Genomic Manipulation

Aims:
Understand why genome manipulation is useful
Understand what instruments are available to scientists and which instrument is most appropriate to each question
Be able to discuss pros and cons of each techniques
Understand what is the direction the field is taking
From outside readings:
Be able to provide some real-life examples of applications of discussed techniques

Genetic manipulation in vivo:

a) interfering with a random gene (unbiased)
b) interfering with a specific gene (ad hoc approach)

"Interfering" includes:

modifying the expression level of a gene (under / over expression)
modifying the expression pattern of a gene (in time and space)
modifying the protein product of a gene (CDS mutations)

Target a specific gene: when you know the what the gene is, go from gene to phenotype
Target a random gene: do not know the gene and perform a random unbiased screening, go from phenotype to gene

a) interfering with a random gene (unbiased)

Random mutagenesis screening using chemical mutagens
Random mutagenesis screening using transposons
Large scale RNAi screening (to be covered in the next lectures)

b) interfering with a specific gene (ad hoc)

Exogenous expression of a gene via transgenesis
Targeted disruption of a gene via HR (knock out)
Targeted modification of a gene via HR (knock in)
RNAi

New frontiers of genetic modification

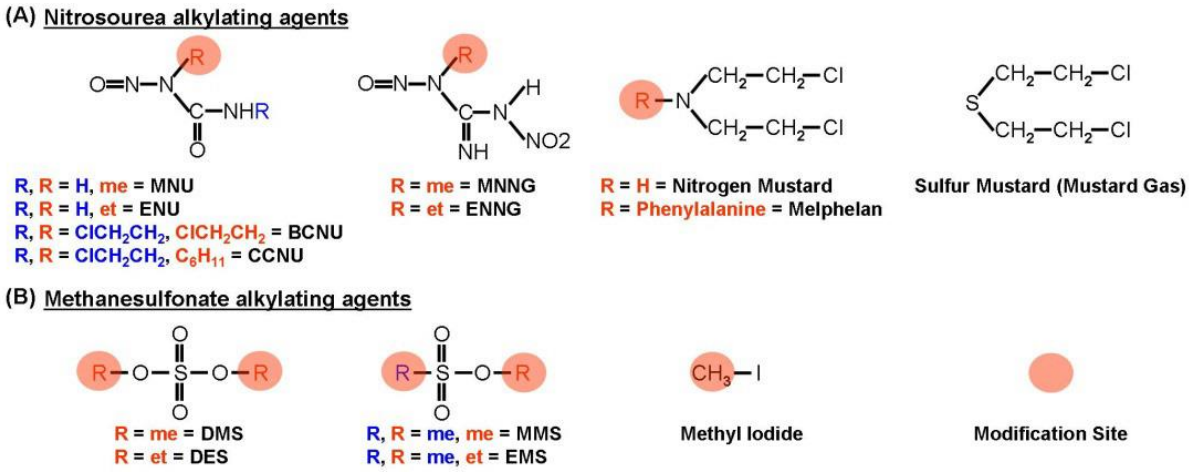
Zinc fingers
TALENs
CRISPR

Targeting random genes: genetic screening

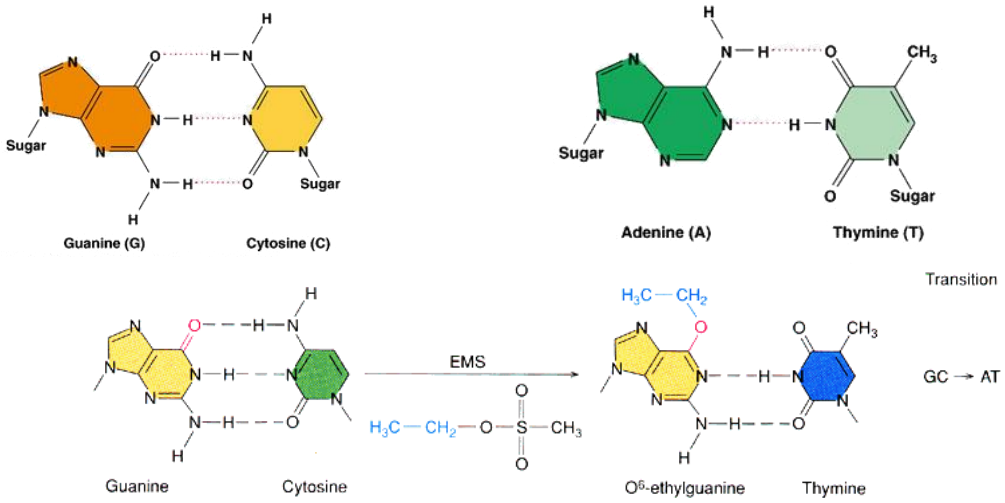
Start from WT or mutant phenotype and end up to modified phenotype. Must have a phenotype you are interested in. (e.g. eyeless mutant → open the door to master gene regulators, a single gene that is responsible for the development of an entire pattern→ genes ca work in a pyramidal way, start with master regulators and then start cascade of signals that regulate expression patterns). They also found some genes are sufficient (antennopedia).

It is possible to mutate randomly by:

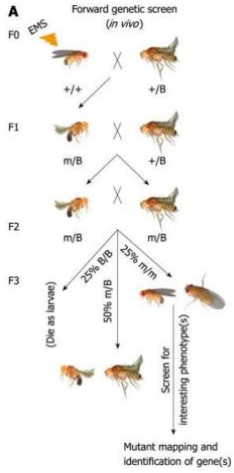
Chemical Mutagenesis



Historically to random mutate was using chemicals (EMS)--> mutagens.
EMS acts on the guanine; change transition from GC to AT pairing. G becomes an A first (A-C pairing forms and so repairing machine intervenes and fix this mess and removes C and replace a T; sometimes there is also complete removal of the bp or also to neighbouring sites)→ deletions are good for random screening.

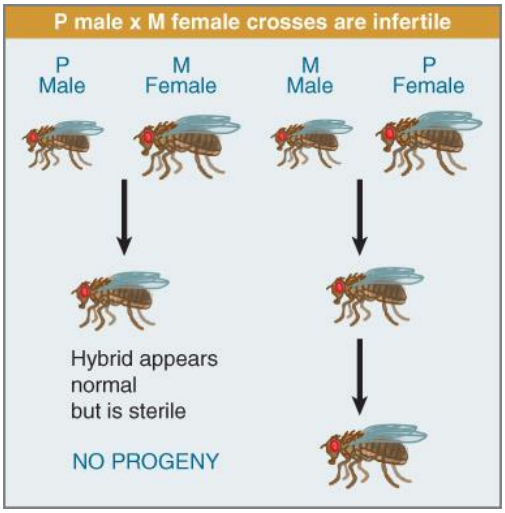


Tighter drug→ enough drug to have mutations but not too much to kill the progeny. Put drug in the food. And then go fro 3 generations (in worms 2 generations are enough). And then you obtain the mutants. For this kind of work you like recessive mutants. You also want to screen 10000-20000 mutants (stocks) per chromosome. It is a lot of work.
Weakness of this system is that you may disrupt genes that are responsible for the development of the organism.

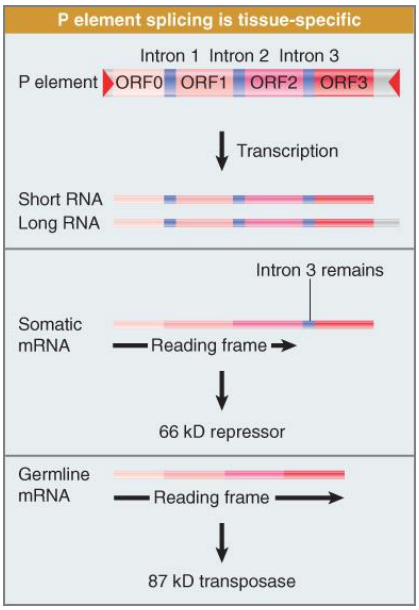


Transposon based random mutagenesis

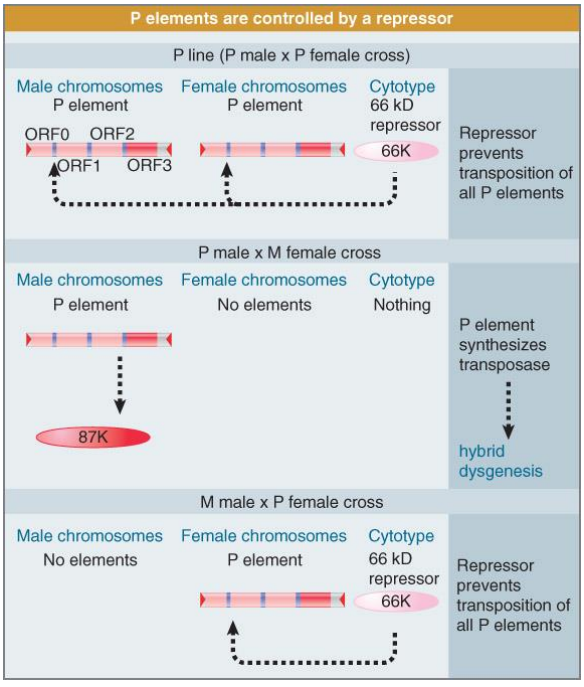
Chemical mutagens they are efficiently crating mutants but good luck in finding where the mutation is (need to do genetic mapping)→ cross flies , look for combinations and visible markers; the problem however is that it is really time consuming to map genes. Random screens looking for mutagens have many drawback: takes a lot to see what the gene is.
A solution is to create mutations through transposable element.



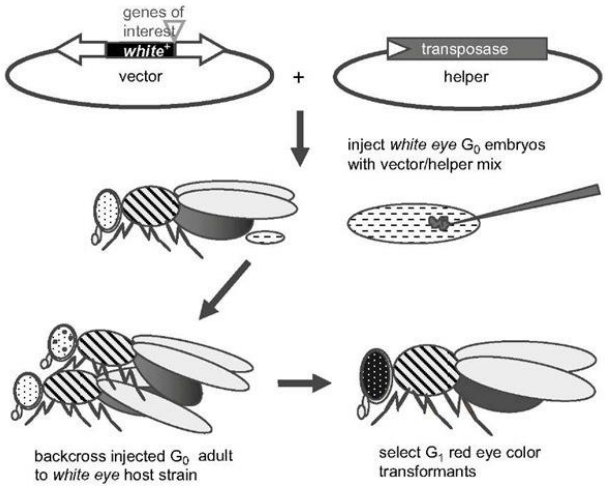
- P strains are found in the wild
- M strains are found in the lab
- Lab strains were initially isolated in 1905 and propagated since then
- In the meanwhile, after 1905, P element transposon have colonized virtually every Drosophila in the wild
- Crossing P flies to M flies results in sex selected hybrid dysgenesis
- The gene for the transposon can also encode for its own repressor via a mechanism of alternative splicing
- The alternative splicing is sex specific



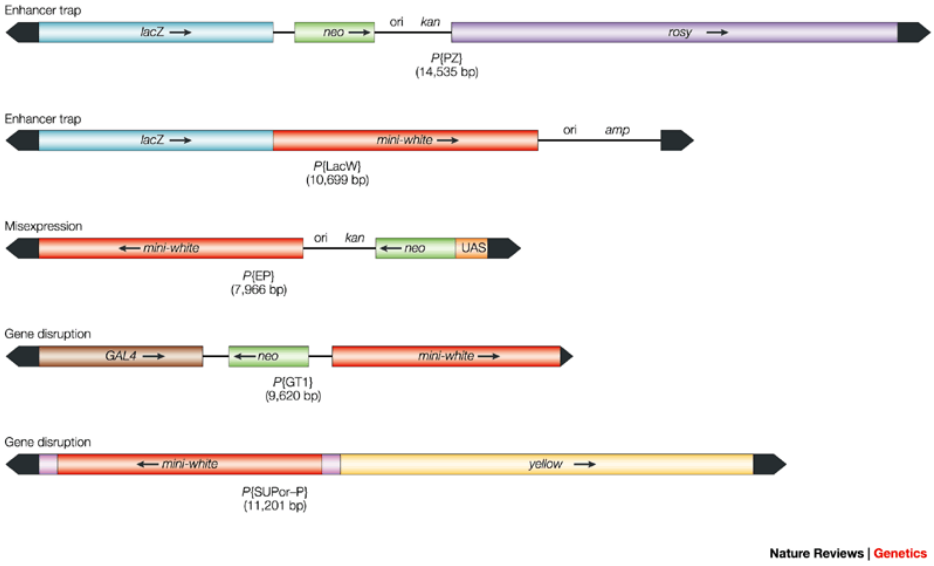
Bottom line is that the P elements is a single gene that has flanking regions and composed of 4 exons. Differential splicing of the exons will produce 2 proteins. If 2 3 exon spliced together get the actual transposase (bind LTR of flanking regions and make element jump around). If you don't have full splicing happening, you get a repressor (no jumping occur). Jumping of the P element depends on transposase. Then scientist engineer 2 pieces. One plasmid encodes for the transpose only (helper plasmid). Inject this in embryo and then at the same time inject other plasmid (with flanking regions but not P element but gene of choice)→ transposase recognise flanking regions and will integrate target gene randomly in the host genome.



Transposable element will be integrated randomly in the genome. Produce fully transgenic animals, if transposon end in gametes. If the vector inserts into a gene it will disrupt the function of a gene.



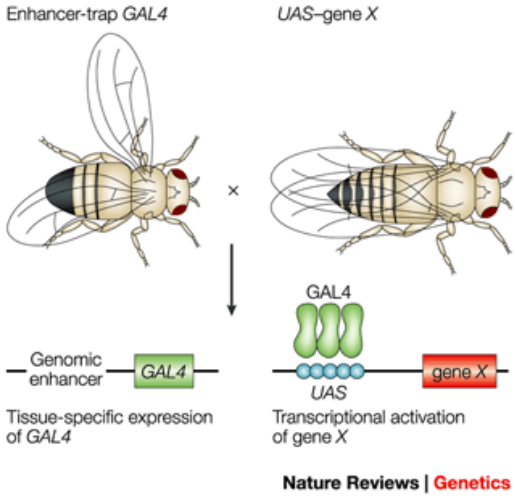
Use of P-elements in Drosophila Genetics



Mark expression of the gene depending on the landing site (trap). Misexpression : insert Uas
Miniwhite is a gene that changes the colour of the eye→ inject the of a fly that has white eye color →if transgenesis is successful the miniwhite will change colour of the eye from white to red.
UAS and Gal4 are two most commonly used in Drosophila genetics. Systems set up in the 80s.

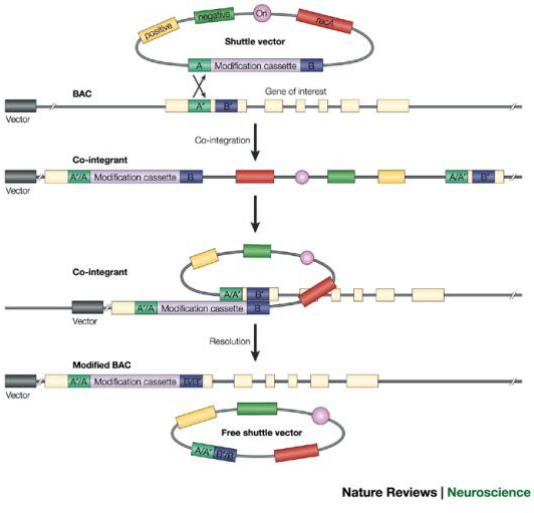
Transgenic *Drosophila:* The GAL4 UAS system

Transfer transcriptional regulator from yeast and put into drosophila. GAL4 is a gene and UAS is a dna sequence, recipient for GAL4 binding. GAL4 binds UAS sequence and activate transcription of what is downstream of UAS. If you pick a genomic enhancer that is expressed only in the eye, then have GAL4 only in the eye (with trapping system). Create libraries of thousands of many gal4, each gal4 for every gene. UAS is another transgenic fly, sometimes generate in another lab. Downstream of UAS there is a gene of interest. With UAS fly need to mate it with GAL4. After 12 days you get progeny of the cross, children will have gal4 and UAS and those fly you do the experiments. Whenever gal4 is expressed it will drive transcription of the downstream UAS.



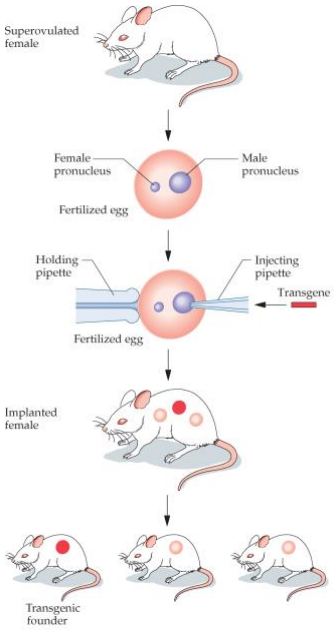
- P-elements insert themselves in Drosophila genome
- They can be hijacked by scientists so that will also insert our gene of interest
- Transposase must be present at moment of injection
- Transposase must NOT be present after injection
- Two independent plasmids are injected in the Drosophila embryo, one encoding for the transposase and the other having the gene of interest flanked by recognition sequences
- Gene of interest will then insert itself and progeny will inherit it
- A stable transgenic is made
- Transgenic individuals can be recognized by a collateral dominant marker (usually mini white gene conferring red eyes)

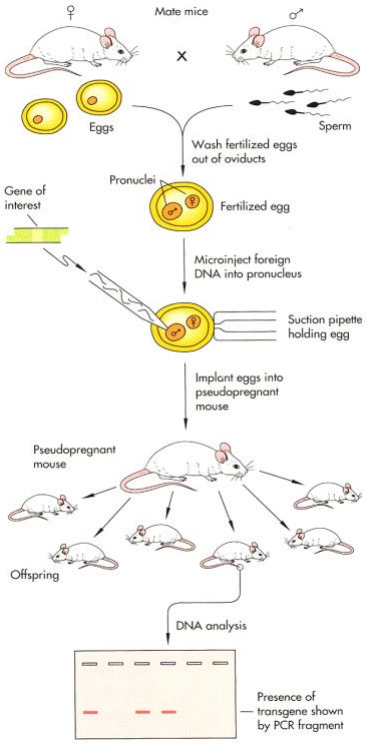
- Outside reading:
  - AttP/attB insertion are the new frontier in terms of transgenesis (phiC31 system)



Targeting Specific Genes

⇒ Transgenic Mice

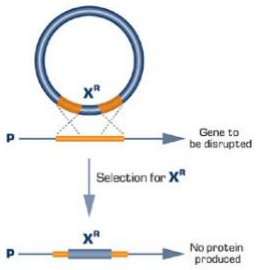




- Most commonly used method
- Only 5% or less of the treated eggs become transgenic progeny
- Need to check mouse pups for DNA (by PCR or Southerns), RNA (by northerns or RT-PCR), and protein (by western or by some specific assay method)
- Expression will vary in transgenic offspring: due to position effect and copy number

Gene Targeting by Homologous Recombination

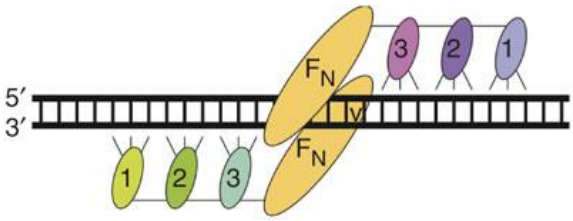
- Exploits mechanism of DNA repair that are endogenous to the cell
- Works on many organisms, from bacteria to mice
- Requires access to a "stem cell" that will become an embryo
- HR can be used to create knock ins and knock outs
- It is the most important genetic procedure in mice genetics
- Can target virtually any kind of DNA (e.g. BAC HR)
- Because it's a mechanism of cell repair, can be improved by artificially coupling a DNA break to it



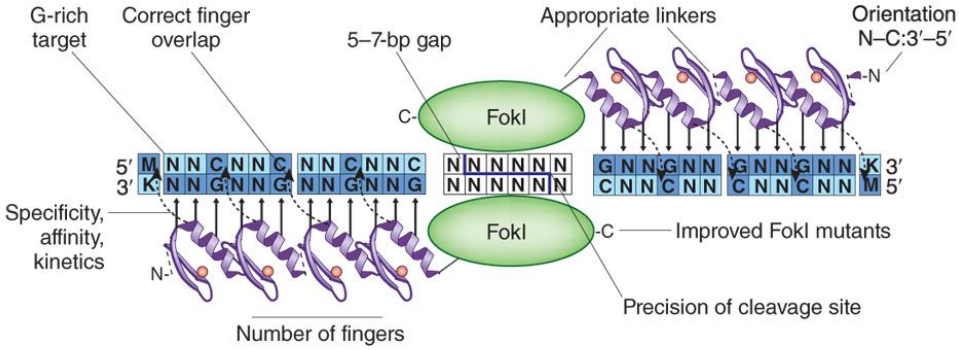
Making transgenic allows us to express sequences to study the function of genes. Often however you want to remove something. Use homologous recombination. Homologous recombination is the natural machinery exploited to repair DNA damage. Create two homology arms (orange), identical to the flanking regions, serve as homology target. In the middle put the gene of interest or disrupt the gene of interest. Here put antibiotic resistance so you can at the same time disrupt gene function and mark the transgenic→ to artificially select only the cells where this event happened.

Gene targeting by HR - Improved by ZFN, TALEN, CRISPR

Zinc fingers:
Zinc finger nucleases are proteins that couple zing domain with nuclease activity. Use zn to bind dna specifically. To obtain ds breaks need two zing finger proteins.
TALEN selectively get proteins domains that binds dna sequences. TALEN works like a code (modular).



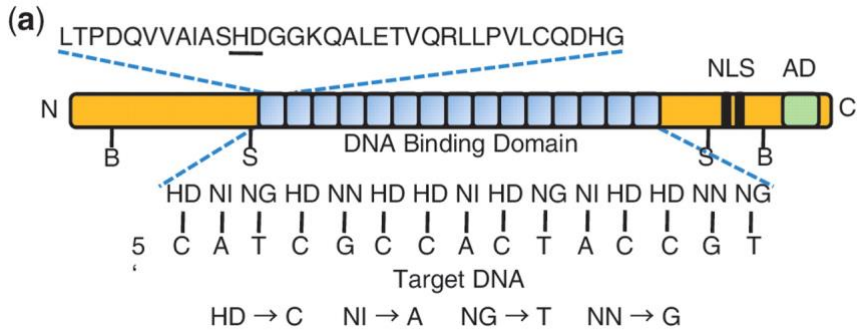
ZFPs functional as a pair



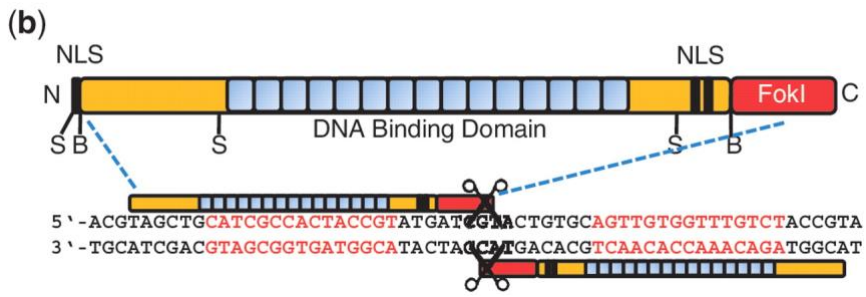
Unique DNA target in genome, no related sites, accessible chromatin

TALEN:

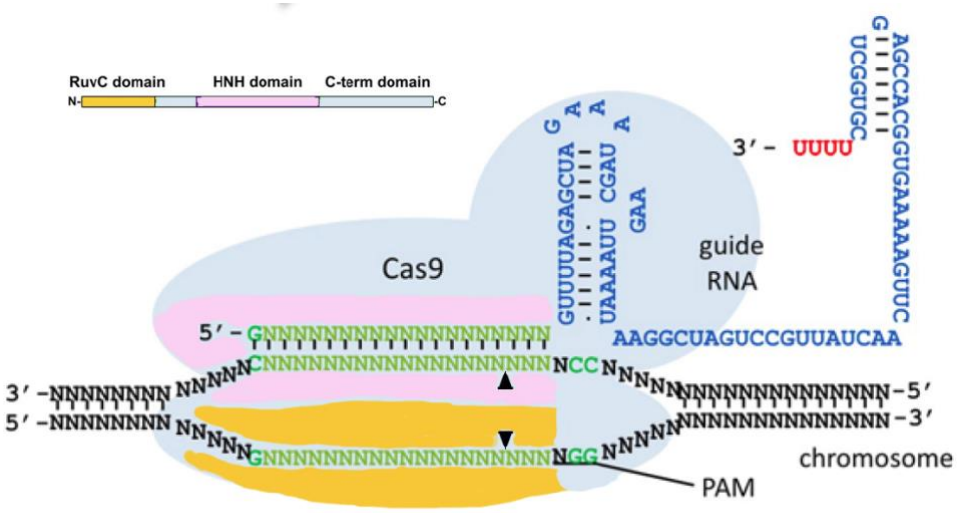
Transcription activator-like effector nucleases (TALEN) are restriction enzymes that can be engineered to cut specific sequences of DNA. They are made by fusing a TAL effector DNA-binding domain to a DNA cleavage domain (a nuclease which cuts DNA strands). Transcription activator-like effectors (TALEs) can be engineered to bind to practically any desired DNA sequence, so when combined with a nuclease, DNA can be cut at specific locations.



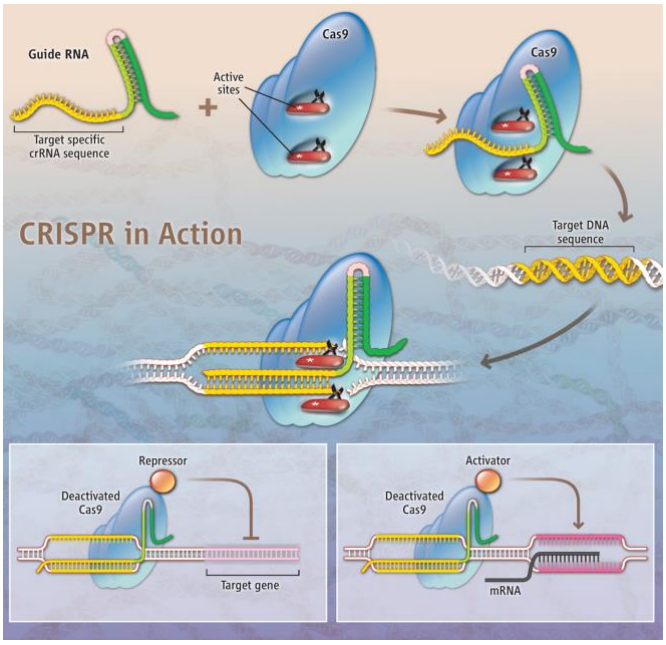
CRISPR



The key step in editing an organism's genome is selective targeting of a specific sequence of DNA. Two biological macromolecules, the Cas9 protein and guide RNA, interact to form a complex that can identify target sequences with high selectivity.
The Cas9 protein is responsible for locating and cleaving target DNA, both in natural and in artificial CRISPR/Cas systems. The Cas9 protein has six domains, REC I, REC II, Bridge Helix, PAM Interacting, HNH and RuvC (Jinek et al. 2014; Nishimasu et al. 2014).
The Rec I domain is the largest and is responsible for binding guide RNA. The role of the REC II domain is not yet well understood. The arginine-rich bridge helix is crucial for initiating cleavage activity upon binding of target DNA (Nishimasu et al. 2014). The PAM-Interacting domain confers PAM specificity and is therefore responsible for initiating binding to target DNA (Anders et al. 2014; Jinek et al. 2014; Nishimasu et al. 2014; Sternberg et al. 2014). The HNH and RuvC domains are nuclease domains that cut single-stranded DNA. They are highly homologous to HNH and RuvC domains found in other proteins (Jinek et al. 2014; Nishimasu et al. 2014).



Must design something that is adjacent to a PAM motif: usually a GG motif (NGG consensus). Preferentially end of of guide RNA need to be GC bp so it is more stable. If no PAM no CUT. Cas9 has two nuclease domains, each one cuts a strand. Can engineer a cas9 that cuts just one strand and not the other: can create a nuclease that is nickase (one strand cuts). To increase specificity of the action.



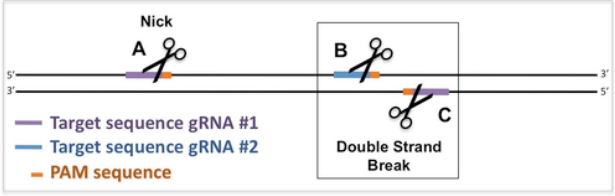
The Cas9 protein remains inactive in the absence of guide RNA (Jinek et

al. 2014). In engineered CRISPR systems, guide RNA is comprised of a single strand of RNA that forms a T-shape comprised of one tetraloop and two or three stem loops (Jinek et al. 2012; Nishimasu et al. 2014). The guide RNA is engineered to have a 5′ end that is complimentary to the target DNA sequence.
This artificial guide RNA binds to the Cas9 protein and, upon binding, induces a conformational change in the protein. The conformational change converts the inactive protein into its active form. The mechanism of the conformational change is not completely understood, but Jinek and colleagues hypothesize that steric interactions or weak binding between protein side chains and RNA bases may induce the change (Jinek et al. 2014).
Once the Cas9 protein is activated, it stochastically searches for target DNA by binding with sequences that match its protospacer adjacent motif (PAM) sequence (Sternberg et al. 2014). A PAM is a two- or three-base sequence located within one nucleotide downstream of the region complementary to the guide RNA. PAMs have been identified in all CRISPR systems, and the specific nucleotides that define PAMs are specific to the particular category of CRISPR system (Mojica et al. 2009). The PAM in *Streptococcus pyogenes* is 5′-NGG-3′ (Jinek et al. 2012). When the Cas9 protein finds a potential target sequence with the appropriate PAM, the protein will melt the bases immediately upstream of the PAM and pair them with the complementary region on the guide RNA (Sternberg et al. 2014). If the complementary region and the target region pair properly, the RuvC and HNH nuclease domains will cut the target DNA after the third nucleotide base upstream of the PAM (Anders et al. 2014).

Double Nicking Technique

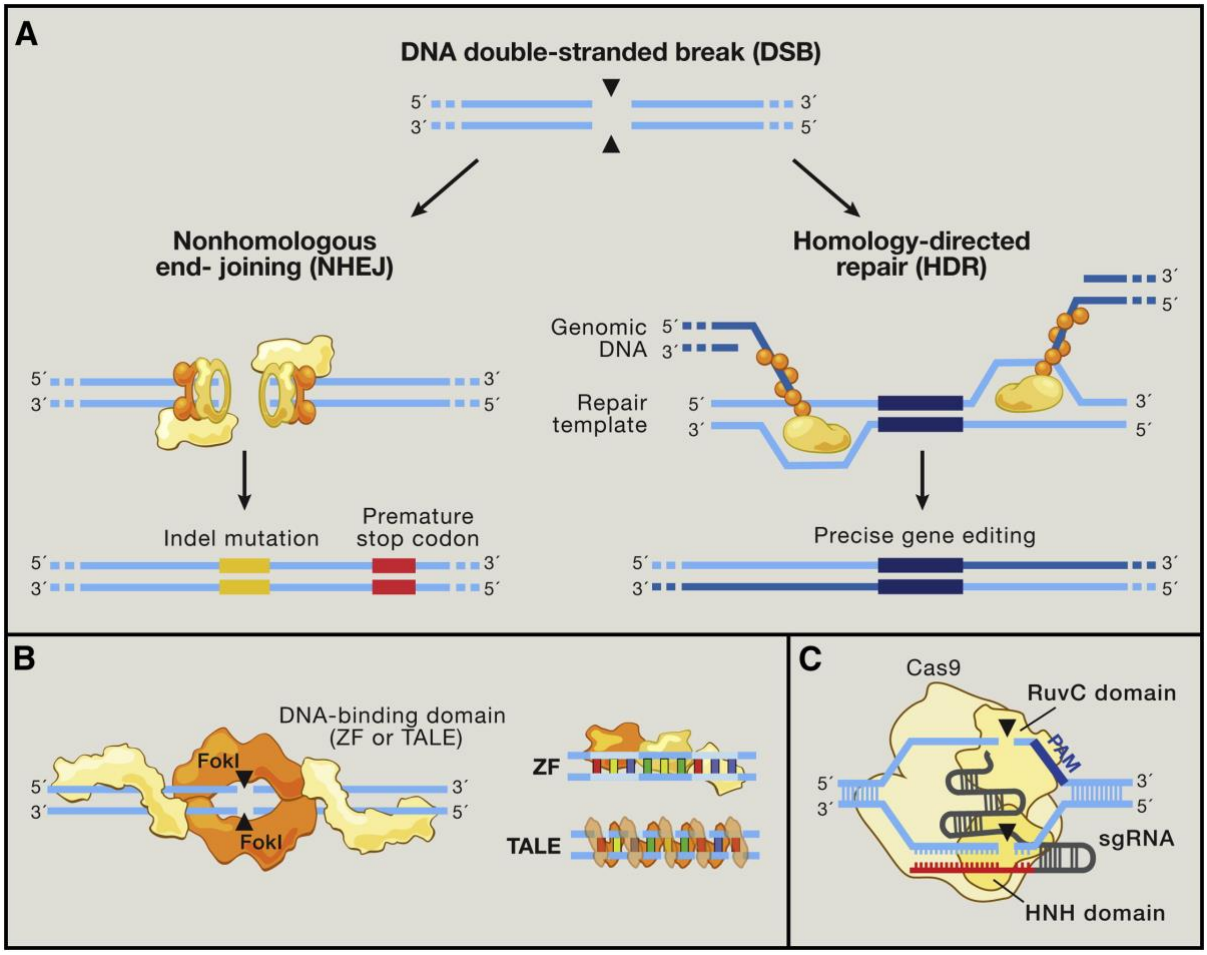
Targeted genome editing technologies have enabled a broad range of research and medical applications. The Cas9 nuclease from the microbial CRISPR-Cas system is targeted to specific genomic loci by a 20 nt guide sequence, which can tolerate certain mis- matches to the DNA target and thereby promote undesired off-target mutagenesis. Here, we describe an approach that combines a Cas9 nickase mutant with paired guide RNAs to introduce targeted dou- ble-strand breaks. Because individual nicks in the genome are repaired with high fidelity, simultaneous nicking via appropriately offset guide RNAs is required for double-stranded breaks and extends the number of specifically recognized bases for target cleavage. We demonstrate that using paired nicking can reduce off-target activity by 50- to 1,500-fold in cell lines and to facilitate gene knockout in mouse zygotes without sacrificing on-target cleav- age efficiency. This versatile strategy enables a wide variety of genome editing applications that require high specificity.

A single-strand break, or nick, is normally quickly repaired through the HDR pathway, using the intact complementary DNA strand as the template. However, two proximal, opposite strand nicks introduced by a Cas9 nickase are treated as a Double Strand Break (DSB), in what is often referred to as a 'double nick' or 'dual nickase' CRISPR system. A double-nick induced DSB can be repaired by either NHEJ or HDR depending on the desired effect on the gene target. See the sections on Gene Editing (HDR) or Gene Disruption (InDel) for more information.



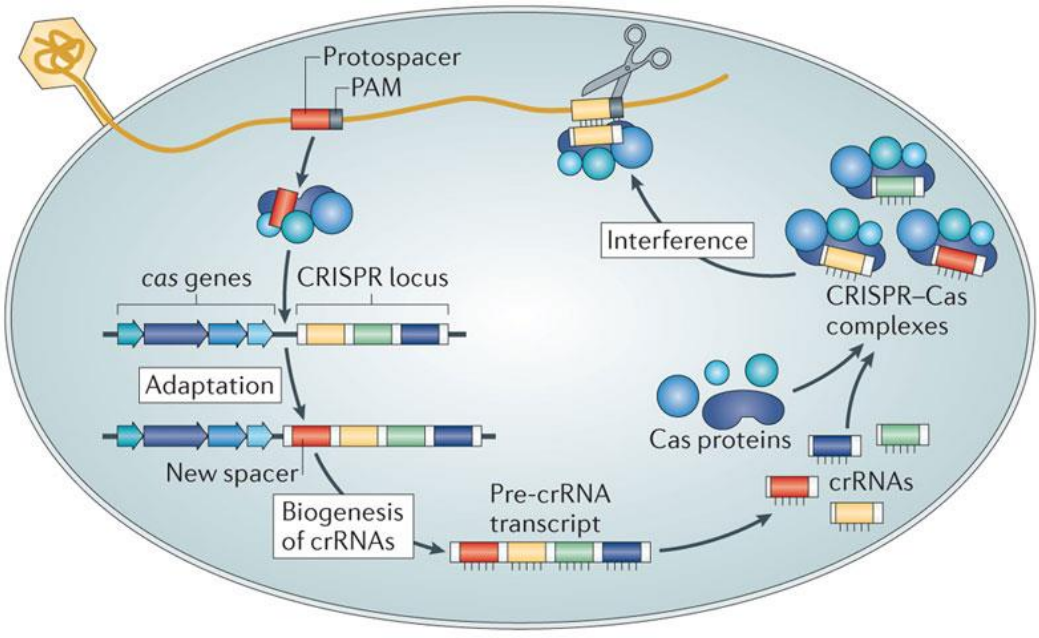
In the figure, two different gRNAs (#1 and #2) bind in a particular genomic region. When gRNA #1 and #2 are co-expressed with a Cas9 nickase, single-strand nicks are created in the DNA at (A), (B) and (C). The nick created at (A) is quickly repaired by HDR using the intact compliment strand as a template and no change occurs. The nicks at (B) and (C) are in close proximity (and on opposite strands) and together behave as a DSB.

If specificity and reduced off-target effects are crucial, consider using the Cas9 nickase to create a double nick-induced DSB. By designing two gRNAs with target sequences in close proximity and on opposite strands of the genomic DNA, off-target effects of either gRNA alone will result in nicks that will not change the genomic DNA. Only at the target location where both nicks are proximal, will the double nicked sequence be considered a DSB. For more information on the double-nickase technique, read Ran et al. *Cell*. 2013 Sep 12. PubMed.

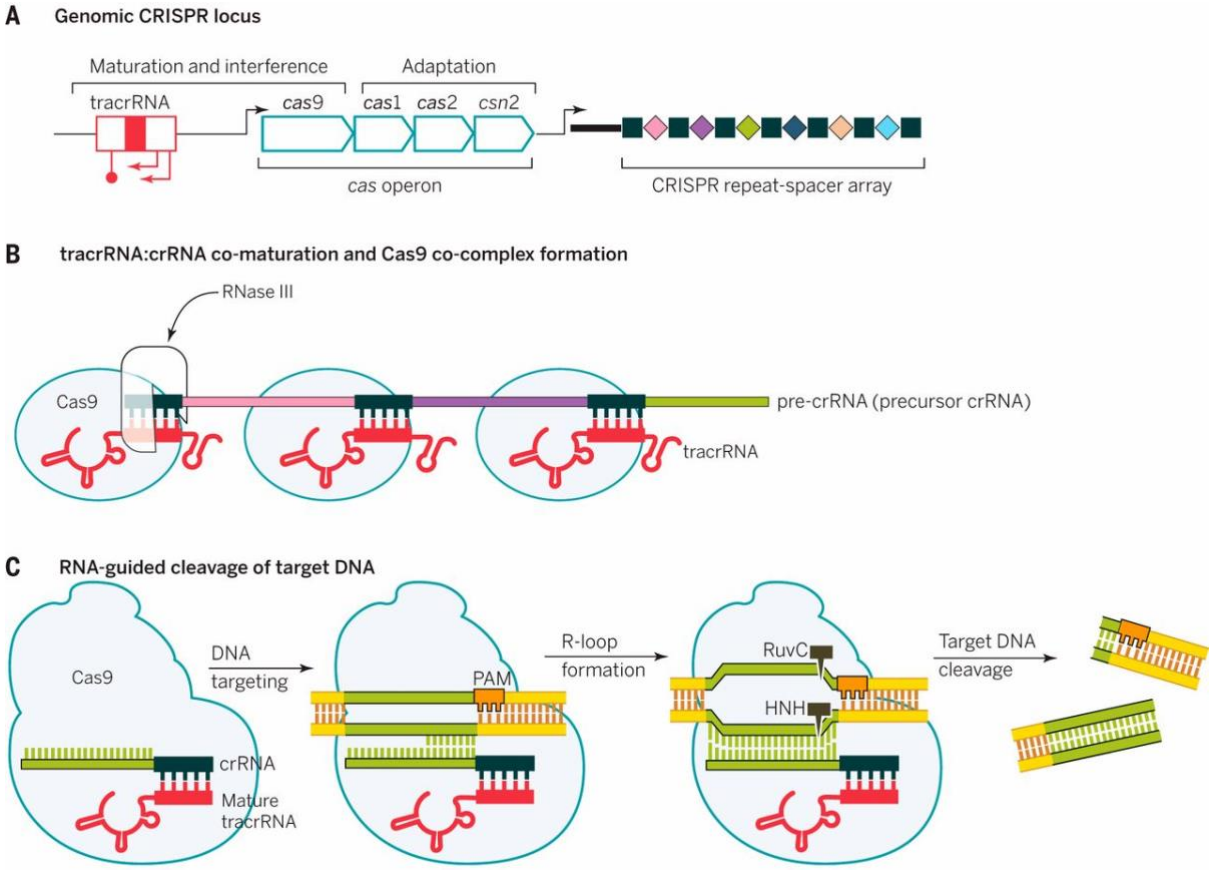


CRISPR is a natural process

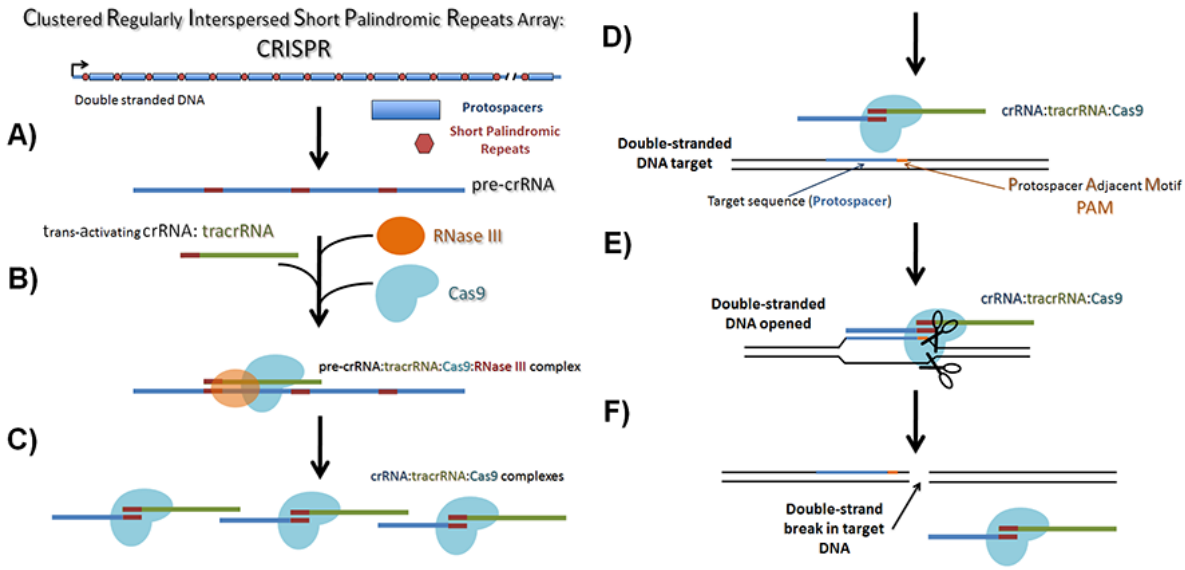
Palindromic repeats are regularly spaced. They are in cluster (belong to same genomic locus). The spacers (coloured rectangles) are interesting because they are the same length but different from each other. Spacer is memory component, elongate the CRISPR locus.



tracerRNA is the component that binds to the repeated regions and creates a complex. In nature the guide rna is not a single molecule but a complex of two molecule (tracerRNA and the spacer+palindromicrepeat)→ 2 things together they create what in the lab is the guide rna.



Protospacer → target sequence on the foreign DNA intruder
PAM → Protospacer Adjacent Motif
Spacer → the bacterial region complementary to the target
Pre-crRNA → a long transcript containing multiple "spacers + short palindromic repeats"
crRNA → the mature pre-crRNA; made of a spacer and a short palindromic repat
TracrRNA → the RNA molecule required for maturation fo pre-crNA
gRNA → an artificial molecule mimicking the crRNA :: tracrNA pairing



Different types of Crispr