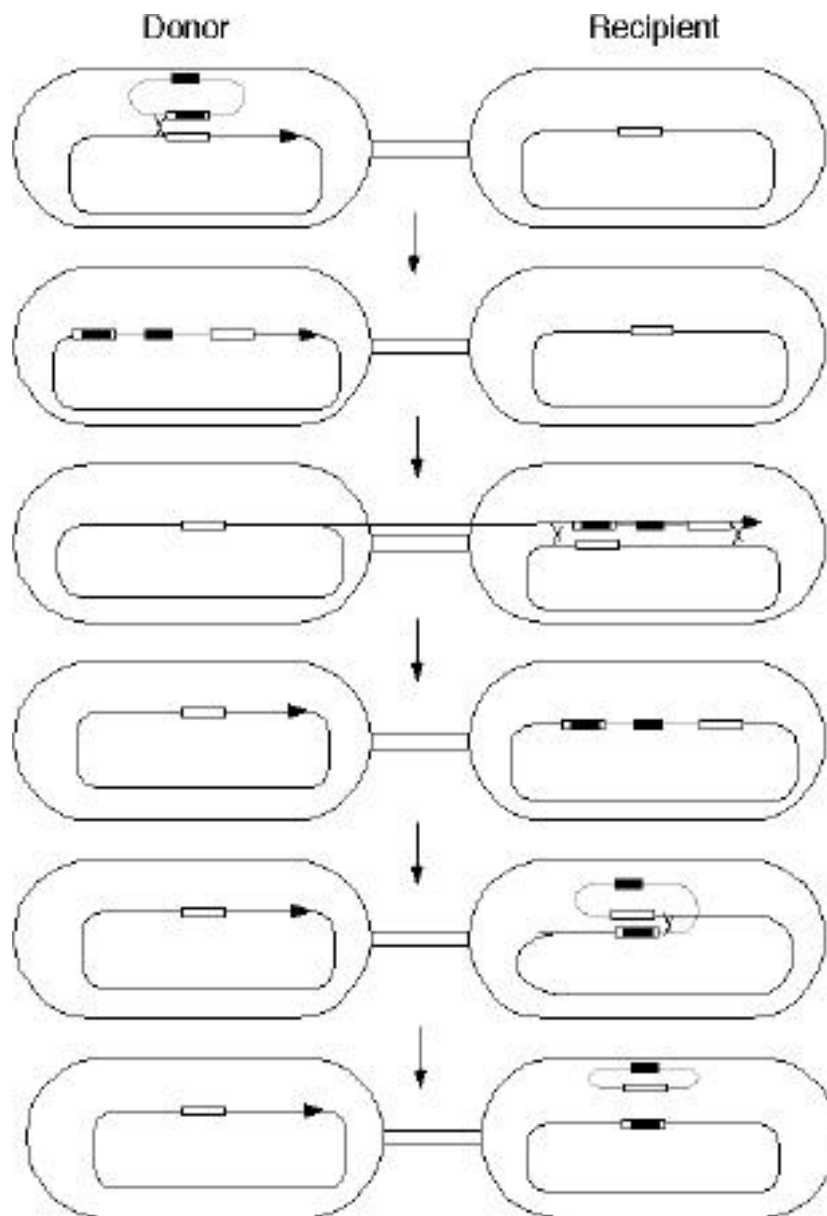

Figure 8-2 Mating to select for integrative recombination

Gene replacement can also be done in yeast or *E. coli* by a 1-step method (Figure 8-3). In both of these organisms, transformation with linear DNA stimulates recombination between the transforming DNA and chromosomal DNA.   Traditionally, a clone is constructed where the deleted DNA is replaced by DNA for a selectable marker. The DNA containing the deletion-replacement is removed from the clone by cutting it out, and the linear DNA is transformed into a recipient strain. Two crossovers are required to replace the chromosomal gene; although this seems like it should be rare, it works (The reasons why it is less rare than might be expected have to do with the mechanisms of recombination and are beyond the scope of this discussion).
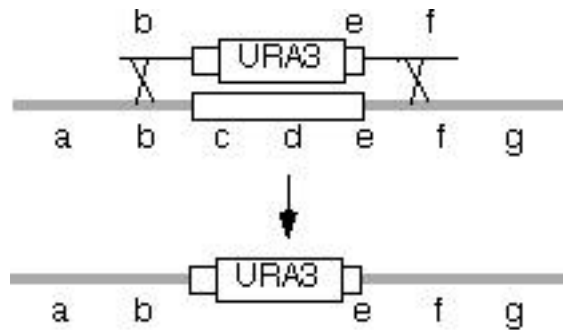

Figure 8-3 Constructing knockouts by recombination with linear DNA.

PCR provides an alternative method to generate the linear DNA (Figure 8-4). Primers are synthesized that have 3' ends that match the ends of the selectable marker, and 5' ends that match the sequences flanking the gene of interest. Amplification generates a linear DNA molecule where the selectable marker is flanked by short sequences that can direct recombination to the gene of interest. It turns out that only a small amount of homologous sequence is sufficient to direct recombination to the desired site, at least in yeast.

In wild-type *E. coli*, linear DNA is usually degraded before it has a chance to recombine. Mutations in the *recBCD* system are used to inactivate the degradation system and enable recombination between linear fragments and chromosomal DNA.
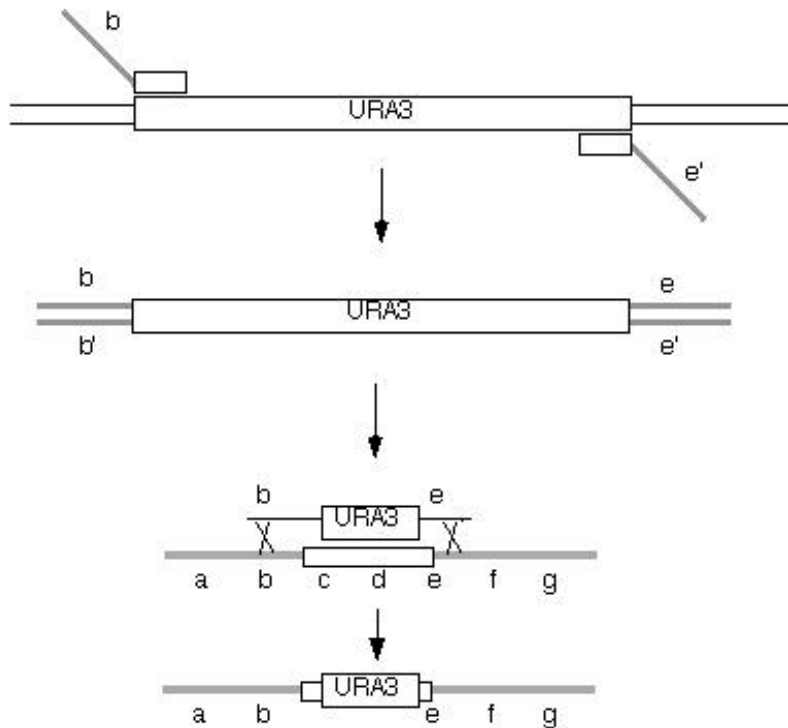
Figure 8-4. Generating linear deletion-replacement molecules by PCR

Generating knockouts by recombination works well in *E. coli* and yeast. In many other organisms, a series of related factors make it harder to generate knockout mutations by recombination. First, the numbers of organisms that can be transformed is limited if getting DNA in involves microinjection of individuals one at a time. Second, it is impractical to select for markers that kill untransformed cells; in a mosaic you'd kill the whole organism. Third, DNA transfected into many higher organisms gets incorporated into chromosomes randomly. Even if homologous recombination occurs, it has to be found against a background of random integration that can be >100X higher than the formation of recombinants by homologous crossovers.

Engineering knockouts by homologous recombination in mice

The review by Melton describes how these problems have been overcome to create knockouts in mice. The problems of getting sufficient numbers of transfected cells and of needing selectable markers to find rare recombinants are both solved by using cultured cells instead of whole mouse embryos. The important breakthrough came from the recognition that when culturable embryonic stem

cells (ES cells) are injected into a host embryo, the ES cells are incorporated into the host and divide to become adult tissues in a mosaic mouse. Sometimes the ES cells are incorporated into the germ cells, so that any genes that were introduced into the ES cells can be passed on to later generations (See figure 1 in Melton).

Mice can be regenerated from ES cells transfected with DNA that becomes stably integrated. This can be used to introduce foreign genes into mice. Randomly integrated DNA can be used in these cases, since one doesn't care what happens to endogenous genes or where the DNA goes. This kind of technology is also used in the genetic engineering of other mammals, which can be engineered to produce commercially important proteins in their milk.

Deletion-replacements similar to those used in yeast or *E. coli* can be constucted and used to transform ES cells. The transfected DNA can have two fates (Figure 8-5). Most of the time it will integrate randomly; often in arrays of two or more copies. Occasionally, homologous recombination can occur, replacing one of the endogenous copies of the targeted gene.
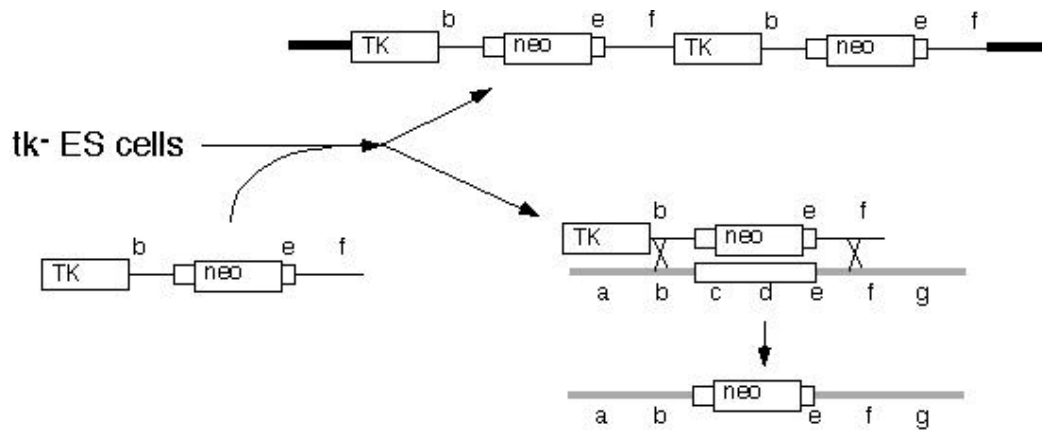

Figure 8-5. Fates of transfected DNA.

In order to find the cells where homologous recombination has occurred among the many cell that have integrated, the transfecting DNA is constructed to carry a counterselectable marker outside of the region of homology needed for recombination. If the DNA integrates randomly, the counterselectable marker is incorporated into the genome. In contrast, the double crossover needed to replace the gene does not bring along the counterselectable marker. Selection against the external marker will kill the cells that carry random integrations.

Conditional knockouts by Cre - Lox

Note that homologous recombination can be used to place other kinds of alleles into the mouse genome. The properties of a phage-encoded site-specific recombination system can be exploited to construct knockouts in specific tissues.

Cre is a site specific recombinase encoded by phage P1 that recognizes a 34 bp site called Lox. Cre catalyzes efficient recombination between pairs of Lox sites. Starting with a cloned gene, Lox sites can be placed on both sides of the gene of interest; this is called **floxing** the gene because you are introducing **fl**anking **Lox** sites. The Lox sites are introduced in such a way that expression and activity of the gene are unaffected. Using the homologous recombination methods described above, the floxed gene can be introduced into the normal chromosomal location. By breeding the resultant mice to get homozygotes, it can be determined whether or not the floxing had any detectable phenotypic effects. When mice containing the floxed gene are bred with mice that contain a construct that expresses Cre, the recombinase activity of Cre will delete anything that is between the Lox sites.

S. Tonegawa's lab at MIT 1996 (Tsien et al 1996 Cell 87:1317) used the Cre-Lox system to show the effect of a tissue-specific deletion of a gene (Figure 8-6). First, they created a construct where Cre is expressed from a promoter for   -CaMKII (- Ca calmodulin dependent kinase II) which is only expressed in the forebrain in mice. Next, they used gene targeting to create a homozygous mouse where both copies of the NMDAR1 gene, which encodes a receptor in postsynaptic ion channels, was floxed. The two mouse strains were crossed, and then F1 mice that carried the Cre construct were backcrossed to the homozygous NMDAR(floxed) parent. Half of progeny of the backcross that contained the Cre gene were also homozygous for the floxed NMDAR gene.

In adult mice, most of the tissues will contain the floxed construct, because Cre is not expressed. In the forebrain, however, expression of Cre will delete sequences between the two Lox sites and inactivate the gene. Using this method, Tsien showed that mice with defects in NMDAR in the forebrain had impaired spatial memory. This would not be possible by conventional knockouts, since the knockout of NMDAR is lethal if it happens in all tissues.
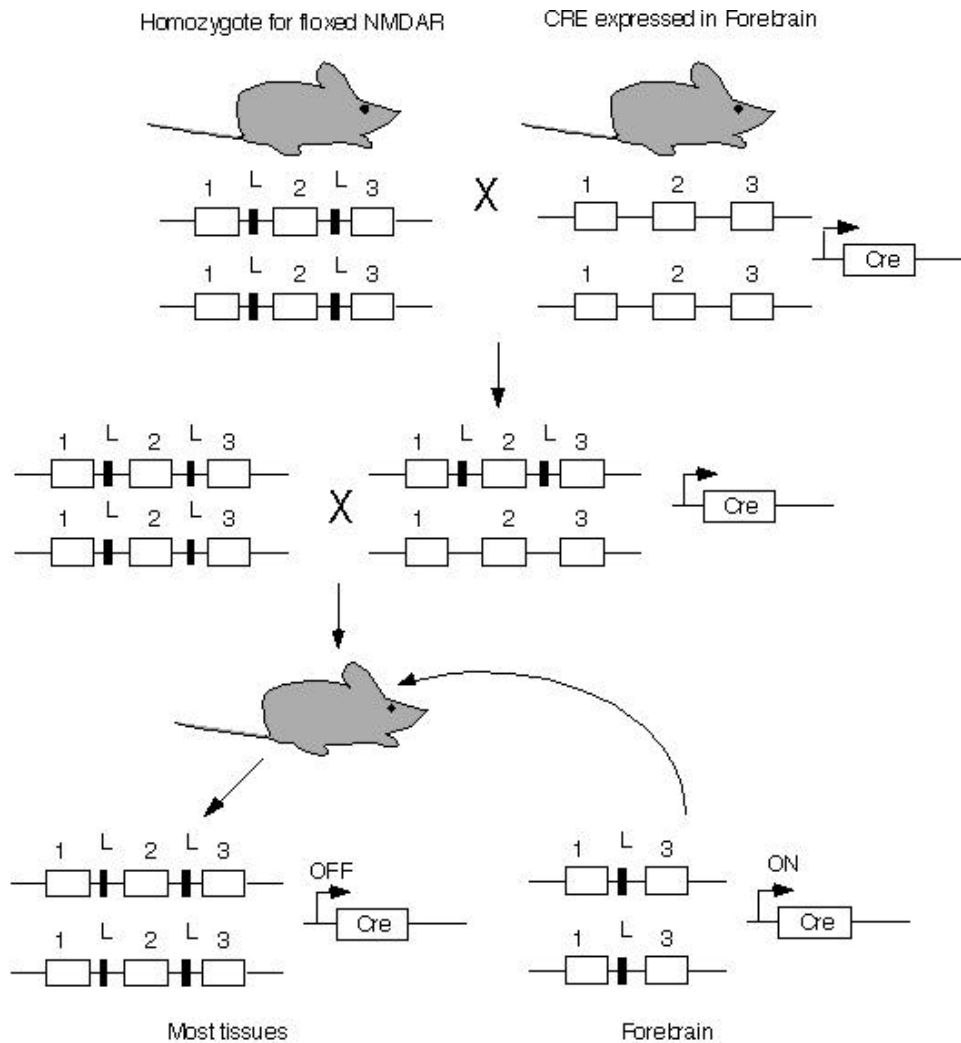
Figure 8-6. Tissue-specific loss of the NMDAR gene. Lox sites were placed in the introns between exons 1 and 2 and between exons 2 and 3 of NMDAR.

Transposon mutagenesis

Ironically, some of the organisms that have been chosen as good systems for genetic analysis are not difficult if not impossible to use for gene replacements by recombination. For flies, worms, and fish, it is possible to get DNA into the organism by microinjection, but hard to find any products of homologous recombination. As discussed above, whole organisms make it impractical to use selections to eliminate background, and the numbers of transformants are too low to isolate rare events.

One alternative approach to making knockouts is to search for transposon insertions in the gene of interest. Transposons are a class of transposable element, DNA sequences that use a variety of mechanisms to move or copy themselves to new chromosomal locations, a process called transposition. Transposons have been found in a wide variety of organisms ranging from bacteria to yeast to mammals and plants (Table 8-1)

Table 8-1. Some transposable elements from different organisms		
Organism	Transposon	Comments
E. coli	Tn3	Transposition leads to formation of a cointegrate, which is resolved be a resolvase
	Tn5	Carries a gene for kanamycin resistance; used in transposon mutagenesis in many other bacteria.
	Tn7	Has both site-specific and random integration mechanisms. The random pathway is one of the most random known. *In vitro* transposition system available as a commercial kit.
	Tn10	Widely used in *E. coli* genetics.
	Mu	Phage that uses transposition as a mechanism for replication. First *in vitro* transposition system.
S. cerevisiae	Ty	Retrotransposon
Drosophila	P element	Retrotransposon responsible for hybrid dysgenesis. P elements brought into an egg from a strain lacking P are activated for transposition. High levels of insertions can be lethal. Widely used as a tag in *Drosophila* genetics.
	copia	
	mariner	
mammals	LINE	
	retroviruses	Integration of cDNA intermediate is mechanistically similar to transposition.
Plants	Ac-Ds	Some of the first transposable elements ever discovered. Found by Barbara McClintock in maize. Transposition of Ds depends on transposase synthesized in trans from Ac.

Transposon insertions are not as clean as deletions or deletion-replacements because there are instances of insertion alleles retaining some gene function. Nevertheless, transposon insertions are much better than nothing.

Transposon mutagenesis cannot be targeted in the same sense that homologous recombination is targeted. The idea in transposon mutagenesis is that if you look at enough transposon insertions, sooner or later one will hit the gene of interest; the question then is how to find the right one. When using transposons as a mutagen in a classical mutant hunt, one looks for insertion mutants that have a desired phenotype after whatever inbreeding steps are needed to make the mutation homozygous. One can then use markers in the transposon to clone the insertion allele and identify the affected gene.

If you want to use transposon mutagenesis to determine the phenotype of knocking out a gene you already have cloned, the phenotype you will be screening for initially is simply the presence of the transposon in the gene. In a diploid organism, the initial insertion will usually have no phenotype because it will be heterozygous and recessive. In addition, although transposons move more frequently than other DNA sequences, transposition into the gene of interest will still be a rare event. A transposon insertion in the desired gene can be detected by PCR using one primer that corresponds to an internal sequence in the transposon, and another primer that correspondds to part of the gene of interest. Amplification will only occur if both primers hybridize near one another. This can only happen if the transposon inserts near the site where the gene-specific primer binds. However, screening a population that contains transposon integrations all over the genome one at a time will be impractical, since the transposition events are random and the target is small. How does one find the desired insertion?

It is possible to find individuals with a specific transposon insertion by a screening method analogous to bulk segregants analysis. **Sib selection** allows testing pools of individuals from a population containing different insertion mutations (Figure 8-7). To see how this works, we will set up an example using *C. elegans.* Imagine you have an initial pool where 1 in 10^4 worms have insertion near your gene and the total population size is $>10^4$. In a sample of 5 x 10^4 worms, the average number of worms per sample will be 5. Suppose we divide this sample into 10

subpopulations, where each pool gets 5×10^3 worms. The average number of mutants per pool is

0.5, but you can't have half a worm in each subpool. The distribution of worms in pools can be

predicted using the Poisson equation; some will get no mutants, some will get one and some might get

more than one. On average, from the Poisson distribution 6 pools will have no desired insertions and
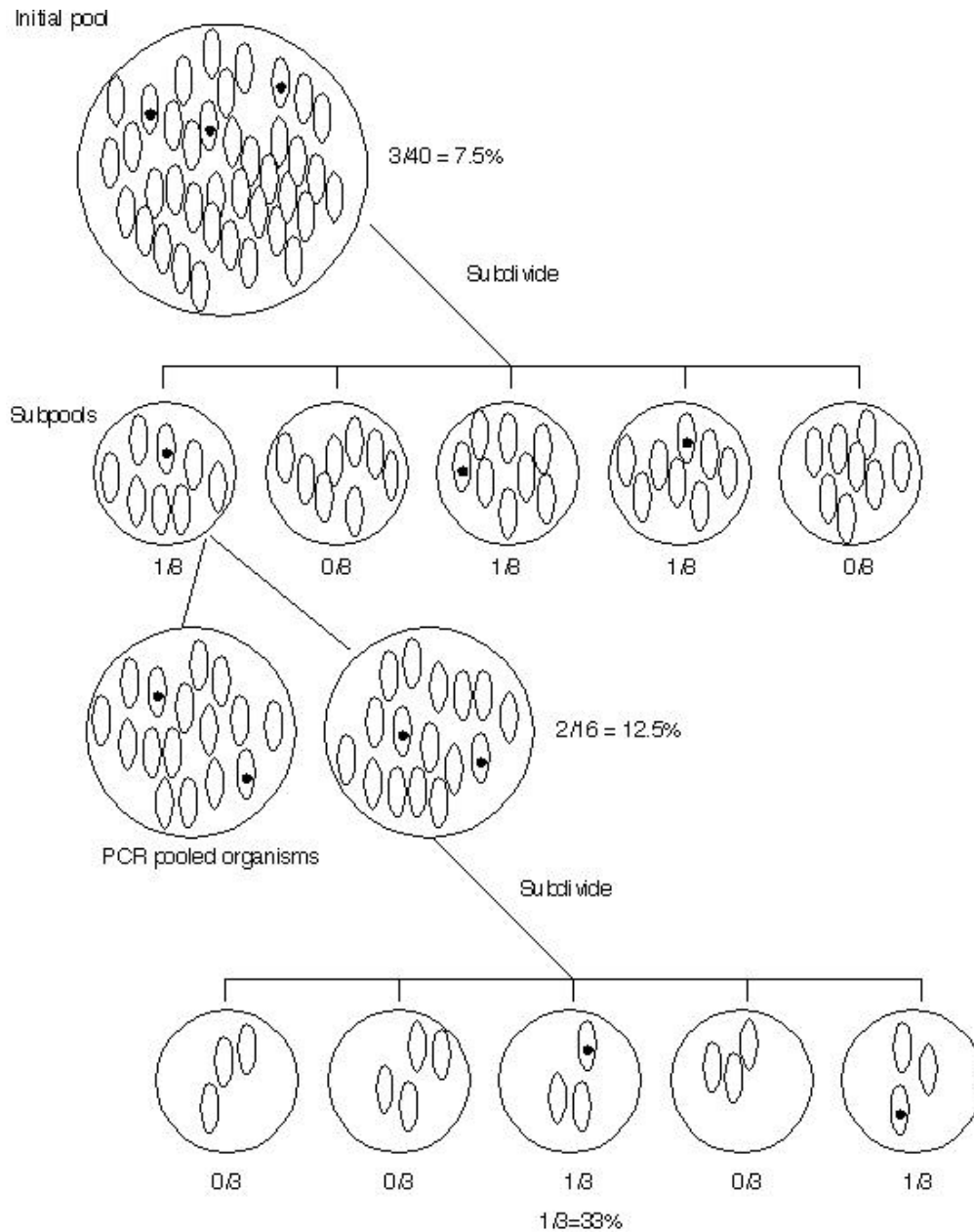
4 will have ≥ 1.

Figure 8-7. Sib selection. Individuals with the desired insertion mutation are indicated by a black dot.

Each pool is allowed to grow up to a large population size, so that if there was one mutant in the original subpool, there are now many clones of that mutant. If there were no mutants in the pool, replication will not create any new mutations. DNA from part of the subpool is extracted and used as a template to assay for the presence of the desired mutation.

In the subpools that give a positive result, the fraction of insertions is $1/5000 = 2 \times 10^{-4}$; the fraction of individuals in this popultation with the desired mutation has been enriched 2X relative to the starting pool. This pool can be subdivided again and the cycle can be repeated. At each step the fraction of mutants is enriched 2X. As the cycles are repeated, the fraction of the population that is mutant increases exponentially. Note that the fold enrichment at each step can be manipulated by how many subpools are used at each round.

Genomic footprinting

David Botstein's lab has developed a method to use transposon mutagenesis and PCR to analyze the consequences of knocking out many genes in parallel, which they call genomic footprinting. The idea is to start with a population of thousands cells, each containing a different insertions. Gene specific primers and transposon specific primers are used to amplify DNA from the pool to generate a pattern of PCR products that reports what insertions are present.

Then subject the population to selection for many generations; any insertion that reduces fitness is lost. By examining the PCR products from the original population that go away, they can find identify those genes that are important in that selective condition.

(XX develop this better next year!)