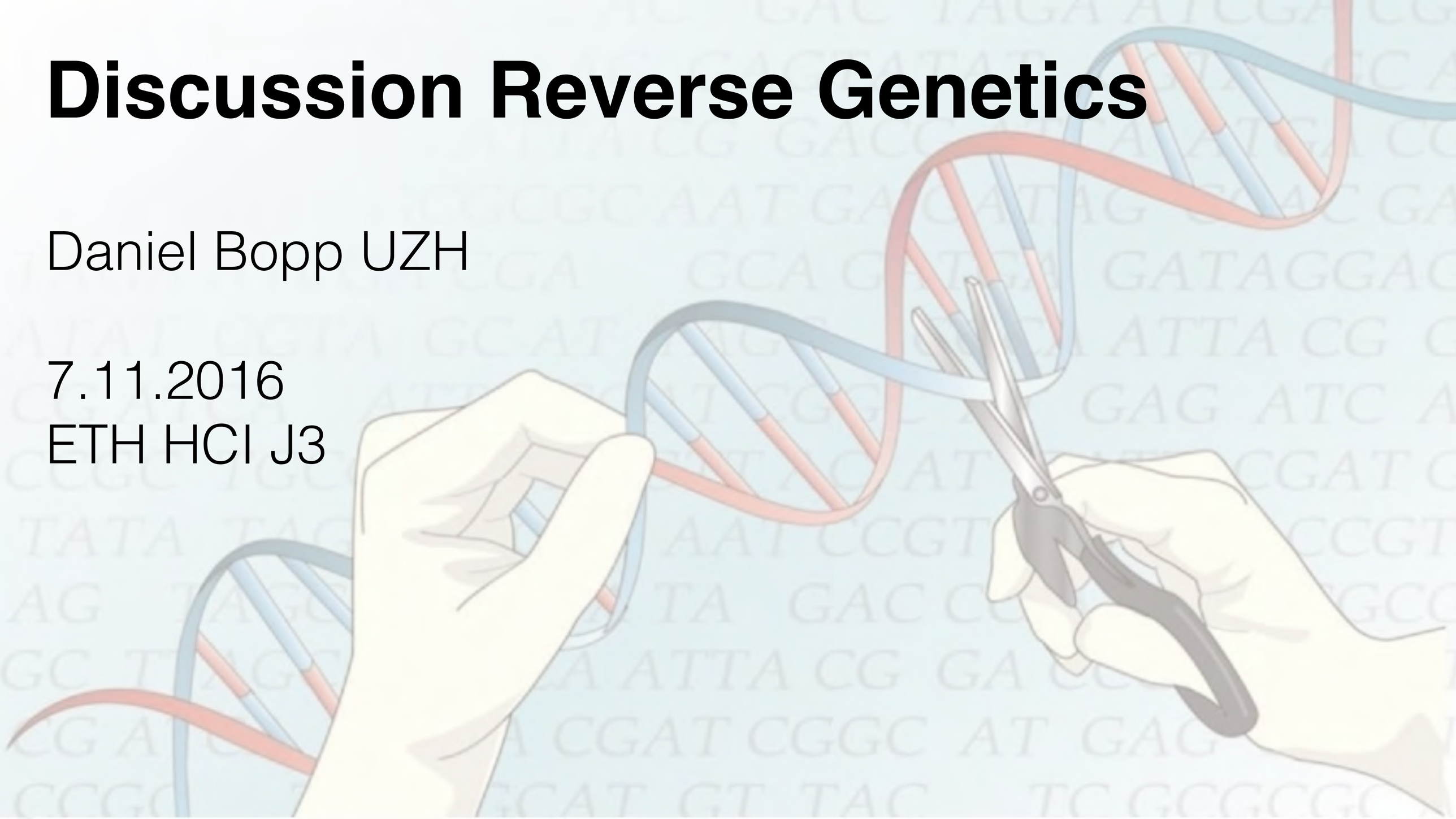
Discussion Reverse Genetics

Daniel Bopp UZH

7.11.2016
ETH HCI J3



concepts

Next Generation Sequencing (different platforms)

Metagenomics

Functional Genomics

Transgenesis (different technologies)

Transcriptional and translational reporter (what is used for?)

Enhancer trap (what is used for?)

overexpression or misexpression studies

somatic recombination

Homologous recombination (positive and negative selection)

¨floxing¨ a gene (example for temporal and spatial control)

Genome editing strategies (different technologies)

Cell autonomy versus non-autonomy (how can this be tested?)

RNAi (mechanism, various applications)

¨old school¨ sequencing

Sanger – Chain-termination method - 1977

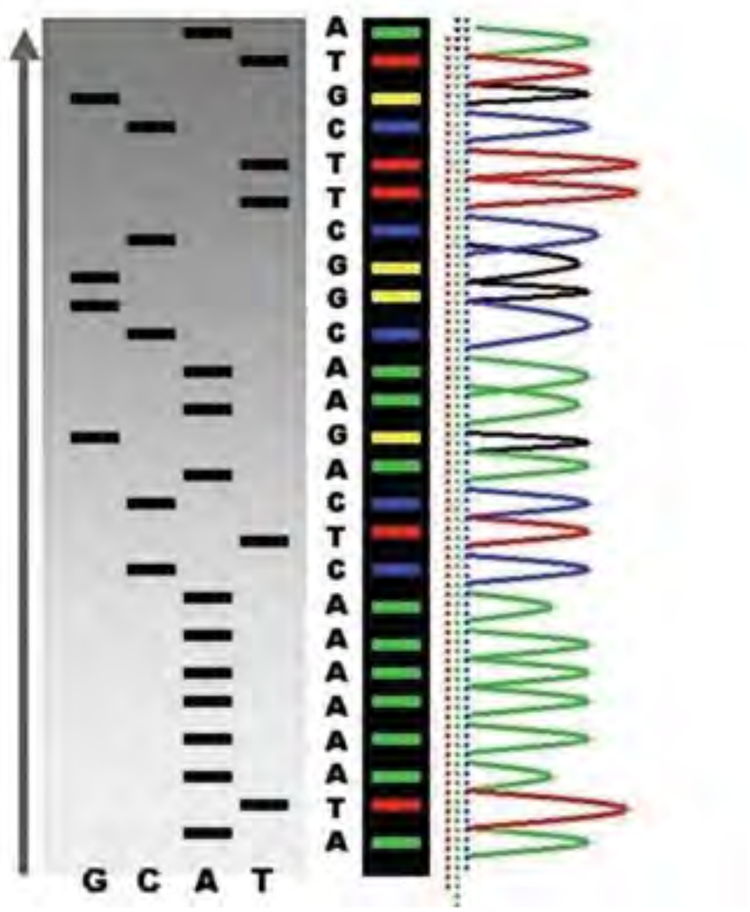
The chain-termination method developed by Fred Sanger became the method of choice.

The method requires:

- a single-stranded DNA template,
- a DNA primer,
- a DNA polymerase,
- radioactively or fluorescently labeled nucleotides, modified dideoxy-nucleotides terminating DNA strand elongation

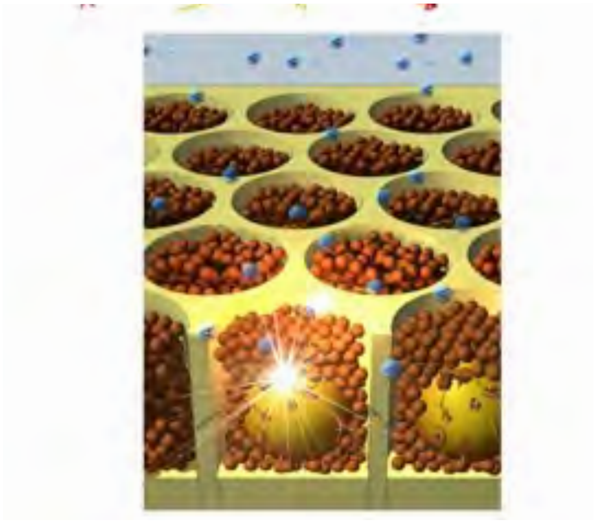
- The result is a mixture of DNA fragments of varying length
- The fragments are separated by size (with a resolution of just one nucleotide) by electrophoresis (gel or capillary).
- The read length is 700-900 bases

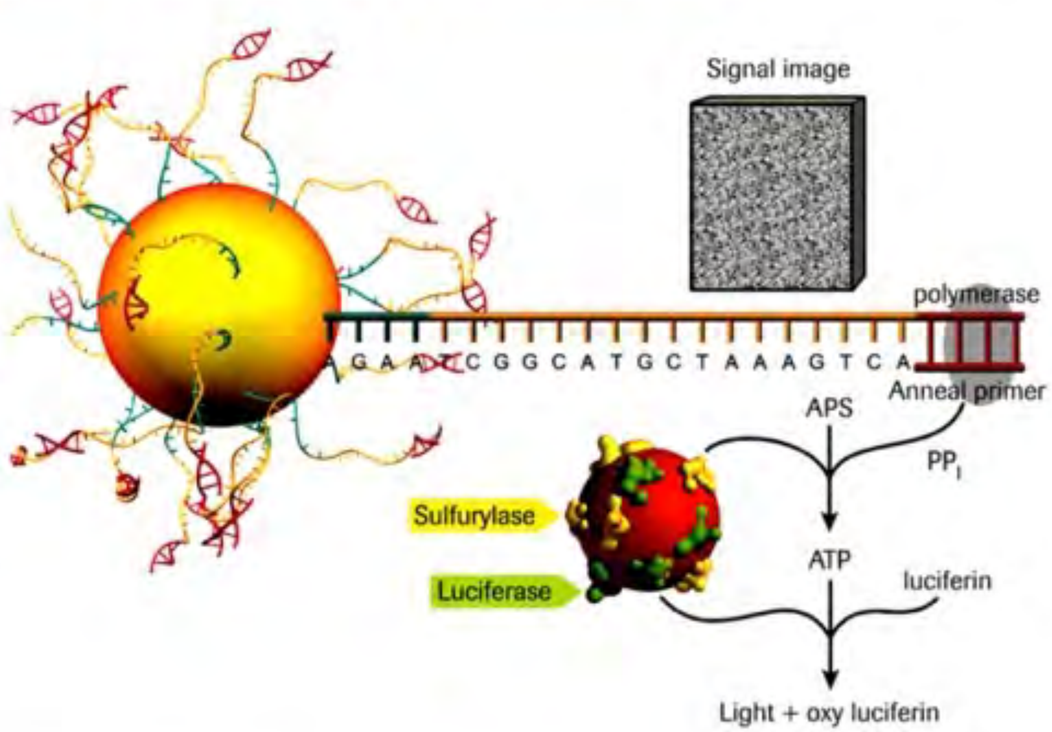
- Human genome (and many others) were sequenced using the Sanger method
- Still gold standard in sequencing



2nd generation sequencing

high throughput by massively parallel sequencing





e.g. Roche 454 System
(pyrosequencing)

Universität Zürich[UZH]

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Sequencing-by-synthesis is the dominant second-generation sequencing

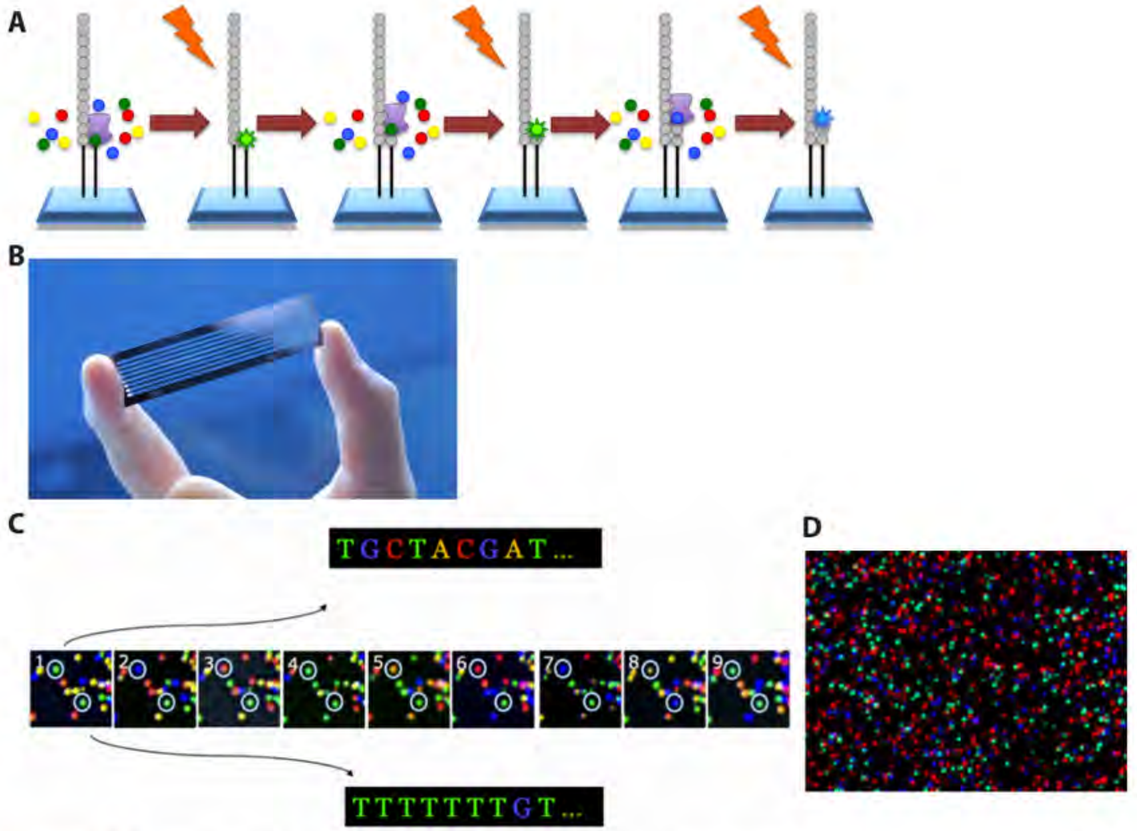


Figure 1-6 Sequencing-by-synthesis. (A) Three cycles of a sequencing-by-synthesis reaction. The polymerase (purple) extends the primer by a single nucleotide. Unincorporated nucleotides are washed away and the fluorescence signal is detected (arrow). The fluorophore and the chain-terminating group are cleaved off and washed away. Then the cycle is repeated. **(B)** Flow cell used in the sequencing-by-synthesis instrument. The DNA to be sequenced is chemically attached to the interior surface of the flow cell's transparent channels. **(C)** Reading a sequence from a series of fluorescence images. Each colored dot corresponds to a DNA sequence of the sequencing library. The images are recorded after each cycle of the sequencing reaction. **(D)** Zoomed up view of a small portion of a flow cell surface. Each dot corresponds to a different DNA molecule undergoing sequencing. The highest performing SBS instruments are able to sequence several billion molecules in a single flow cell.

Illumina HiSeq2000

c-Bot
(automated cluster generati

Cluster density
750-850/mm^2

HiSeq Flow Cells

- 2 flow cells (can be run independently)
- Up to 320Gb mapped sequence per FC
- 64Gb sequence per day (2 flow cells)

3rd Generation

PacBio

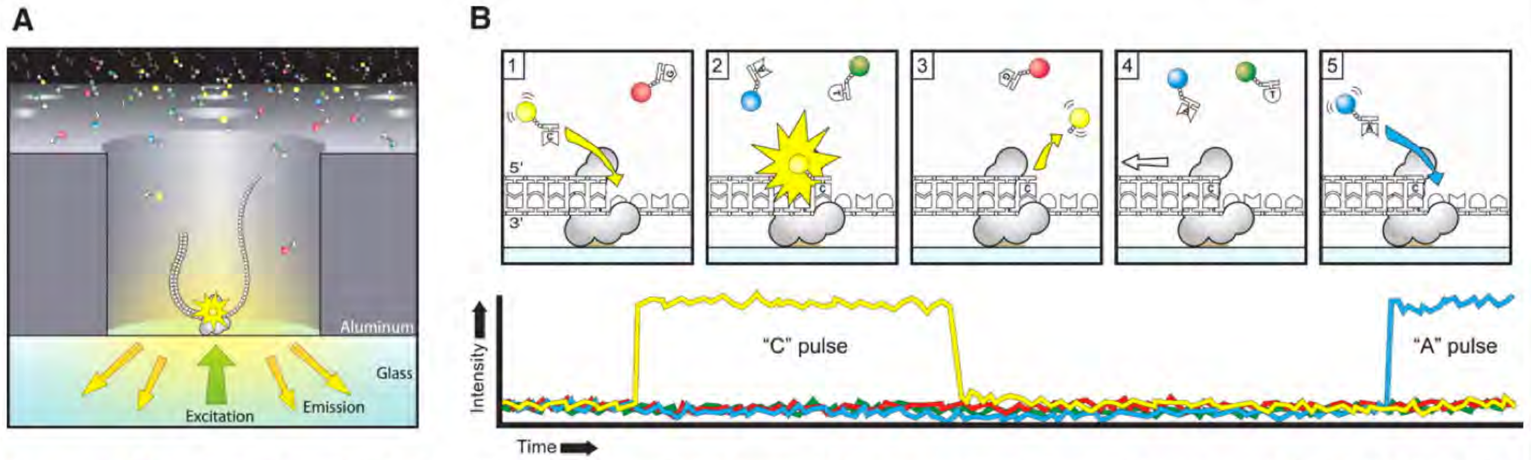
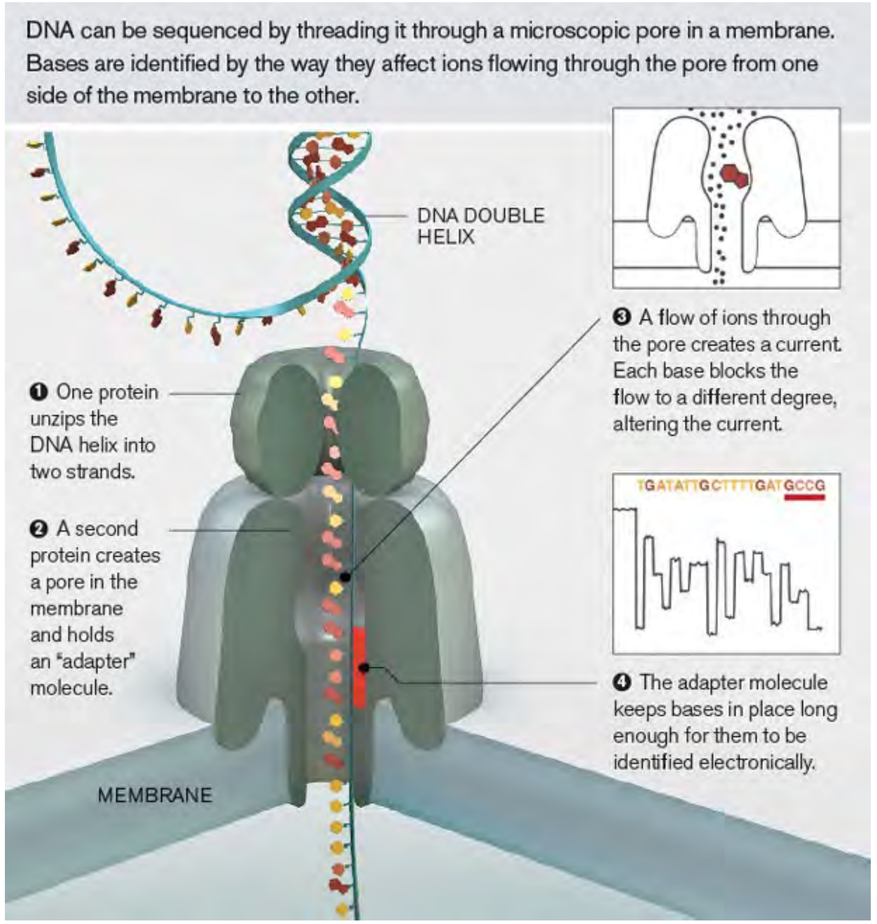


Figure 1-7 Principle of the PacBio single-molecule real-time sequencing technology. (A) Individual DNA polymerase molecules are immobilized at the bottom of 100 nm diameter nanowells etched into the bottom of the reaction chamber. **(B)** Individual fluorescently labeled nucleotides diffuse in and out of these nanowells and may be incorporated into the growing strand when they are complementary to the next unpaired nucleotide on the template strand. During the time between successful binding and the formation of the phosphodiester bond, the gamma phosphate-linked fluorophore is excited and emits an optical signal that is picked up by the instrument's optics. The color of the fluorescence (trace bottom right) reveals which base is incorporated. (from Eid *et al., Science,* 2016)

single molecule in real time!!

Nanopore

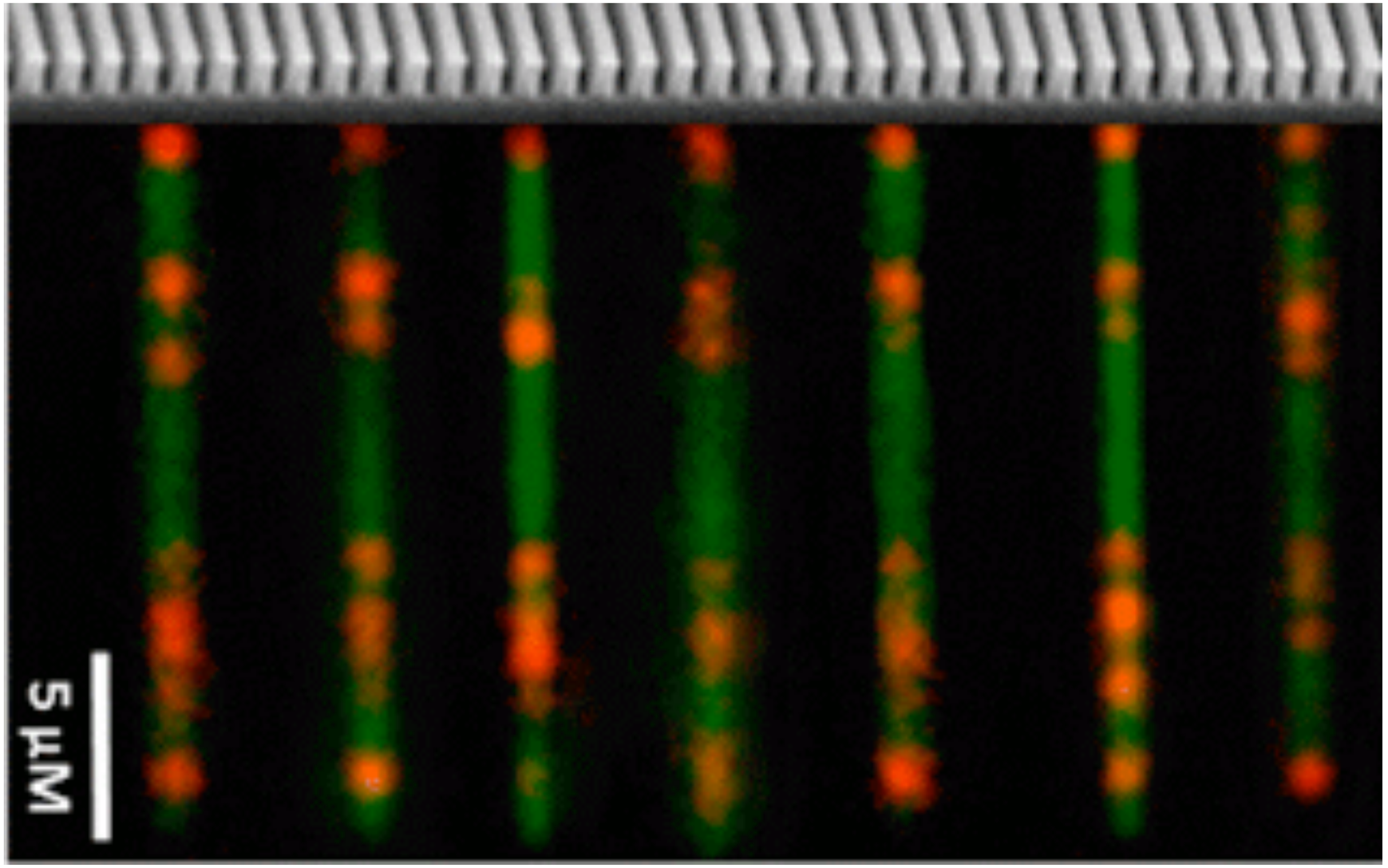


Universität Zürich[UZH]

ETH

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

optical mapping of single DNA strands

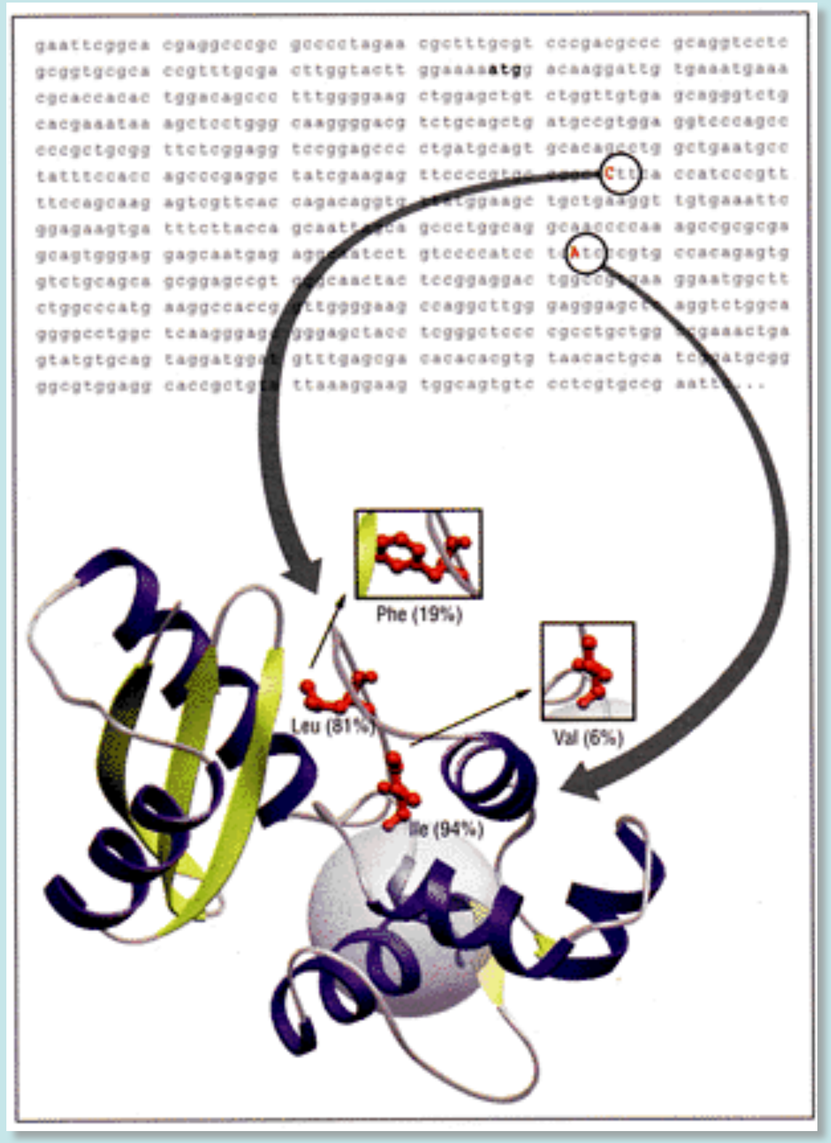


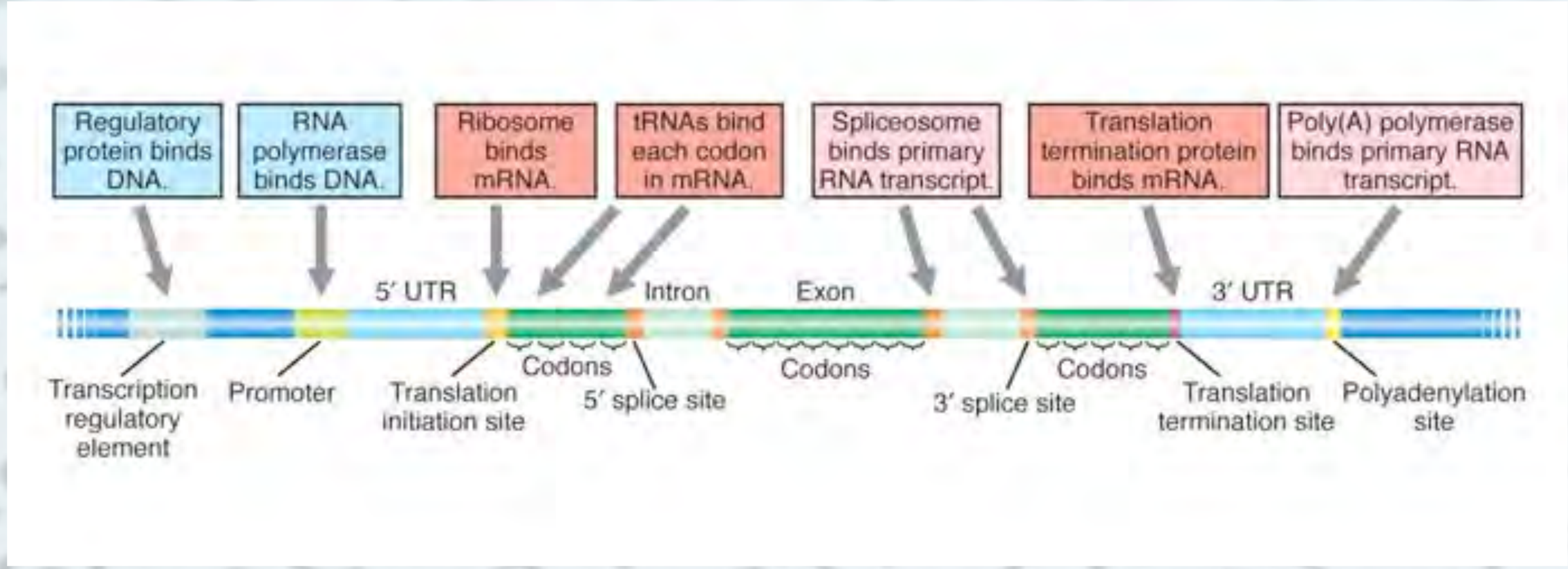
"Bioinformatics is the application of computer science to the management and analysis of biological information."

Bioinformatics

Gene annotation and structure

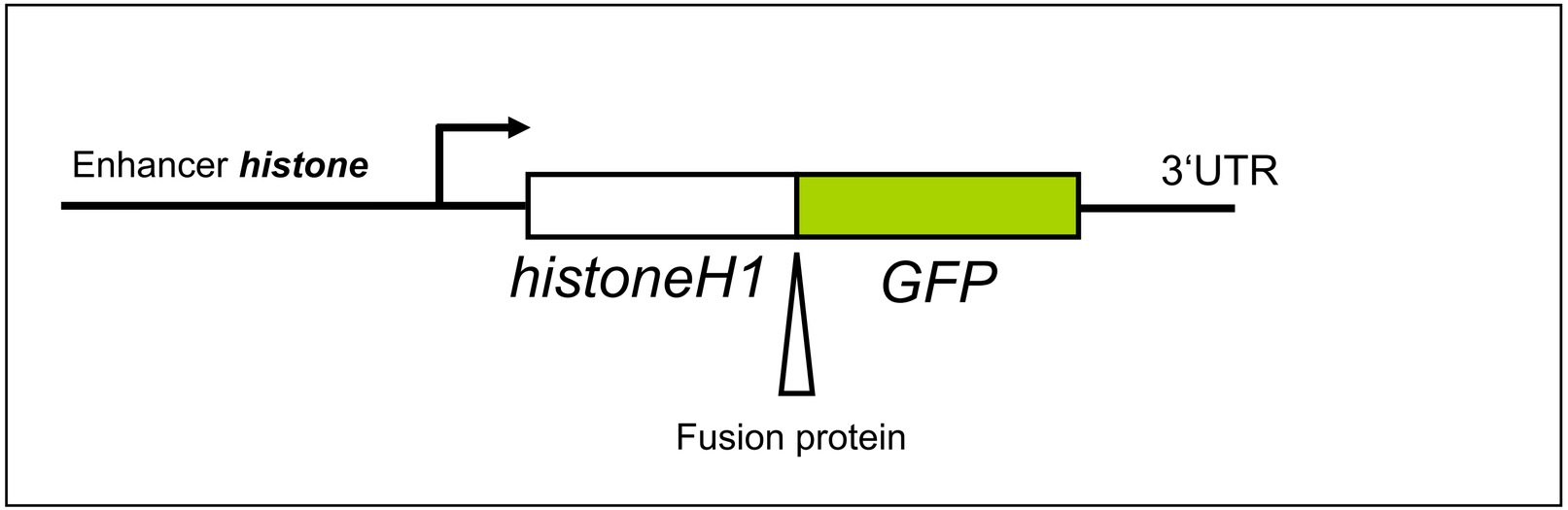
predicted proteins structure and function

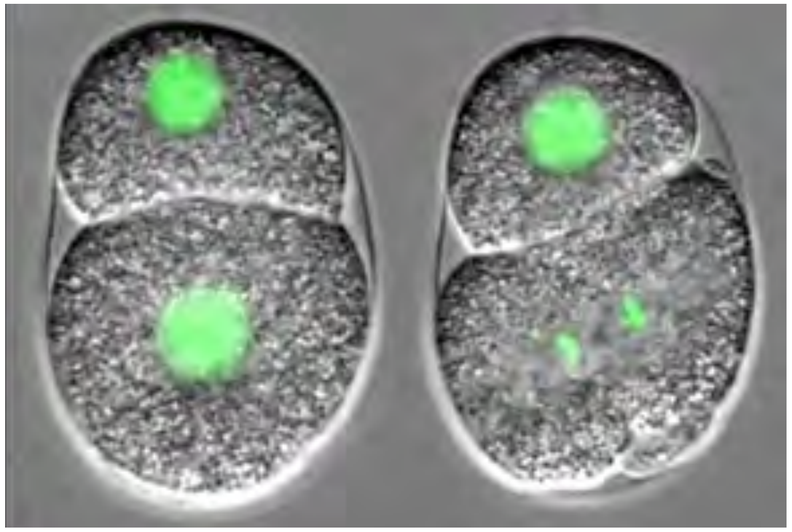




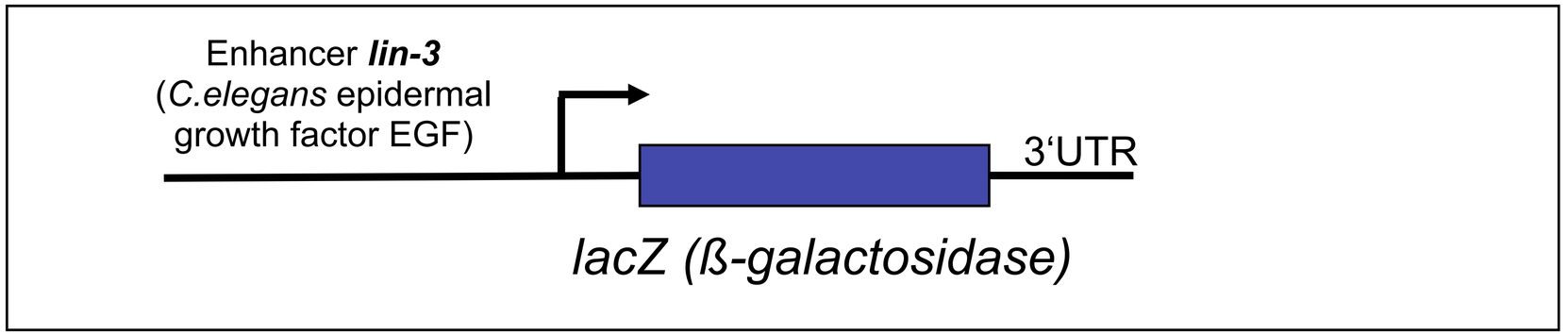
- gene prediction algorithms
- homology finding algorithms (BLAST)
- EST database (partially sequenced cDNAs)
- RNAseq data

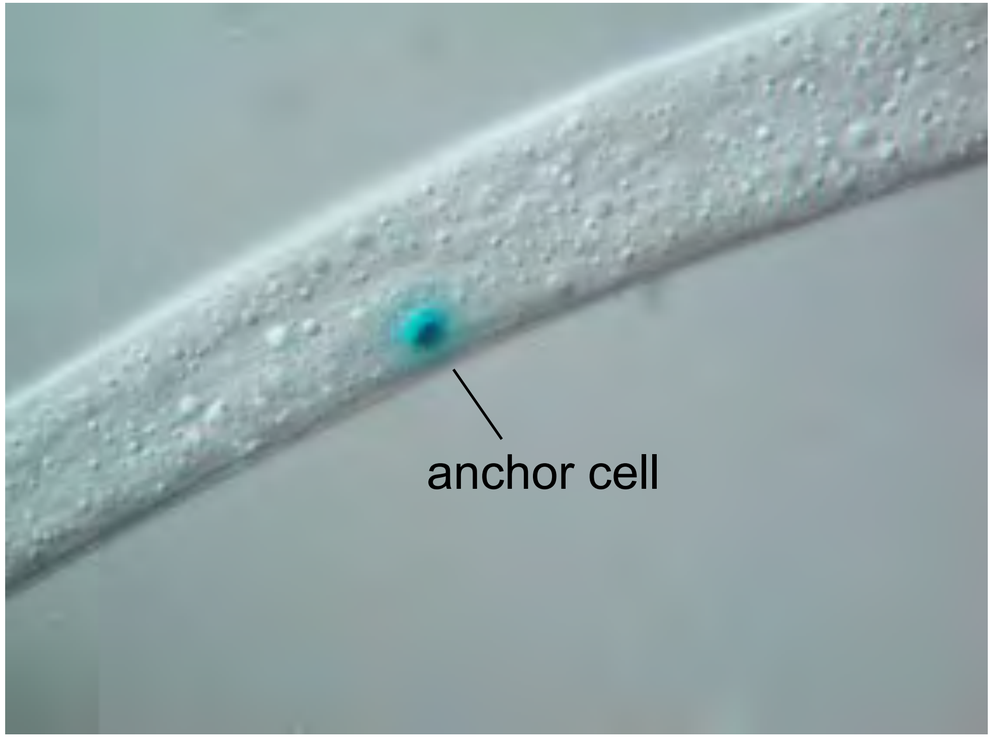
Translational reporter transgene

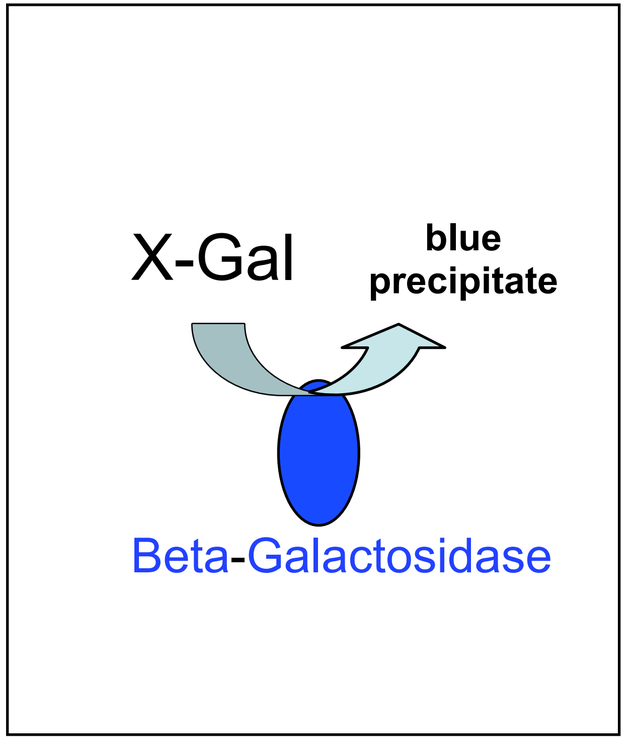




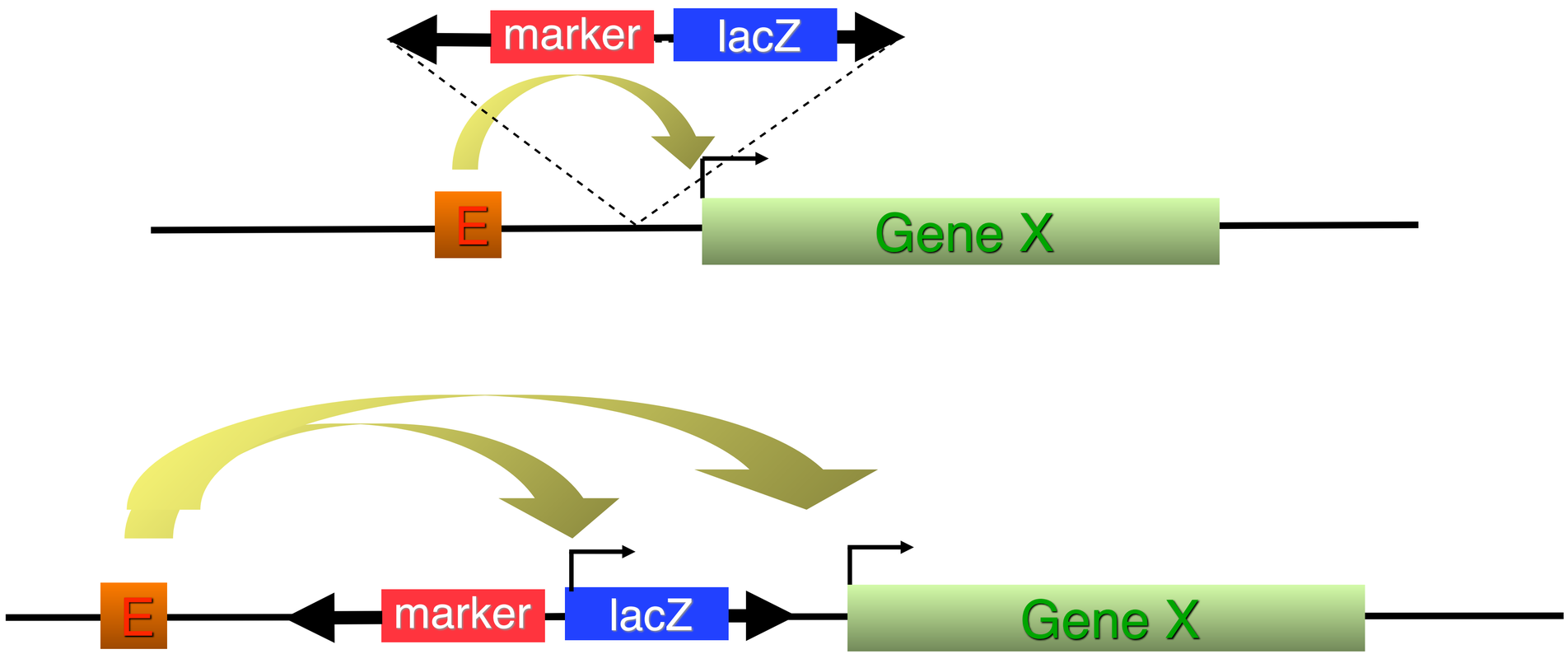
Transcriptional reporter transgene

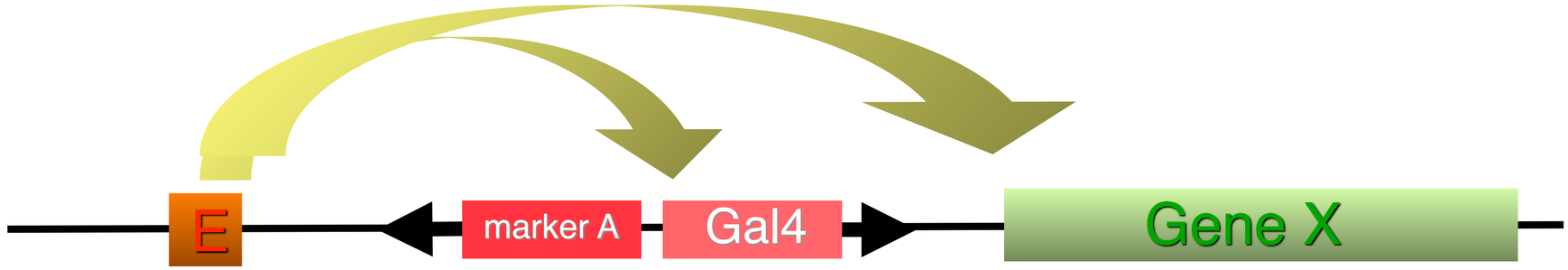




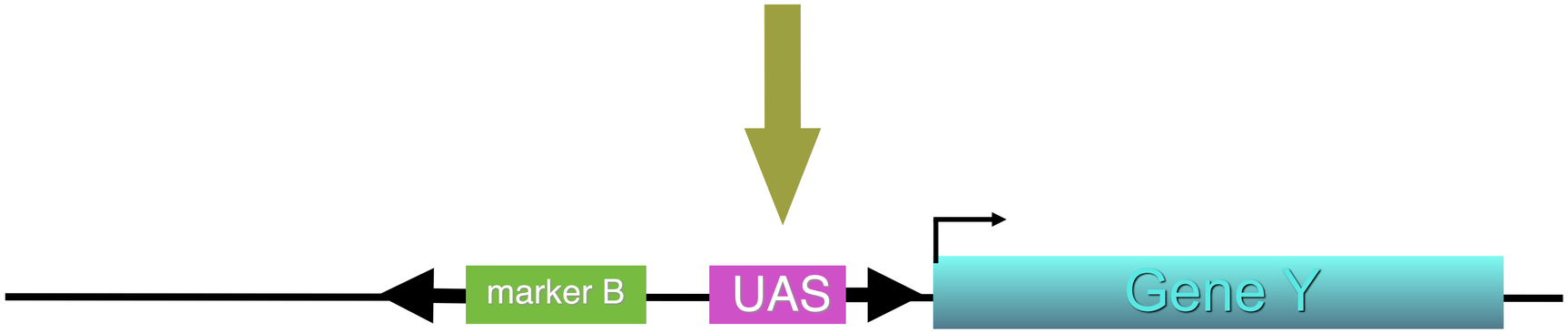


Enhancer Trap





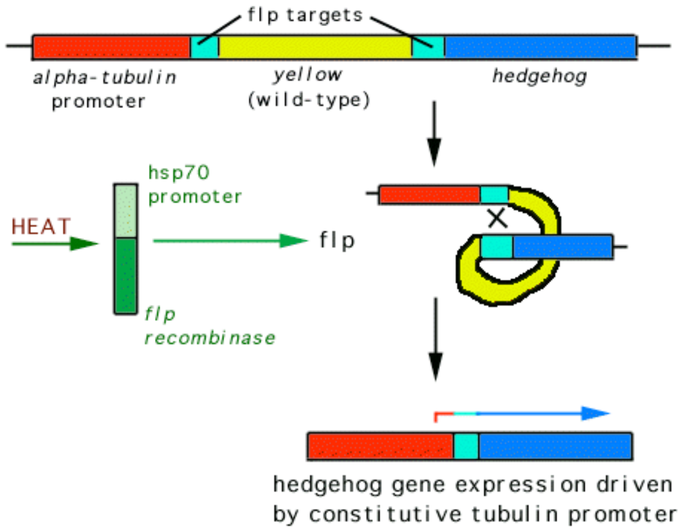
enhancer of gene X drives expression of Gal4 (driver line)

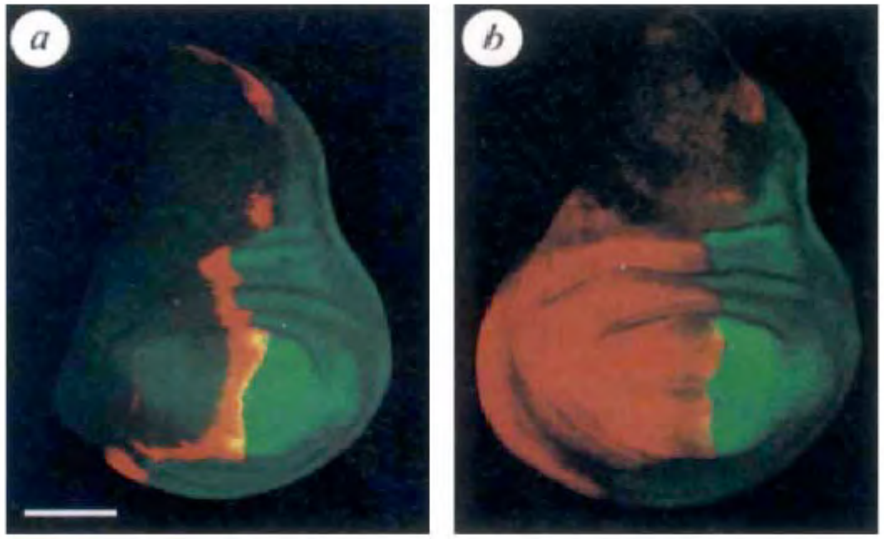


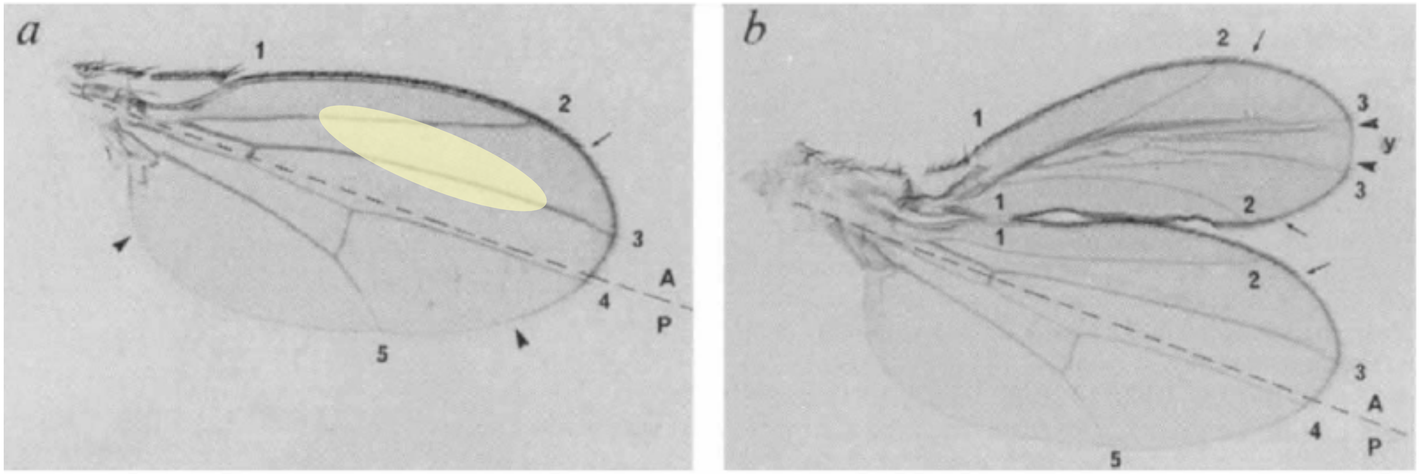
GAL4 Protein interacts with UAS and activates transcription of Gene Y (tester line)

over and/or ectopic expression → gain-of-function mutation in gene Y

example: ectopic expression of hedgehog

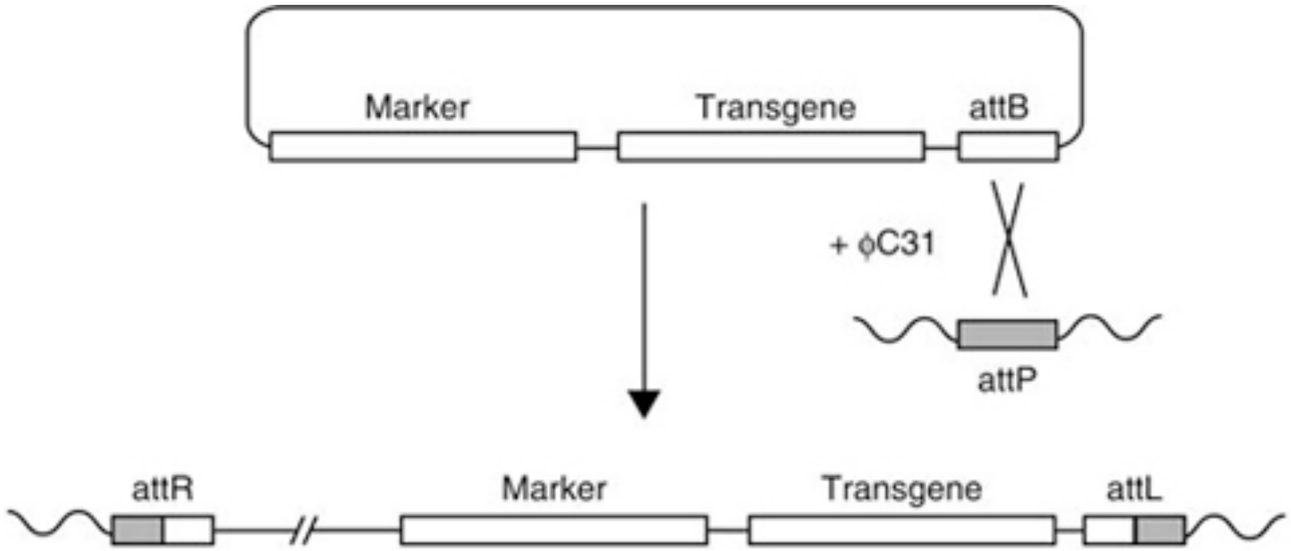






Basler, K. and Struhl, G. 1994. Compartment boundaries and the control of Drosophila limb pattern by hedgehog protein. *Nature* 368: 208-214.

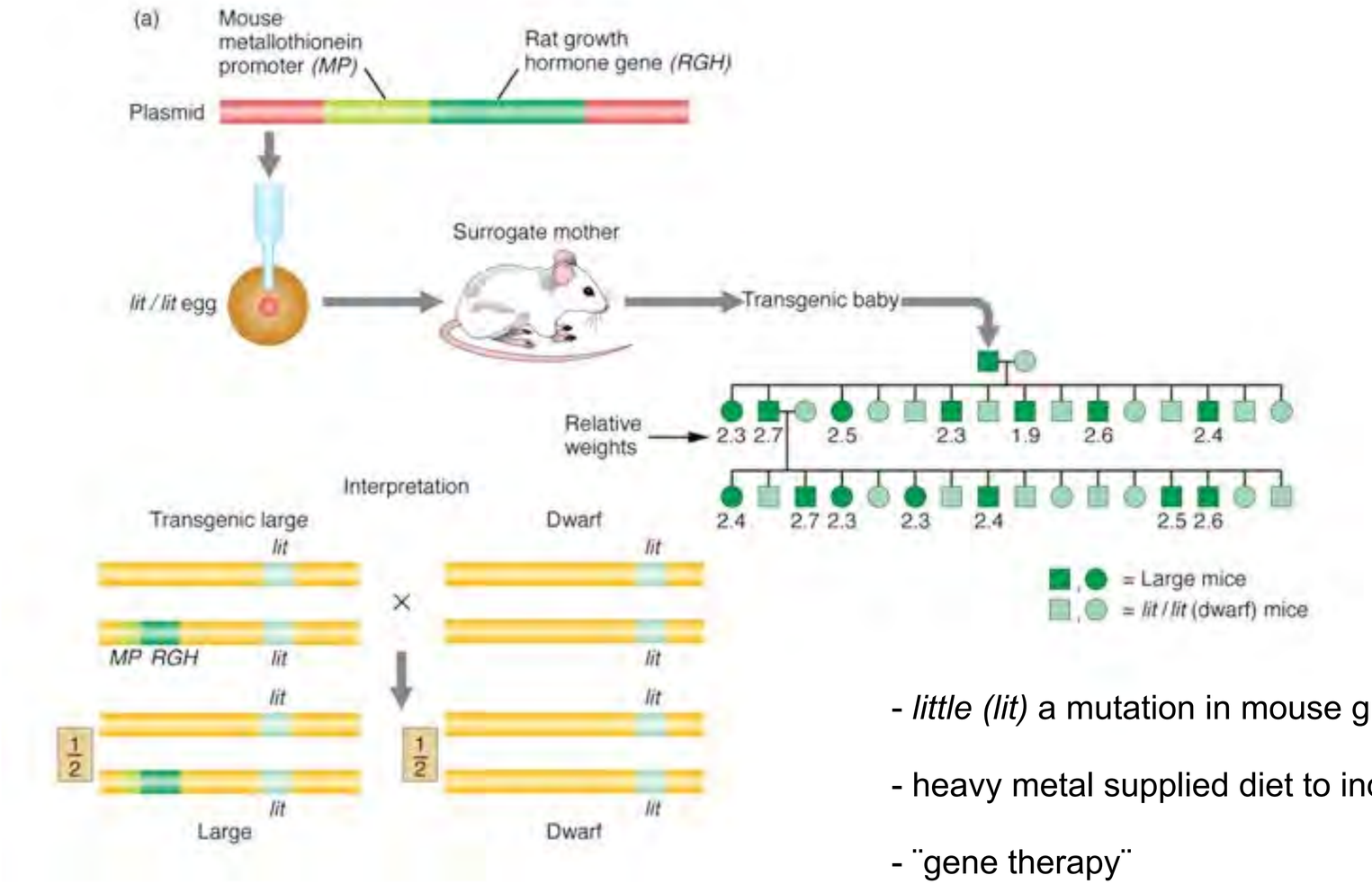
Site-directed transgenesis



φC31 integrase System

- φC31 integrase can only catalyse integration no excision
- attP sequence serves as a fixed docking site in the genome
- identical genomic surrounding controls for possible position effects

transgene-mediated phenotypic rescue in the mouse



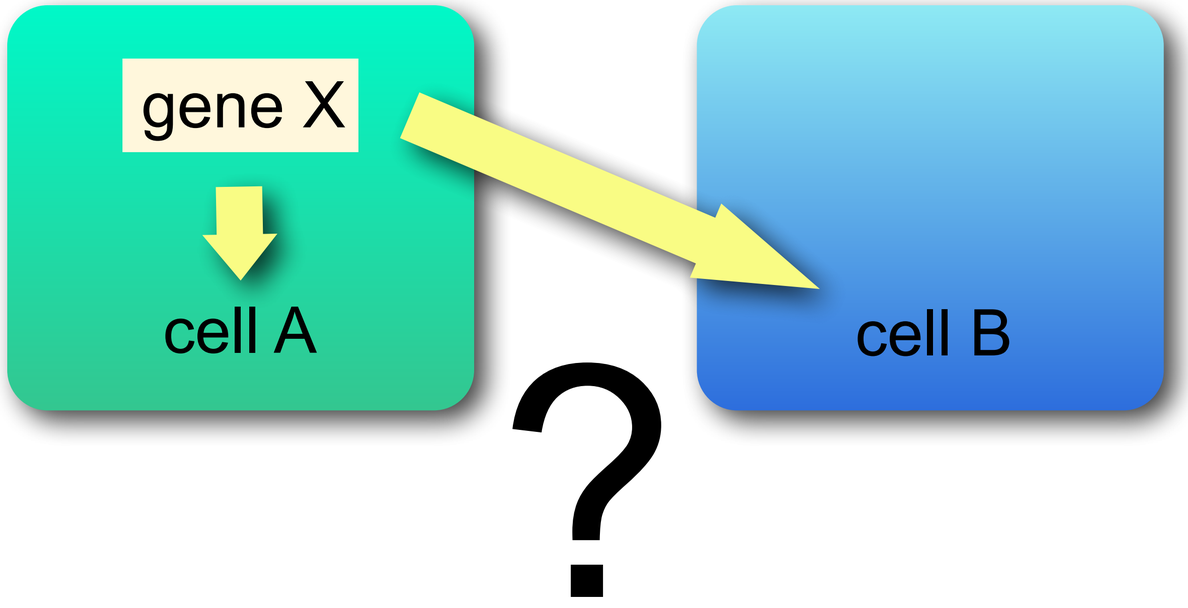
(b)

- *little (lit)* a mutation in mouse growth-hormone gene
- heavy metal supplied diet to induce transgene
- ¨gene therapy¨

why using genetic mosaics?

test autonomy versus nonautonomy

Is the function of gene required in the cells where it is expressed or in neighboring cells?



tools for somatic recombination

Cre / loxP system (*phage P1*) ¨lambda integrase¨ family

- integrates bacteriophage genome into host genome
- very efficient, when high recombination rate required
- preferentially used in mammalian systems

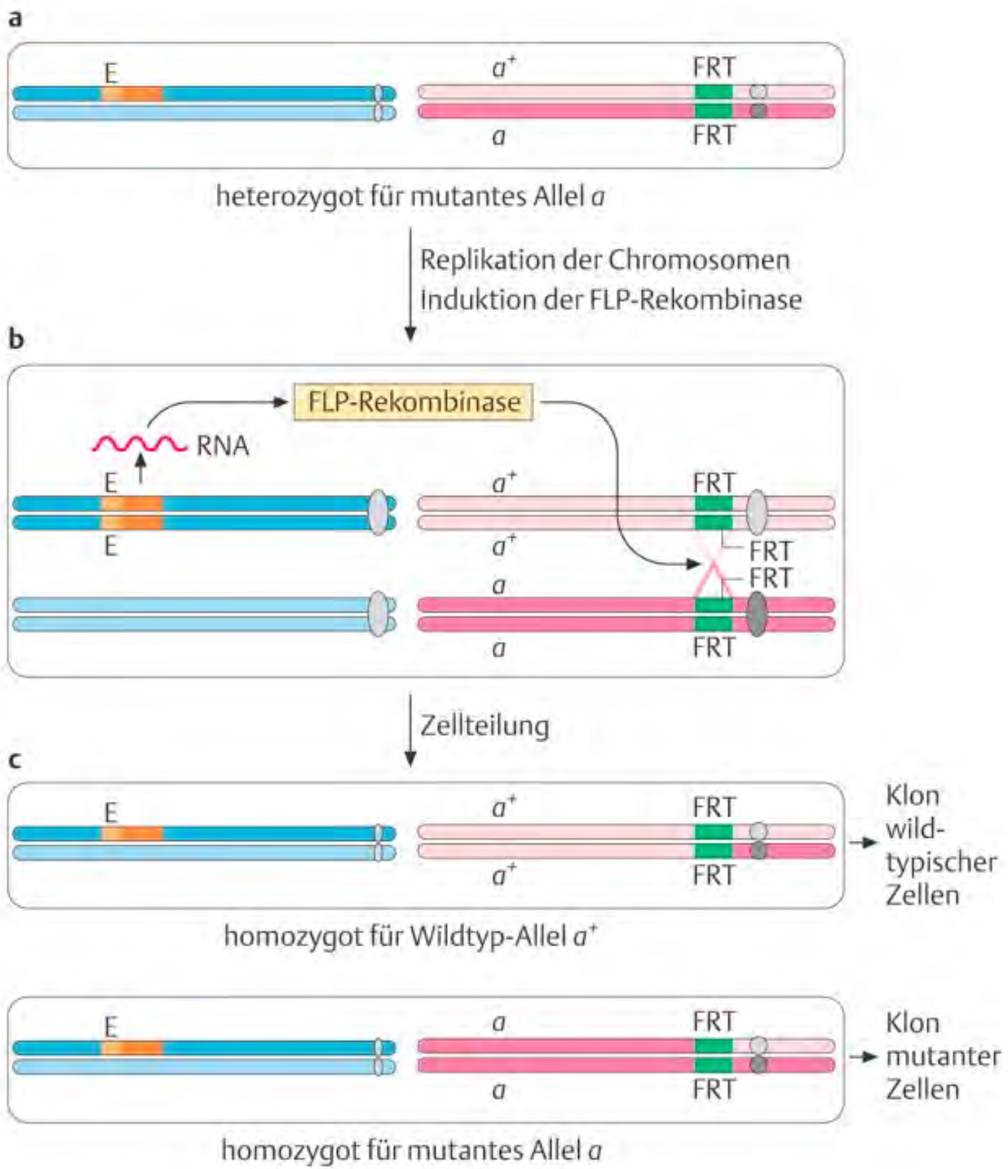
FRT / FLP system (*S. cerevisiae*) ¨lambda integrase¨ family

- maintains copy number of 2µm plasmid
- less effective, when low recombination rate required
- dominant tool in Drosophila

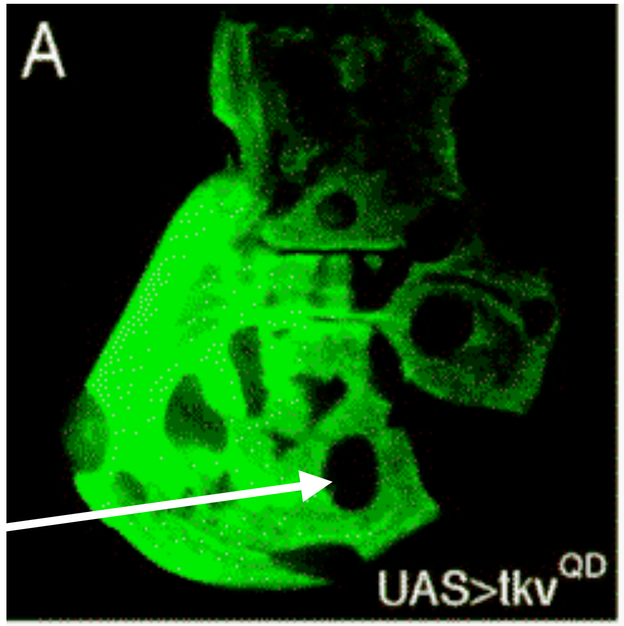
- asymmetric spacers give direction
- catalyses integration and excision

	Recognition sequence				
Site	Direct repeat	1 bp	Inverted repeat 1	Spacer	Inverted repeat 2
FRT	5′GAAGTTCCTATTC	C	GAAGTTCCTATTC	TCTAGAAA	GTATAGGAACTTC 3′
	3′ CTTCAAGGATAAG	G	CTTCAAGGATAAG	AGATCTTT	CATATCCTTGAAG 5′
loxP			5′ATAACTTCGTATA	ATGTATGC	TATACGAAGTTAT 3′
	non essential		3′ TATTGAAGCATAT	TACATACG	ATATGCTTCAATA 5′

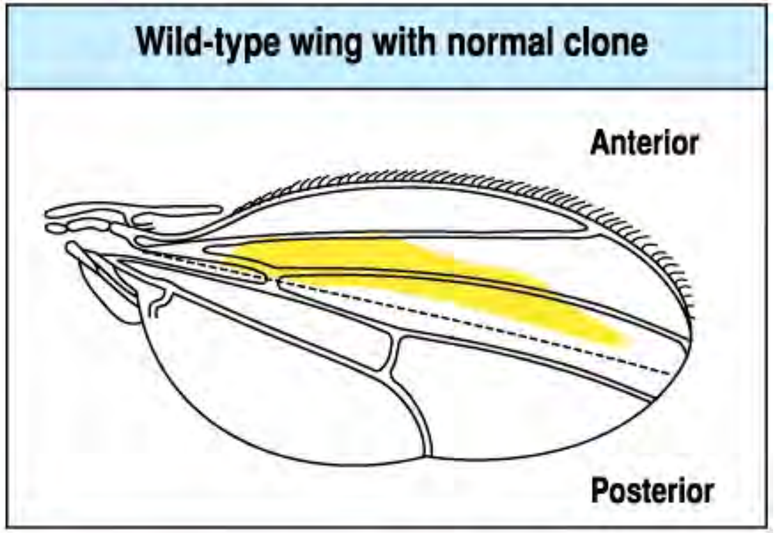
The FRT-FLP technique applied

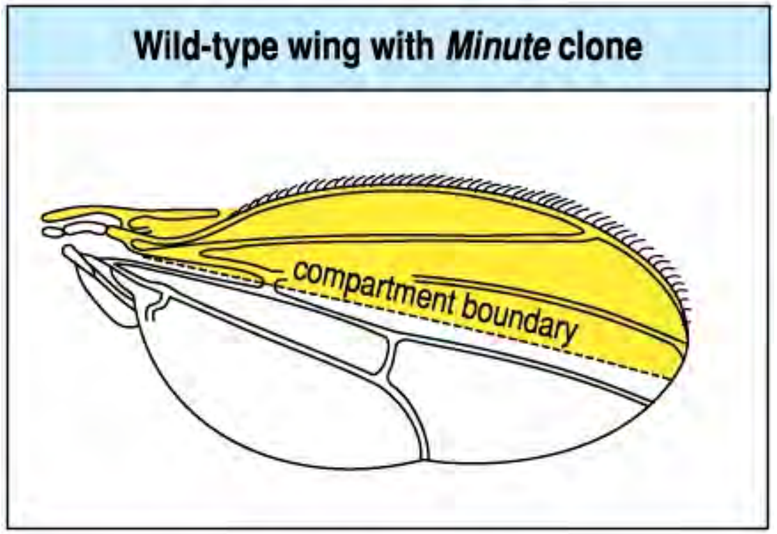


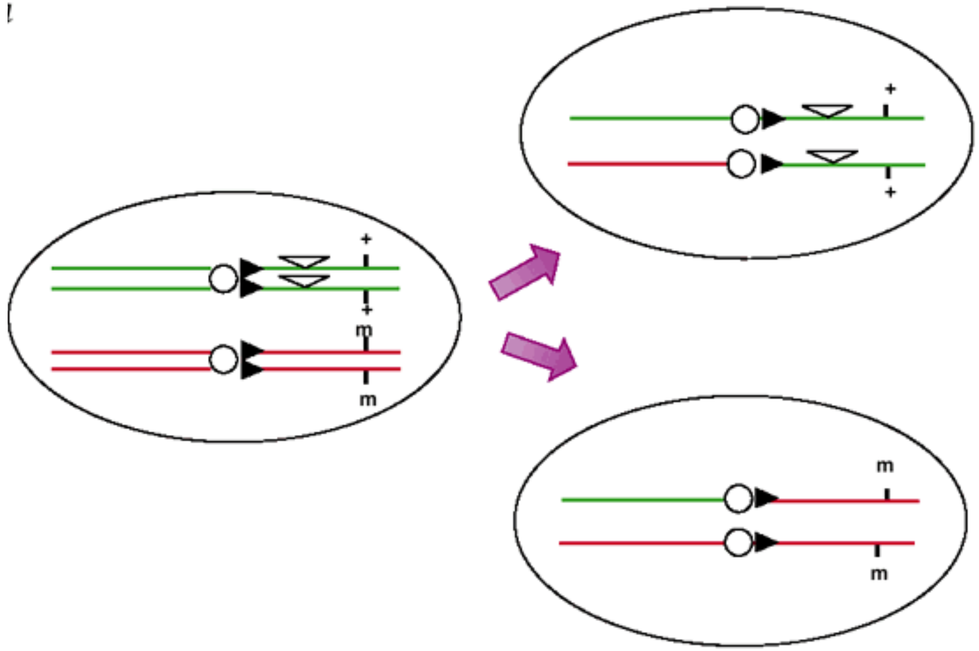
- very efficient
- not random but targeted to FRT sites
- inducible (temporal control)
- tissue-specific (spatial control)
- clone can be marked (e.g. GFP reporter on wt chromosome)



Minute technique expands clonal area

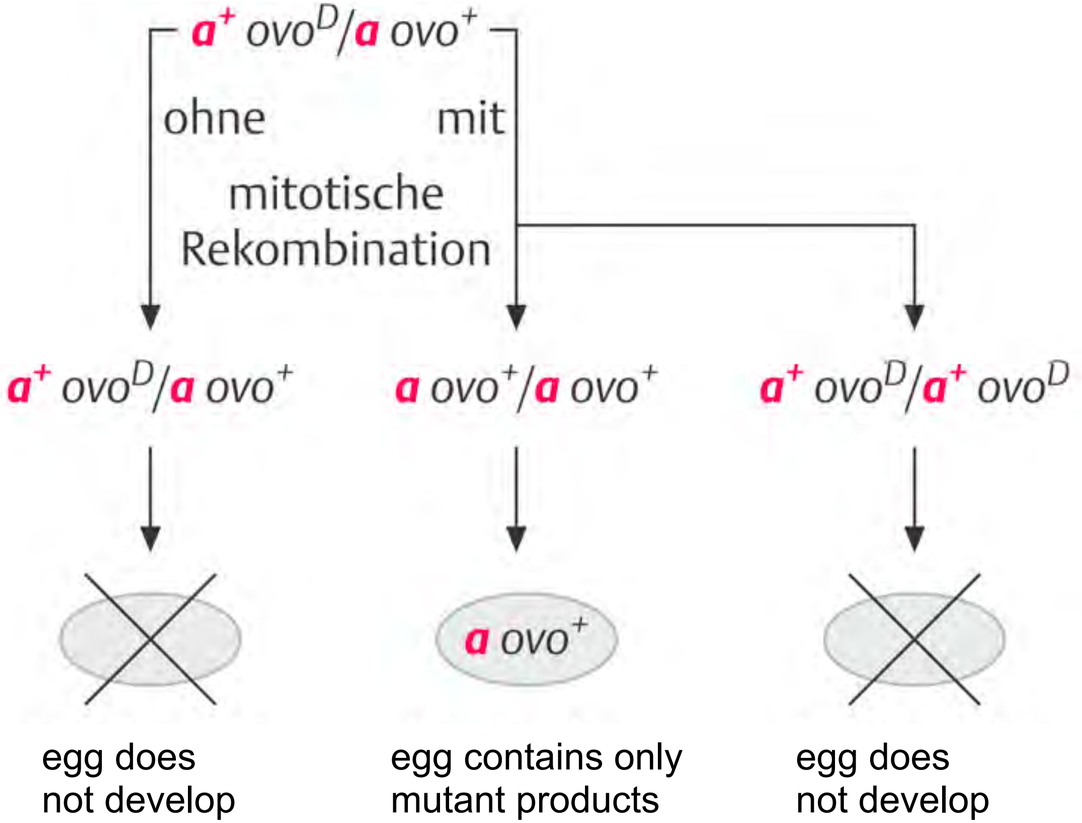






- m^+ have a growth advantage over m^- mutant clones

maternal contribution?: germline clones



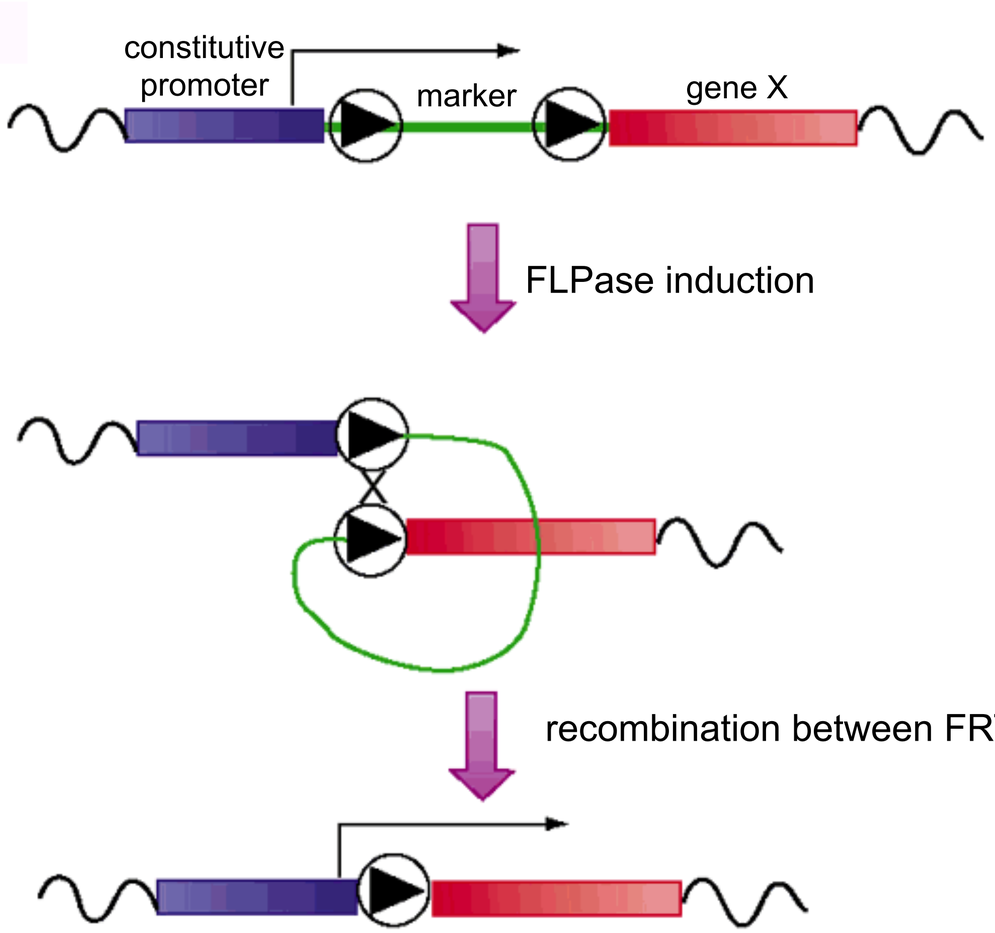
- strategy to mutate maternal activity of a gene (a) in the zygote

- ovo^D is a dominant sterile mutation

phenotype of the zygote?

constitutive expression (gain-of-function)

The FLP out cassette technique

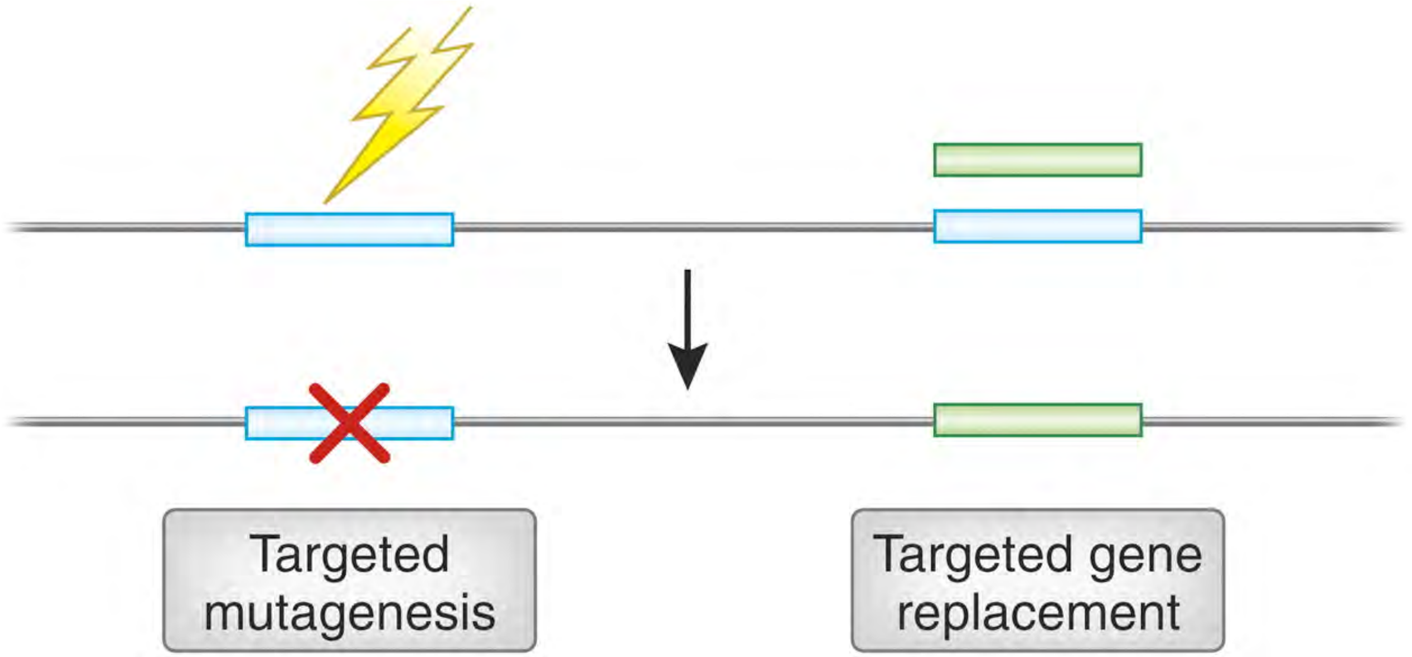


expression of marker (e.g. GFP)

gene X not expressed

expression of gene X

Genome engineering

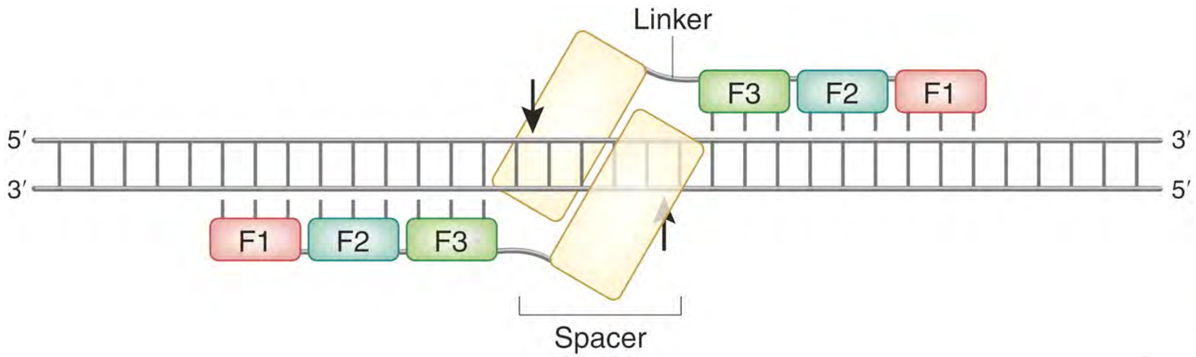


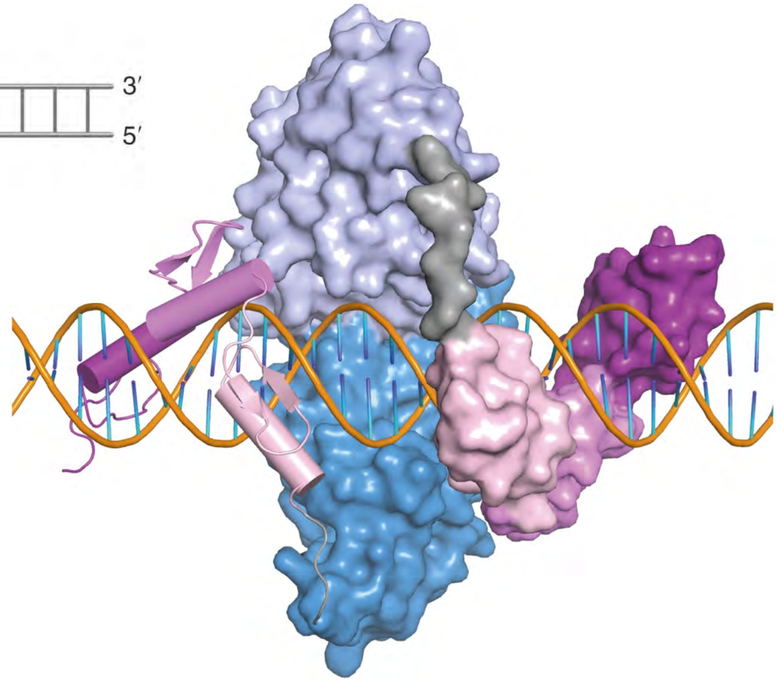
Problem: absolute frequency of homologous recombination ranges between 10^{-4} and 10^{-7} in yeast and mice

Solution: doube stranded breaks substantially increases frequency of local recombination

*Fok*I type IIS restriction enzyme

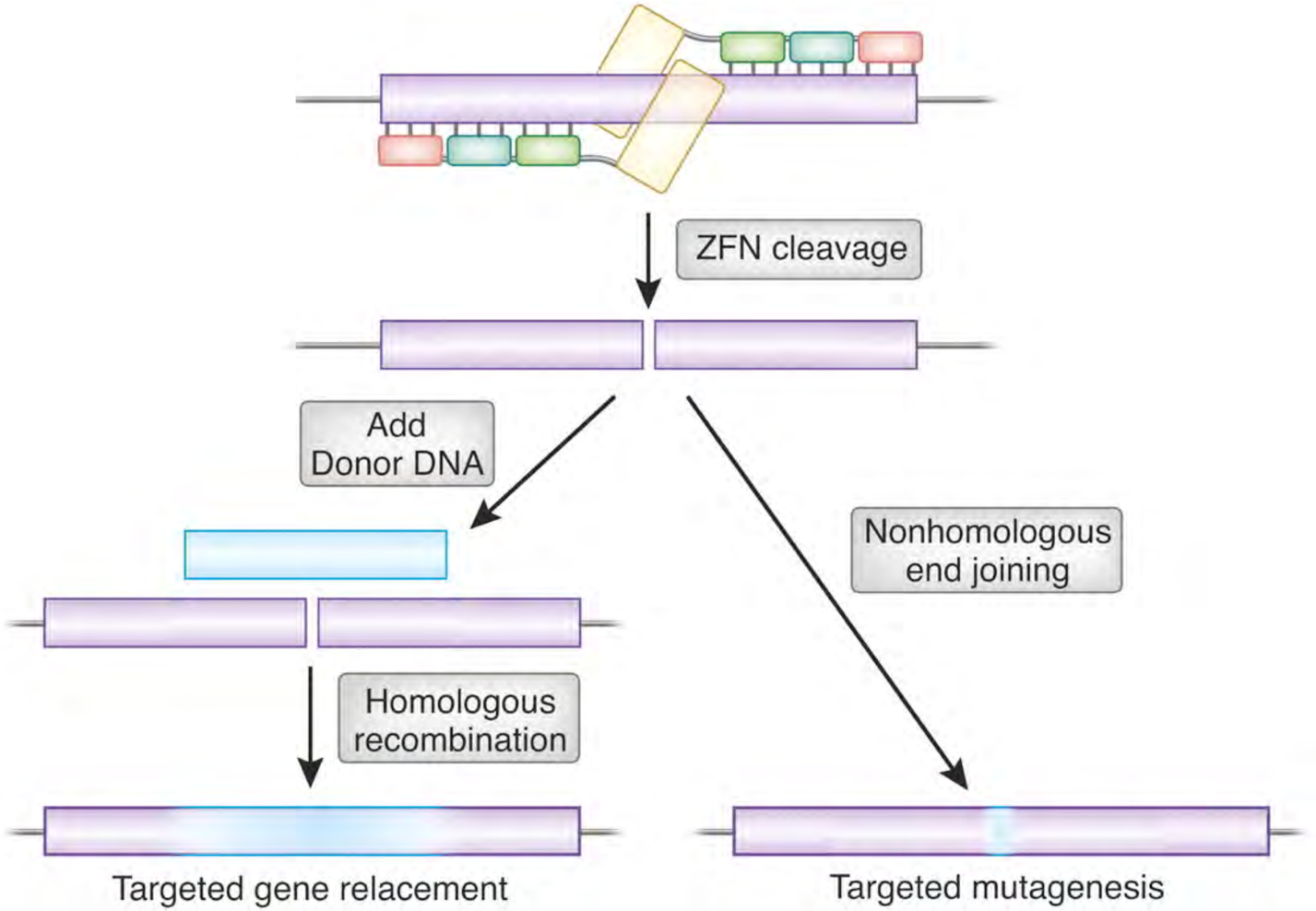
- separable DNA binding and cleavage domain
- DNA binding can be replaced by Cys_2His_2 zinc fingers
- modular character, each finger contacts three nucleotides
- cleavage requires dimerization





Carroll, D. (2011) *Genetics, Vol. 188, 773-782*

ZFN: Zinc Finger Nucleases



HDR NHEJ

Carroll, D. (2011) *Genetics, Vol. 188, 773-782*

ZFN

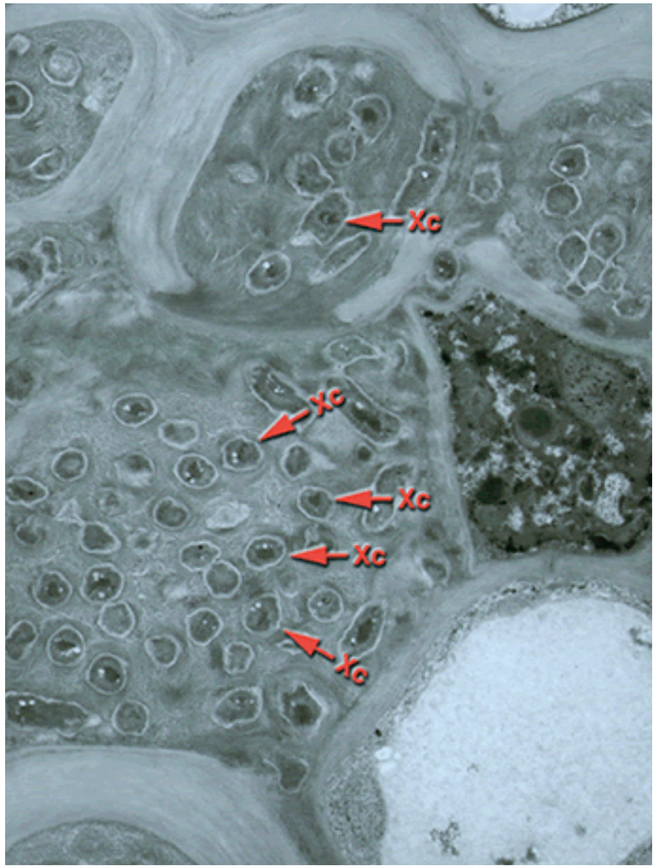
Prospects for ZFN-based gene targeting:

- for model organisms with no effective gene-targeting procedure (e.g. zebrafish, rat)
- for non model organisms
- alternative for ES-based homologous recombination
- genome modification in crop plants
- gene therapy (e.g. knockout of CCR5 in T cells)

Limitations of ZFN-based gene targeting:

- substantial portion of ZFN pairs fails in vivo (limited target capacity)
- context dependent effects (neighboring fingers can alter specificity)
- off target effects, when extensive causes cell-lethality
- expensive and time-consuming assembly

native function of TALEs



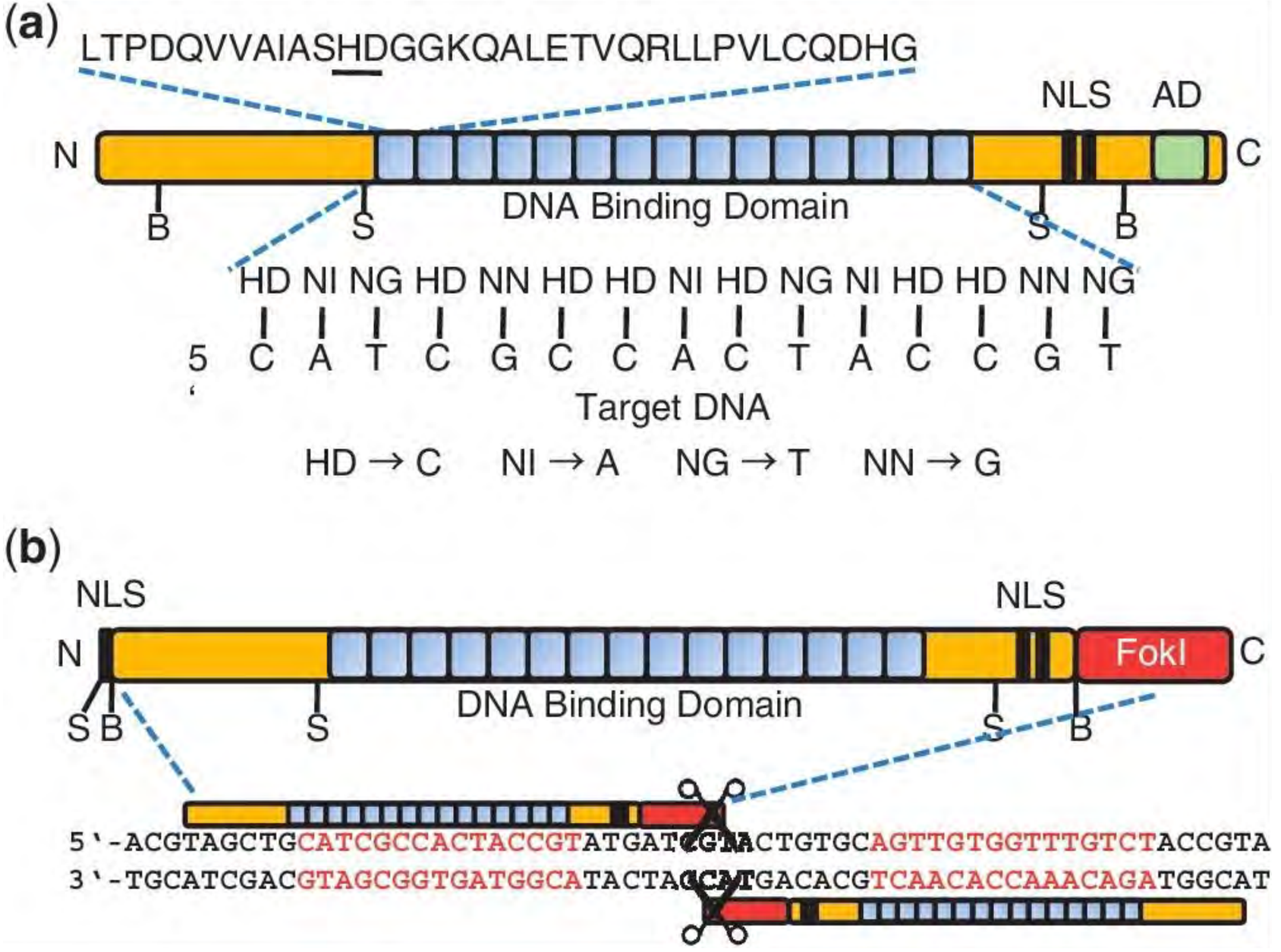
Gram-negative rod-shaped flagellated bacteria

TAL effectors are a family of virulence factors produced by a genus of plant pathogens, ***Xanthomonas spp.***, which, when injected into a host plant, bind to specific host promoter sequences that regulate genes affecting the disease process, both positively and negatively. (Some plants have resistance genes whose promoters bind TAL effectors.)

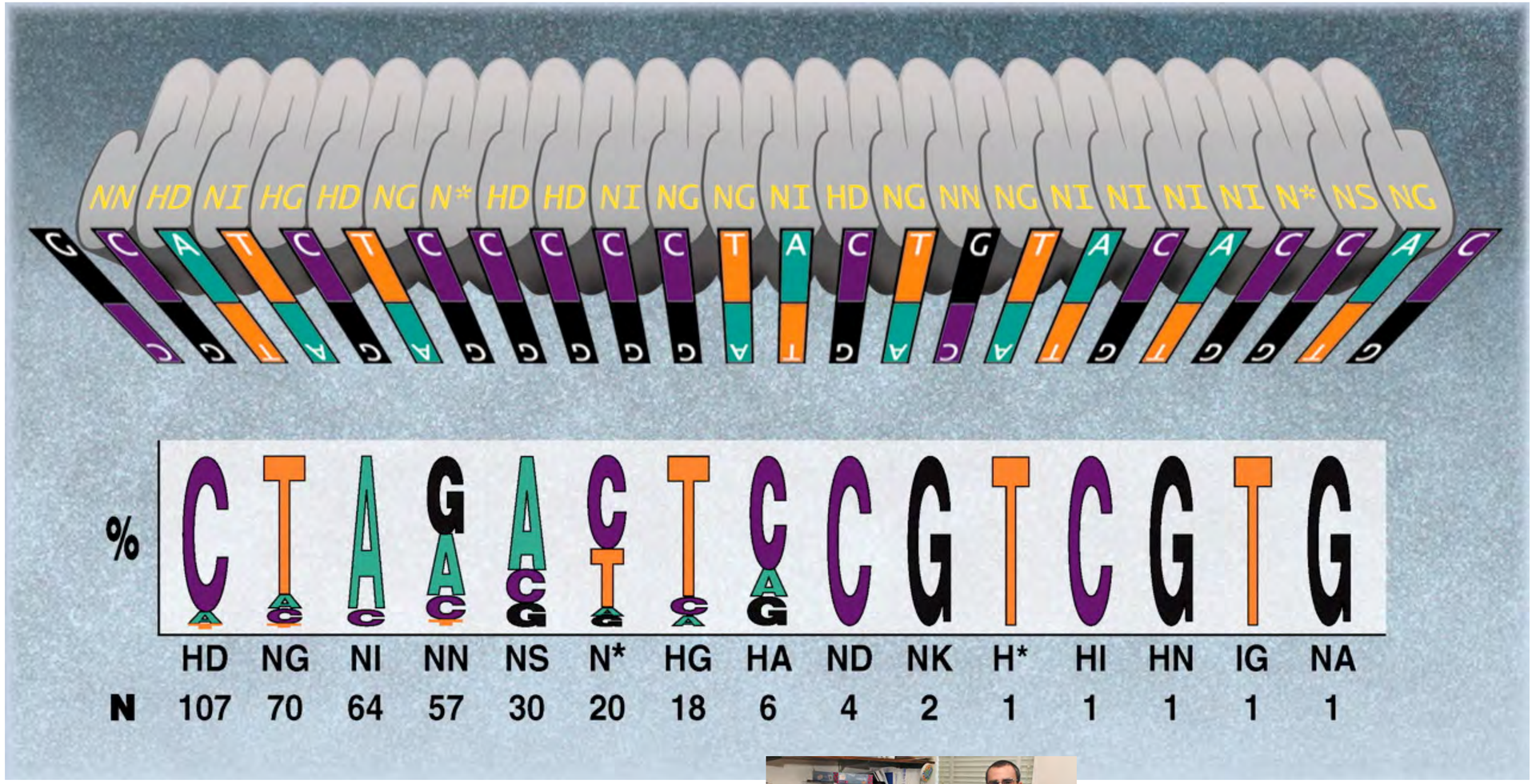
coneflower

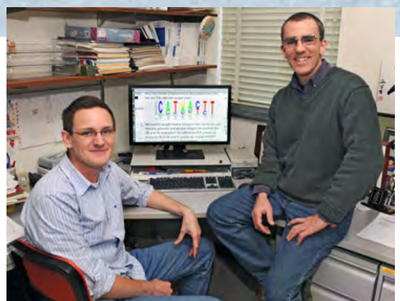
grapefruit

Xanthomonas causes small brown angular to circular spots with yellow halos. In some plants these bacteria can cause dead spots in foliage and or fruit and sometimes cankers in stems.


(a)
LTPDQVVAIASHDGGKQALETVQRLLPVLCQDHG
NLS
AD
N
C
B
S
DNA Binding Domain
S
B
HD NI NG HD NN HD HD NI HD NG NI HD HD NN NG
5 C A T C G C C A C T A C C G T
Target DNA
HD → C
NI → A
NG → T
NN → G
(b)
NLS
NLS
N
FokI
C
S B
S
DNA Binding Domain
S
B
5'-ACGTAGCTGCATCGCCACTACCGTATGATCGTACTGTGCAGTTGTGGTTTGTCTACCGTA
3'-TGCATCGACGTAGCGGTGATGGCATACTAGCATGACACGTCAACACCAAACAGATGGCAT

First reports describing the DNA binding capacity of TAL effectors appeared in 2007 (Romer et al.), and only a year later two groups **cracked the code** (Boch et al., Moscou adn Bogndanove), each showing a one-to-one correspondence between particular amino acid pairs and single nucleotides.





Romer, P. et al. Science 318, 645–648 (2007)
Boch, J. et al. Science 326, 1509–1512 (2009)
Moscou, M.J. & Bogdanove, A. Science 326, 1501 (2009)

CRISPR/CAS SYSTEM

CRISPR: Clustered, Regularly Interspaced, Short Palindromic Repeats

Adaptive immune system in bacteria and archea against foreign viral and plasmid DNA

Three types of CRISPR systems known (type II used for genome engineering)

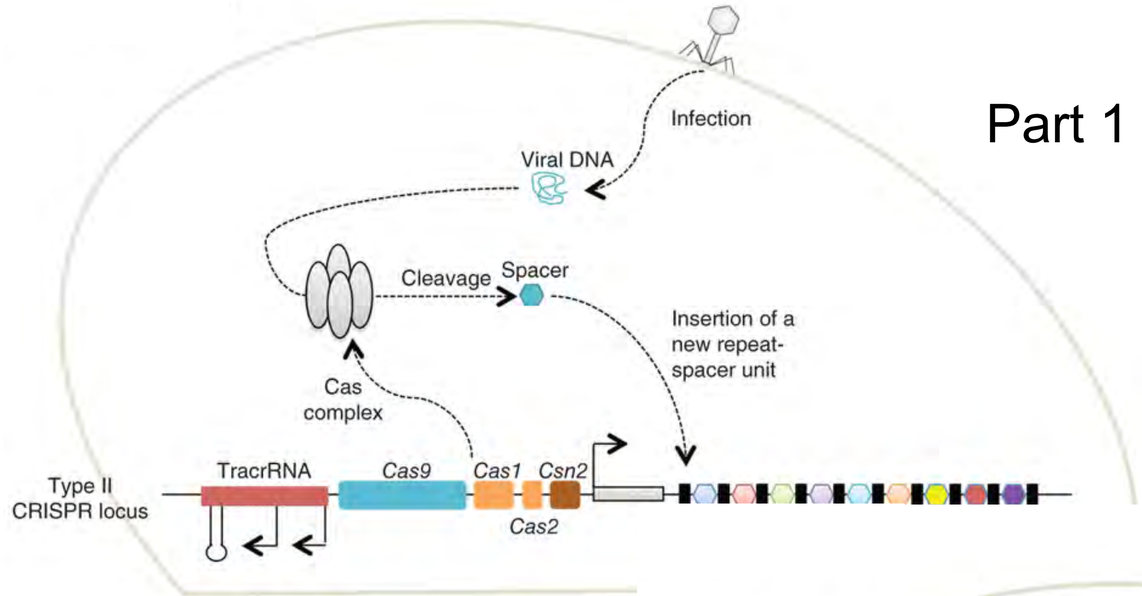
Three component system:

1) **Cas9**: provides enzymatic activity
2) Transactivating CRISPR-RNA: **tracrRNA**, triggers Cas9 activity
3) CRISPR-RNA: **crRNA**, short guide that targets the genomic sequence

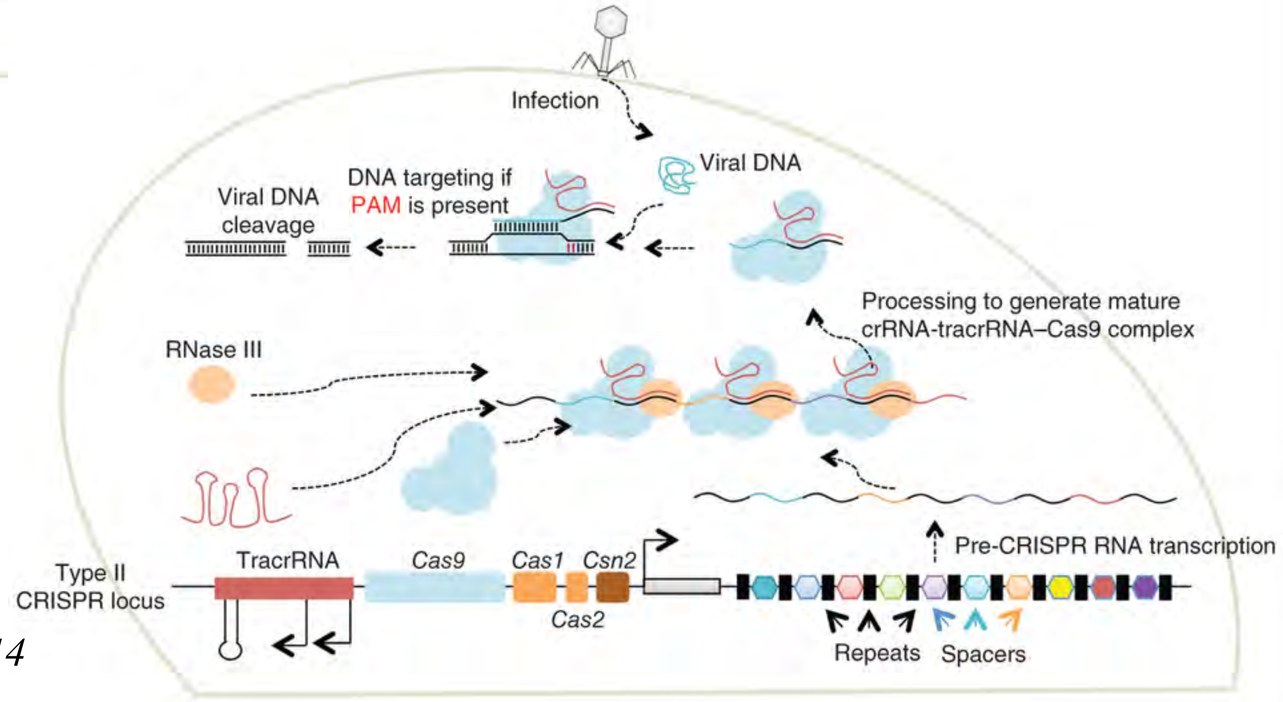
tracrRNA and crRNA can be combined into one precursor

Bacterial Defense

Part 1 Immunization



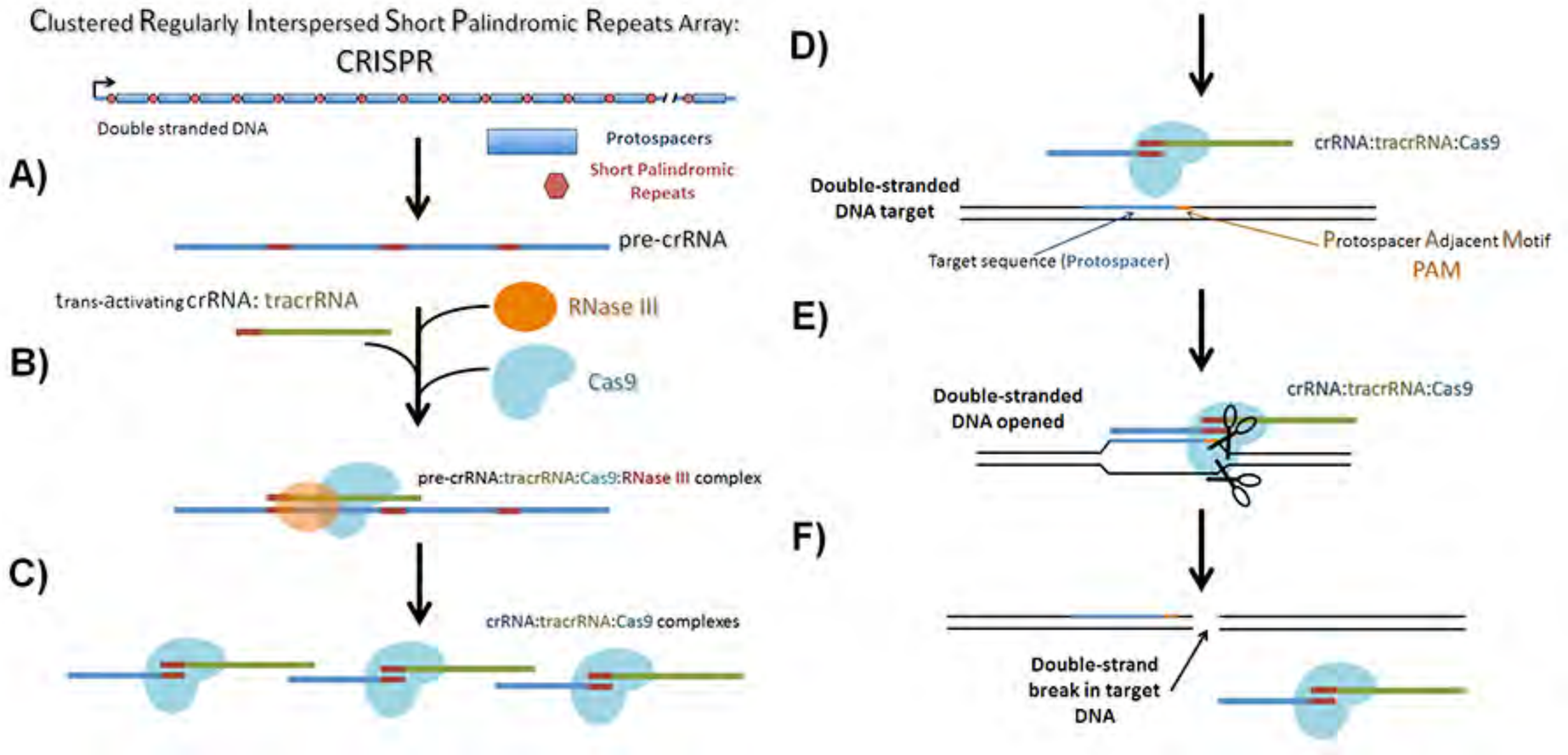
Part 2 Immunity



¨Concepts of Modern Genetics¨ HS 2014

Universität Zürich[UZH]
ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

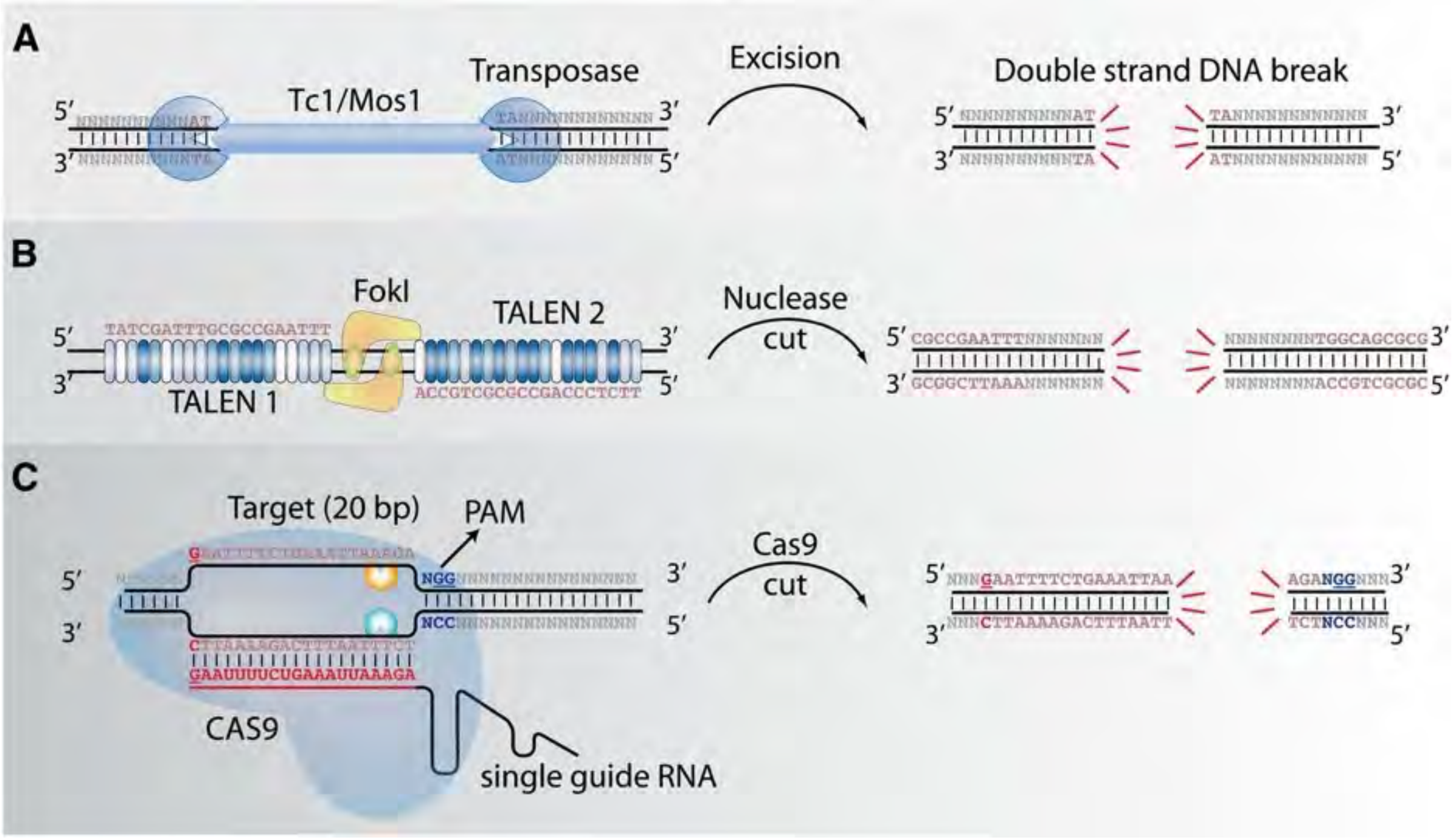
CRISPR/CAS type II Mechanism



Universität Zürich[UZH]

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Different methods to generate ds DNA breaks



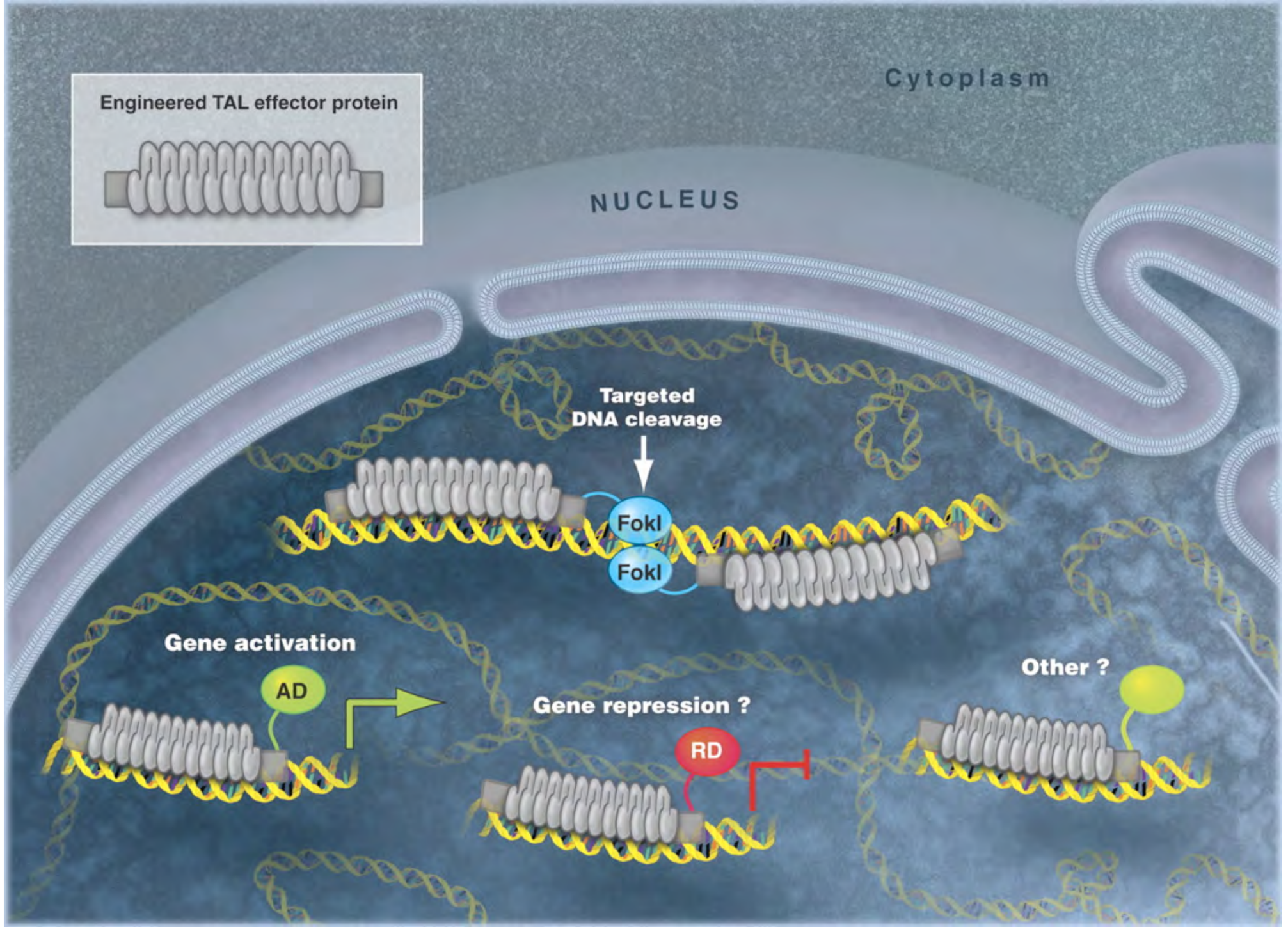

Engineered TAL effector protein
Cytoplasm
NUCLEUS
Targeted
DNA cleavage
FokI
FokI
Gene activation
AD
Gene repression ?
RD
Other ?

Table 1 Cost of designer nucleases

Source	Resource	Price	Time required
ZFNs			
Sigma (Sangamo)	Premade or custom ZFNs	$25,000 custom; $12,000 premade	2 months
Zinc Finger Consortium and Addgene	OPEN Protocols, reagents, including plasmids and bacterial strains	$5,000 for hundreds of Zfns	6 months–1 year
Zinc Finger Consortium	CoDA	Not available as a kit; $600–900 to assemble reagents for one Zfn	1 month
Addgene	Modular assembly	$650 for hundreds of Zfns (not all work)	2 months
TALENs[a]			
Cellectis	Premade or custom TALENs	$5,000 custom; $10,000 mammalian cell validation	1.5 months
Addgene	Golden Gate method cloning kit	$350 for full kit (one-time cost)	1 week

[a]Life Technologies is a licensed provider of TAL effectors. As *Nature Biotechnology* went to press, the company was not making them broadly available.

Source: Dave Segal, UC Davis

Genome editing

ZFN

less flexible modules

targetability 500 bp

context dependent

off-target effects

long study (10 years)

low immunogenicity

high costs

TALEN

unlimited, fast assembly

target range every 35 bp

context independent

less off-target effects?

largely unexplored

immunogenicity?

affordable

some important abbreviations

TALE: transcription activator-like effector

TALEN: transcription activator-like effector nuclease

RVD: repeat variant di-residue

NHEJ: nonhomologous end joining

HDR: homology directed repair