

# Another approach to gene insertion: targeted recombination induced by DNA breaks

DNA can be inserted by homologous recombination (gene targeting)

- Inefficient: won't happen in most cells receiving DNA
- Requires selection in culture (e.g. stem cells)

Increase efficiency of homologous recombination by breaking the genome at the region you want to engineer

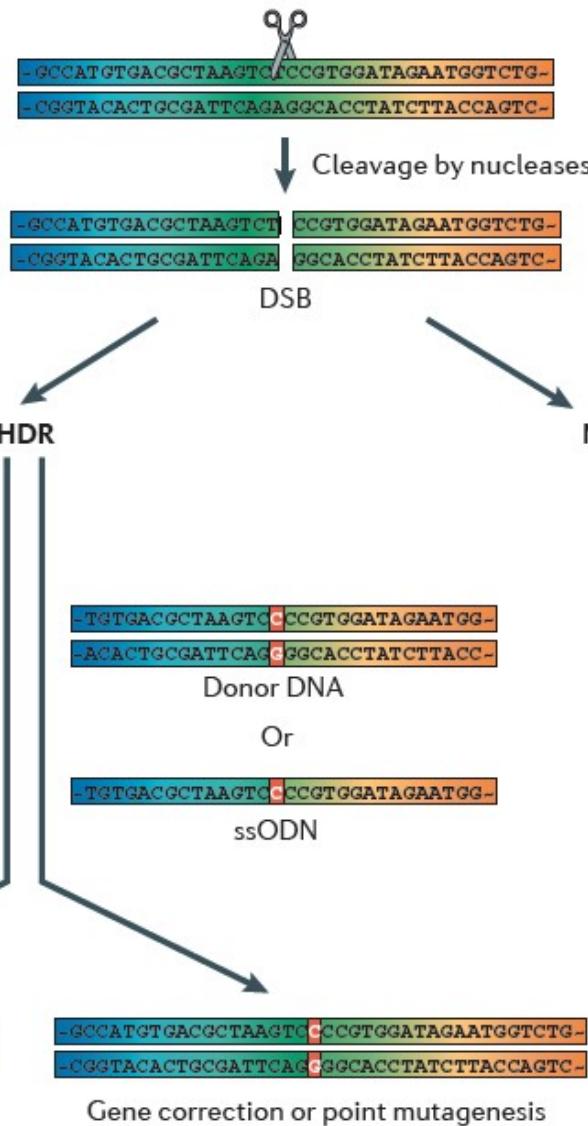
Make a machine that can:

- Find specific DNA site with high selectivity
- Cut both strands of DNA after finding its site

# What happens after a double stranded break in DNA?

a

HDR: Homology Dependent Repair

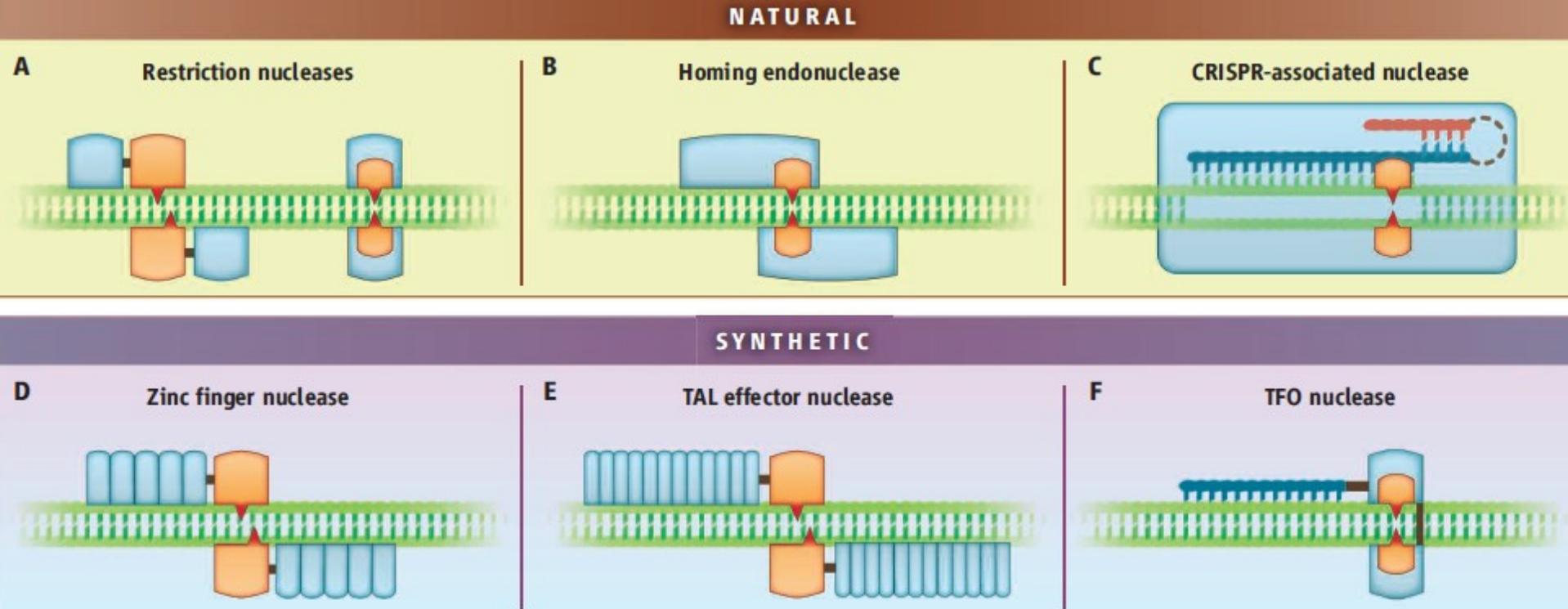


NHEJ: Non Homologous End Joining

A guide to genome engineering with programmable nucleases

Hyongbum Kim<sup>1</sup> and Jin-Soo Kim<sup>2,3</sup>

# Several strategies for defining specific DNA cut sites



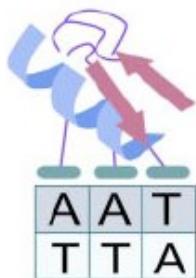
Blue: DNA recognition, Orange: DNA nuclease

**New Tool for Genome Surgery**  
John van der Oost  
*Science* **339**, 768 (2013);  
DOI: 10.1126/science.1234726

To introduce mutations, co transfect with

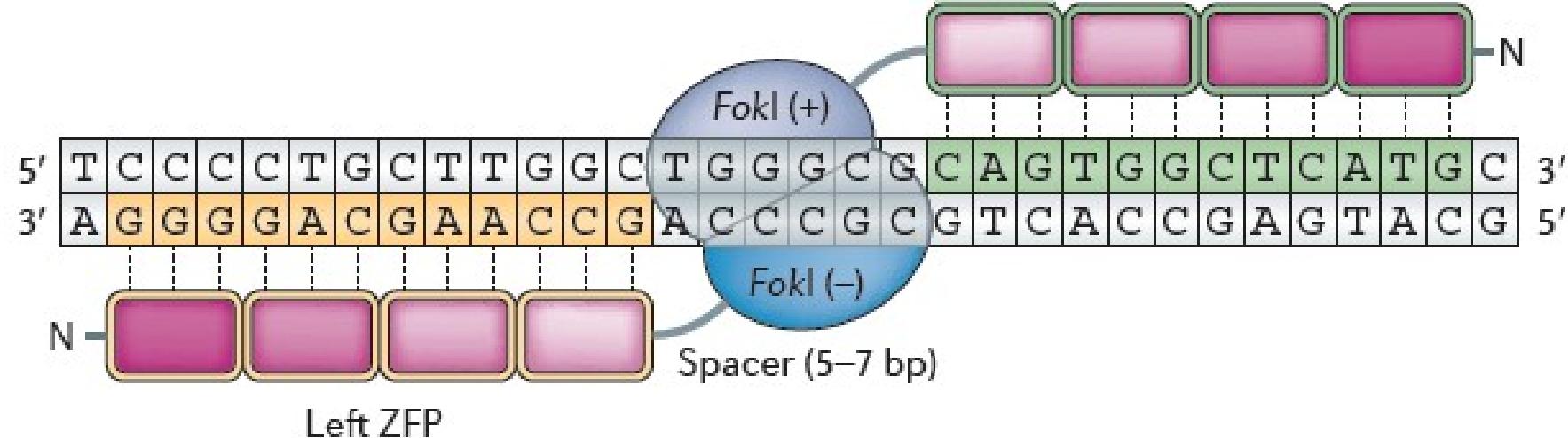
- A) programmable nuclease
- B) Targeting vector (containing desired DNA)

# Zinc finger nuclease (ZFN) mechanism



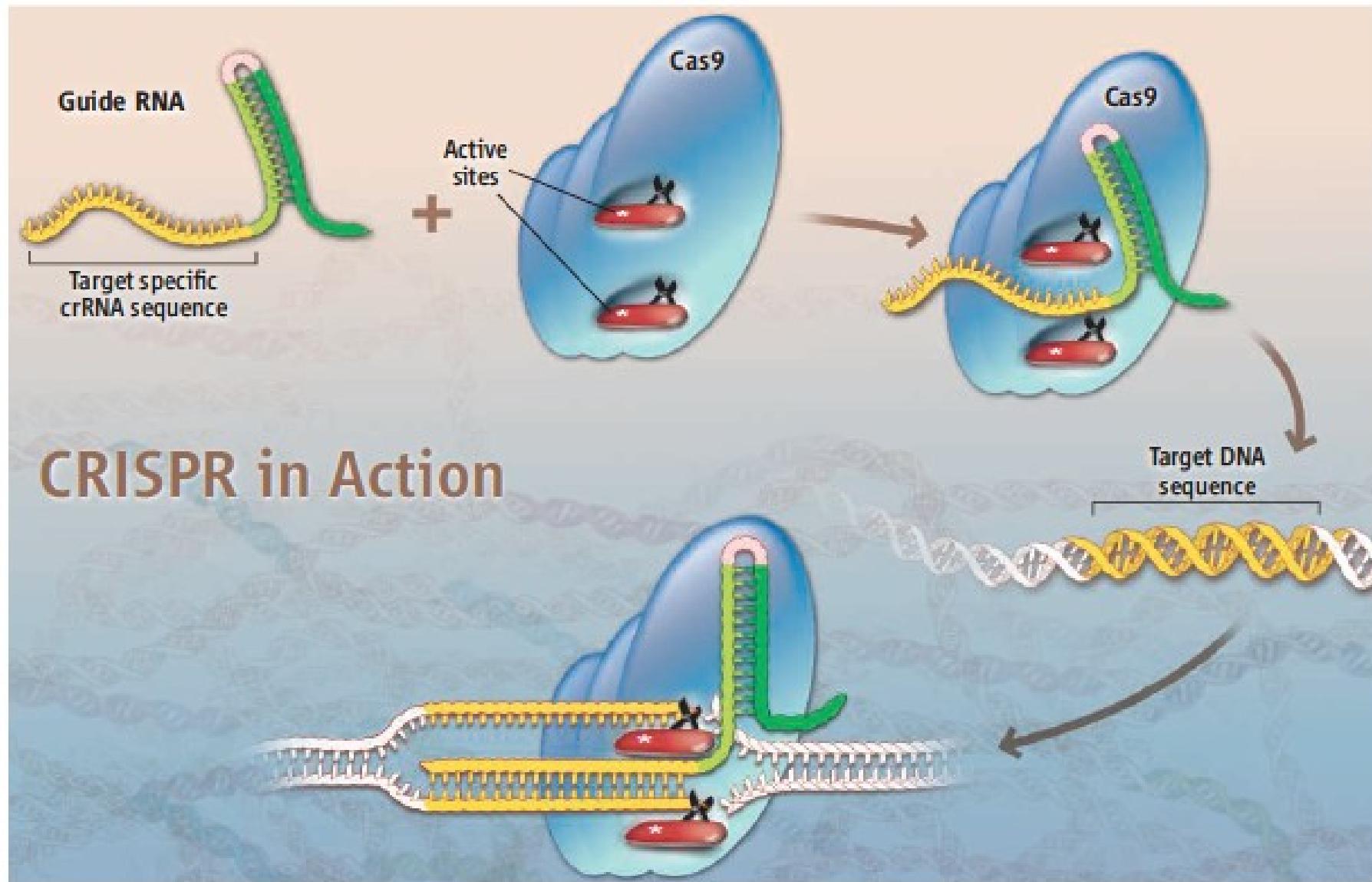
A single  
zinc finger  
motif

A	A	T
T	T	A



Zinc fingers can be hypothetically tailored to any DNA site;  
sequence must be long enough to ensure one site/genome

# CRISPR: clustered regularly interspaced short palindromic repeats



'Genetic microsurgery for the masses.' 20 DECEMBER 2013 VOL 342 SCIENCE  
www.sciencemag.org

# Summary of the main nucleases for genetic surgery

	ZFNs	TALENs	RGENs
<b>DNA targeting specificity determinant</b>	Zinc-finger proteins	Transcription activator-like effectors	crRNA or sgRNA
<b>Nuclease</b>	FokI	FokI	Cas9
<b>Success rate<sup>‡</sup></b>	Low (~24%)	High (>99%)	High (~90%)
<b>Average mutation rate<sup>§</sup></b>	Low or variable (~10%)	High (~20%)	High (~20%)
<b>Specificity-determining length of target site</b>	18–36 bp	30–40 bp	22 bp (total length 23 bp)
<b>Restriction in target site</b>	G-rich	Start with T and end with A (owing to the heterodimer structure)	End with an NGG or NAG (lower activity) sequence (that is, PAM)
<b>Design density</b>	One per ~100 bp	At least one per base pair	One per 8 bp (NGG PAM) or 4 bp (NGG and NAG PAM)
<b>Off-target effects</b>	High	Low	Variable
<b>Cytotoxicity</b>	Variable to high	Low	Low
<b>Size</b>	~1 kb × 2	~3 kb × 2	4.2 kb (Cas9 from <i>Streptococcus pyogenes</i> ) + 0.1 kb (sgRNA)

A guide to genome engineering  
with programmable nucleases

ZFN: zinc finger nuclease  
TALEN: TAL effector nuclease  
RGEN: RNA guided endonuclease

## Gene therapy summary

- 1) Gene therapy targets, types of vectors
- 2) Retroviral gene therapy, advantages and drawbacks
- 3) Precision genome engineering site-specific genome cleavage and recombinational repair

# CRISPR Cas9

- Where CRISPR comes from
- How CRISPR works
- Cas9 variants and their utility
- Delivery of CRISPR Cas9

# Guide to readings

- 1) *Crispr review 2014.* Overview of Crispr/Cas9 in historical context.
- 2) *Gene editing 2020.* A summary of current gene editing approaches based on CRISPR, with focus on base editing
- 3) *Base editing and sickle cell 2021.* A base editor is used to treat sickle cell disease in mice

# What is CRISPR, and where did it come from?

CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats

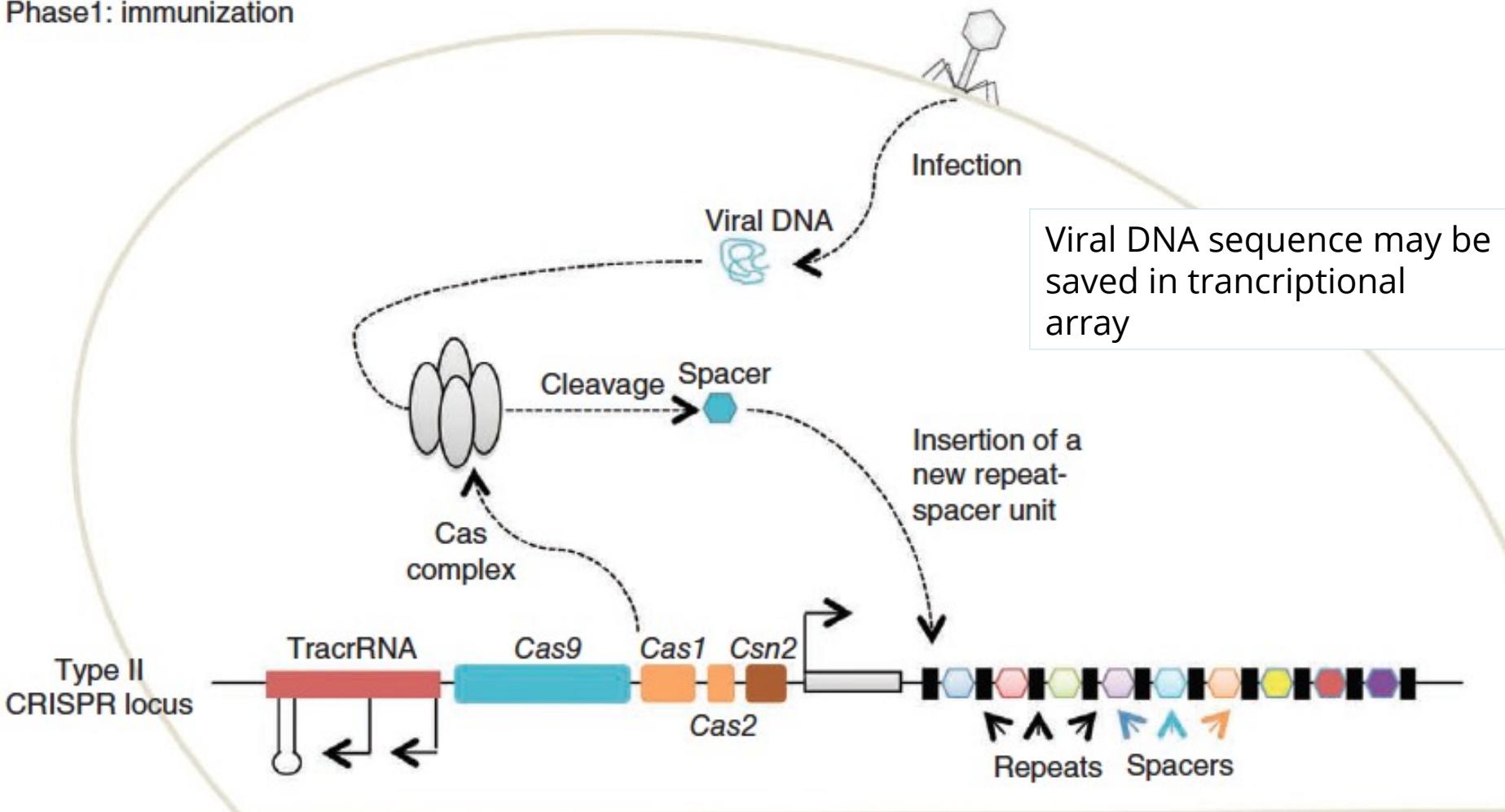
Originally discovered as an unusual sequence element in some sequenced bacterial and archaeal genomes

The repeat sequences are separated by spacers that match bacterial and archaeal virus sequences

The CRISPR sequences are accompanied by a set of CRISPR Associated (Cas) proteins

# How Type II CRISPR Cas works in bacteria: adaptive immunity

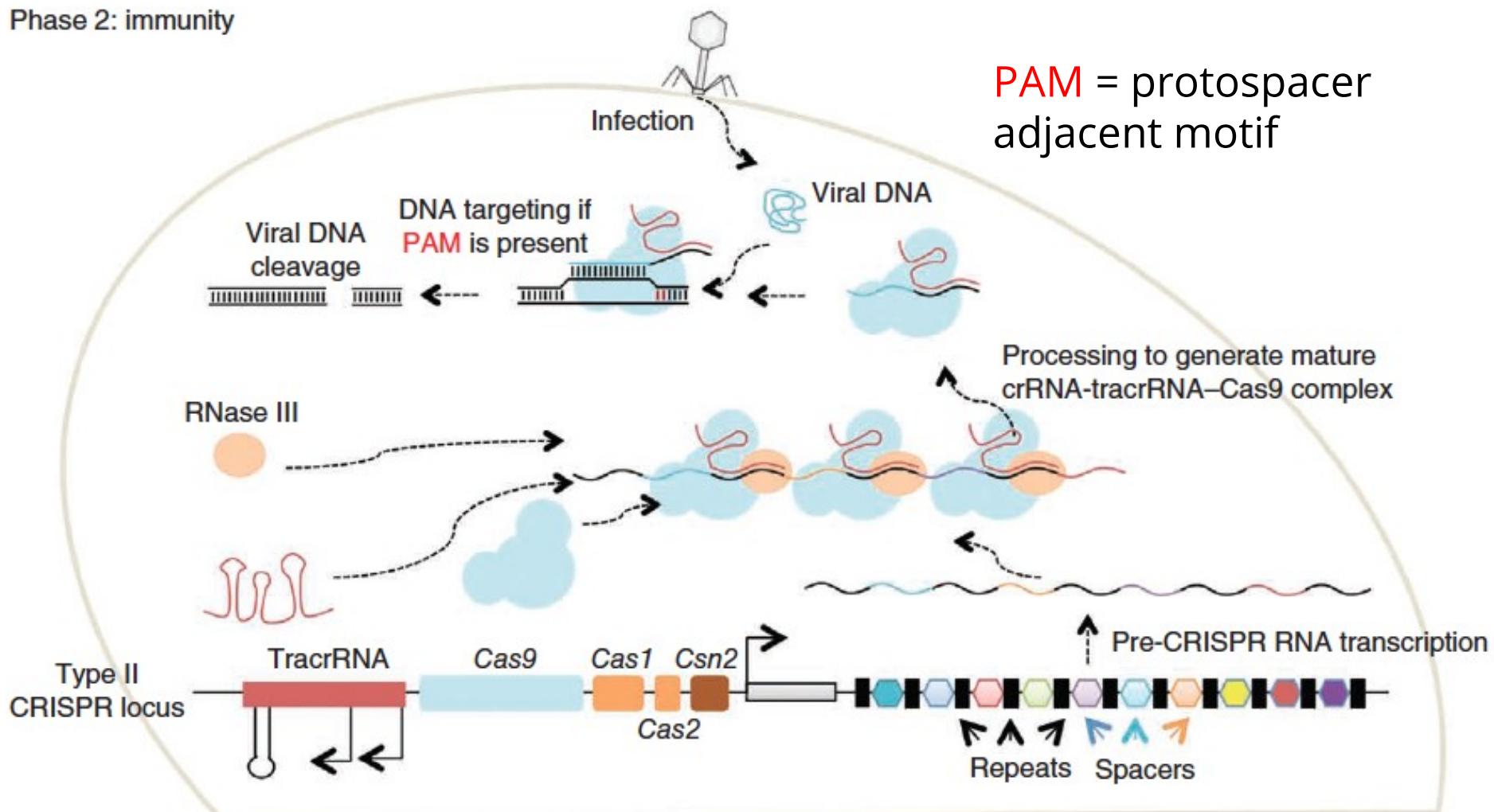
Phase1: immunization



Cas9 as a versatile tool for engineering  
biology

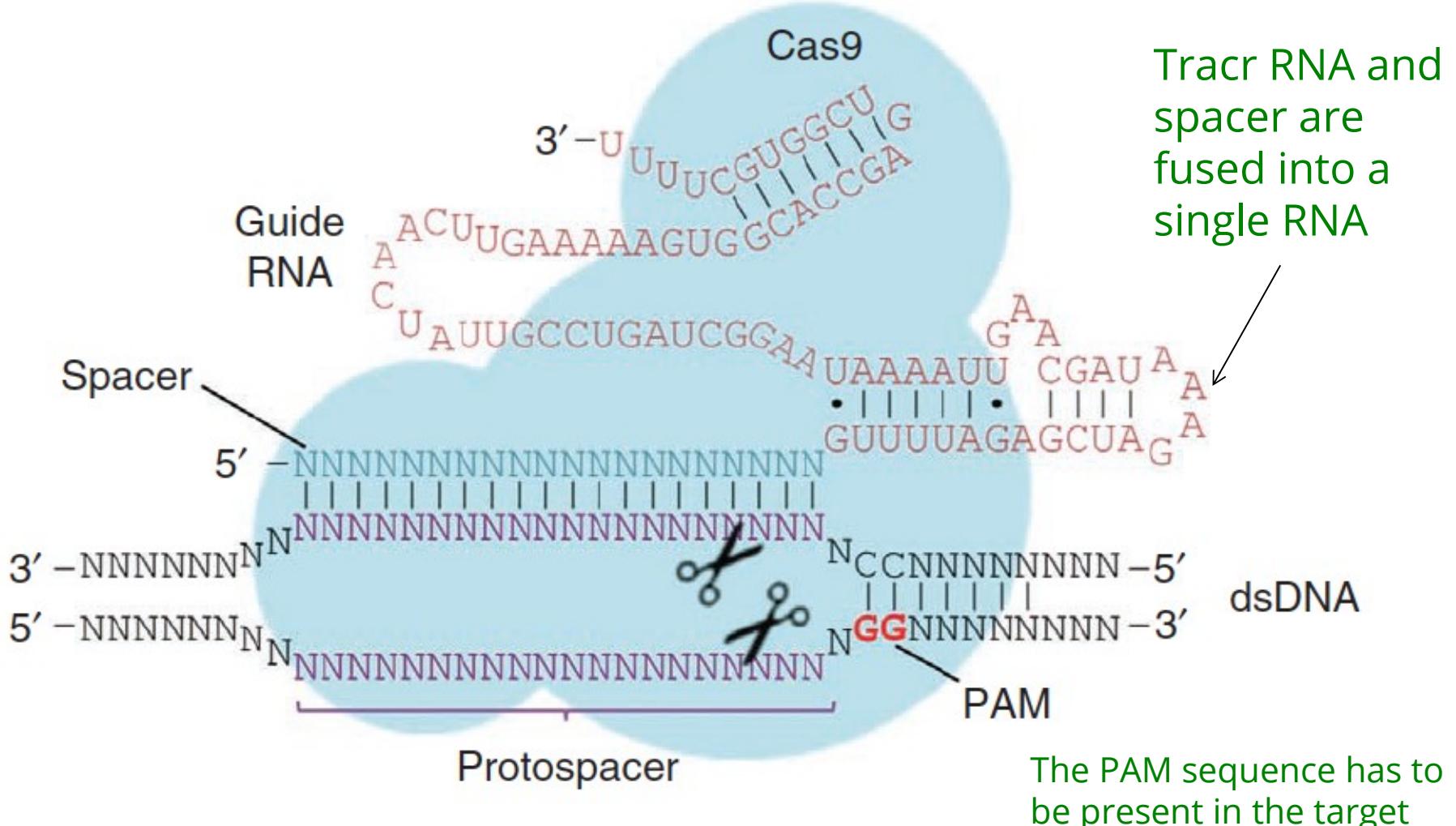
# How Type II CRISPR Cas works in bacteria

Phase 2: immunity



Cas9 as a versatile tool for engineering biology

# Basic engineering with *S. pyogenes* Cas9 RNA-guided nuclease



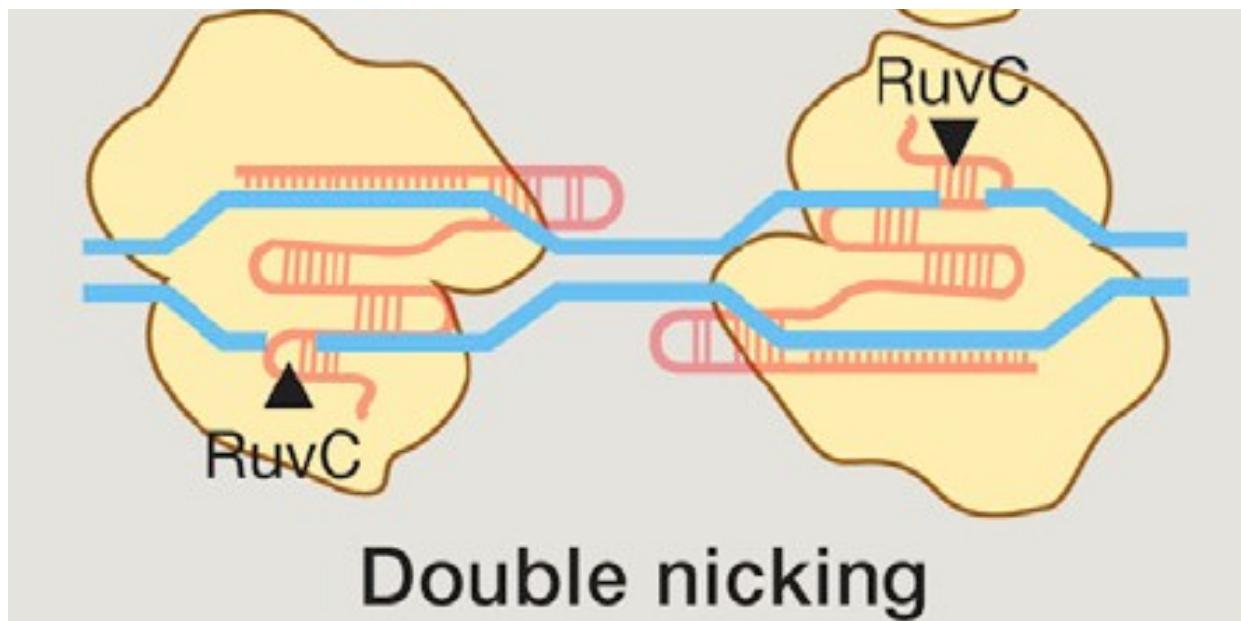
Co-express Cas9 and the guide RNA for a guided nuclease

# The basic Cas9-guide RNA complex

- Has 2 nuclease active sites
- Recognizes 20 nt protospacer sequence target
- Specificity of sequence recognition is not absolute
  - Can tolerate 1 to 3 mismatches (sometimes more) and will still cut DNA (off target effects)
  - Mismatches at 5' end are better tolerated than those at 3' end (because of 'zipper' mechanism of base pairing)

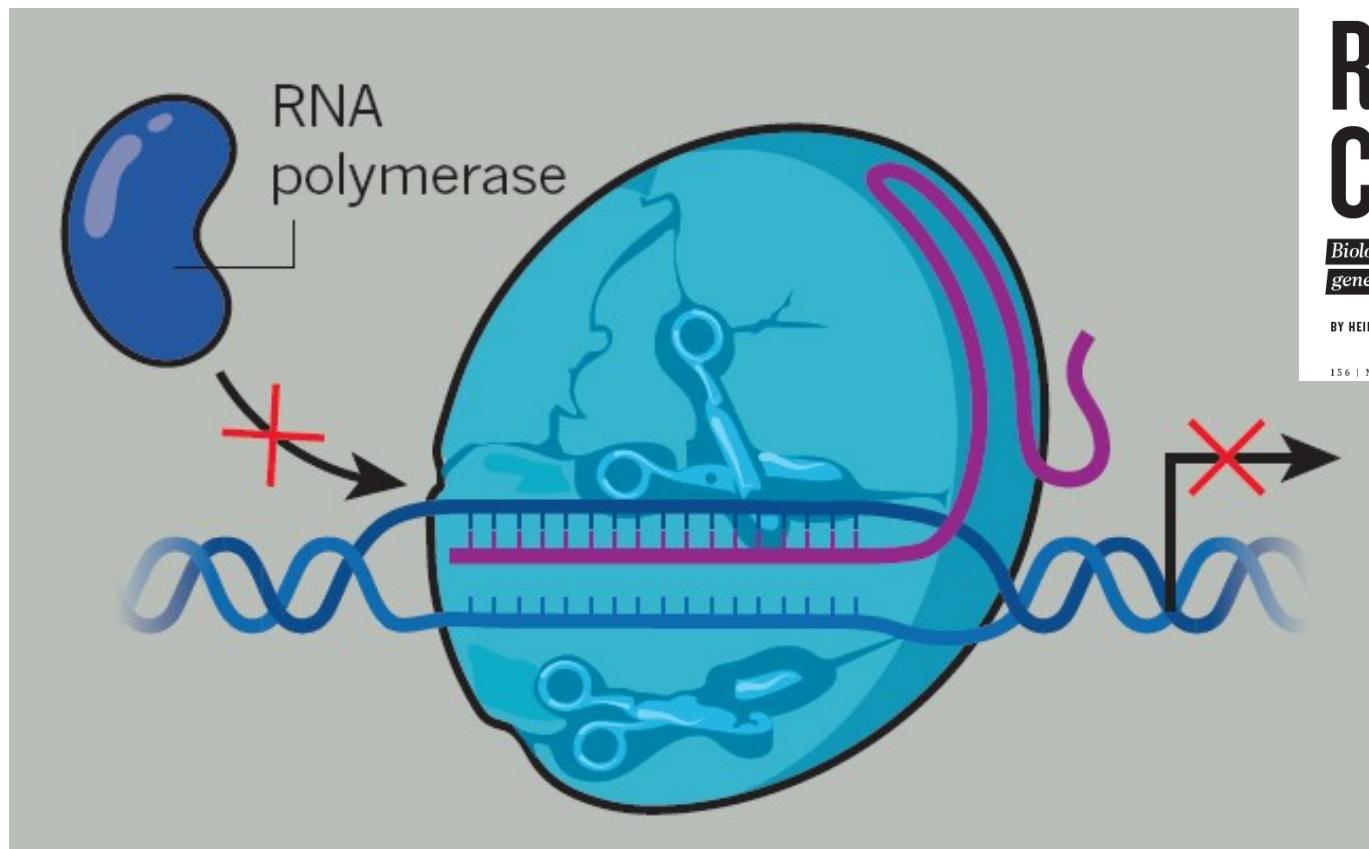
# Mutations in nuclease domains can be useful: 'nickase'

- 'Nickase' version of Cas9: cuts one strand of DNA only
- Use two nickases simultaneously for offset nicks, doubling specificity and reducing off target toxicity



# Mutations in nuclease domains can be useful: 'nuclease null'

- Cas9 with no nuclease activity: guide RNA finds its target with high specificity
- Can be used as a repressor of transcription



## RIDING THE CRISPR WAVE

*Biologists are embracing the power of gene-editing tools to explore genomes.*

BY HEIDI LEDFORD

156 | NATURE | VOL 531 | 10 MARCH 2016

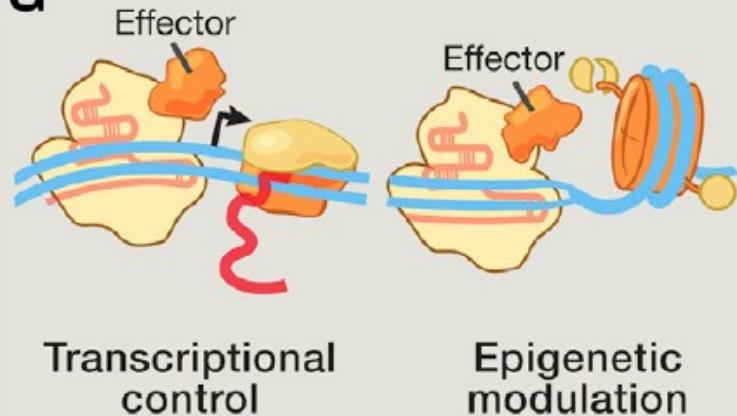
# More uses for nuclease null Cas9

## Gene expression

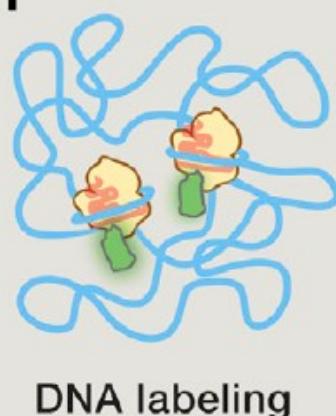
## DNA labeling

## Controllable cutting

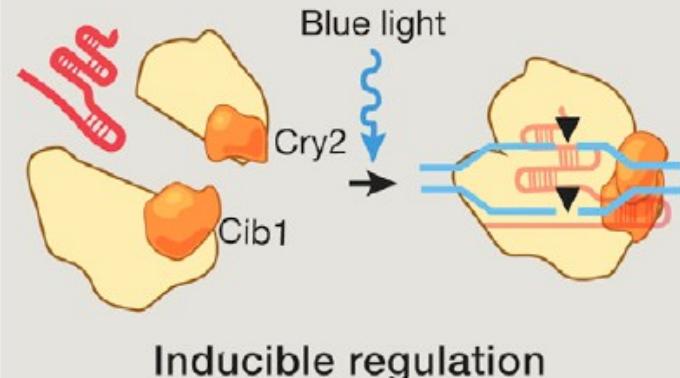
G



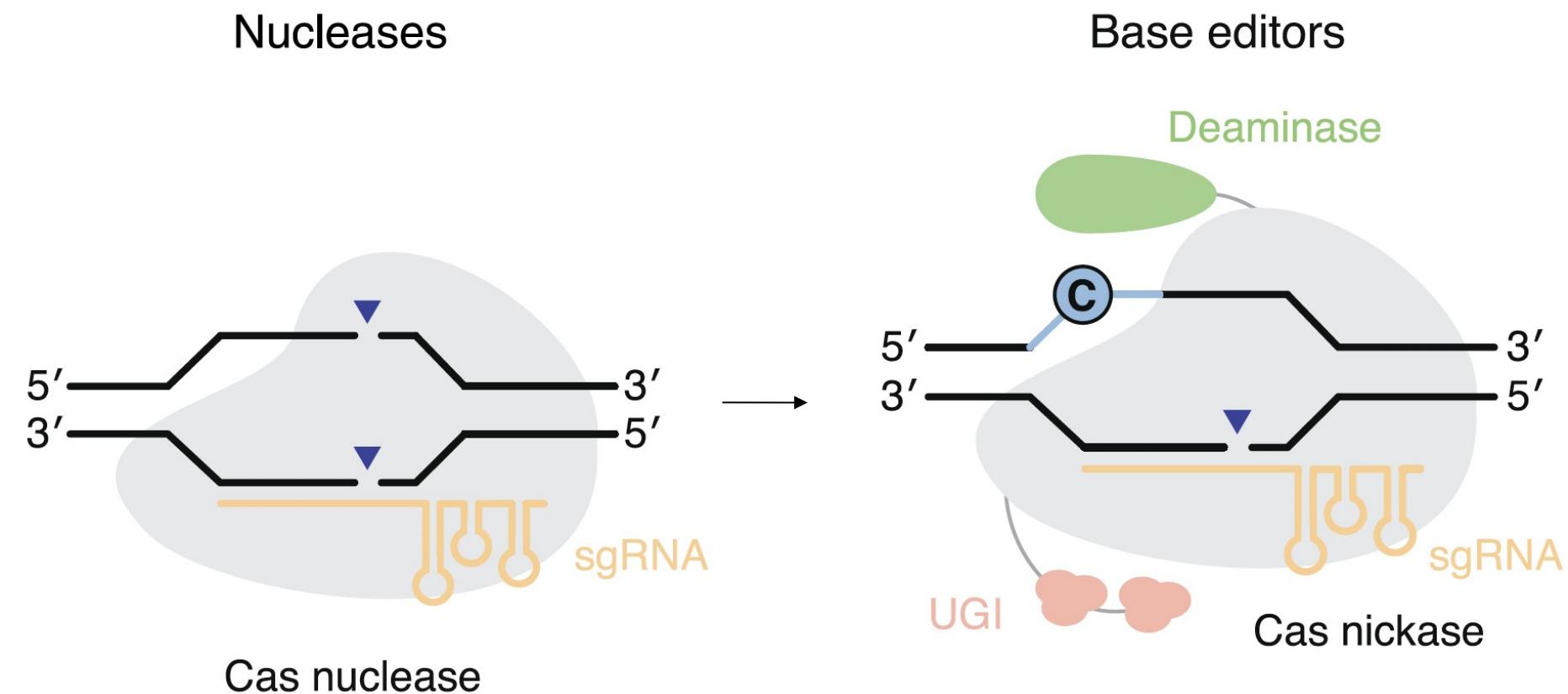
H



I



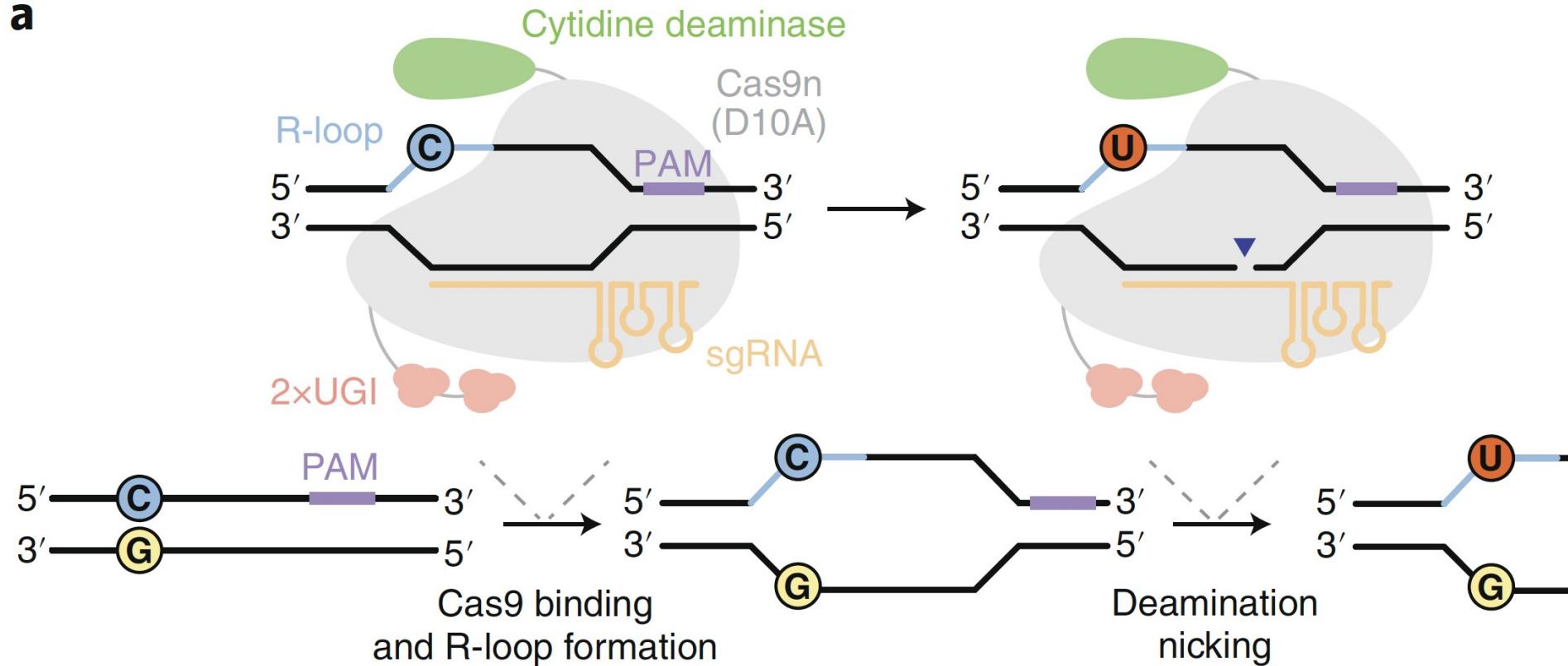
# Cas9 can also be engineered to be a base editor



- Deaminase converts single-stranded C to U
- UGI (uracil glycosylase inhibitor) prevents uracil excision

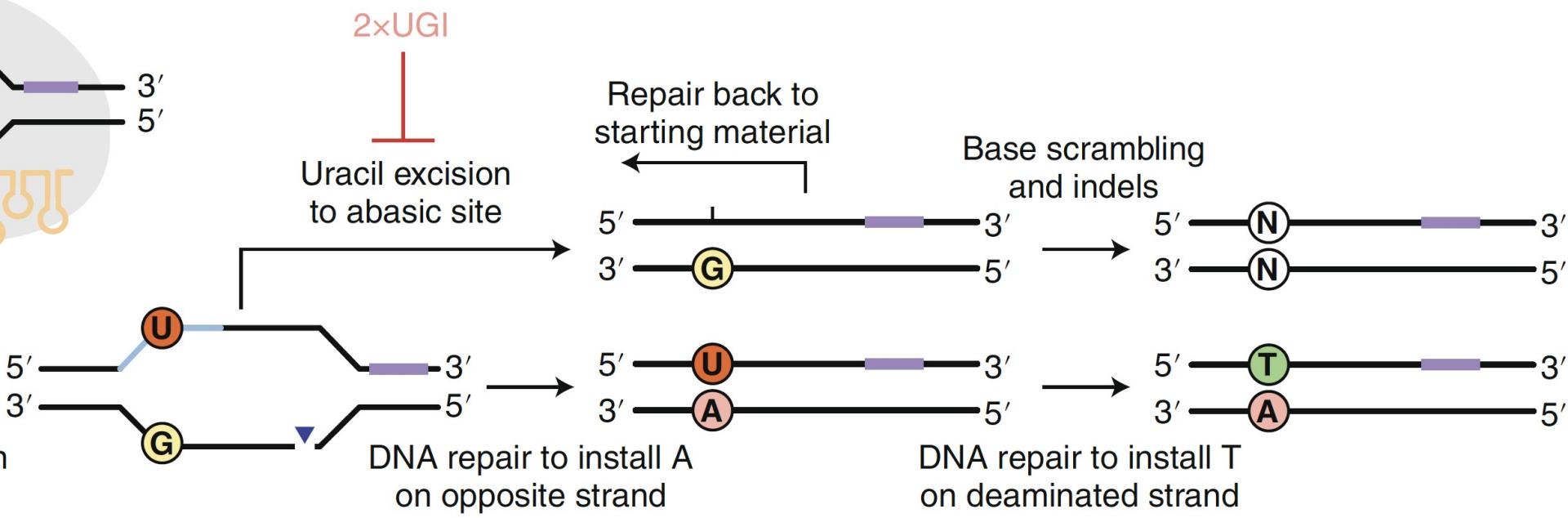
# Example of a C-G to T-A base editor in action, part I

a



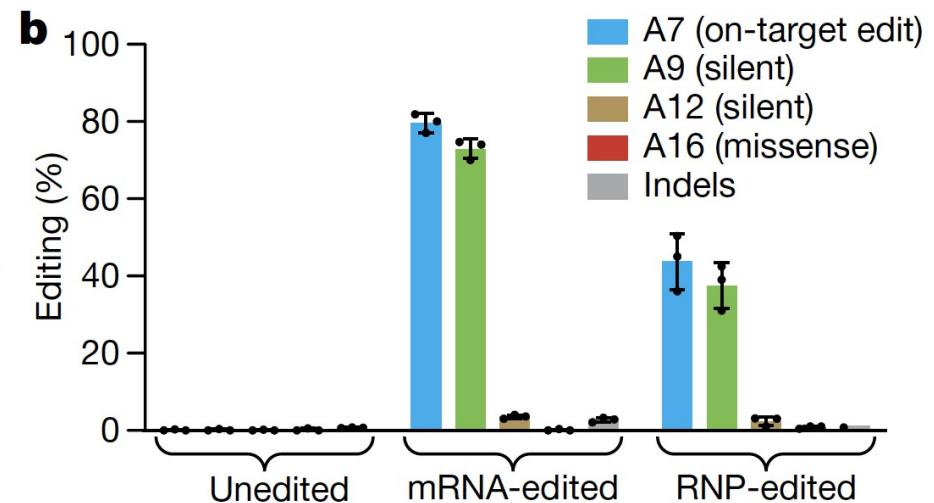
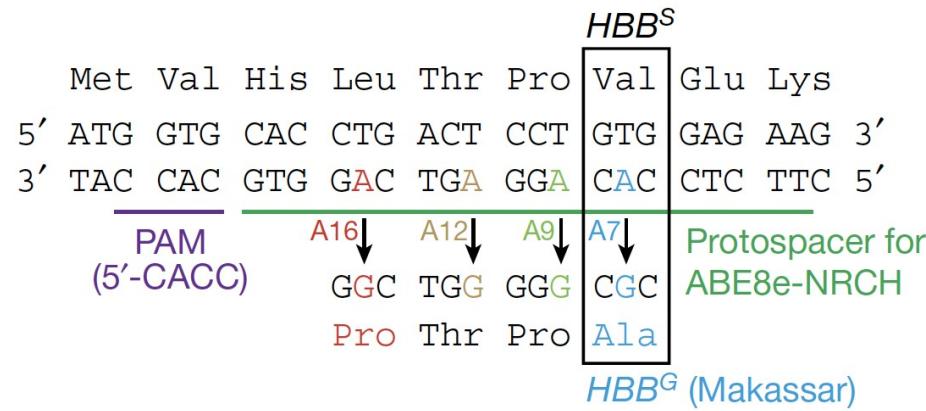
C to U change is catalyzed, and a nick is introduced

## Example of a C-G to T-A base editor in action, part II



- DNA repair pathway operates on nicked strand
- UGI prevents uracil excision pathway (which leads to repair or 'base scrambling'/indel formation)

# Use of a base editor to correct sickle cell mutation in hematopoietic stem and progenitor cells (HSPCs)

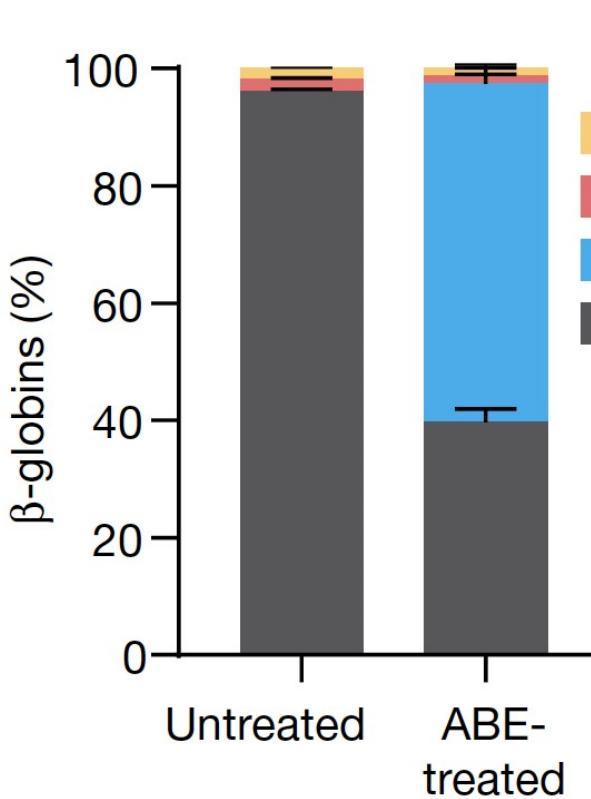
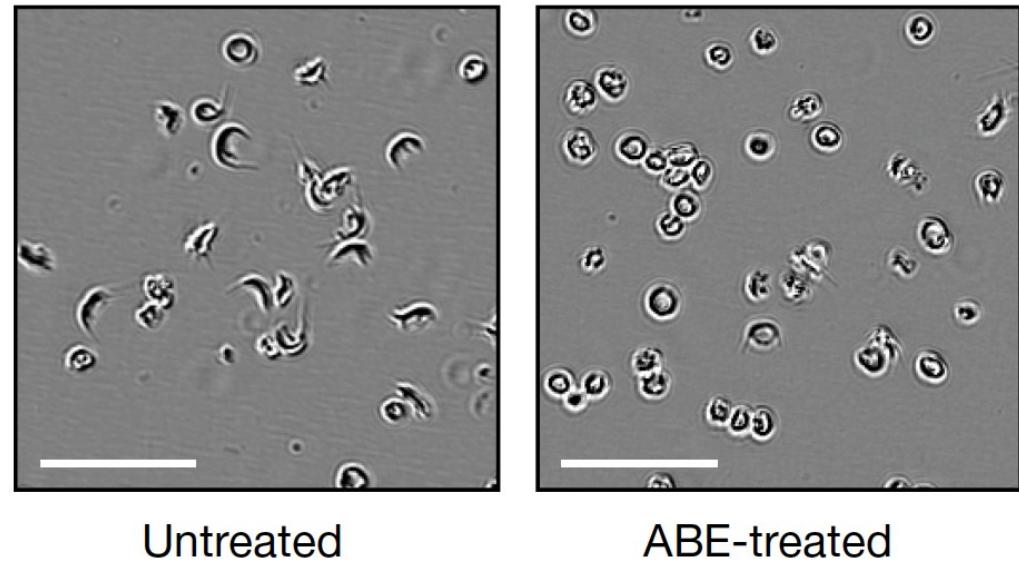


Targeting the 'A' in CAC (val) to convert to CGC (ala)

Off-target 'A' conversions also occur, but are either silent or very rare

# Engineered cells returned to sickle cell mice

Much more normal beta globin (blue), and far less sickling

**g****h**

Untreated

ABE-treated

## Article

# Base editing of haematopoietic stem cells rescues sickle cell disease in mice

<https://doi.org/10.1038/s41586-021-03609-w>

Received: 10 January 2021

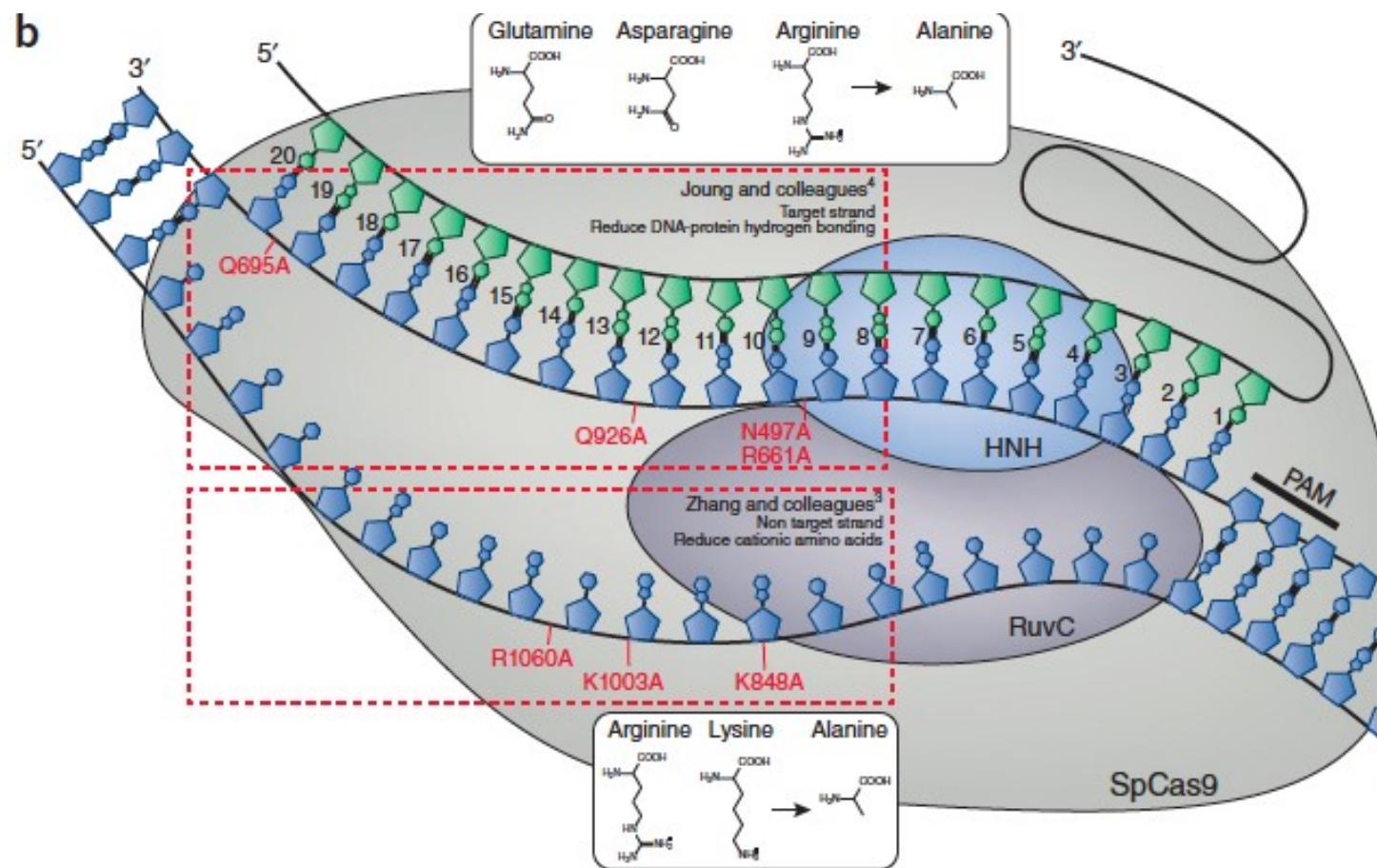
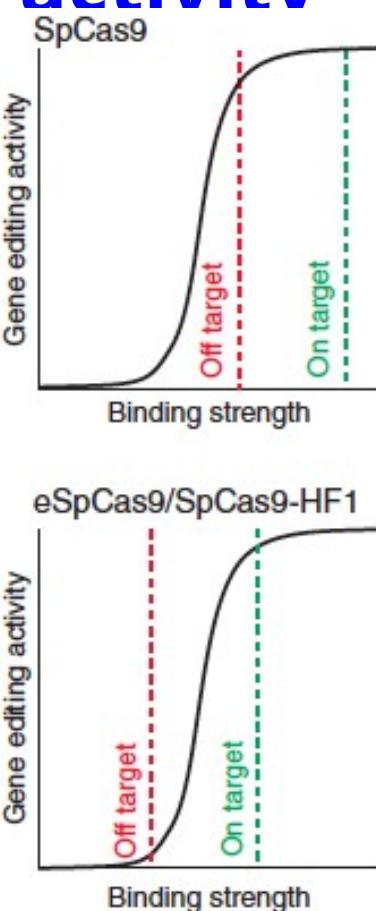
Accepted: 4 May 2021

Published online: 02 June 2021

Check for updates

Gregory A. Newby<sup>1,2,3,7</sup>, Jonathan S. Yen<sup>4,12,8</sup>, Kaitly J. Woodard<sup>4,12</sup>, Thiagaraj Mayurathan<sup>4,12</sup>, Cicera R. Lazzarotto<sup>4</sup>, Yichao Li<sup>4</sup>, Heather Sheppard-Tillman<sup>5</sup>, Shaina N. Porter<sup>6</sup>, Yu Yao<sup>4</sup>, Kalin Mayberry<sup>4</sup>, Kelcee A. Everett<sup>1,2,3</sup>, Yoonjeong Jang<sup>4</sup>, Christopher J. Podrucky<sup>1,2,3</sup>, Elizabeth Thaman<sup>7</sup>, Christophe Lechavre<sup>8</sup>, Akshay Sharma<sup>9</sup>, Jordana M. Henderson<sup>9</sup>, Michelle F. Richter<sup>1,2,3</sup>, Kevin T. Zhao<sup>1,2,3</sup>, Shannon M. Miller<sup>1,2,3</sup>, Tina Wang<sup>1,2,3</sup>, Luke W. Koblan<sup>1,2,3</sup>, Anton P. McCaffrey<sup>9</sup>, John F. Tisdale<sup>10</sup>, Theodosia A. Kalfa<sup>7,11</sup>, Shondra M. Pruitt-Miller<sup>6</sup>, Shenghra Q. Tsai<sup>1</sup>, Mitchell J. Weiss<sup>4,12</sup> & David R. Liu<sup>1,2,3,12</sup>

# Mutation of Cas9 to reduce DNA binding strength can help reduce off-target activity



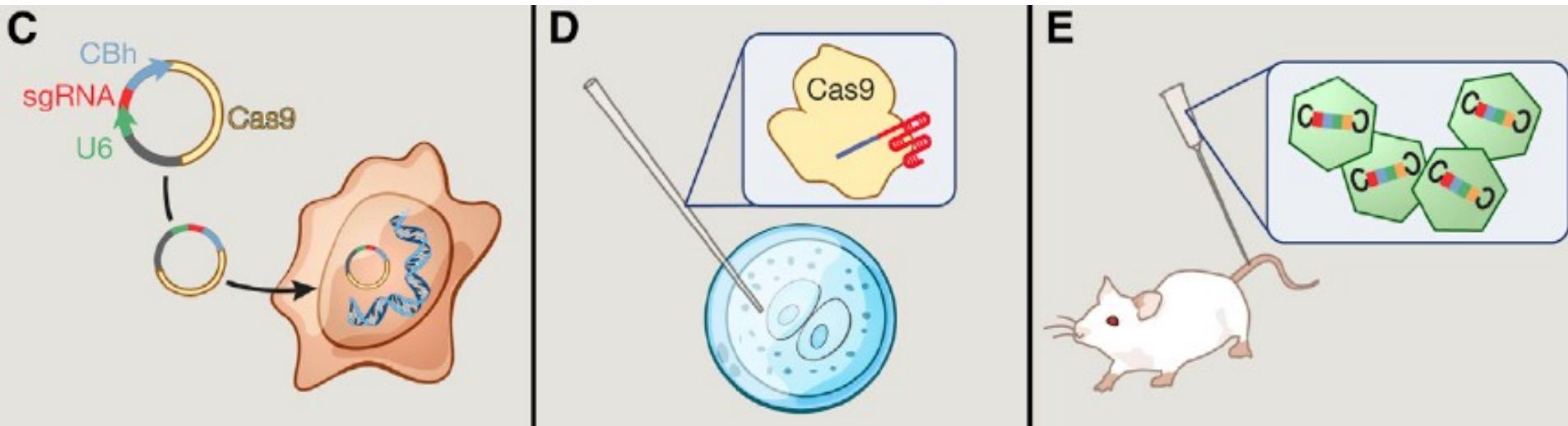
Cas9 loosens its grip on off-target sites

# Delivery of Cas9 and gRNA to cells

Expression in target cells (C and E)

- Transformation of cultured cells with plasmid
- Viral vectors for expression in animal

Direct injection of Cas9 protein with sgRNA (D)



# CRISPR Cas9

- Where CRISPR comes from
- How CRISPR works
- Cas9 variants and their utility
- Delivery of CRISPR Cas9

# **Vectored and nucleic acid vaccines**

- 1) Attenuated bacterial vector
- 2) Recombinant virus vaccines
- 3) DNA vaccines
  - a) Mode of delivery
  - b) How antigens generate immunity

## Guide to readings: DNA vaccines

- 1) DNA vaccines review (2018). Summary of the current state of the art
- 2) Nucleic acids therapeutic landscape (2021)
- 3) National Cancer Institute guide to CAR T-cell and related adopted cell therapies (ACT):  
<https://www.cancer.gov/about-cancer/treatment/research/car-t-cells>

# Vaccination to prevent viral and bacterial diseases

Exposure to a pathogen exposes antigens to the immune system

The immune response results in **humoral** [B-cell] and/or **cell-mediated** [T-cell] immunity, so any future infection by the pathogen is less severe

- humoral immunity: production of antibodies that mark antigens for destruction
- cell-mediated immunity: T-cells recognize and kill cells presenting foreign antigens on their surfaces)

Vaccination: generate immunity without a full-blown infection

- Antigens by themselves can contribute to future immunity
- Vaccinating antigens are often protein components from the pathogen, for example surface-exposed ligands
- When antibodies generated by the immune system bind to antigen, the pathogen should be neutralized

“Vaccines historically represent one of the most established and cost-effective procedures in medicine, having perhaps the greatest impact on human health of any medical intervention.”

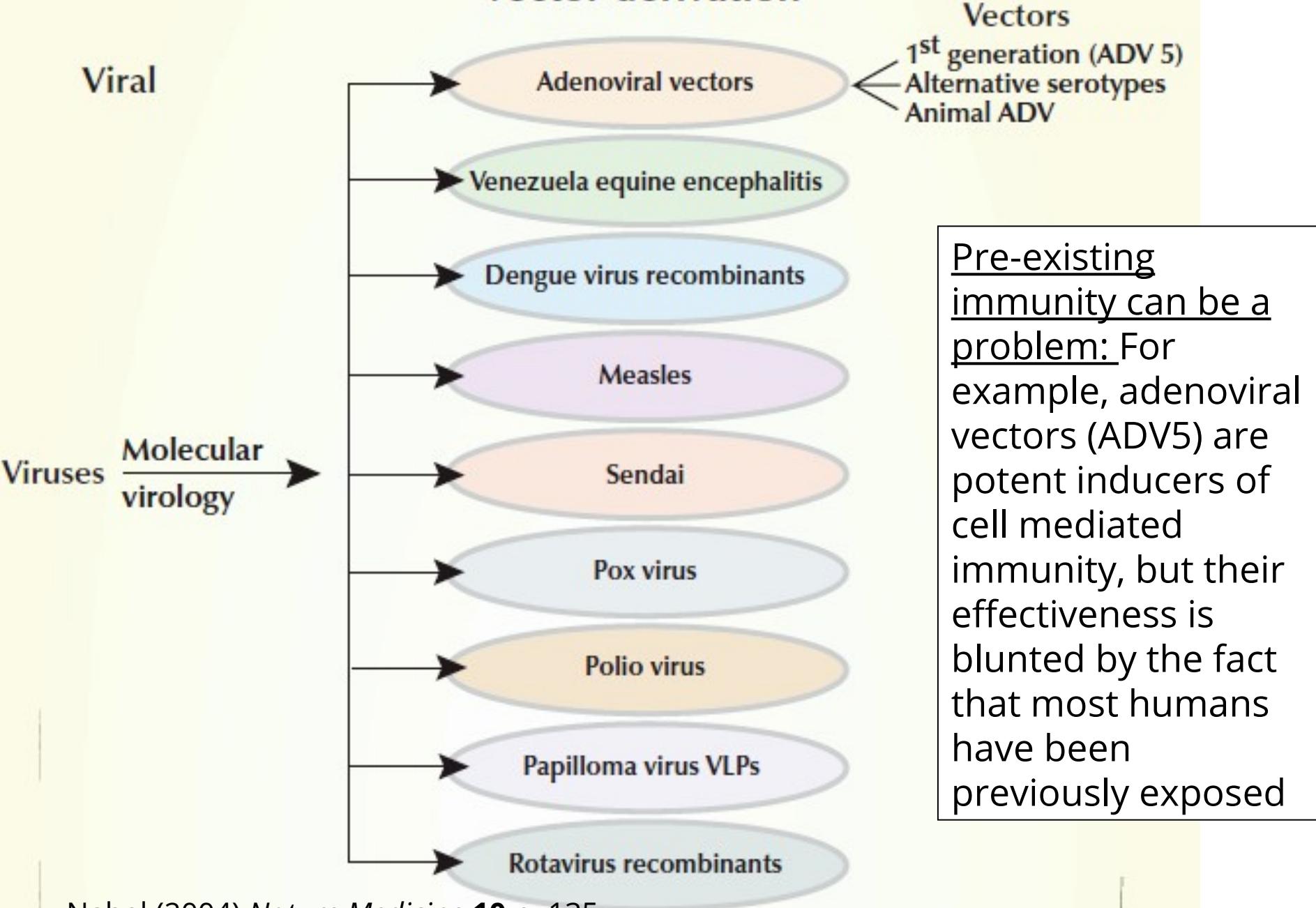
(Nabel, 2004)

# Bacterial vectors for vaccination

Use “attenuated” pathogen

- *Salmonella typhimurium* with mutated virulence factors
  - Avirulent:  $10^{10}$  cells won’t kill mouse
  - (Virulent:  $10^4$  cells kill)
- Avirulent *S. typhimurium* strain causes “self-limiting” infections, elicits immune response
- Clone antigen protein genes into attenuated pathogen, confer immunity to pathogen of choice

# Viral delivery of antigens



# Nucleic acid

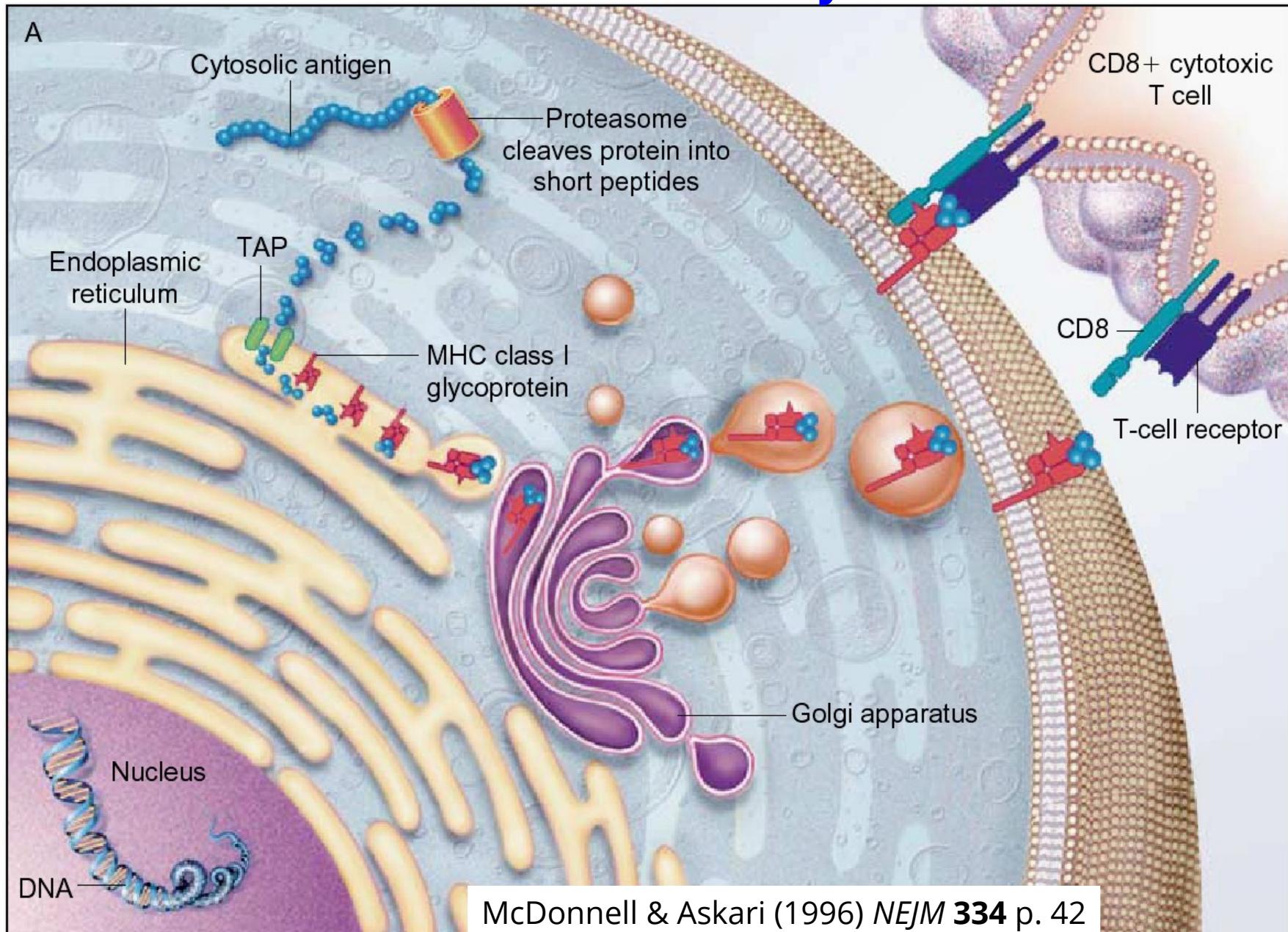
## vaccines

- Introduced DNA or RNA can bring immune response if transcribed/translated, & protein is immunogenic
- Example : a non-replicating vector containing an **antigen gene** (e.g. a viral coat protein)
  - Antigen mRNA made from a highly active promoter
  - mRNA has splicing and polyadenylation signals
- Antigenicity is best if protein is secreted or membrane-bound in an antigen presenting cell (APC)
- DNA/RNA may be delivered by direct injection, liposome nanoparticles, biolistics, etc.
- Many vaccine candidates in progress: SARS CoV2, flu, measles, HIV, Ebola, *M. tuberculosis*, prions, etc.
- Many in clinical trials, SARS CoV2 likely to be first to gain full approval (currently: EUA, emergency use)

# mRNA vaccines: giant steps

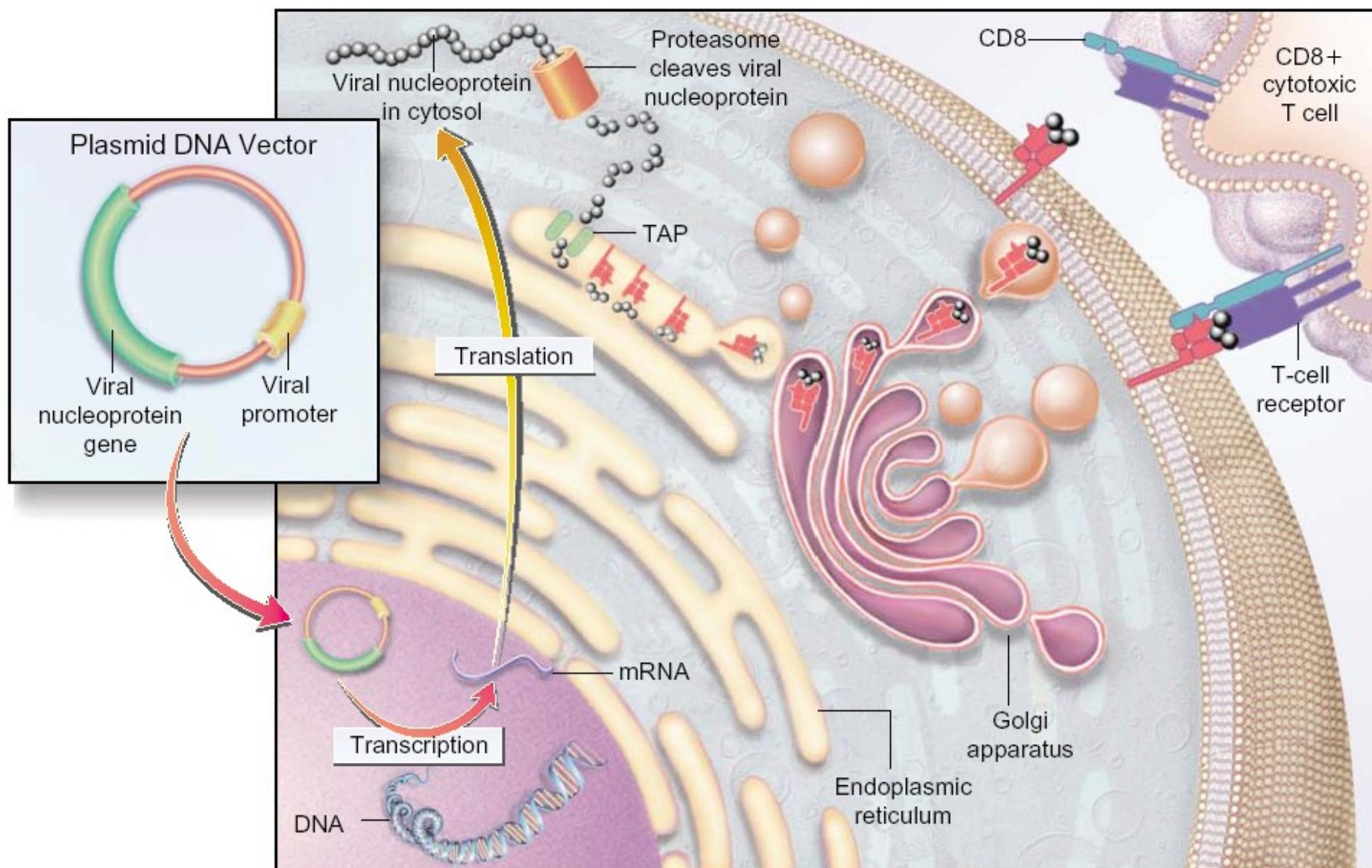


# MHC class I presentation of antigens to T-cells – stimulates T cell-mediated immunity



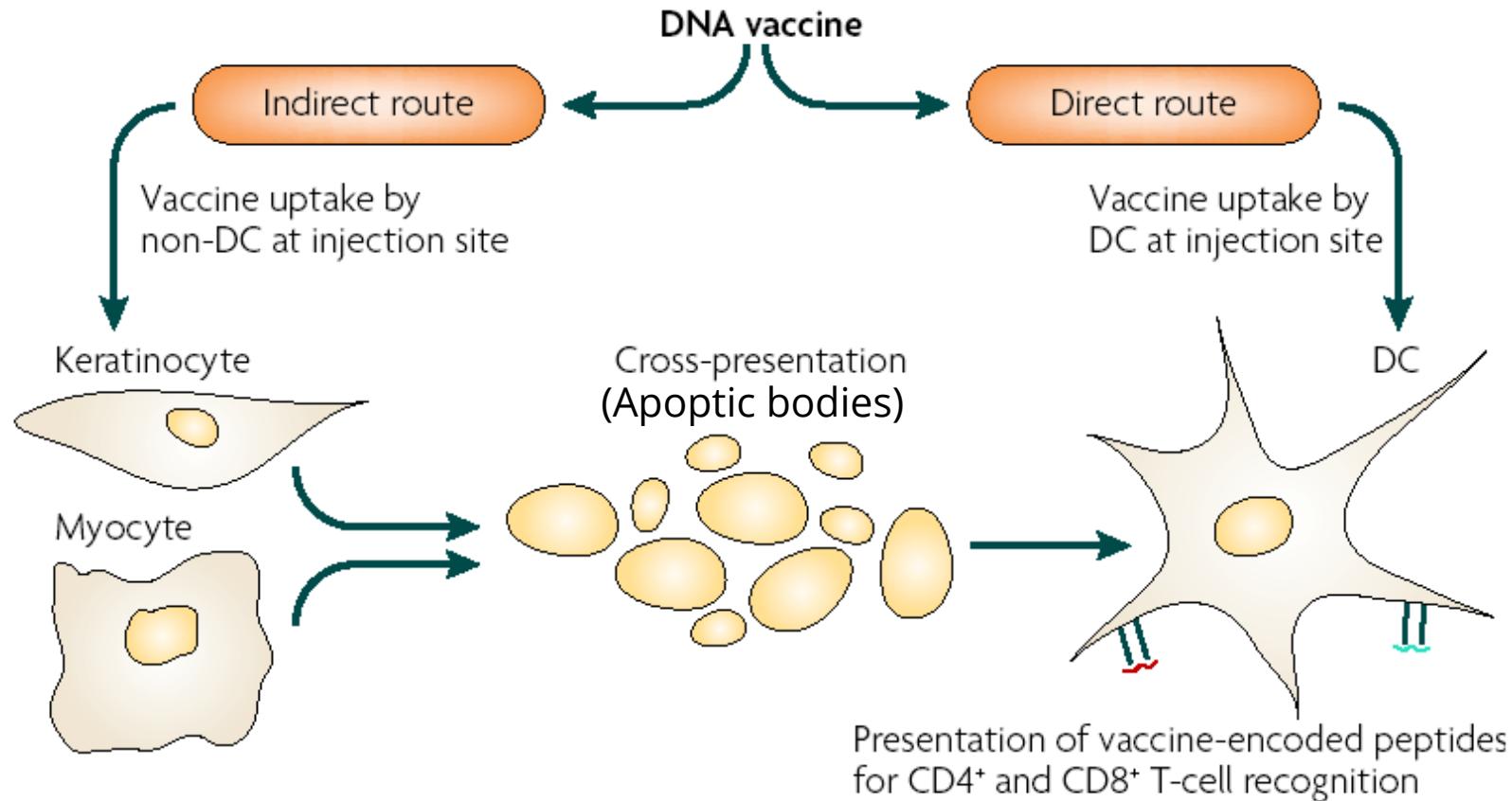
# A route for gene expression and T-cell activation for a DNA vaccine

McDonnell & Askari (1996) *NEJM* 334 p. 42



The activated T-cells provide immunity

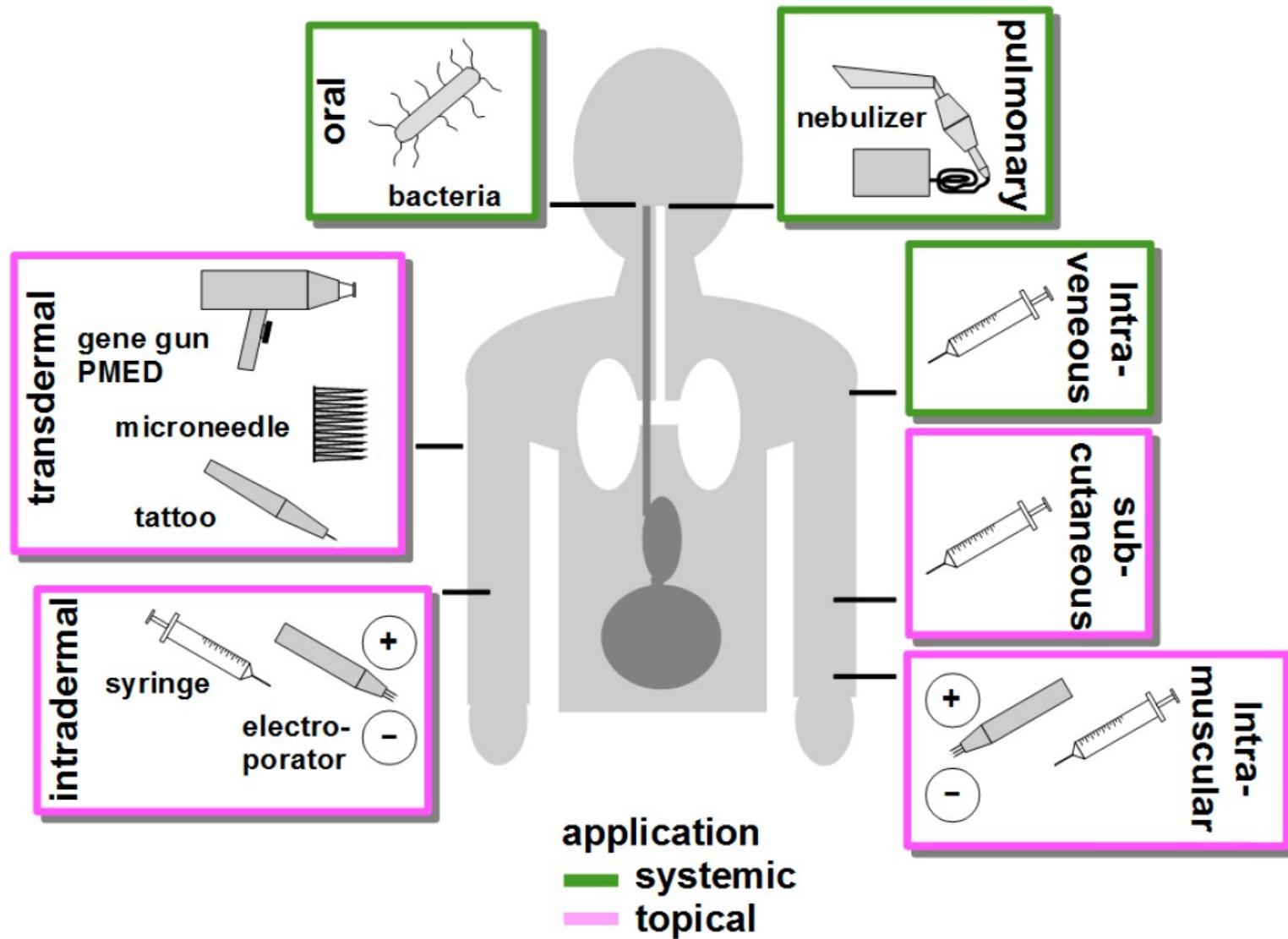
# Which cells express the antigen? DC: Dendritic Cells (also called APCs, Antigen Presenting Cells)



**Figure 1 | Routes of antigen presentation.** DNA vaccines are generally injected into muscle or skin. Encoded antigen is then expressed in myocytes or keratinocytes. For activation of T cells, antigen must be transferred to a 'professional' antigen-presenting cell, usually a dendritic cell (DC). This indirect process of transfer of antigenic material, possibly as apoptotic vesicles, is termed cross-presentation. A small proportion of DNA is also taken up directly by DCs and the encoded antigen can then be processed and presented endogenously.

Rice et al. (2008) *Nat. Rev. Cancer* **8** p. 108.

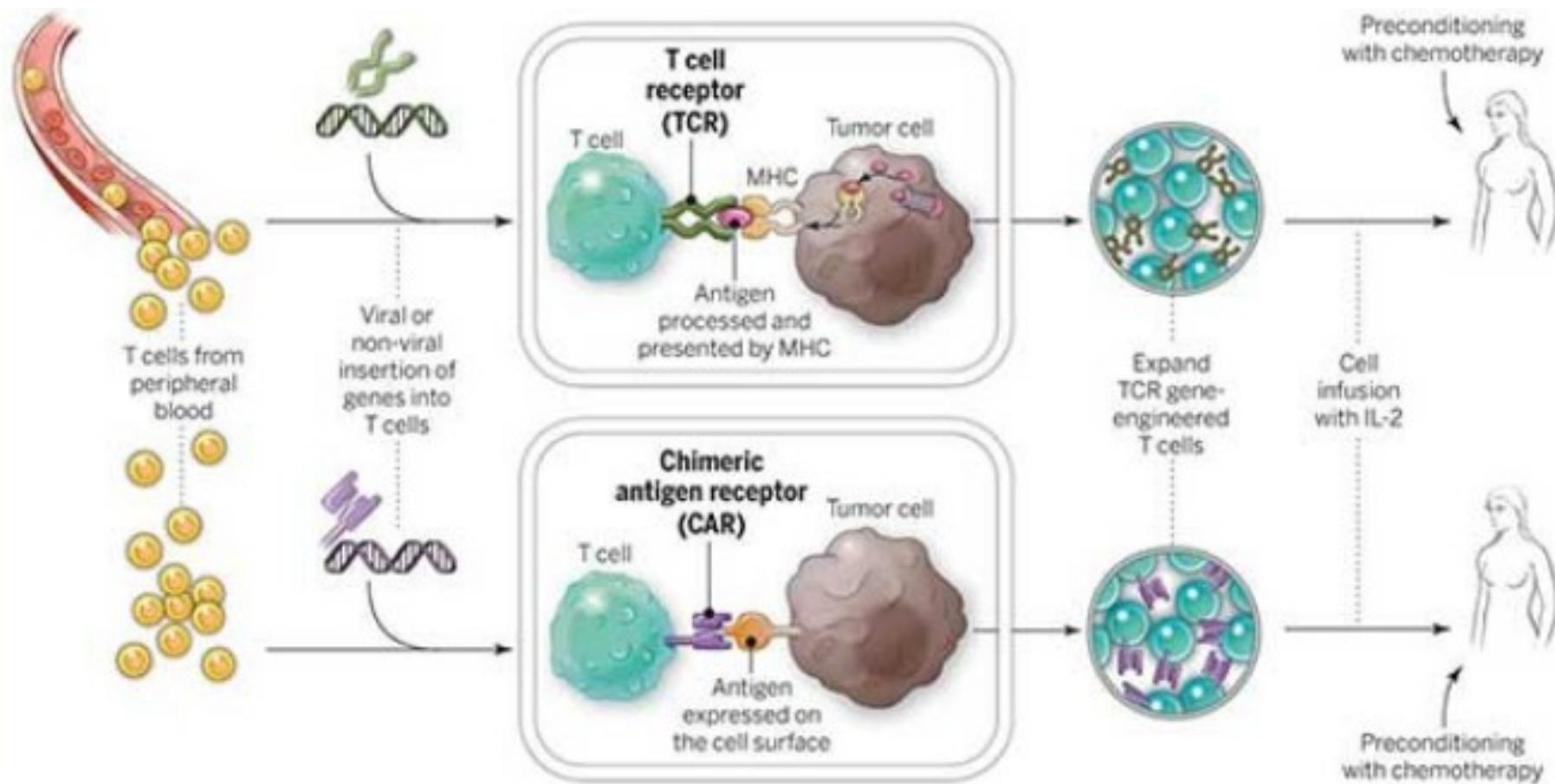
Nucleic acid vaccines may be given systemically, or topically



# Using the immune system to control/prevent cancer

- Aim: destroy or inhibit tumor cells
- Tumor antigens: expressed on tumor cell surfaces or complexed with MHC class I or II and presented on the surface of cancer cells
- Adoptive Cell Transfer (ACT): collect and use patients' own immune cells (T cells) to treat their cancer
- ***DNA intervention:*** activate or engineer antibody/T-cell responses towards selected tumor cell antigens
- Engineered antibodies/T-cells can recognize the tumor cell antigens and kill the tumor cells

T cells removed, engineered with either a new TCR or a CAR (Chimeric Antigen Receptor), expanded, added back



CAR T cells and TCR T cells are engineered to produce special receptors on their surfaces. They are then expanded in the laboratory and returned to the patient.

Credit: National Cancer Institute

# CAR (Chimeric Antigen Receptor) T-cell therapy

- This is a type of Adoptive Cell Transfer (ACT)
- Engineer T-cell with 'chimeric antigen receptor' that is specific for a tumor antigen
- The antigen receptor contains a fusion protein:
  - synthetic antibody domains (recognize tumor antigen)
  - signalling domains (stimulate killing of target cells)
- No need for T-cell activation by antigen presentation by MHC on tumor cell
- <https://www.cancer.gov/about-cancer/treatment/research/car-t-cells>

# Gene therapy cures leukaemia in eight days

- › 26 March 2013 by [Andy Coghlan](#)
  - › Magazine issue 2910. [Subscribe and save](#)
  - › For similar stories, visit the [Cancer](#) and [Genetics](#) Topic Guides
- 

WITHIN just eight days of starting a novel gene therapy, David Aponte's "incurable" leukaemia had vanished. For four other patients, the same

<http://www.newscientist.com/article/mg21729104.100-gene-therapy-cures-leukaemia-in-eight-days.html>

- Acute lymphoblastic leukemia: B-cells (which produce antibodies) become malignant. B-cells express CD-19 on their surfaces
- Adoptive Cell Transfer (ACT): T cells *engineered* ex vivo to contain CD19-specific receptors. They target B-cells exclusively (alternative to expressing CD19 as an antigen)
- Immune response from these T cells destroys all B cells  
<https://www.fredhutch.org/en/news/center-news/2016/02/immunotherapy-remission-blood-cancer-AAAS-riddell.html>

## Advantages for nucleic acid vaccines

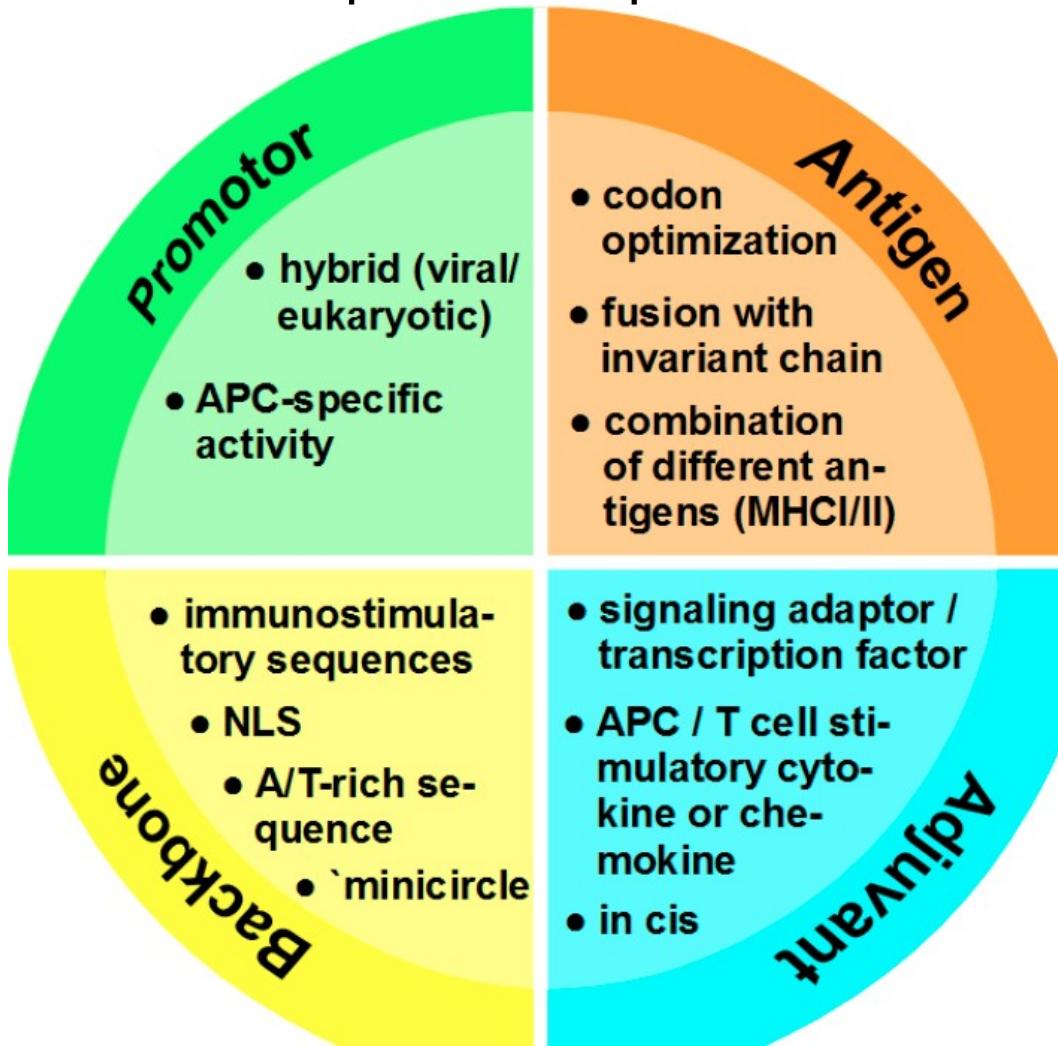
- No risk for infection during vaccination
- Other genes can be added along with the antigen that will influence the immune response (e.g. cytokines, lymphokines, to help generate an immune response)
- Could be useful in boosting the immune system to fight diseases that are already established in a patient

## Disadvantages

- The immune response for a nucleic acid vaccine can be low & thus not strong enough for effective immunity – it seems to depend on the antigen
- Limited to proteins (e.g. no bacterial polysaccharide antigens)

# Optimizing DNA vaccines:

- make sure the antigen gene is expressed at high levels and in the right cell type (APC)
- Make sure that a robust immune response accompanies expression of the transgene



- Since nucleic acids tend to break down quickly, nano carriers (like lipid nanoparticles) help to stabilize and improve efficacy

# Recombinant and nucleic acid vaccines: Summary

- 1) Attenuated, recombinant bacterial and virus approaches
- 2) Nucleic acid vaccines
  - a) Mode of delivery
  - b) How antigens generate immunity
  - c) Cancer vaccines

Introduction:

Dr. Jonathan Nelson  
OHSU Dept. of Nephrology and Hypertension

Employment opportunity

# Basic gene cloning: plasmids and transformation

- Plasmid biology
- How to add DNA to a plasmid
- Transformation of *E. coli*

The key role of *E. coli* and an RNA expression plasmid in mRNA vaccine production (Pfizer/BioNTech, specifically)

<https://www.nytimes.com/interactive/2021/health/pfizer-coronavirus-vaccine.html?action=click&module=Spotlight&pgtype=Homepage>

-

# Readings:

- 1) *22 MC4 Plasmids*. History of plasmid development, plasmid replicons, care of *E. coli* and plasmids, specialized plasmids, transformation, alpha complementation.
- 2) *25 MC4 Antibiotics*. Summary of several antibiotics used for plasmid maintenance, and how they work
- 3) *24 MC4 Cloning in Plasmids*. Practical aspects of cloning and transformation, plasmid dephosphorylation. Discussion of recombinational and ligation independent cloning.
- 4) Gibson assembly 2009. Report demonstrating easy DNA assembly using overlapping sequences.
- 5) IVA (in vitro assembly) 2016. Cloning without ligation.
- 6) Link: [www.addgene.org](http://www.addgene.org), a nonprofit plasmid repository

**Cloning:** isolation and propagation of a specific piece of DNA

**Cloning vector:** a movable DNA element that is modified to contain another piece of DNA

**General features:**

- Easy to isolate and purify
- Can be manipulated in the test tube
- Can be moved into a new cell
- Selectable: easy to propagate and test for presence or absence

**Types of cloning vectors:**

- Plasmid
- Virus
- Artificial chromosome
- transposon

# Bacterial plasmid: Non-chromosomal DNA molecule

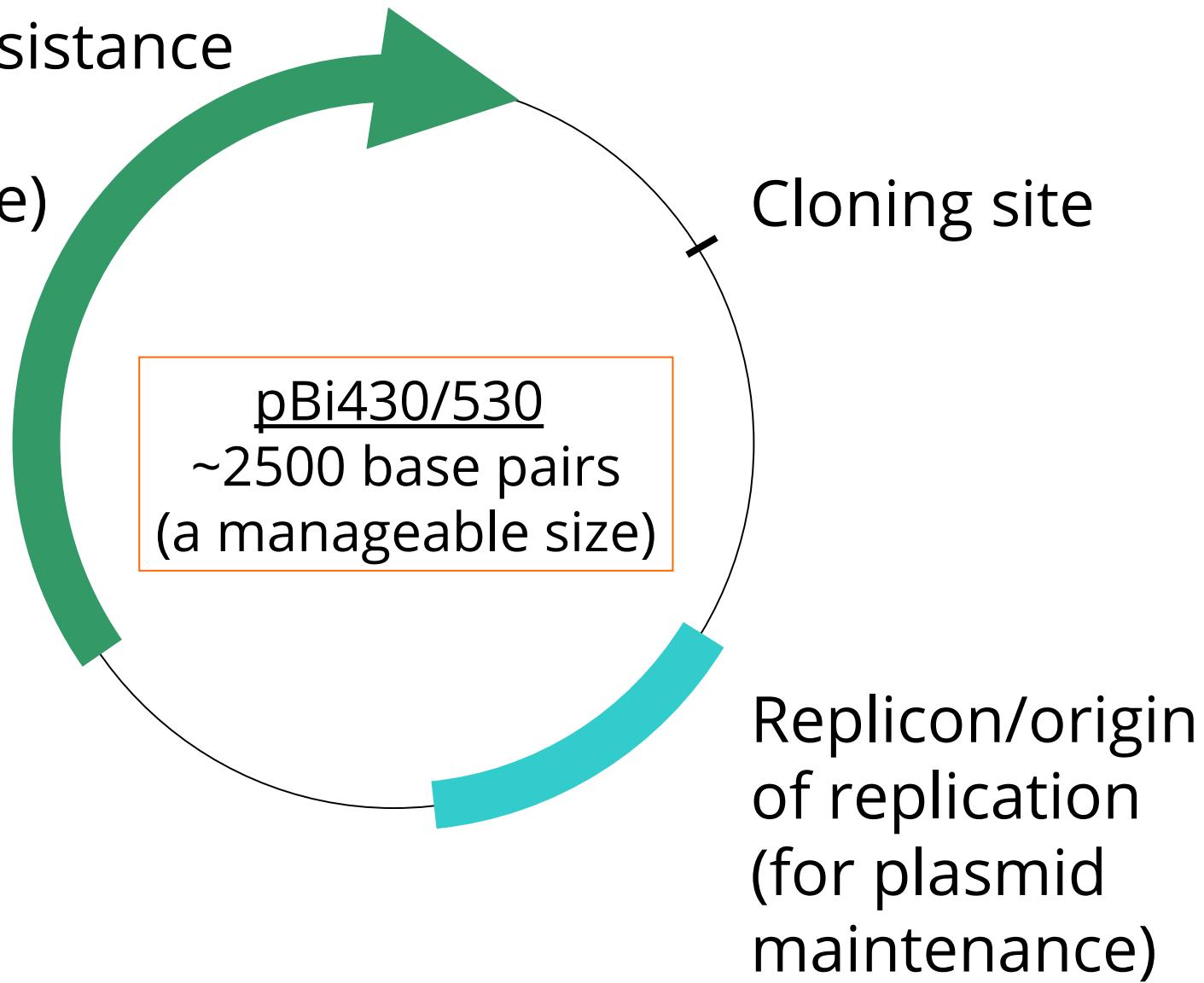
- Small, circular, supercoiled
- Replicates independently of the bacterial chromosome
- Copy number in cell is maintained through an origin of replication (replicon)
- Selectable through a gene that is beneficial for the host bacterium, eg. antibiotic resistance

# A useful plasmid

1. Is relatively small
  - Gives higher copy #, stability, and transforming efficiency
  - Can accept larger pieces of DNA
  - Easier to handle (less susceptible to breakage)
2. Has a selectable marker (e.g. antibiotic resistance)
3. Has single sites for many restriction enzymes

# Plasmid: basic parts list

gene that encodes  
antibiotic resistance  
(for plasmid  
maintenance)



# Replicon: how the plasmid replicates

- A DNA sequence and associated factors
  - origin of replication, ori: a site on the DNA
  - protein and RNA factors that manage replication
- The replicon helps define the “ copy number ” , the number of plasmid copies per cell

# A few common plasmid replicons

<u>PLASMID</u>	<u>REPLICON</u>	<u>COPY #</u>
pBR322	pMB1	15-20
pUC19	Modified form of pMB1 (RNA II mutation)	500-700
pACYC	p15A	18-22
pSC101	pSC101	~5

# Plasmid copy number is an important consideration

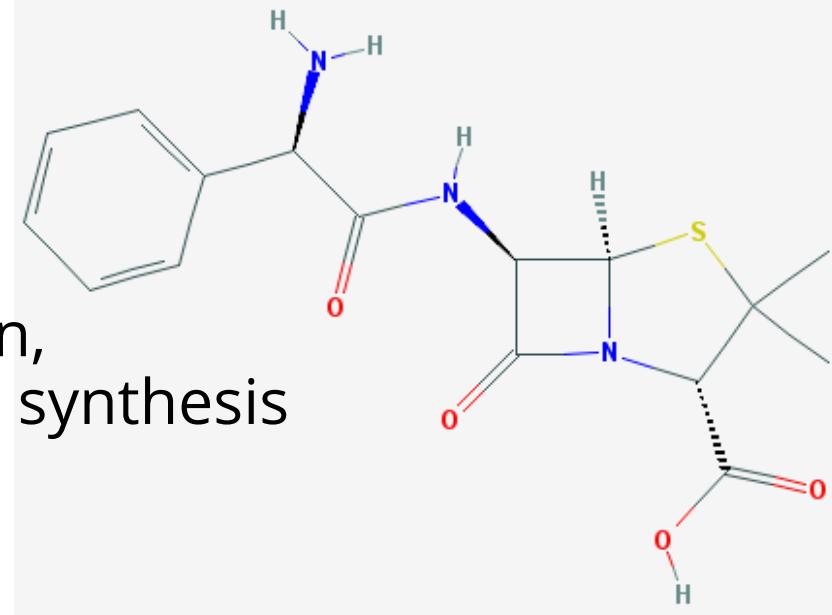
- High copy number plasmids
  - More copies/cell = easy to make a lot of DNA in a plasmid prep
  - Used for routine manipulation of small (<15 kb) recombinant DNAs
- Low copy number plasmids
  - For genes that are lethal or unstable in high copy number plasmids
  - Bacterial Artificial Chromosomes (BACs) that can propagate large (>100 kb) recombinant DNAs. Lots of DNA is a drag on cell physiology

# How to make sure a plasmid is maintained

- Plasmids contain **selectable markers**: genes carried by the plasmid that confer functions required for host survival
- Selection: only those cells with the plasmid will survive
  - Allows transformation (a rare event) to be easily detected
  - Cells cannot lose the plasmid, even if it causes a selective disadvantage (eg. slow growth or toxic gene product)

# Antibiotic resistance genes

- Antibiotic
  - **ampicillin** (related to penicillin, carbenicillin) inhibits cell wall synthesis
- Resistance gene
  - beta lactamase (bla) breaks down ampicillin, so cells carrying this gene are called **amp<sup>r</sup>**

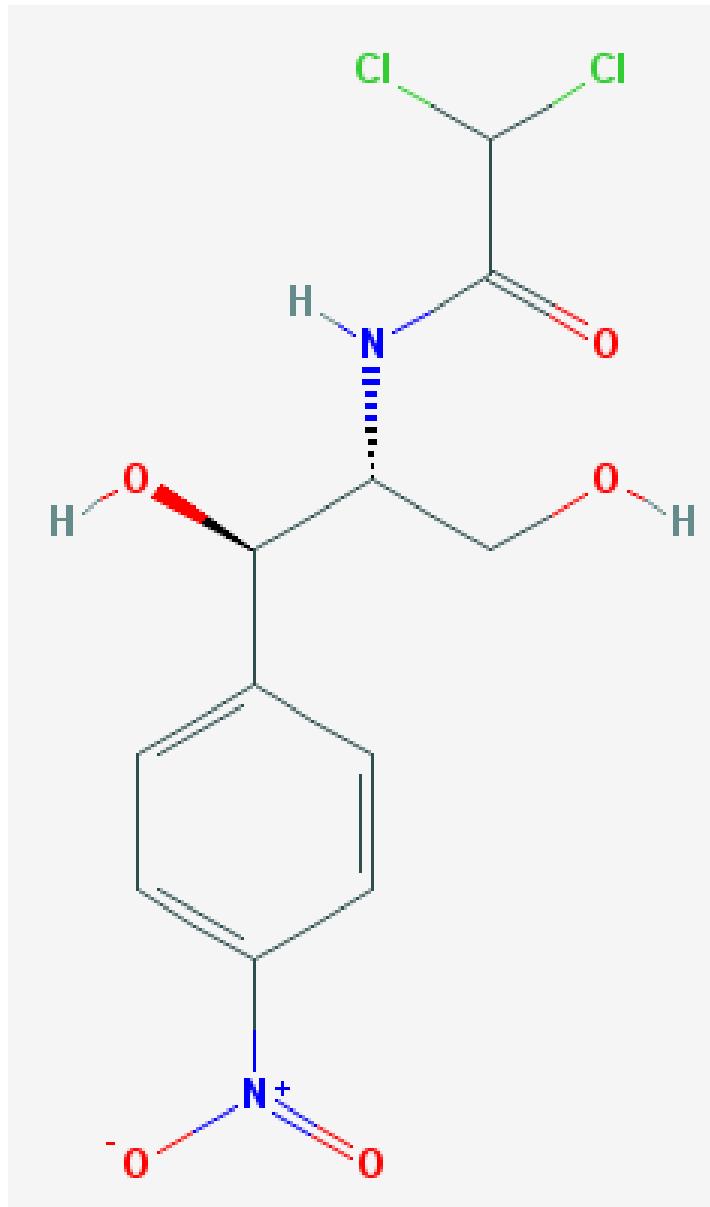


## Keep in mind

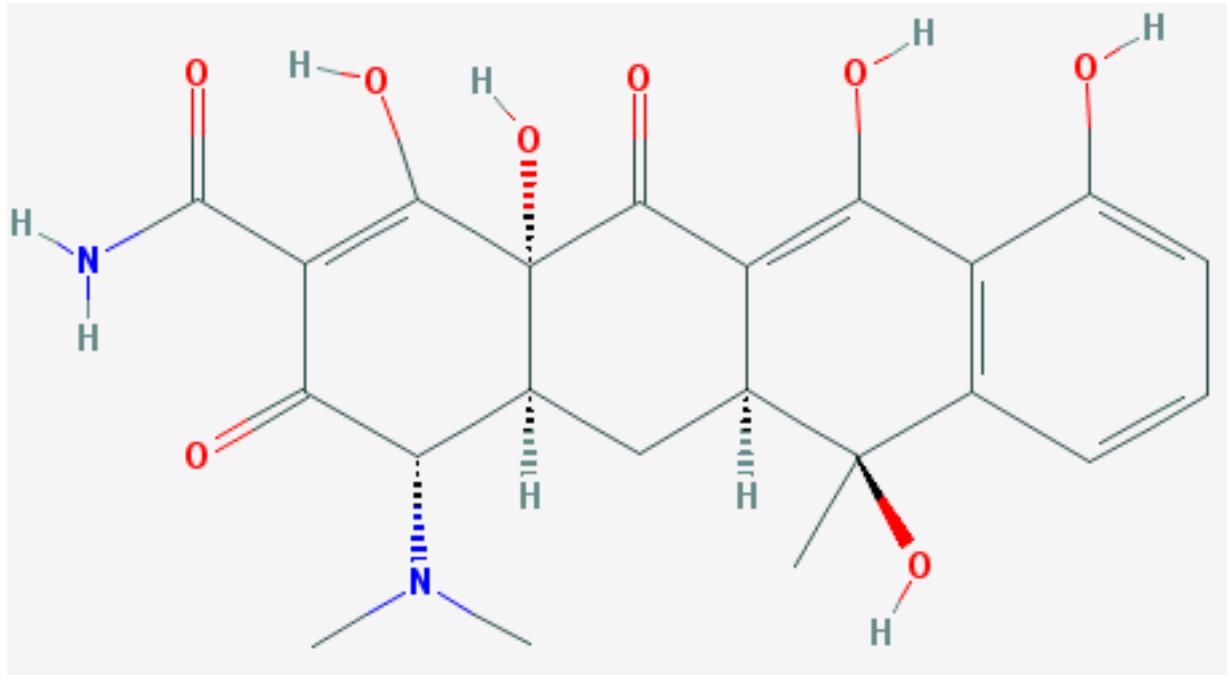
- beta-lactamase is secreted into the medium where it breaks down the antibiotic and depletes it
- ampicillin sensitive cells can grow, defeating the selection

# Antibiotic resistance genes

- Antibiotic
  - chloramphenicol (cm) inhibits translation (peptidyl transferase activity of the ribosome)
- Resistance gene
  - chloramphenicol acetyl transferase (CAT) inactivates chloramphenicol



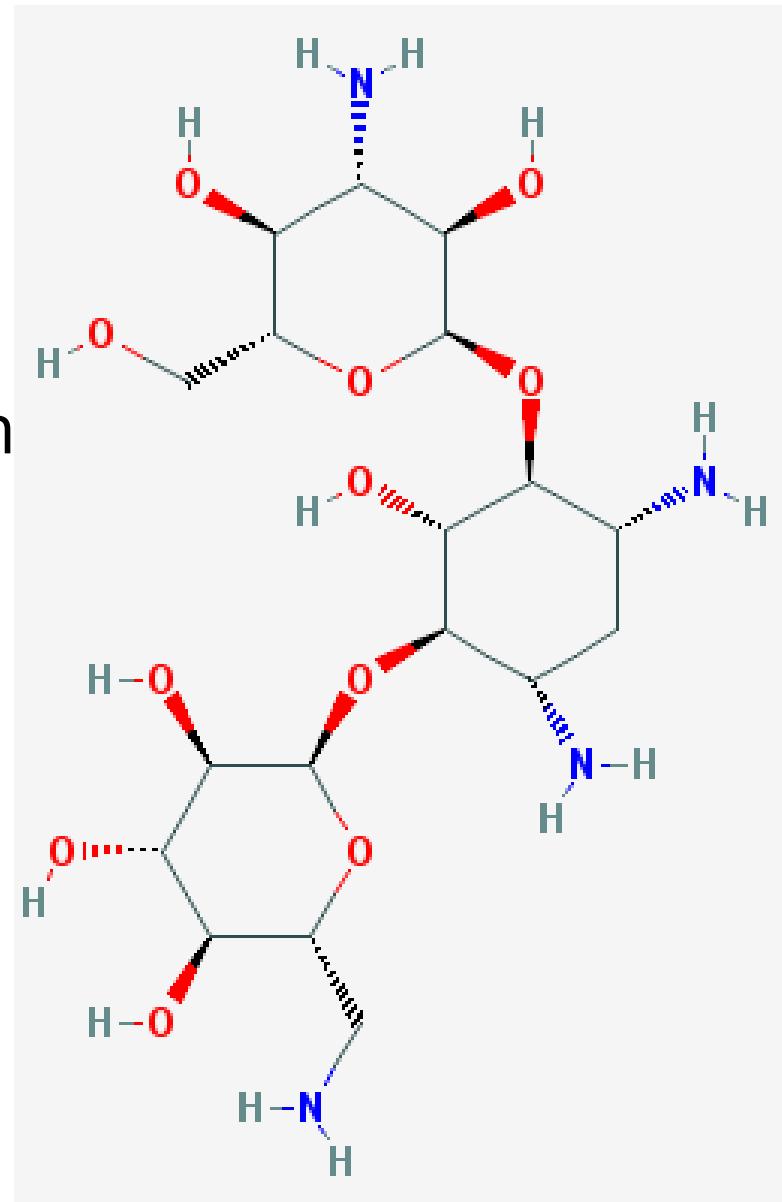
# Antibiotic resistance genes



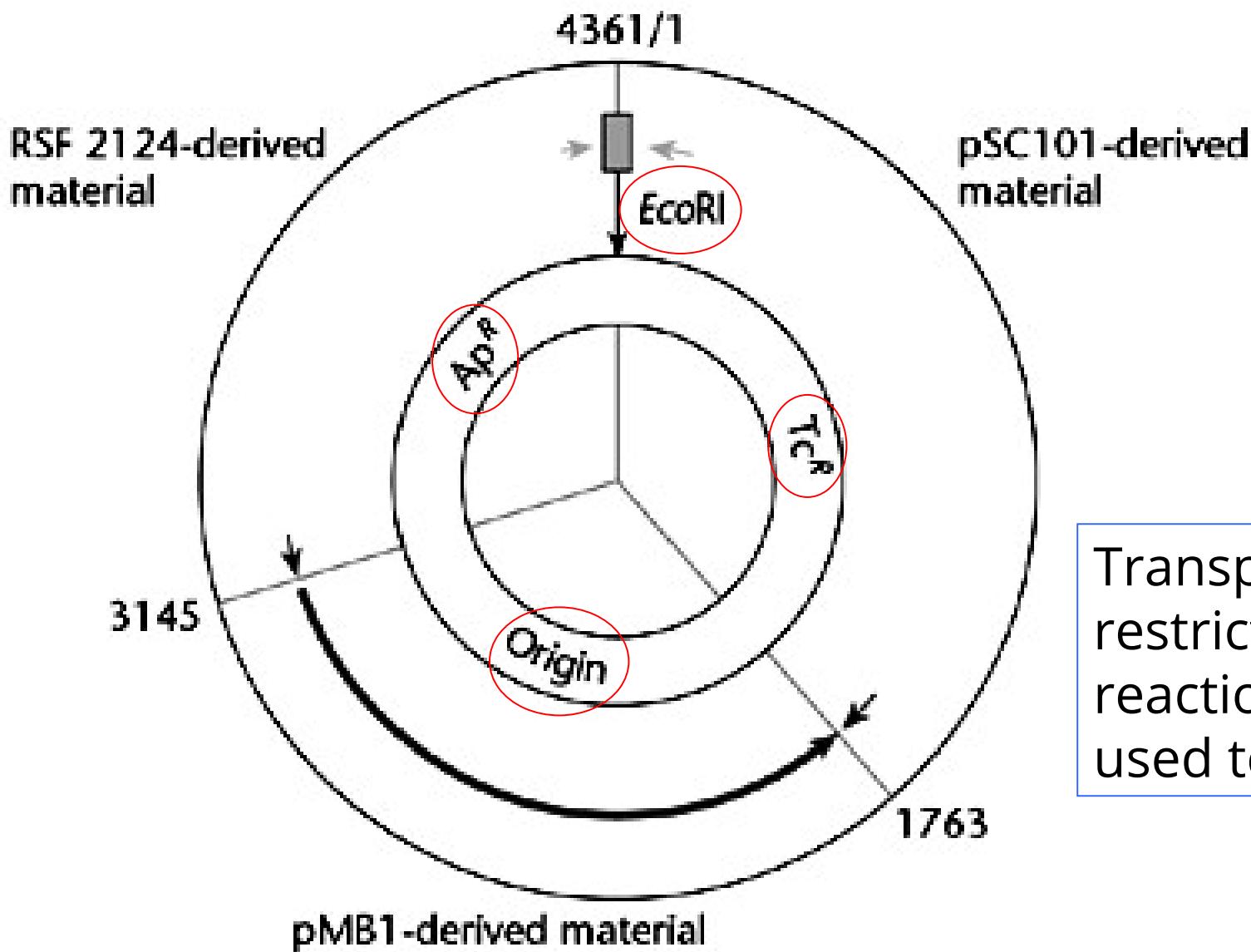
- Antibiotic
  - **tetracycline** inhibits translation (interacts with ribosome 30S subunit)
- Resistance gene
  - Tet A (C ) protein confers resistance to by moving this antibiotic out of the cell

# Antibiotic resistance genes

- Antibiotic
  - kanamycin: aminoglycoside antibiotic, inhibits translation (interacts with ribosome 30S subunit)
  - others of this class include G418, neomycin
- Resistance gene
  - bacterial aminophosphotransferase inactivates kanamycin by phosphorylation

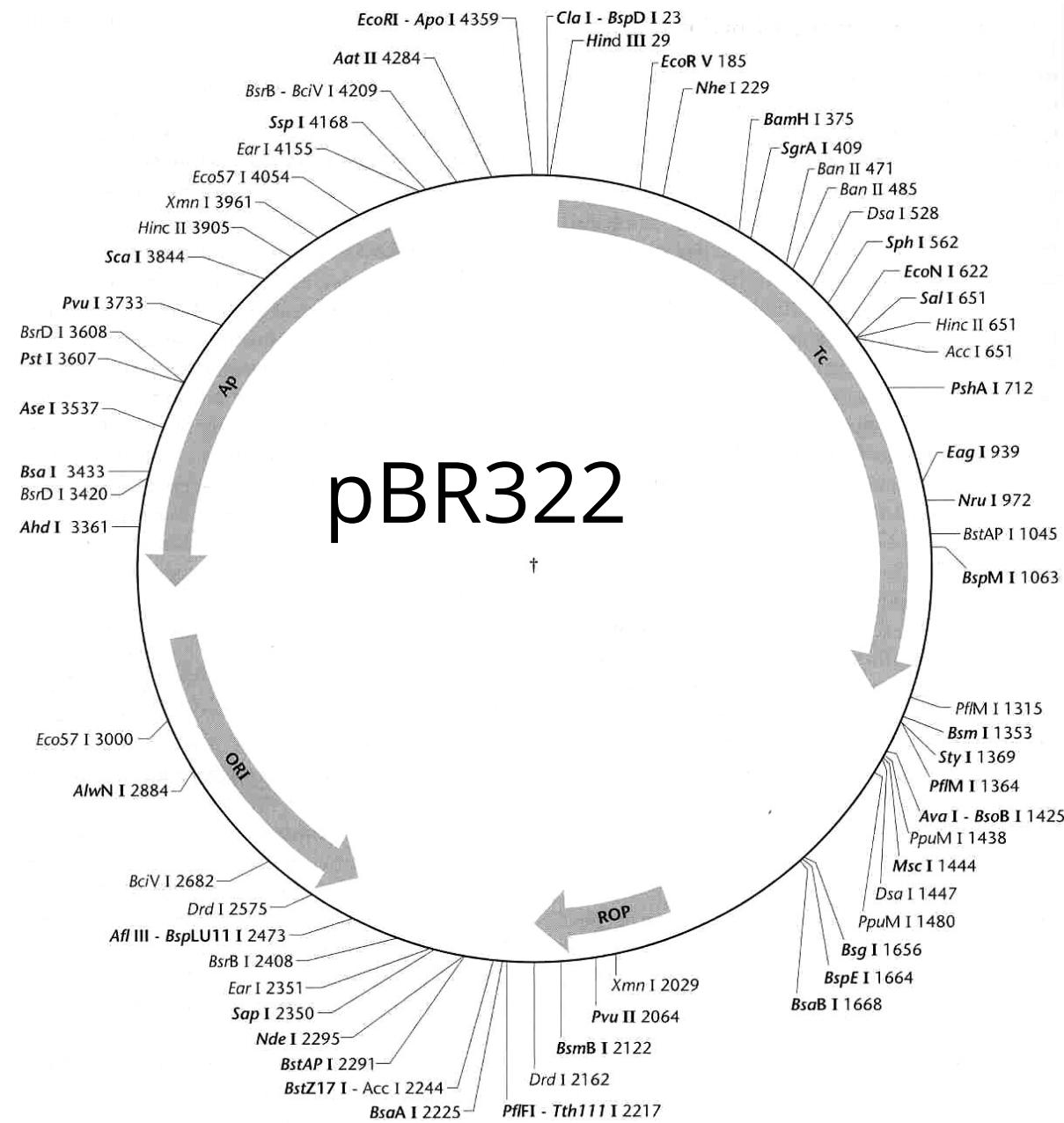


# pBR322: the first widely used plasmid



Transposition and  
restriction/ligation  
reactions were  
used to make it

# Using pBR322



Clone into sites in the Tc<sup>r</sup> gene,

Plate on ampicillin,  
then 'replica plate' on  
tetracycline.

Recombinants are amp  
resistant, tet sensitive

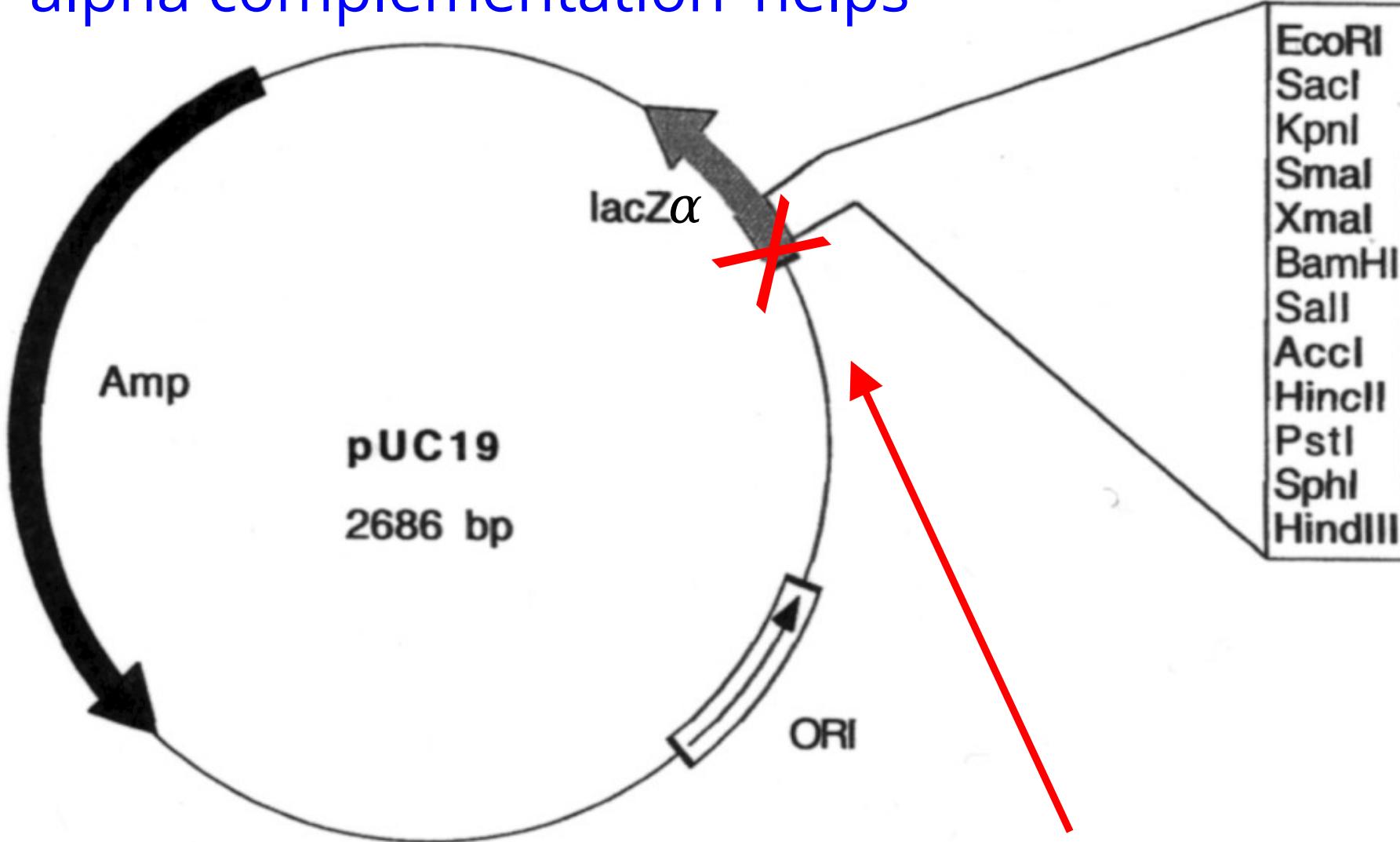
But: pBR322 has **low copy number, large size**

## pUC plasmids: improved cloning vectors

- Reduced size (about 2000 bp), easier for cells to handle
- Multiple cloning site (MCS, also called “ poly-linker” ): unique sites for lots of different restriction enzymes
- Very high copy number (mutation in RNA II)
- A “ blue-white” screening tool for recombinants (“ alpha complementation” is disrupted by foreign DNA in the MCS)

# Identifying recombinant plasmid: 'alpha complementation' helps

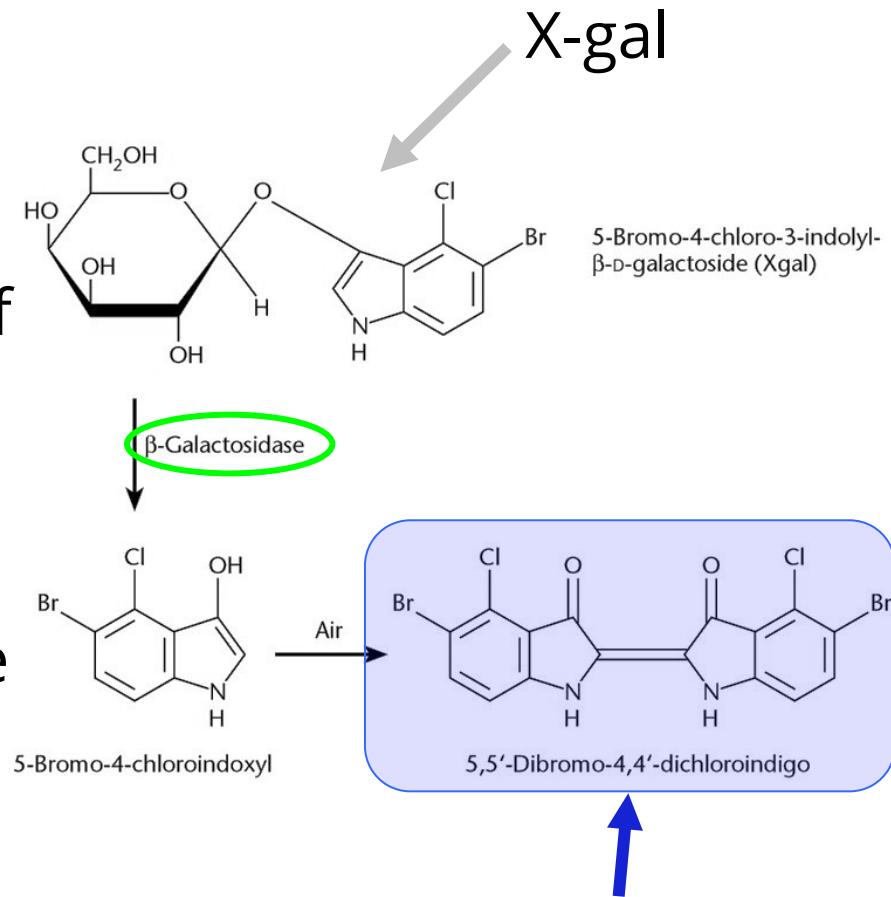
Polylinker  
(MCS)



DNA cloned in the MCS interrupts  
the lacZ gene alpha fragment

# Alpha complementation

- Plasmid has the N-terminus of beta galactosidase (alpha fragment)
- Host *E. coli* has the C-terminus of beta galactosidase (omega fragment)
- Both the N- and C-terminal fragments are needed for enzyme activity
- if beta galactosidase is present, X-gal is cleaved, producing a bright blue product

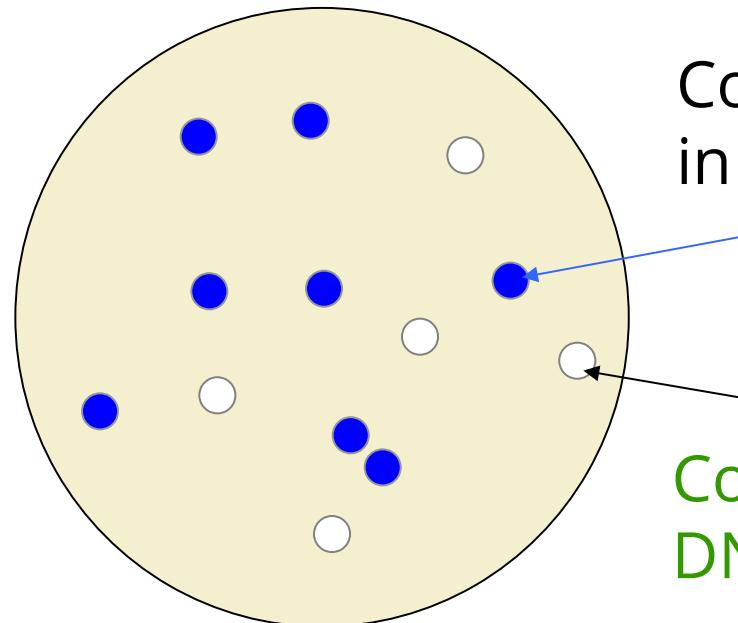


Bright blue

# Alpha complementation on a petri plate

- Plasmid has alpha fragment of beta galactosidase with an MCS, bacterial host has omega fragment
- Clone DNA into the MCS, no alpha fragment
- No alpha fragment, no functional B-galactosidase
- No B-gal, no blue color
- white colonies

pUC19  
transformation  
plate



Colony with no DNA  
in MCS. Don't want.

Colony with cloned  
DNA in MCS. Want.

# Many specialized plasmids are available

- For expression of specific genes
  - Plasmids are engineered to include “tags” on expressed protein, to assist in purification or tracking
- For cloning of unstable or toxic genes
  - Low copy number
- For cloning gigantic pieces of DNA
  - Low or single copy number

Want a new plasmid? Addgene.org, a plasmid repository

- <https://www.addgene.org/>

Keeping track of your plasmid sequence and structure

- A plasmid editor (ApE)
  - Save plasmid sequence
  - Include annotations to the plasmid
  - <http://biologylabs.utah.edu/jorgensen/wayned/ape/>

# How to create a recombinant plasmid (general protocol)

## 1) Prepare the plasmid vector

- a) Make the plasmid linear (PCR, or restriction enzyme)
- b) Purify, quantify the DNA

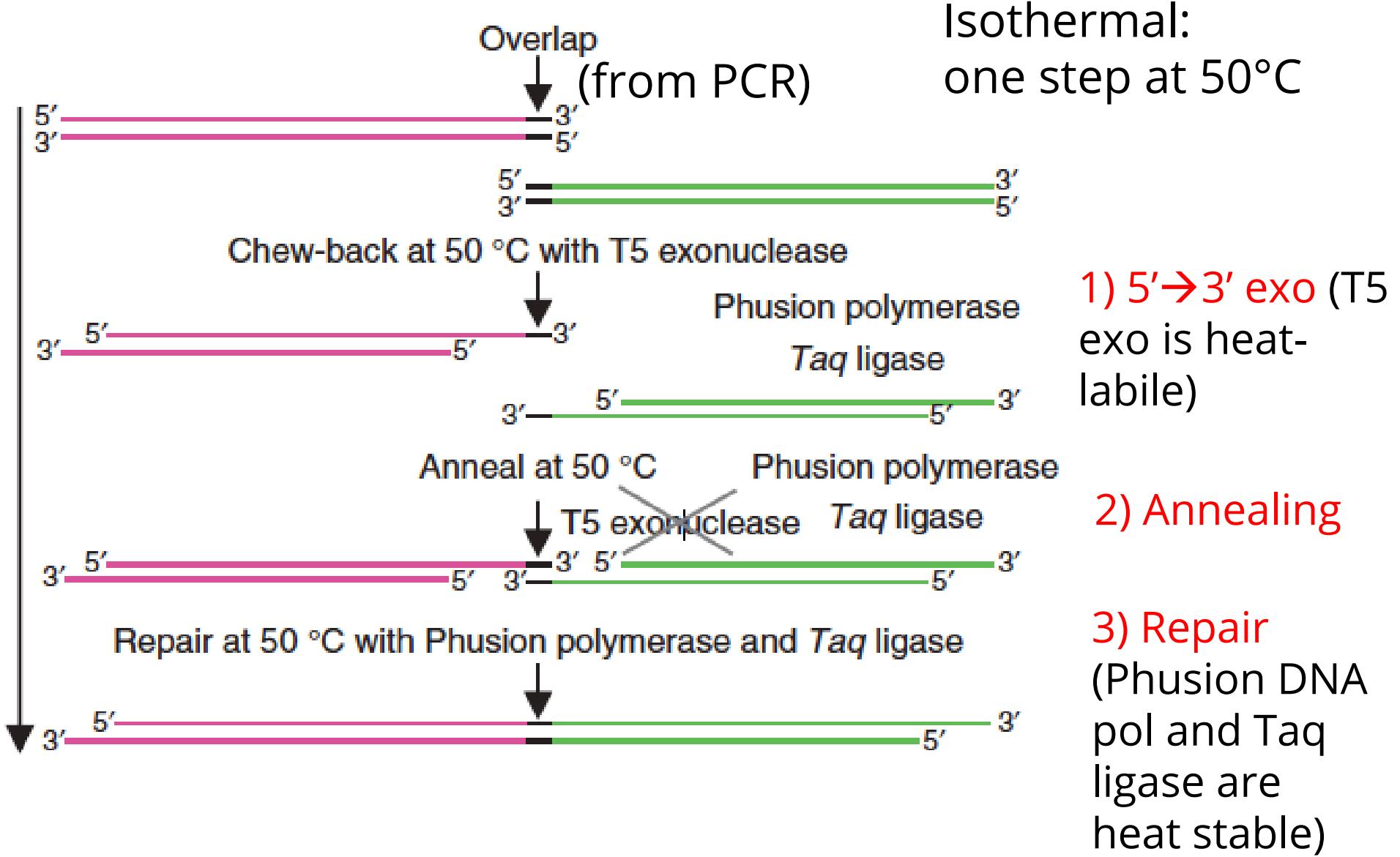
## 2) Prepare the DNA to be cloned

- a) Obtain the DNA
  - PCR
  - Small pieces of genomic DNA
- b) Treat the ends of the DNA to make them clonable
  - Primers with modified 5' ends
  - Add sequence to 3' ends with terminal transferase
  - Ligate adaptor sequences to ends

## 3) Combine the plasmid and clone DNA

- a) Standard cloning: ligase
- b) Gibson Assembly: Exonuclease, DNA pol, ligase
- c) In vivo assembly (IVA): the cell assembles pieces with overlapping ends

**Gibson DNA assembly:** make synthetic genes, pathways, or *entire genomes*.



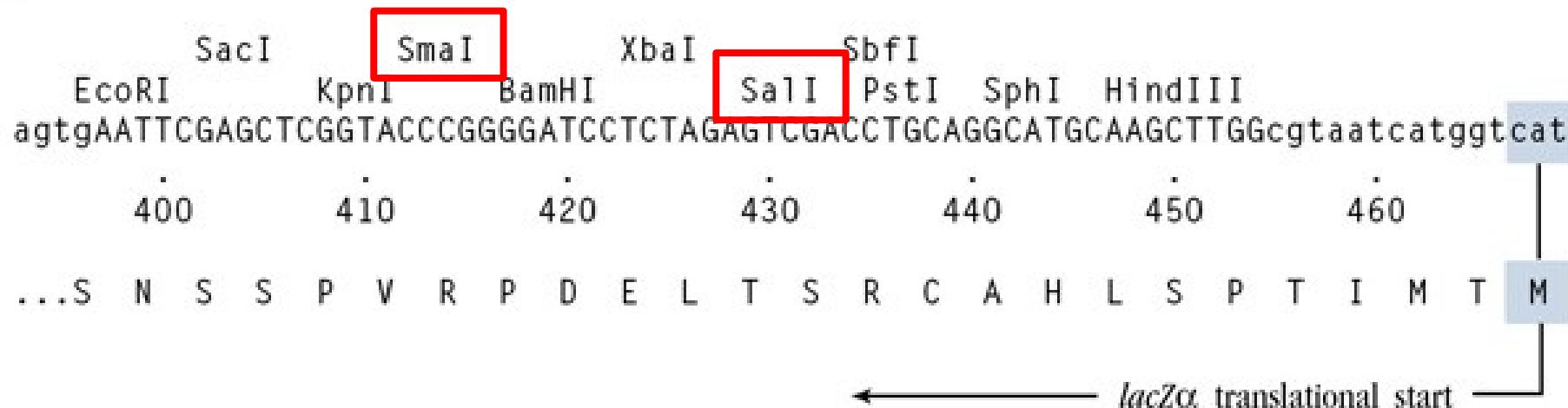
## Gibson assembly example

# Cloning vector: plasmid pUC19

## Cloned DNA: Pf1831 (*Pyrococcus furiosus* histone protein)

ATGGGAGAATTGCCAATTGCCCGAGTTGACAGGCTTATAAGAAAGGCAGG  
TGCTGAAAGAGTTAGCGAGCAAGCAGCCAAAGTCCTAGCAGAGTACCTCG  
AAGAGTACGCTATTGAGGTCGCAAAGAAGGCAGTAGAGTCGCAAGGCAC  
GCAGGTAGAAAGACTGTTAAGGTTGAAGACATTAAGCTCGCAATTAAGAG  
CTGA

pUC19 MCS



Cloning strategy: <http://nebuilder.neb.com/>

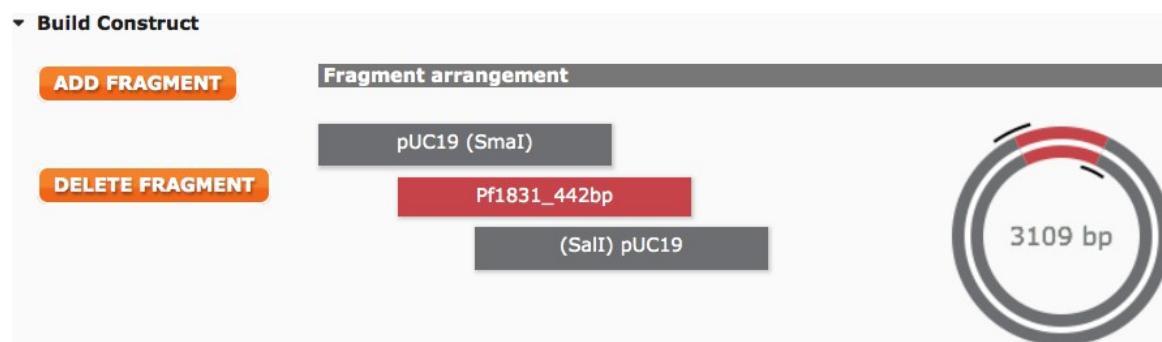
## Parameters for cloning by Gibson assembly:

The final product has to be circular

Need at least a 15 bp overlap at the cloning junctions (this is accomplished by extra sequences at the 5' end of the PCR primers)

Picking primers: can be done manually, or with an automated online tool

Overlaps	Oligo (Uppercase = gene-specific primer)	Anneals	F/R	3' Tm	3' Ta *
pUC19	tcgagctcggtaccc <b>GGCAATCTATTGGAAATC</b>	Pf1831	Fwd	56.0°C	<b>56.0°C</b>
pUC19	ttgcatgcctgcagg <b>GATATACTTTAATTCGGGAGG</b>	Pf1831	Rev	57.2°C	<b>56.0°C</b>

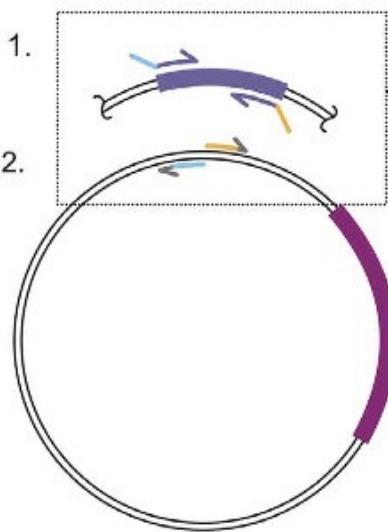


## "IVA" cloning: In Vivo Assembly

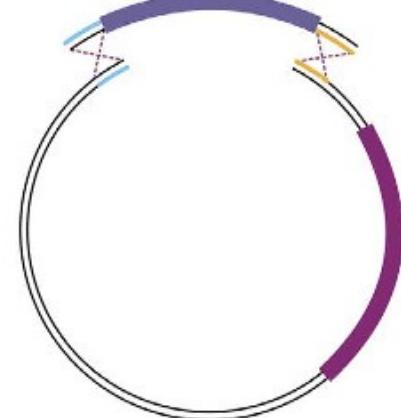
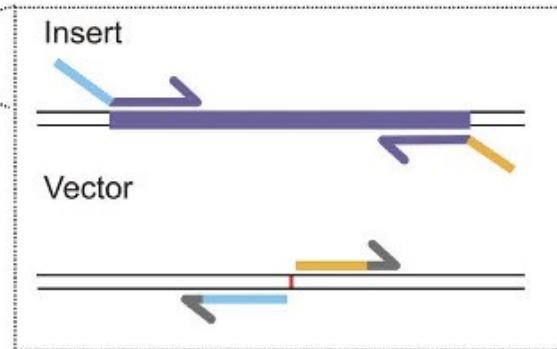
- *E. coli* has a recombination machinery that only requires short regions of homology (15-20 base pairs)
- Transform cells with DNA with overlaps, the recombination machinery will put them together automatically, and with reasonably high efficiency

b

Sub-cloning



You do



Cell does

# Moving DNA: transformation of *E.coli* with plasmid DNA

Cells acquire DNA by

- Chemical competence
- Electroporation
- Natural competence (in some lucky cases)

*E.coli* strain: must be antibiotic sensitive, and should not have restriction enzyme systems

Following transformation, the cells are cultured on a petri plate

## Making chemically competent *E. coli*

- Grow cells to “ mid-log” phase ( $A_{600}$  of 0.4) spin to get cell pellet
- Suspend cells in  $\text{CaCl}_2$  (100 mM), pellet again
- Suspend in small volume of  $\text{CaCl}_2$ /glycerol
- Freeze cells (-80°C) or go straight to transformation protocol

# Transformation of chemically competent cells

DNA binds to cells



- Mix DNA and competent cells, on ice for 30 min.

DNA uptake by cells



- Heat shock (42°C) for 1.5 minutes
- Add growth media, 37°C for 1 hour

Cells recover



- Plate on growth medium plus selection (antibiotic) for the plasmid

Selection occurs



Efficiency:  $\sim 10^5 - 10^8$  antibiotic resistant colonies/microgram plasmid DNA

# Transformation by electroporation

- Grow cells to  $A_{600}$  of 0.4
- Centrifuge and resuspend in water + 10% glycerol (do this 4 times to reduce conductivity)
- Place cells with DNA in electrode-containing cuvette, deliver electrical pulse
- If there is arcing (sparks) transformation efficiency will be poor (uneven transfer of charge). To avoid this make sure the ion concentration is very low (less than 10 mM salt)
- $> 10^9$  transformants/microgram DNA (ideally)

# Setting up a transformation – treatments and controls

1. *No plasmid* (negative control, nothing should grow on this plate)
2. *Supercoiled plasmid* of a known concentration (to determine efficiency of competent cells)
3. *Plasmid DNA* backbone: without clone DNA (background transformants)
4. *Plasmid DNA with clone DNA* (desired products)

# Ideal results from a transformation

- 1) No DNA--No colonies
- 2) 2 nanograms ( $2 \times 10^{-3}$  micrograms) supercoiled plasmid DNA--500 colonies (efficiency of cells:  $2.5 \times 10^5$  transformants per microgram DNA)
- 3) Plasmid alone--small number of colonies
- 4) Plasmid + insert—lots more colonies than #3

# Identify recombinant plasmids

- Screen colonies/plasmids for cloned DNA by PCR
- Alpha complementation: white colonies represent presence of insert DNA blocking functional beta galactosidase

Confirm clones by sequencing

# When cloning a piece of DNA consider:

- 1) Choice of plasmid: cloning sites? antibiotic? replicon?
- 2) Adding DNA to plasmid: how will the reaction be set up to give the desired product?
- 3) Moving DNA by transformation: what strain of *E. coli* will you transform into? Which method for transformation?
- 4) Screening for successful ligation products (recombinant plasmid DNA): how will the recombinant plasmids be identified?

# Basic gene cloning: plasmids and transformation, cutting and pasting

- Plasmid biology
- Adding DNA to a plasmid
- Transformation of *E. coli*

## Vectors for *E.coli* part II

- I. Bacteriophage (bacterial viruses): lambda and M13
- II. Moving and storing large DNA molecules: PACs, and BACs

## Readings:

- 1) 28 *MC4 Lambda and M13*. Short introduction to these historically important molecular biology tools, also two M13 protocols.
- 2) 33 *MC4 High capacity vectors*. Summary of vector types, with an emphasis on bacterial artificial chromosomes (BACs)
- 3) 29 *MC4 Cre/Lox*. Discussion of the cre-lox system for forcing site-specific recombination

# Bacteriophages: viral vectors for molecular cloning

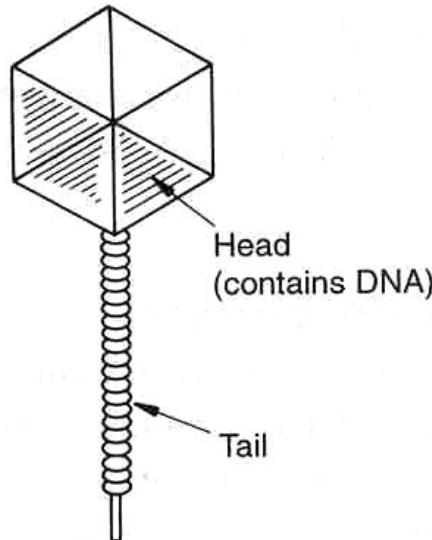
- I.     M13: a filamentous phage
  - Life cycle
  - genome structure
  
- II.    Lambda: a “head and tail” phage
  - Life cycle
  - Basic cloning in lambda

# Bacteriophages

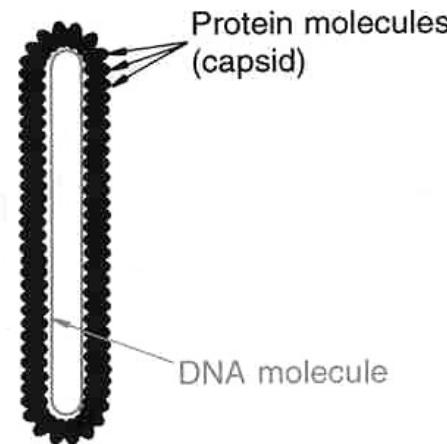
Viruses that infect bacteria

Morphologies:

- a) Head and tail
- b) Filamentous



(a) Head-and-tail



(b) Filamentous

- Nucleic acid molecule
  - Carrying genes for infection and replication
  - Surrounded by a protective protein coat (capsid)
- Infection (instead of transformation):
  - Phage attaches to outside of bacterium, injects DNA
  - Phage DNA is replicated, capsid proteins are made
  - Phage assembled and released

# Use of bacteriophages in cloning:

## M13 -- single-stranded DNA genome

- Then:
  - Sequencing
  - Site-directed mutagenesis
- Now:
  - Phage display of foreign peptides

## Lambda -- large-ish DNA fragments (25 kilobases)

- For gene cloning (large eukaryotic genes)
- Good selection schemes for recombinants
- Clone lots of precisely-sized DNA fragments for library construction

# M13: a filamentous bacteriophage

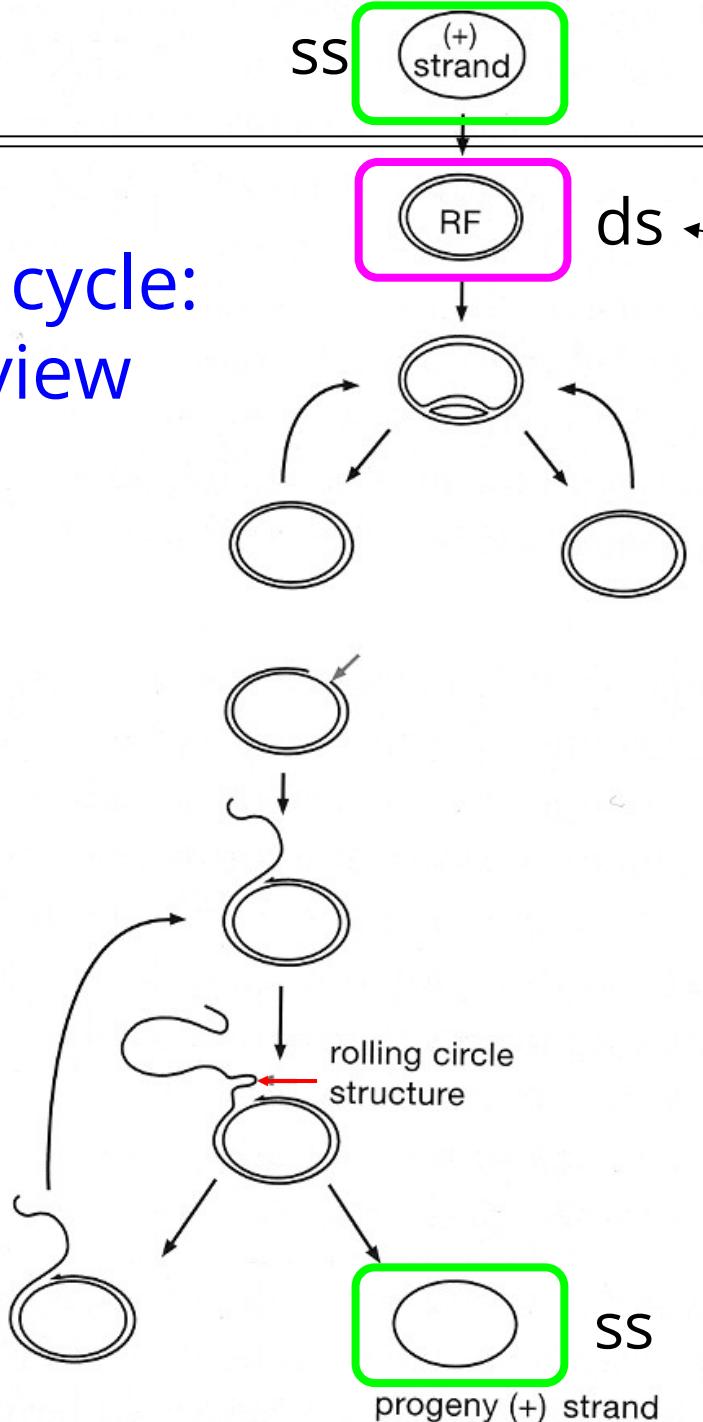
- Single-stranded, circular genome, 6.4 kb
- Can carry foreign DNA up to 6X the M13 genome size (36 kb) -- **the larger the DNA, the less stable the clone**
- Drawback: **foreign DNA can be unstable** (slow host cell growth – deletions confer a selective advantage)

M13 doesn't lyse cells, but it slows them down



M13 infections form 'turbid' plaques

# M13 life cycle: an overview



Infecting single-stranded circular viral DNA is converted to double-stranded replicating form (RF) DNA by host-encoded enzymes.

Isolate for cloning

Several rounds of replication occur through θ structures.

The (-) strand of the RF DNA is transcribed into viral mRNAs.

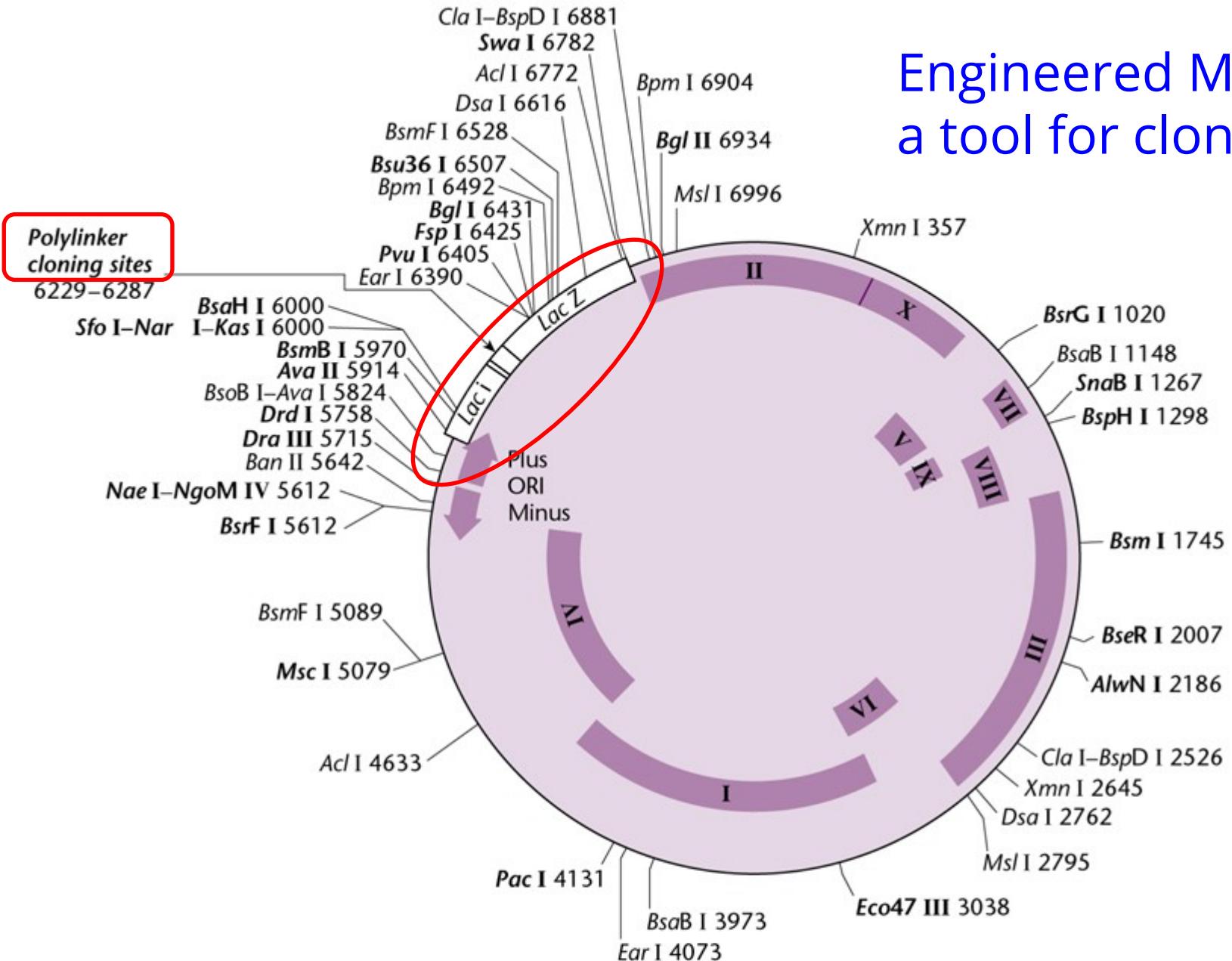
The viral gene II product introduces a nick at a specific site (red arrow) in the (+) strand of the RF DNA.

Progeny (+) strand is synthesized continuously by movement of the replication machinery around the (-) strand template (rolling circle replication).

The completed progeny (+) strand is cleaved from the rolling circular structure by the viral gene II product (red arrow). The progeny strand then circularizes.

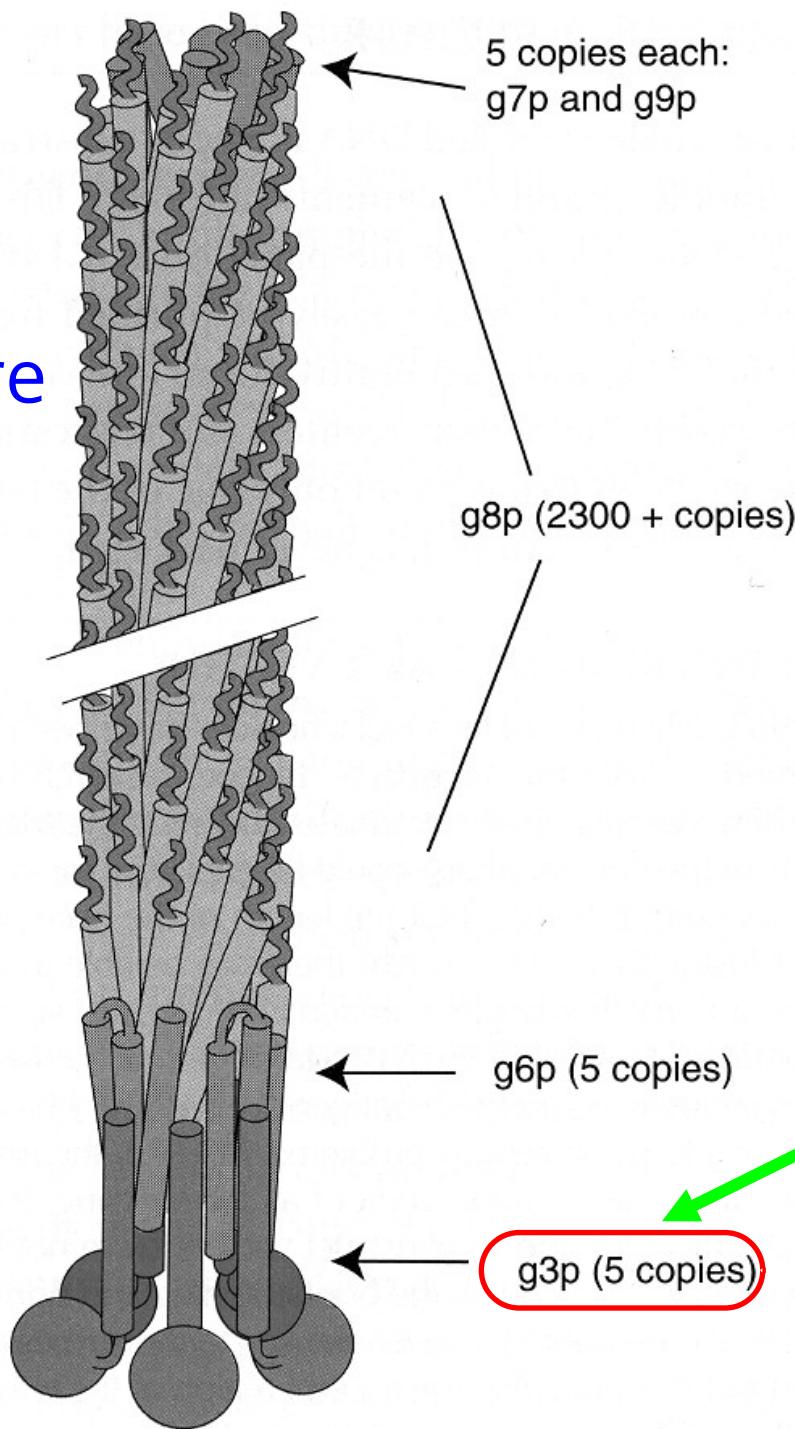
Synthesis of progeny (+) strand continues.

# Engineered M13: a tool for cloning



alpha complementation (like pUC plasmids)

# M13 phage particle structure



Variable length,  
determined by  
DNA size

Used in  
'phage  
display'  
techniques

# Bacteriophage lambda

- “ head and tail” phage, well-studied model
- Large, linear genome--48.5 kb
  - Central region of genome (“ stuffer” ) is dispensable for infectious growth--it can be engineered out
- Two infectious phases:
  - Lytic: replicating phase (killing/lysing cell)
  - Lysogenic: latent phase (integrated, waiting for opportunity)
- Can hold 5-25 kb DNA fragments

# Lambda as a cloning vector

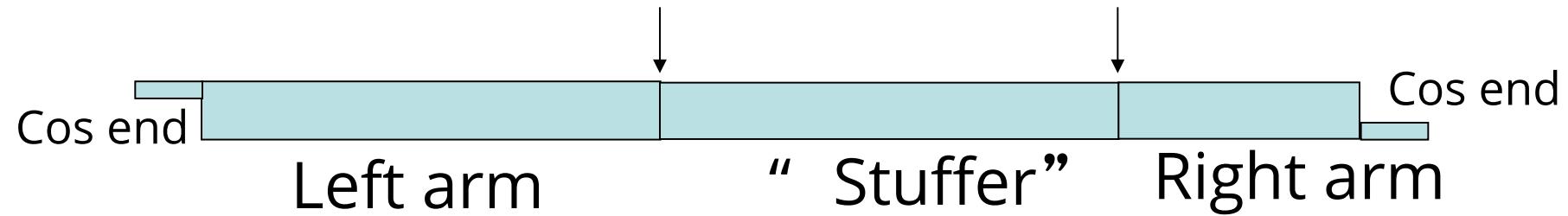
Phage capsid:

- Size of head is invariant
- Capsid must be filled
  - Too little DNA and it's not infectious
  - Too much DNA and the genome won't fit

a) Insertional vectors: can only increase genome size by 5% (about 3 to 5 kb)

b) Replacement vectors: remove “ stuffer” , can clone larger pieces of DNA, 8 to 24 kb (sufficient for many eukaryotic genes)

# Cloning in lambda phage – the genome view



- 1) Cut out stuffer, save right and left arms
- 2) Ligate with foreign DNA
- 3) “ Package” ligation mixture into phage heads
- 4) Plate mixture on *E. coli*, individual plaques represent recombinant clones

# Selection for *recombinant* lambda

There is a minimal size of DNA that can be packaged in lambda phage heads

- If the stuffer is removed the ligated “ arms” cannot be packaged (too small)
- Presence of recombinant DNA makes the phage large enough, so the recombinants are selected

# Cloning large DNA fragments

- 1) Make genomic *libraries*: the larger the DNA fragment, the fewer you need to make a complete *library* (representing the entire genome in fragments)
- 2) Clone DNA large enough to contain an entire eukaryotic gene

To get a functional gene, you need the whole gene and its regulatory regions

- Average human gene (exons and introns): 27 kilobases
- Regulatory regions may add another 10 to 100 kilobases upstream/downstream of gene

# Vectors for large recombinant DNA fragments

- Bacteriophage P1 plasmid: 70-100 kb
- YAC: 250-400 kb (or higher)
- PAC: 130-150 kb
- BAC: 120-300 kb (up to 700 kb is possible)

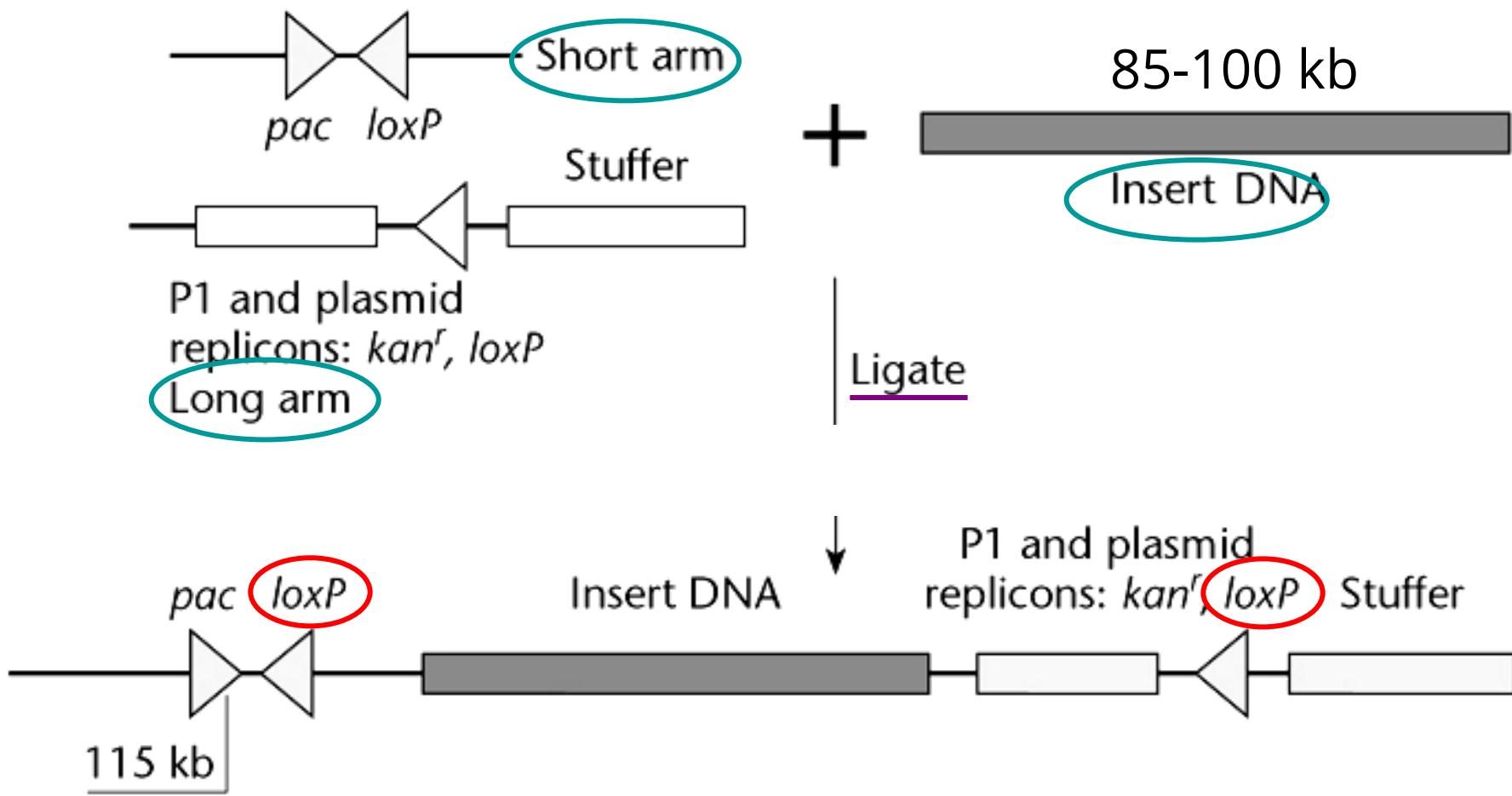
TABLE 1. High-capacity vectors for genomic cloning

Vector	Capacity (kb)	Replicon	Host	Copy number	Recovery of cloned DNA
P1	70–100	P1	<i>Escherichia coli</i>	1 (amplifiable)	Alkaline extraction
YAC	250–400	ARS	Yeast	1	Pulse-field gels
PAC	130–150	P1	<i>E. coli</i>	1	Alkaline extraction
BAC	120–300	F	<i>E. coli</i>	1	Alkaline extraction

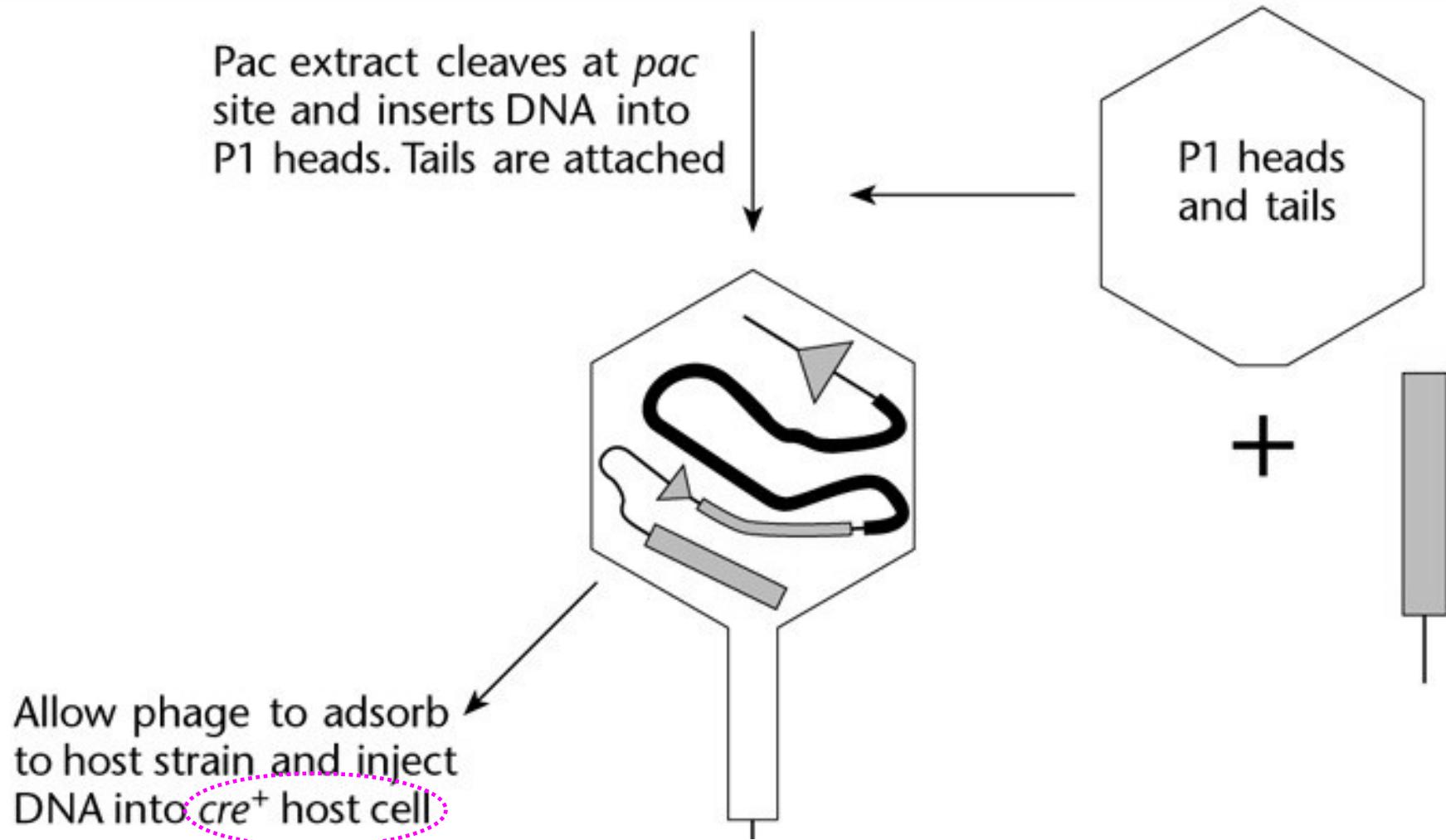
BACs are most commonly used

# Phage P1 vectors:

clone large DNA fragments (85-100 kb)

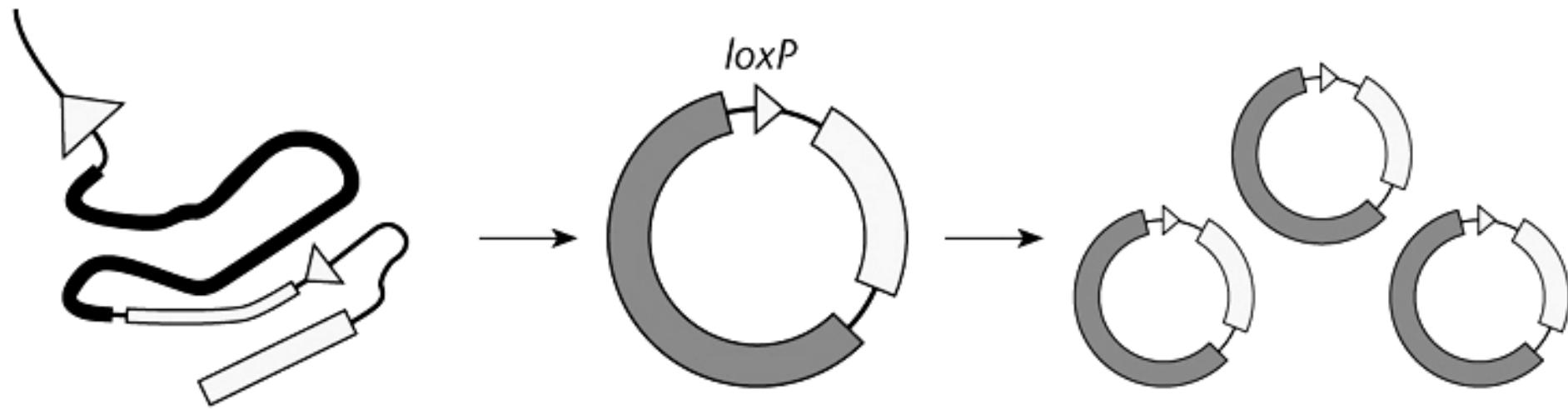


# Phage P1 vectors



Efficiency of packaging is typically low: thus it is not good for making large genomic libraries

# Phage P1 vectors

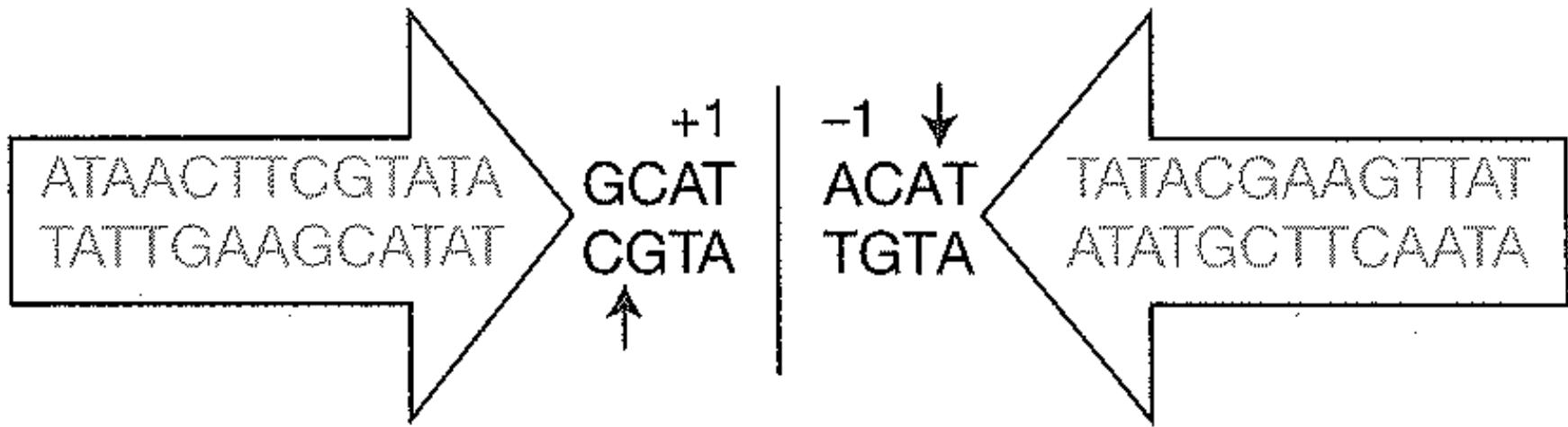


*Cre* recombinase protein circularizes injected DNA at the *loxP* sites. DNA replicates using plasmid replicon. Plasmid copy number is increased by induction of P1 lytic replicon.

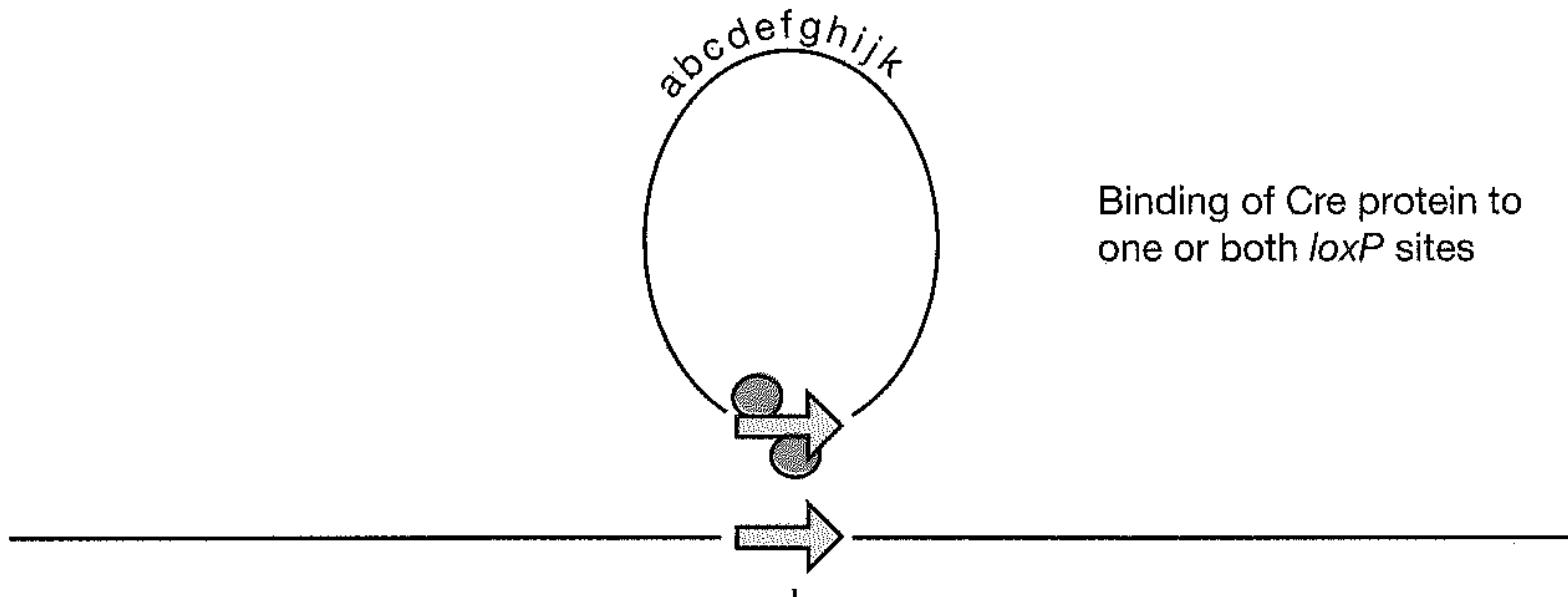
PACs: like P1 vectors but the DNA is not packaged (transfer by electroporation)

# Cre and Lox: a site-specific recombination system

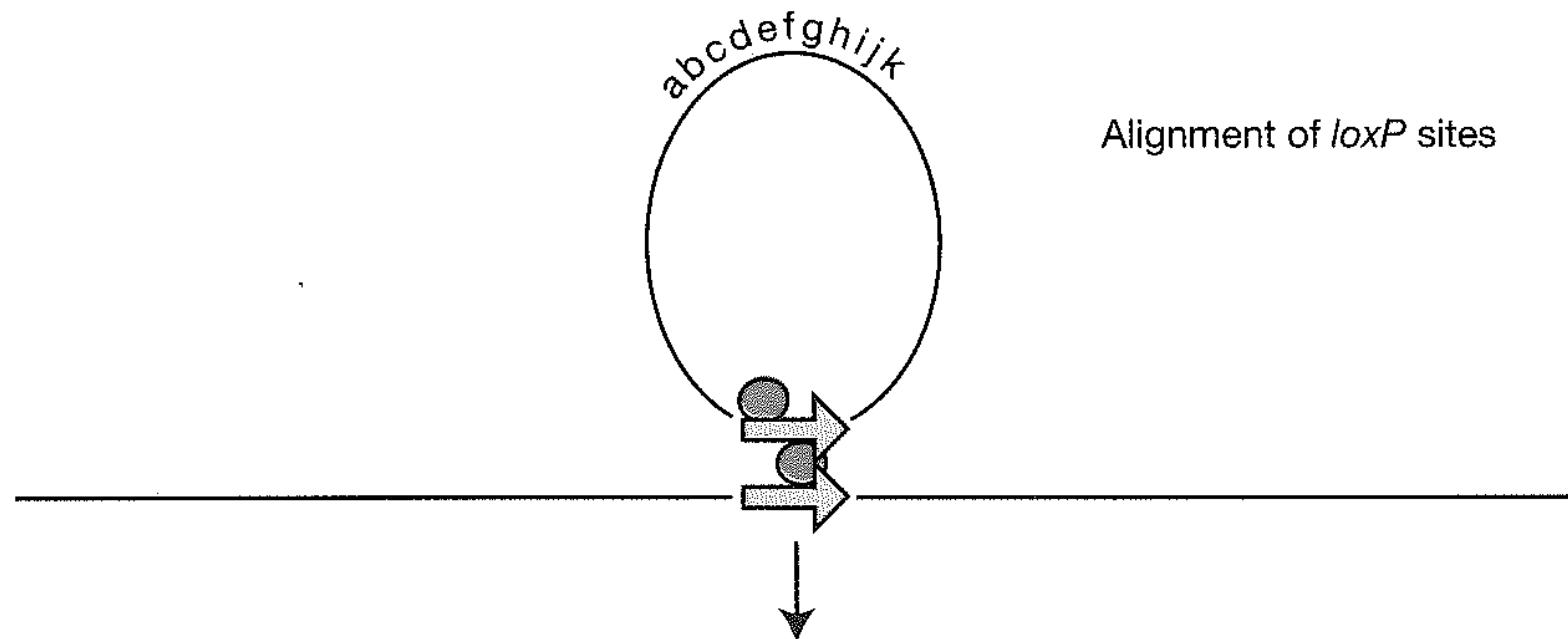
## The loxP site



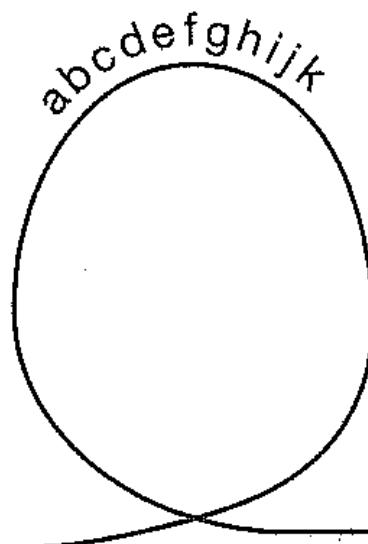
Two 13 bp inverted repeats separated by an asymmetric linker



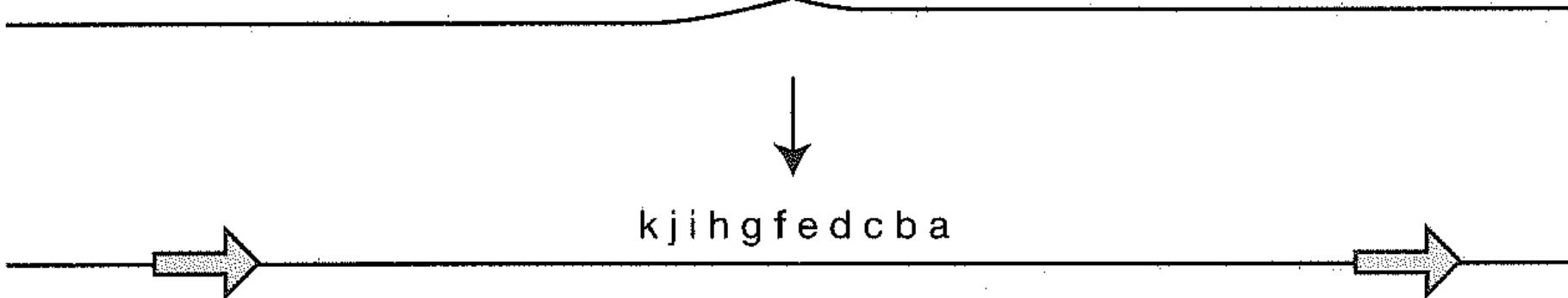
Binding of *Cre* protein to  
one or both *loxP* sites



Alignment of *loxP* sites



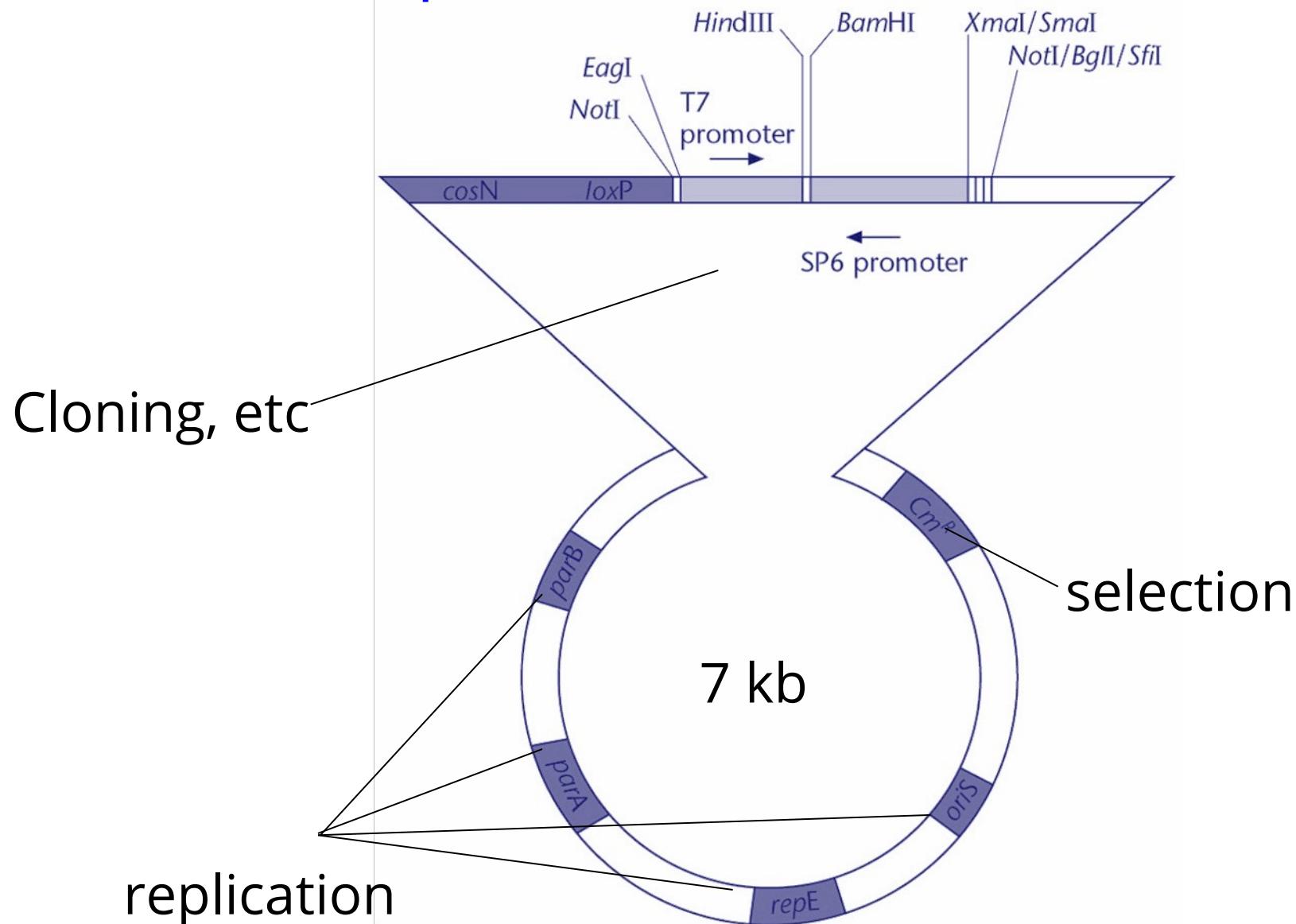
Asymmetric cleavage  
of *loxP* sites followed by  
strand invasion and synapse



# BAC: Bacterial Artificial Chromosome

- F factor of E. coli:
  - 100 kb plasmid, propagates through conjugation
  - low copy number (1-2 copies per cell)
  - 2 genes (parA and parB): accurate partitioning during cell division
- BACs: engineered to have par genes, replication ori, cloning sites, selectable marker
- Holds very large pieces of DNA: **up to 300 kb**
- Fairly easy to manipulate: move into cells by transformation (electroporation)

# BAC vector map



## Vectors for *E.coli* part II

- I. Bacteriophage lambda and M13
- II. Moving and storing large DNA molecules: PACs and BACs

# Gene cloning: Expression of protein and RNA

## Controlled gene expression

### I. proteins

- o Various tags and their functions
- o A protocol for purification of 6xHis tagged proteins expressed in *E. coli*

### II. RNA

- o For in vitro uses
- o For RNA interference (RNAi) in vivo

## Guide to readings:

- 1) *30 MC4 Protein expression.* Intro to protein expression in bacterial and eukaryotic systems, vectors, optimization.
- 2) *32 MC4 Epitope tags.* A list and description of tags/handles that can be added to recombinant proteins.
- 3) *Protein expression tags Malhotra 2009.* More good info on tags.
- 4) *31 MC4 His tag protein purification.* A detailed protocol for production of 6 x His tagged proteins.
- 5) *RNAi delivery methods 2015.* Trends in RNAi –based therapies
- 6) *SARS 1 vaccine 2009.* Using spike protein for immune response
- 7) *Hotez and Bottazzi 2020.* Production of affordable SARS CoV2 protein antigen vaccine.

# Controlled protein or RNA production is important

## 1) Production of proteins & RNA:

- Biochemical studies (enzyme function)
- Structural studies (crystallography, NMR)
- Commercially useful proteins or RNAs
- Protein pharmaceuticals
- Induction of RNA interference (RNAi)

## 2) Living systems

- Cell biology -- protein localization
- Transgenic plants and animals
- Human gene therapy

# Strategy for protein “overexpression”

## (1) Cloning

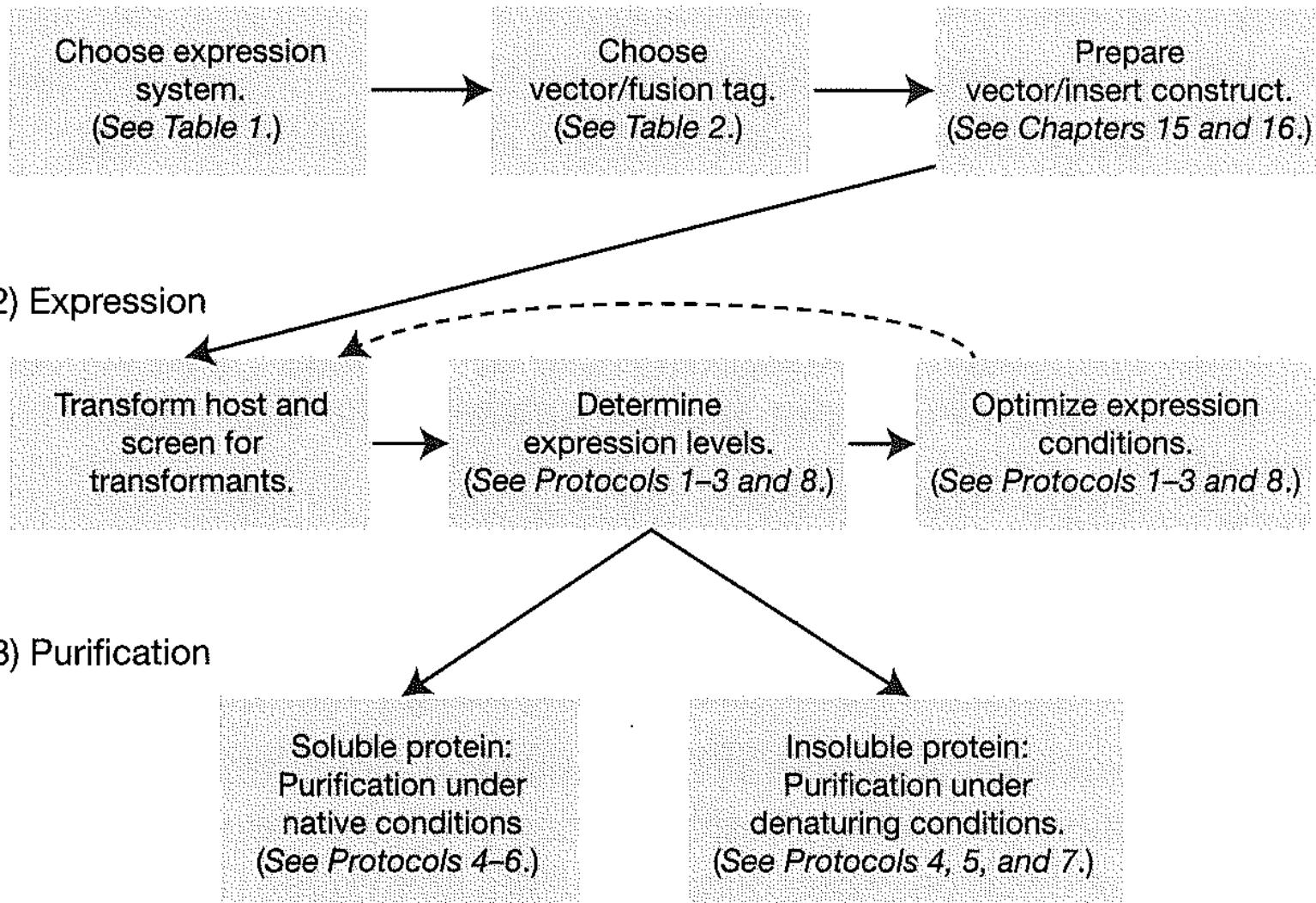


FIGURE 1. Overall strategy for recombinant protein expression and purification.

# **Expression vectors for protein/RNA purification**

- Make lots of protein or RNA
  - Control timing of synthesis
  - Control amount of synthesis
- 
- Optimized for transcription, translation
  - Facilitate post-translational handling by adding “tags”

# The most common expression platform is bacterial, specifically *E. coli*

- Lots of vectors available
- Lots of host *E. coli* strains available
- Large variety of tags for proteins
- Straightforward, easily adaptable protocols
  - Add gene to expression plasmid
  - Transform host strain with expression plasmid
  - Grow cells to mid-log (exponential phase)
  - Induce protein expression
  - Purify the protein

# Alternative expression hosts sometimes needed

Eukaryotic protein expression in bacteria not always successful:

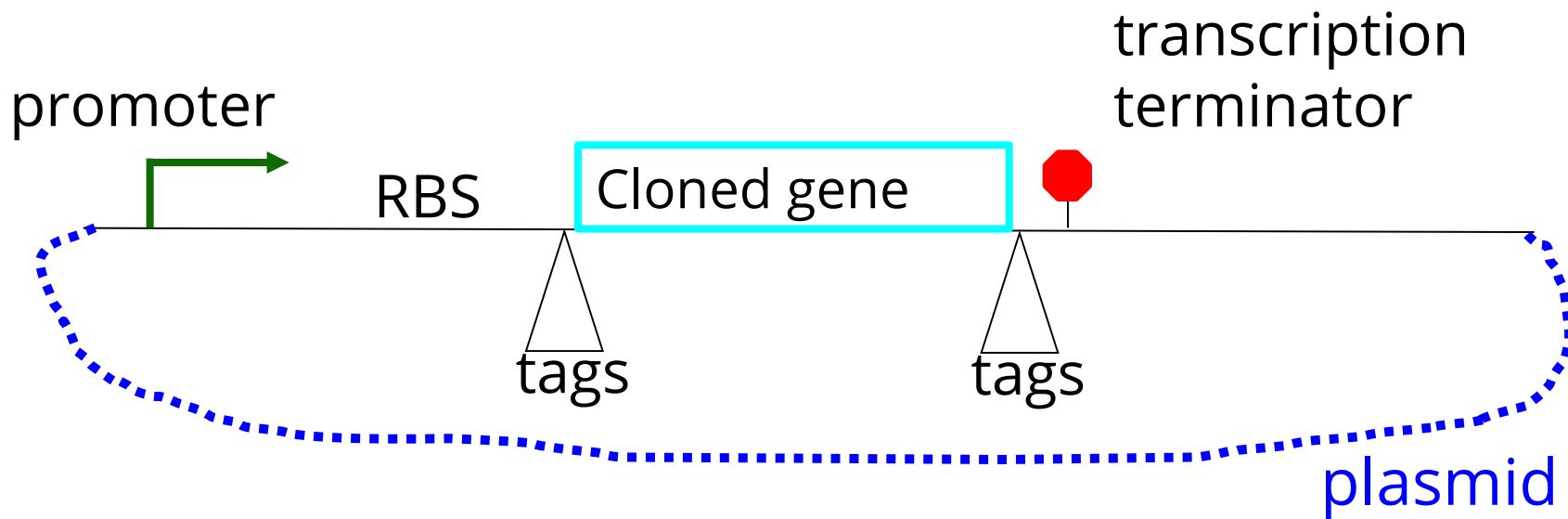
- Some proteins do not fold properly
- May lack post-translational modifications required for activity, especially glycosylation
- Protein might be toxic to bacteria

Some alternatives to bacterial expression:

- 1) Baculovirus-infected insect cell culture
- 2) *Pichia pastoris*, an easily cultured yeast
- 3) Mammalian cell culture
- 4) *Cell-free* translation system

See Table 1 in *30MC4 Protein Expression* for comparison of these methods

# A bacterial protein expression vector



- 1) Promoter for transcription
- 2) Transcription terminator
- 3) Ribosome binding site (RBS) for translation
- 4) Epitope tags for purification or detection of the protein

# To maximize synthesis of gene product:

- 1) Bring cells to mid-log phase (all cells in the process of growth and division) without expressing the gene: cell physiology is optimized
- 2) When conditions are right, “induce” the gene, by turning on transcription/mRNA synthesis
  - lots of ribosomes available
  - protein synthetic machinery is at a maximum to cope with new mRNA for induced gene
  - Ideally, massive amount of new protein will be made

# Inducible promoters for bacterial gene expression

Promoter	Transcription regulation	Induction
Lac	Lac repressor	IPTG (lactose analogue)
Trp/lac	Lac repressor	IPTG
T7	Lac repressor, T7 RNA polymerase	IPTG
pBAD	AraC repressor	arabinose
Lambda P <sub>L</sub>	Temperature sensitive lambda repressor	Shift from 30°C to 37°C

# Fusion protein tags: protein purification/detection

- Tags can be added to the N- or C-terminus, or can be internal to the protein sequence (on a surface loop)
- Tags can be engineered to be removable
- More than one tag can be added to a protein, for “Tandem Affinity Purification”
- Peptide tags. Example: **FLAG tag**, detected by monoclonal antibodies
- Whole proteins. **Biotin carboxylase**: covalently attaches to biotin, biotin binds to streptavidin which can be immobilized on columns or beads

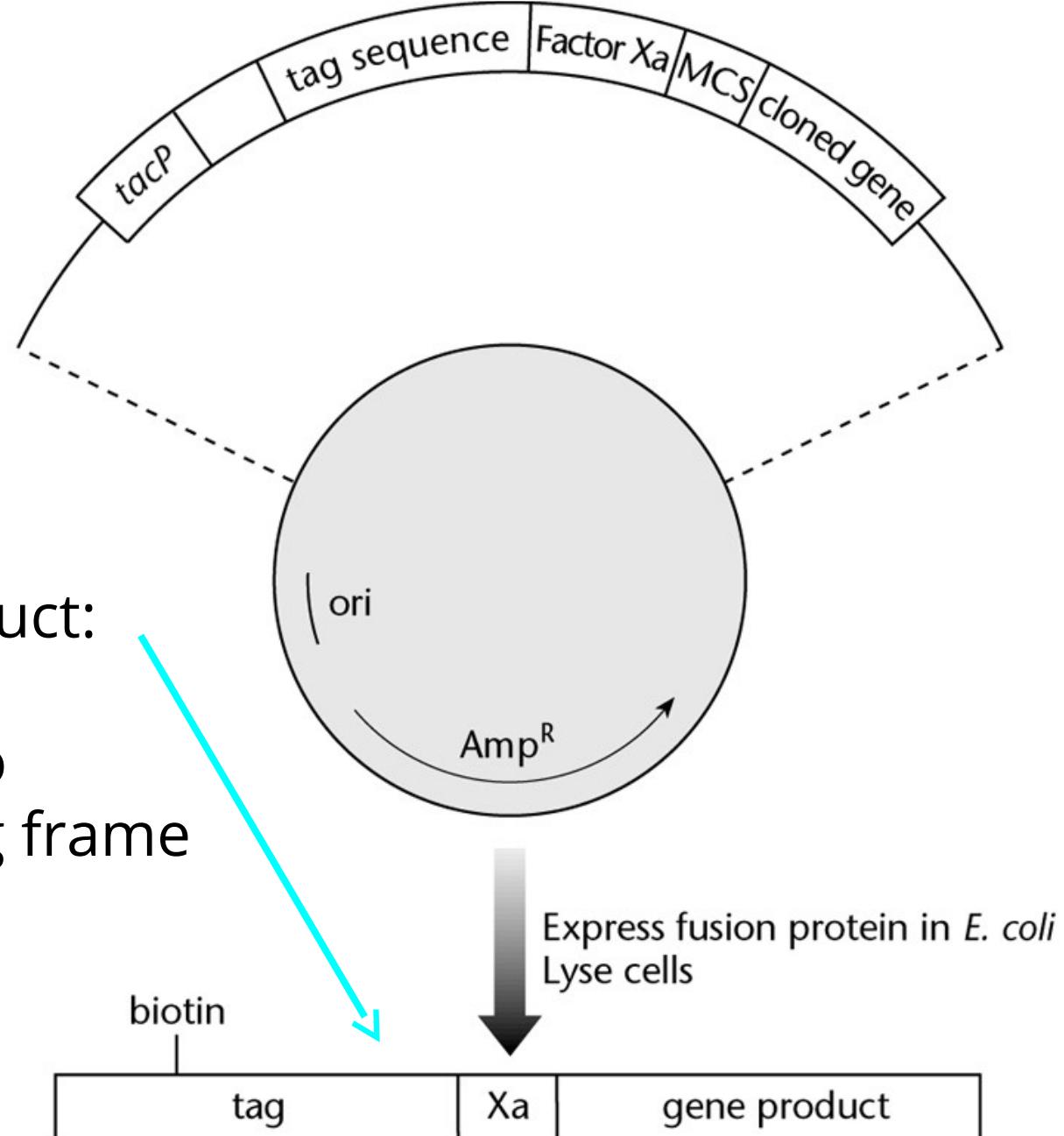
# Epitope tags as purification handles

Protein tag	Size	What the tag sticks to	What causes release
Polyhistidine (6-10xHis)	6-10 a.a.	Immobilized Ni, Co, Cu, Zn ions	Imidazole >100 mM
Glutathione S-transferase (GST)	211 a.a.	Immobilized glutathione	Reduced glutathione
FLAG-tag: DYKDDDDK	8 a.a.	Anti-FLAG antibody	FLAG peptide or low pH
Protein A	280 a.a.	Any immunoglobulin G	Protease
Calmodulin binding protein	26 a.a.	Immobilized calmodulin	EGTA 2mM
Acceptor peptide <b>(biotin</b> added by biotin ligase)	15 a.a.	Avidin/streptavidin	Biotin

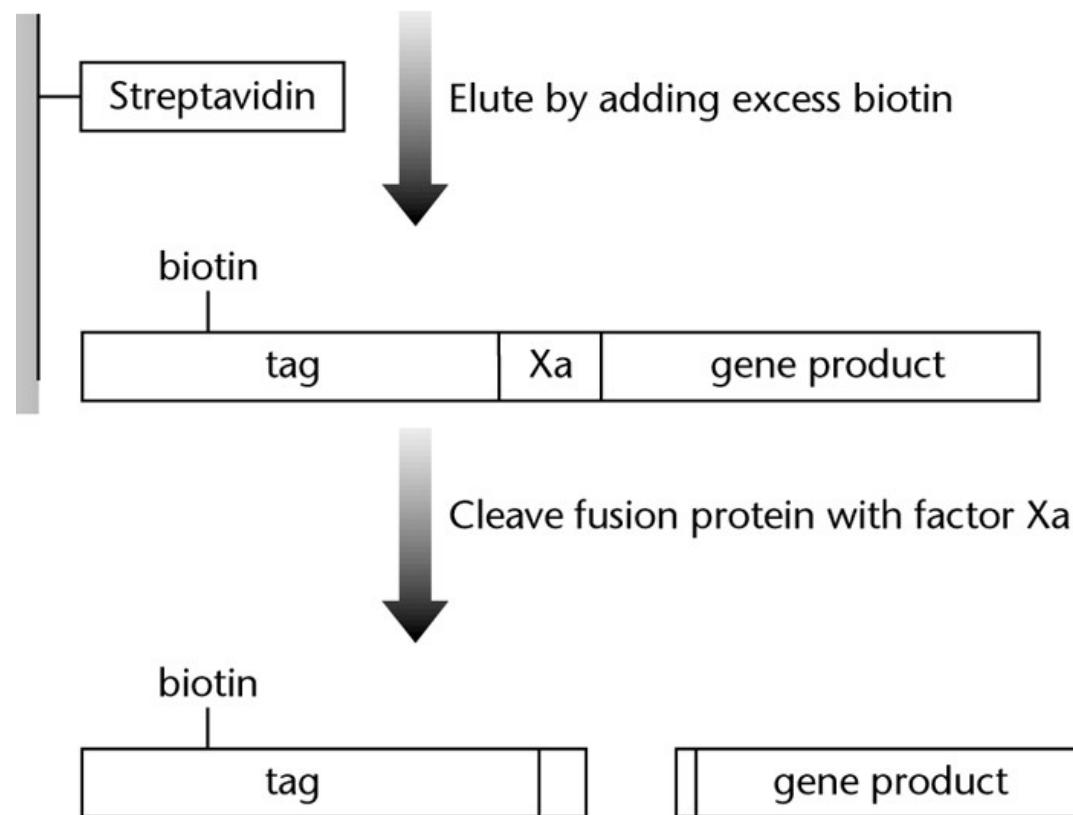
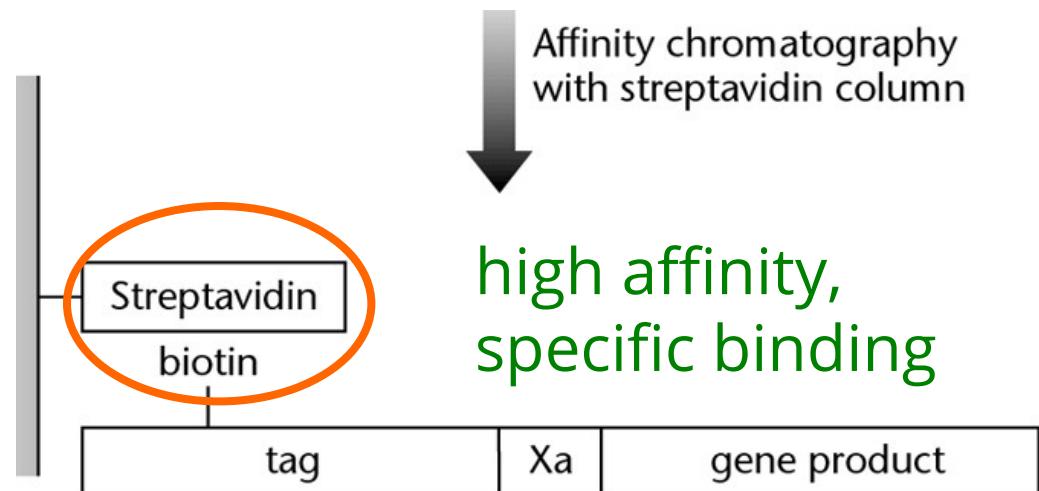
# Tags for protein folding/solubility

<b>Protein tag</b>	<b>Size</b>	<b>What the tag does</b>	<b>Additional facts</b>
SUMO (small ubiquitin-like modifier)	98 a.a.	Improve protein folding/solubility	Can be removed by SUMO protease
MBP: maltose binding protein	396 a.a.	Helps folding/solubility AND sticks to cross-linked amylose for purification	Released from amylose by adding maltose
Trx: thioredoxin	109 a.a.	Improve protein folding/solubility	Extremely soluble, even at 40% of total cell protein, prevents inclusion bodies

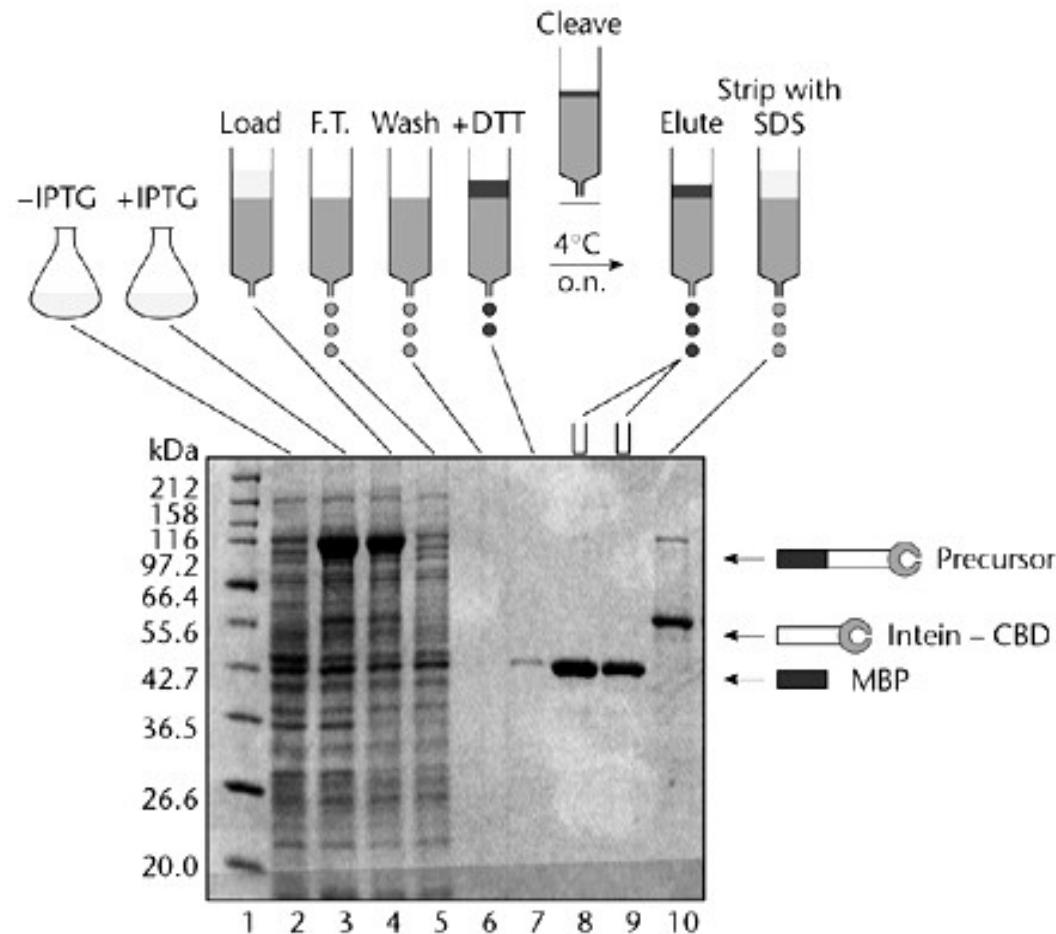
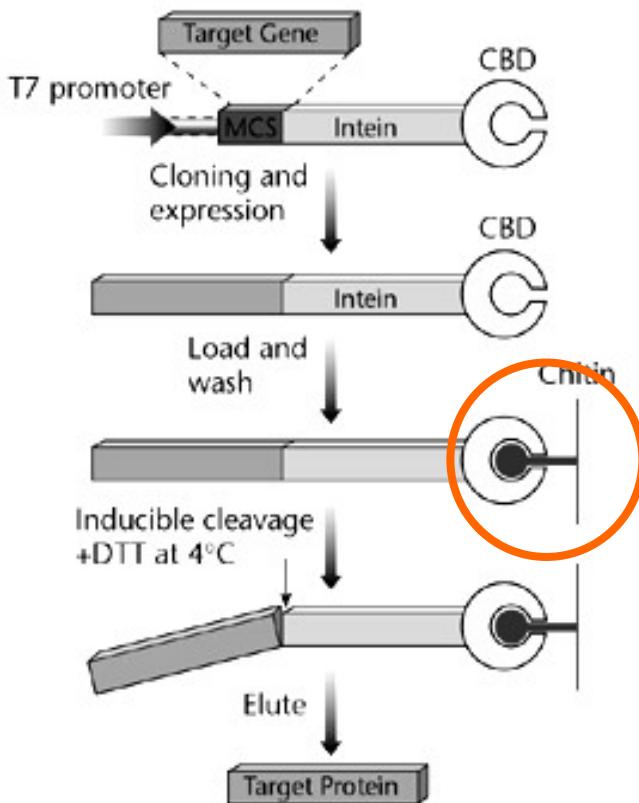
# Biotin tag for protein purification



# Biotin tag for protein purification



# Another protein purification scheme--removable tag



Lane 1: Protein Marker.

Lane 2: Crude extract from uninduced cells.

Lane 3: Crude extract from cells, induced at 15°C for 16 hours.

Lane 4: Clarified crude extract from induced cells.

Lane 5: Chitin column flow through (F.T.).

Lane 6: Chitin column wash.

Lane 7: Quick DTT wash to distribute DTT evenly throughout the chitin column.

Lanes 8-9: Fraction of eluted MBP after stopping column flow and inducing a self-cleavage reaction at 4°C overnight.

Lane 10: SDS stripping of remaining proteins bound to chitin column (mostly the cleaved intein-CBD fusion).

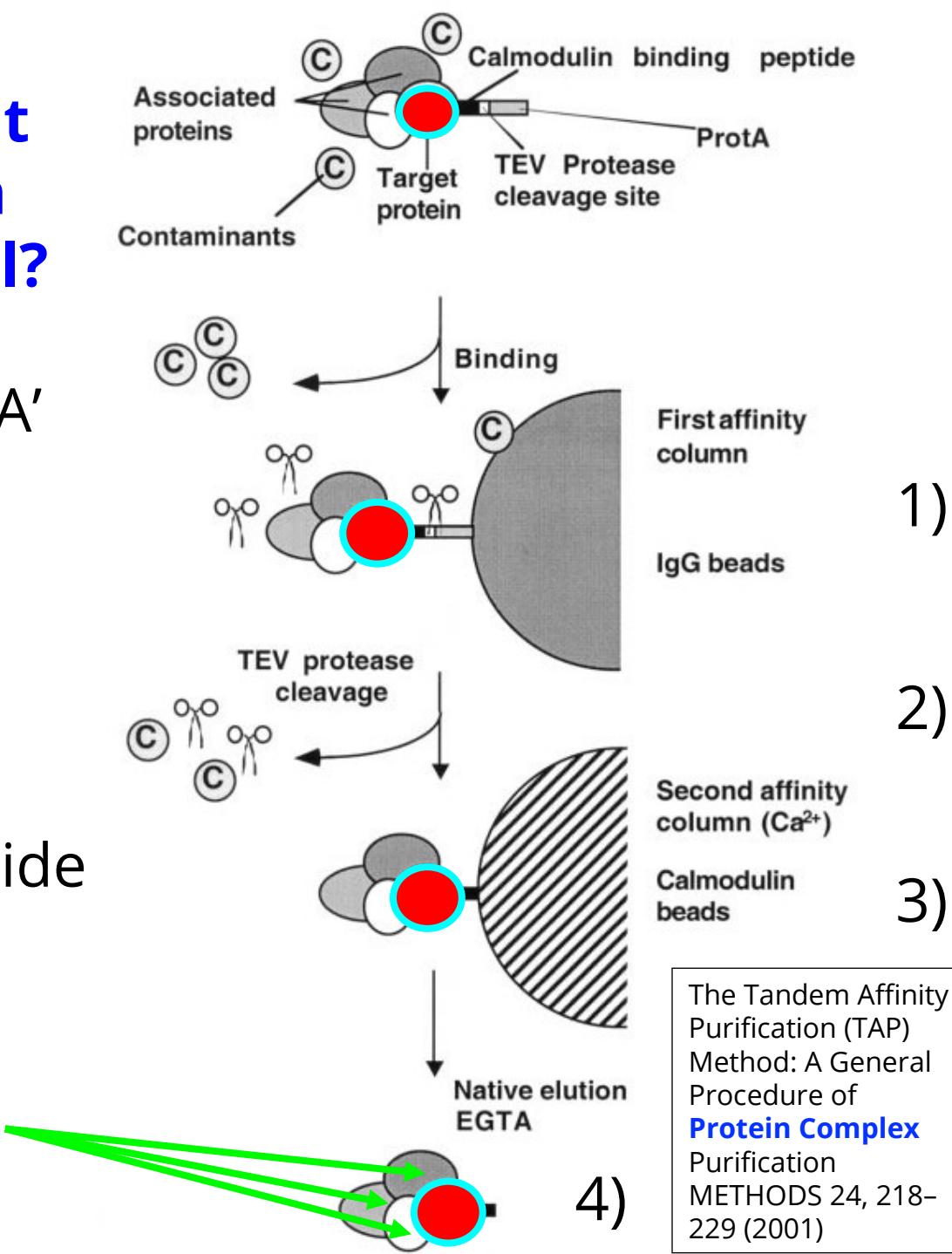
# Tandem affinity purification (TAP): what does my target protein interact with in the cell?

1) IgG beads for 'protein A' tag

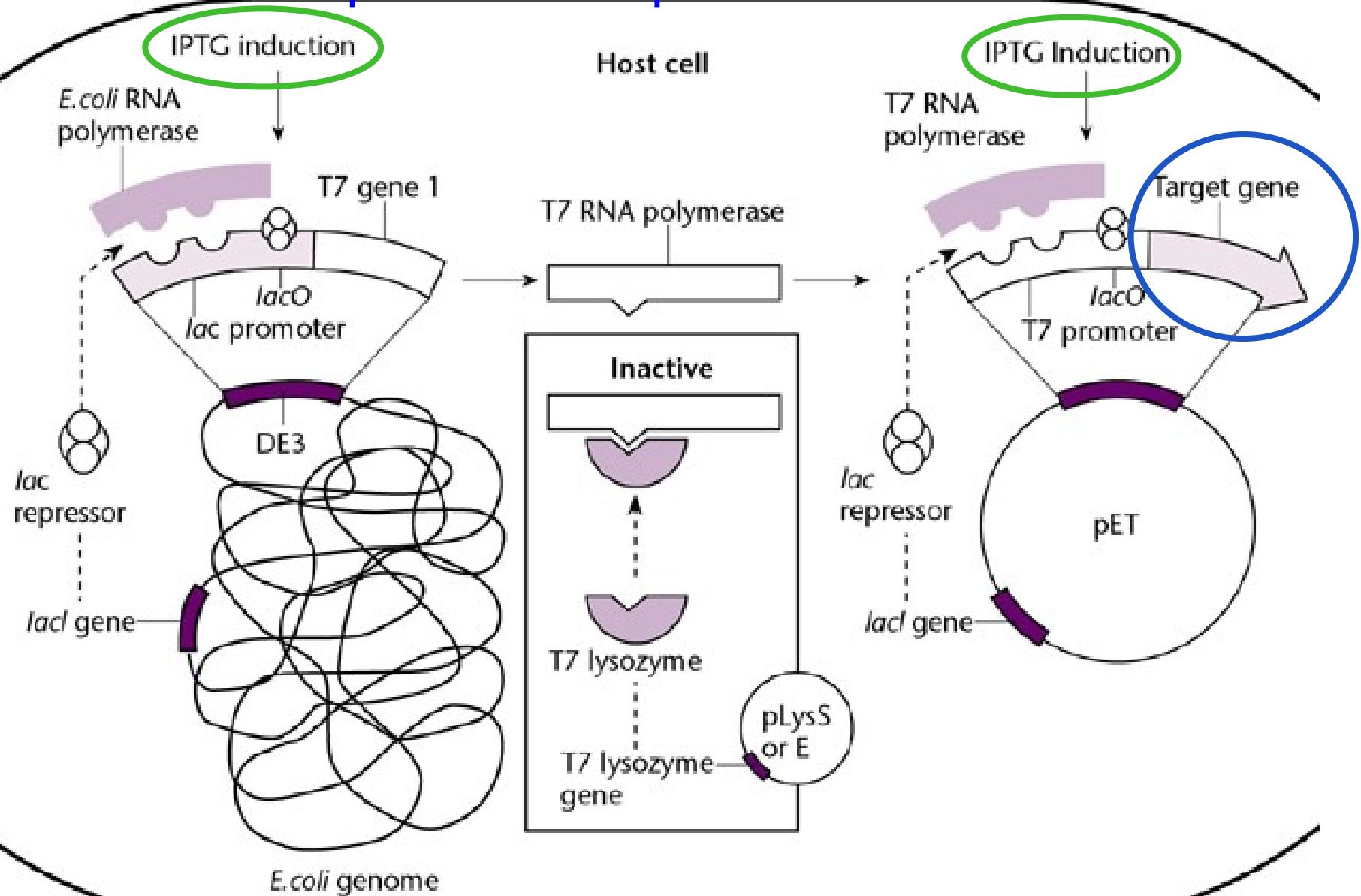
2) TEV protease cleavage

3) Calmodulin beads for calmodulin binding peptide

4) Pure protein and associated complex



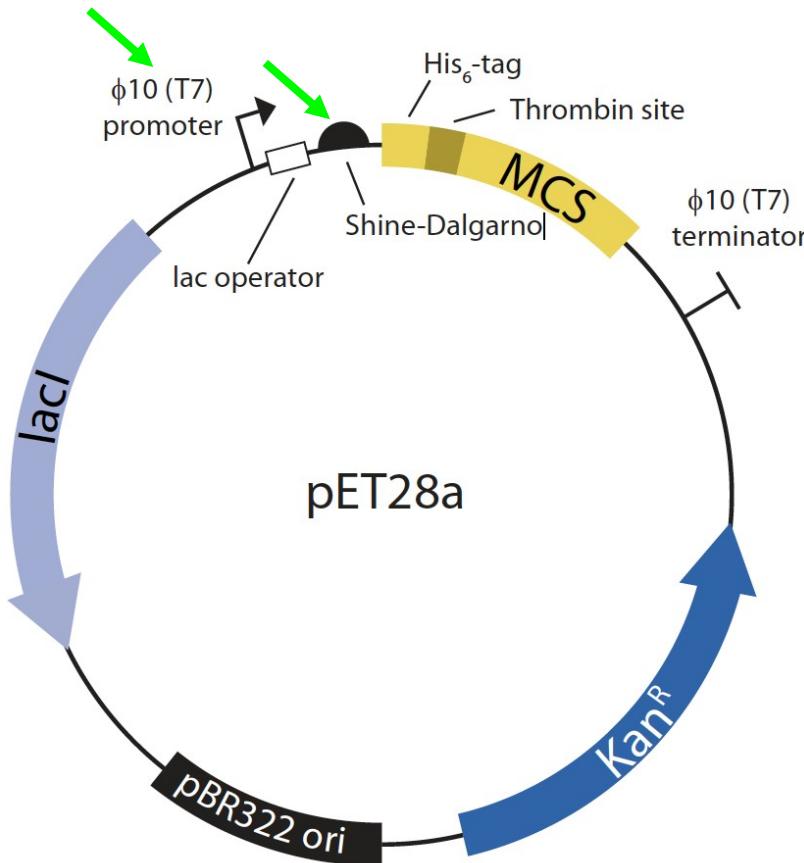
# pET vectors: protein expression



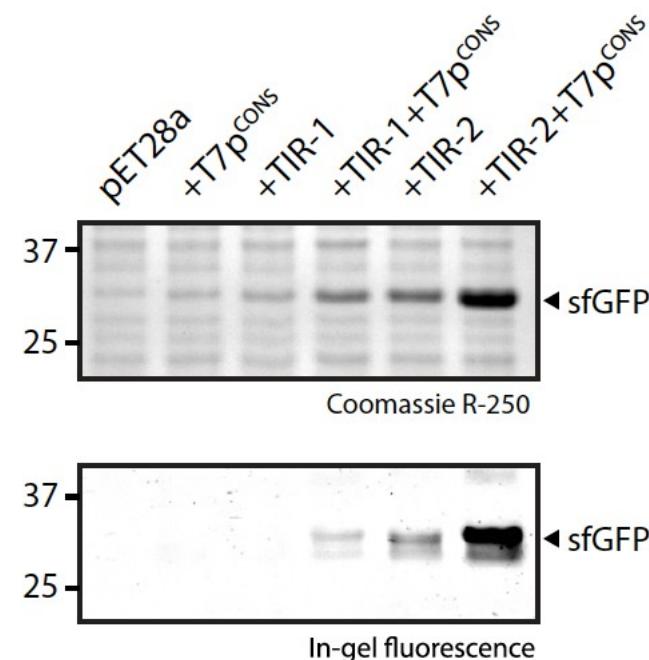
+IPTG = high T7 RNAP, high target gene expression

# Updating pET vectors: new developments in 2020

## How to get even more protein expression



- 1) Change T7 promoter to the  $\phi 10$  consensus sequence
- 2) Alter translation initiation region (TIR) by searching through library of TIR mutants



Improved designs for pET expression plasmids increase protein production yield in *Escherichia coli*

Patrick J. Shilling<sup>1</sup>✉, Kiavash Mirzadeh<sup>1,2</sup>, Alister J. Cumming<sup>1</sup>, Magnus Widesheim<sup>1</sup>, Zoe Köck<sup>1,3</sup> & Daniel O. Daley<sup>1</sup>

COMMUNICATIONS BIOLOGY | (2020)3:214 | https://doi.org/10.1038/s42003-020-0939-8 | www.nature.com/commsbio

# Purification of a 6xHis tagged protein, start to finish

## 1) Clone gene into pET vectors

- Design primers, PCR
- Choose vector, ligate gene in frame to vector
- Sequence to verify the clone
- Transform the plasmid into BL21 strain
  - T7 RNA polymerase gene
  - Lacks proteases

## 2) Overexpress the protein

- Grow the cells to mid-log phase
- Add IPTG to induce overexpression of the gene
  - Induction temperature may vary (15 – 42°C)
  - Concentration of IPTG varies (0.01 – 2.0 mM)
  - Time of induction varies (1 hour to overnight)

# Purification of a 6xHis tagged protein, start to finish

3) Break open the cells

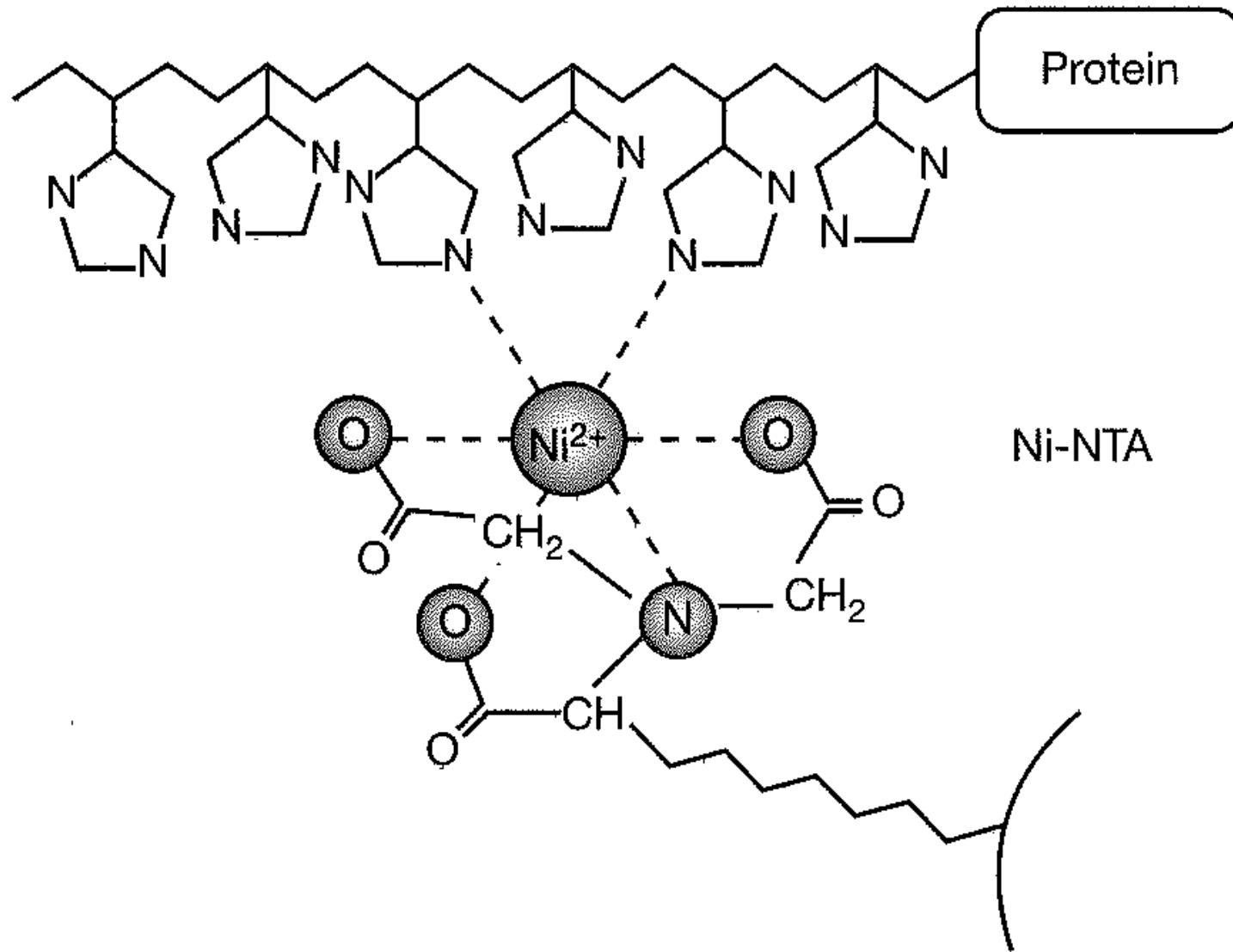
- lysozyme, sonication
- add protease inhibitors
- spin to get a clarified “lysate”

4) Affinity purification of protein

- apply lysate to a Nickel column or to Nickel beads
- wash with low concentration imidazole (20 mM)
- elute with high concentration imidazole (200 mM)

5) Check protein purity and concentration by SDS-PAGE

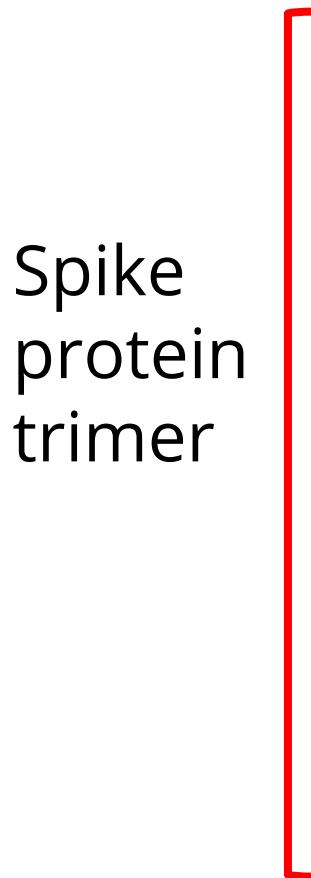
## 6xHis interaction with Ni<sup>++</sup>



# Protein purification and vaccines: SARS CoV-1 in 2009

- o Many different vaccines for SARS Cov-2 are in development now, for example:
  - Inactivated SARS CoV2 virus (Sinopharm, CoronaVac)
  - Vector-driven protein expression (Sputnik V)
  - Nucleic acid (RNA or DNA) vaccine (Moderna, Pfizer/BioNTech)
  - Antigenic proteins (especially the spike **protein**)
- o Production of antigenic proteins?
  - Example: SARS CoV-1, outbreak in 2003, thousands of infections, hundreds of deaths
  - No new cases since 2004
  - Can spike protein epitope (including only the receptor binding domain) confer protective immunity?
  - Du et al. 2009, doi:10.1016/j.virol.2009.07.018

# Structure of SARS CoV-1 spike with ACE2 receptor



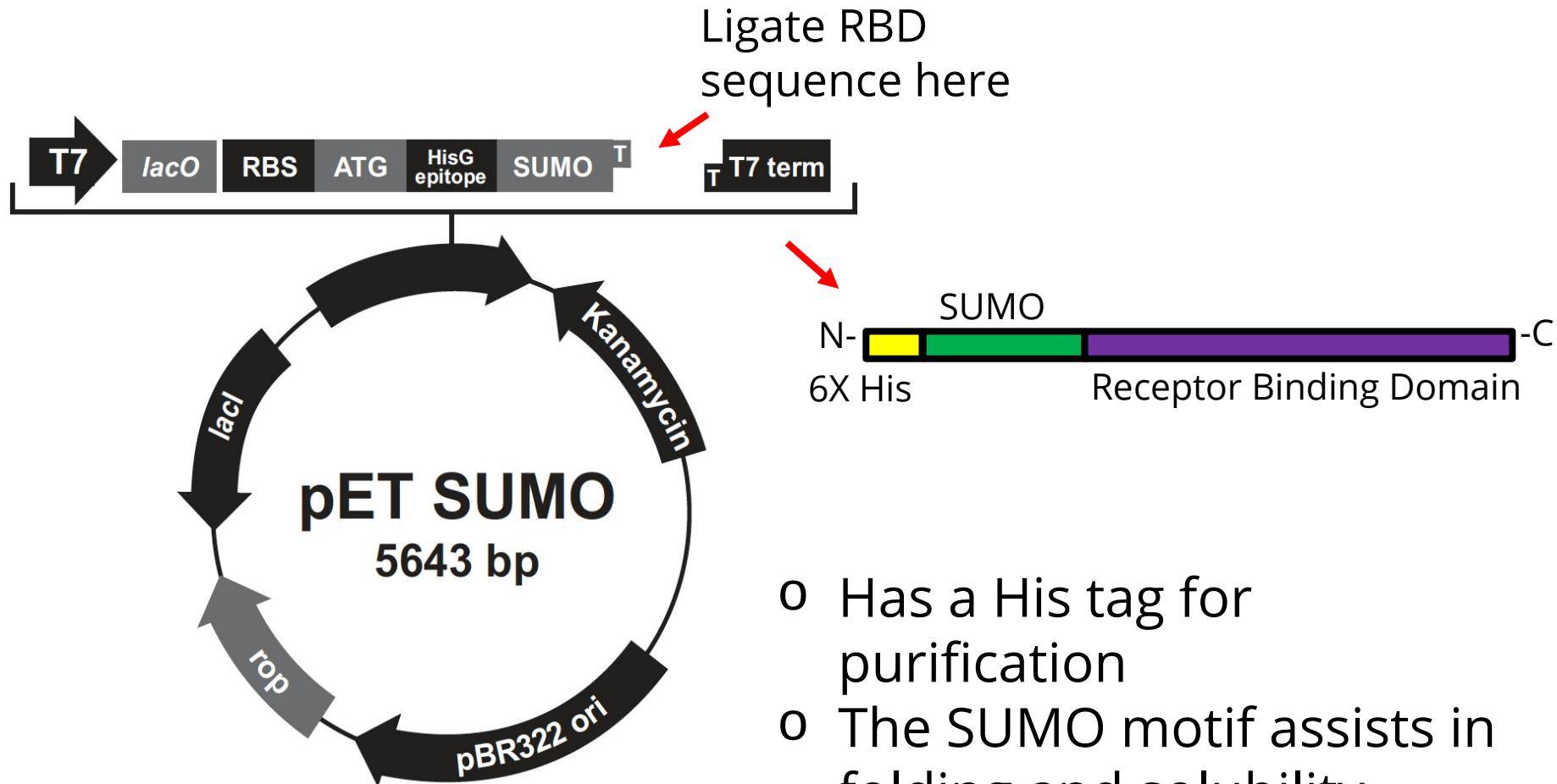
← Human ACE2 receptor

← RBD: receptor  
binding domain, a.a.  
318-510

PDB ID  
6ACG

Can the RBD epitope  
by itself elicit a  
protective immune  
response?

# Making the RBD: protein expression in *E. coli*



- o Has a His tag for purification
- o The SUMO motif assists in folding and solubility
- o SUMO motif gets cut away by SUMO protease

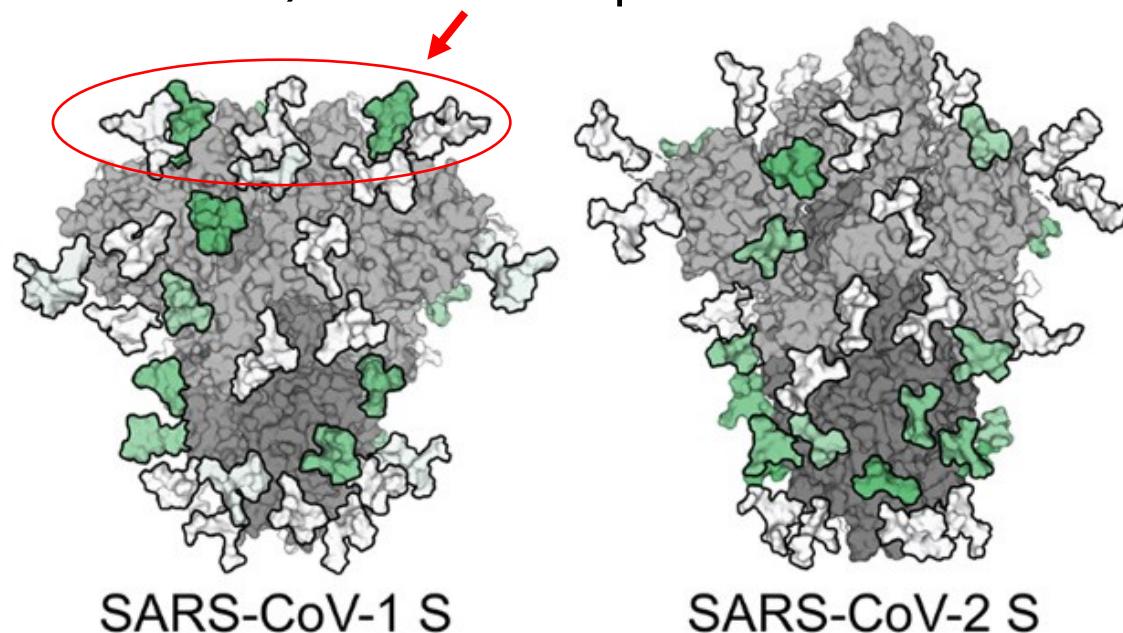
# Comparison of *E. coli* produced RBD to baculovirus and human cell line expressed versions

- o RBD was also expressed in human cell line, as well as in baculovirus infected insect cell line
- o All three versions were used to vaccinate mice
  - 20 micrograms purified RBG, followed by two boosts of 10 micrograms each in 3 week intervals
  - The mice were tested for:
    - humoral immune response (antibodies)
    - protective immunity
- o All three yielded humoral (antibody-mediated) response, although less so with *E. coli*-made RBD
- o All three vaccines conferred similar protective immunity (in mice) to SARS CoV-1 infection

# Advantage and disadvantage to *E. coli* expression?

- Using *E. coli*, lots of doses of the vaccine could be made cheaply & easily
- However, the protein would not be glycosylated in *E. coli*, this may explain the lower humoral response in mice

Glycosylation map shows sites (in green and white) that overlap the RBD



## 2020-21 update:

- Spike (RBD) protein currently being tested as low cost (~\$1/dose) vaccine alternative, see Hotez & Bottazzi 2020
- Vaccine testing in progress in India
- Both SARS CoV and SARS CoV2 RBD proteins being tested
- Proteins produced using the yeast *Pichia pastoris*
  - can do glycosylation
  - Proteins secreted into growth medium, so easier to purify

# Protein overexpression: pitfalls and limitations

## 1) Low or no gene expression

- a) mRNA synthesis and stability:
  - i. Premature transcription terminator
  - ii. Unstable mRNA
- b) Protein synthesis and stability
  - i. Gene has lots of rare codons: if a low abundance codon for *E. coli* is used a lot in gene, translation rate will be slow (limited [tRNA])
    - i. Check codon usage in *E. coli* at:  
<http://people.mbi.ucla.edu/sumchan/calton.html>
    - ii. Protein product is toxic to the cell: gene deletions occur
    - iii. Protein is degraded by cellular protease

E. coli rpoB

ATGGTTTACTCTTACCGAGAAAAACGTATTCGTAAGGATTGGTAA  
ACGCCAACGTTCTGGTGAACCTTATCCTCTTACAGTGACT  
CGTTTCAAGAATTTCGAGCAAGCTGAAAGGGCAGTATGGCTGGAA  
GCTCTTCCGTTCCGTTCCGATTAGAGCTACAGCGGTAACTCGA  
GCTGCAATAGCTGACTGCTGGTGAACCTTGGCAGGGTGTGAC  
AATGTCATAATCGTGGCTGACCTTATCCGACCGCGTGGCTTAAACTG  
CGTCTGGTATCTAGGCGCGAAGCGCCCGAAGGCACCGTAAAGACAT  
TAAAGAACAAAGTCTACATGGGCAAACTTCCGCTCATGACGACA  
GTACCTTTGTTACACCGTACTGGCTGTATCTTCCAGCTGAC  
CGTAGCTGGCGCTCTTGTGACTCGACAAAGGAAACCCACTCTC  
GGGAAAGTGTGATATAACCGGGTATCATCCCTACCGTGGTCTCG  
TGGACTCTGAAATTGATCGGCAAGAACACTGTTGACGGCT  
CGCCGAAACTCGCTGGCACCATCTTCCGCGCCCTGAACTACACCAC  
AGAGCAGATCTCGACCTTCTTAAAGGTTCTTAAAGCTTGG  
ATAAACAGCTCGAGATGGAACGTGGTCCGGAAACGGCTCGGTGG  
GCATCTTGTGACATCGAACGTTAACGGTAAAGTGTACGTAGAAAAGGCC  
CGTATCGACCGCCACATTGCGAGCTGGAAAAGACGACGCTCAAAC  
TGATCGAAGTGGCTTGGTGTGACATCGCAGTAAAGTGTGAAAGAC  
TATATTGATGAGTCTACGGCGAGCTGATCTGCGACGCAAGATGGCT  
GAGCGTGGATCTGCTGGCTAAGCTGAGCCAGTCTGGTCAACAGCTATCG  
AAAGCTGTTACCGAACGATCTGATGACCGGCTATATCGTAAAGAC  
TTACCTGTCGACCCAACTAACGGCTCTGCGACTGGTAGAAATCTA  
CCGATGTCGCCCTGGCAGCCGCGACTCTGTAAGCAGCTGAAAGCC  
TGTGAGGCTTCTTCCTCCGAAAGCGCTTATGACTTGTCTGGTGG  
GGTCTGATGAAAGTCAACCGTCTCTGCTGGCAGAAGAACTGAAAGGTT  
CGGTATCTGAGCAAAGACGACATCATTGATGTTATGAAAGGCTCATCG  
ATATCGGTAACGGTAAACGGGAACTGATGATGATCGAACCTCGTAA  
CGTCTATCGTCTGGTGGGAAATGGCGGAAACCGTGGCTGGTGG  
CTGGTACGTTAGAGCTGCGGTGAAAGAGCGTCTGCTCTGGCGATC  
TGGTACCCCTGATGCCACAGGATATGATCAACCGGAAAGCGGATTTCG  
GAGTGAAGAGTTCTCGGTTCCGACCGAGCTGCTGAGCTTATGGACCA  
GAACACCCCTGCTGAGGATTAACGCAACAGCTGATCTCCGACTCG  
GCCAGGGCTGACCCGTGAACGCTGCAAGGCTCGAAGTTCGAGACGTA  
CCCCGACTACTGGCTCGTGTGCAATCGAACCTGAAAGGGTCA  
GAACATCGGCTGATCAACTCTGCTCGTGTGACCGACAGACTAACGA  
ACGGCTTCTGAGACTCTGATCGAAAGTGACCCGACGGTGTGTA  
GACGAAATTACTACCTGCTGATCGGAAAGGCAACTCGTAGTAAAGAC  
CCAGCGGAACCTAACCTGGATGAAAGAAGGCCACTCTGAGTAAAGAC  
TAACCTGGCTAGCAAAGGGAATCCAGCTTCAAGCCGGACCCAGGTT  
GACTGATGACGATCCACCCAGGGTGTGCTGGTGTGCTGGTGTG  
GATCCGTTCTGGAAACAGCATGACGCCAACGGTCTGATTGGTGG  
ACATGCAACGCTCAGGCCCTGGACTCTGGCGCTGATAAGCCGCTGGTT  
GGTACTGGTATGGTGTGCTGGTGTGACTCTGGTGTACTGGTGT  
AGCTAACCGTGGTGTGCTGGCTAGTACGTTGATGCTTCCGTTATCGTA  
TCAAAGTTAACGAAGACGAGATGTATCGGGTGAAGCAGGTATCGACATC  
TACAACCTGACCCAAACCTGGTCTTACCGAACCTGATCAACCA  
GATGCCGTTCTGGTGTGCTGGTGAACGGTGGCAGCTGG  
CACAGCGTCCGTCACCGCACCTGGTGAACTGGCGTGGTCAAGACATG  
CGCTAGGCTCATGGCTGAACGGTAACTGGTACACTGGGAACT  
CGTATCCGAGCTGTTCTGGTCAAGGACGGTCTTACCCACATCAC  
AGGAACCTGGGTGTGCTGGCTGACCAAGCTGGTGGGAGAGATC  
ACCGCTGACCCGAACGGTGAACGGCTCTGGGAACT  
ATCCGGTATGTTACCTGGCAGAAGGTAACTCAGCTGACCCCAGAAGAAAA  
CTGCTGGCTGCGATCTGGTGAAGGAAAGCCTCTGAGTAAAGACTCTC  
TCTGGCTGACCAACGGTGTACCGTGTGCTGGTCAAGGATCTGGT  
TTACTCGGATGGCTGAGGAAAGACGCTGCTGGTCAAGGAA  
ATGCACTAACACGGCGAACGAAAGACCTGCTGAGAAGAC  
CGAAGCGGGCTGTTCTGGCTGACCTGGCTGGGAGAG  
TTGAAGCTGAGAACGCTGACAAACTGGCGGCGATCTGGCTGGAGCTG  
GGCGTACAGACGAGAACGAAAGAACATCTGAGGAAACGGTGT  
GTATGACGACTGAAACACGGAGTCTGGAGAACGAAAGAC  
GCAAAATCACCGGCGACGATCTGGCACCGGGCTGCTGAGAAGATTGTT  
AAGGTATATCTGGGGTAAACAGCGTATCCAGCGCTGGTACAAGATGCC  
AGGCTGCTGCGGTAACAGGGTGTAACTTCAAGATCGCAGTGGATCGA  
ATATGCTTACGATGAAACAGCTGCGGCTGAGACATCTGACTGAA  
CTGGCGTACCGTCTCGTATGAAACATCGGTAGATCTCGGAAAC  
GGGTATGGCTGCGAAAGGTATCGGGCGACAAGATCAACGCCATGCTGAAAC

AGCAGCAAGAAGTCGCGAAACTGCGCAATTCTCAGCGTGCCTGACGAT  
CTGGCGCTGACGCTCGTCAAGAAGTGTGACTGAGTACCTTCAGCGATGA  
AAGAAGTATCGCTGGTGAACCGTCAAGGTTGCAAGTGGCTGAC  
CGCCGCTGTTGACGCGTGGCAAGAACGCAAGAAATTAAAGGCTGCTGAA  
CTTGGCAGCTGCCGACTCCGGTCAAGATCGCCGCTGACGATGGTGCAC  
TGTTGAAACAGTCTGGCTGGCTGAAAGCTGGTACATGATCGTGA  
AACTGAAACCCACTGTCGACGACAGATGCAACCGCTTACCCAC  
TACAGCCTGGTACTCAGCAGCCGCTGGTGAAGGCAAGTCTGGTGG  
TCAGCTGGTCTGGGAGATGGAAGTGTGGCCCTGGAAAGGATACCGCCGAG  
CATACCTCTGGCAAGAAAATGCTACCGTGAAGCTGACGCTGAA  
CGTACCAAGATGATAAAACATCGTGGACGCCACCATCGATGGAGCC  
GGGATGCGAGAATCTTCAACGTTAGTGTGAAAGAGATTGCTGCTG  
GTATCAACATCGAAGTGGAGACGAGTAA

Archaeal RNAP subunit B

ATGAGAGCTCGACGACTGTTGATGATGTTACTCCGAGATCTTGGATTG  
AATGGAGAGCTATTGGCAAGAGAAAAGGGCTTGTGAGACAGCATCTTGC  
CATAACTGCCTTATTGAGAGAGGACTTCAAGAAGGTTGAAATGAGTT  
GGTGAATAAGACGGGATATTCCAGATTTGAAGAATGTTGCTGAA  
TAGAGTAGGAGAACGGAGTTCAAGAACCTCATGGTCAAAGAAGGCC  
TTATCCAATGGATGCAAGAATAAGAAACCTAACCTACGCCGACCCACTC  
TACCTGAAAGATGTTCAAGTAAAGGAGGTTGAGCAAGATGGCTAGA  
GTTGAGAATGGGAAACTCTTAAATGCTAAATGAAAGTTGTTAGG  
TGTACGGACTTAGCGATGAGAATTAAAGCTTGGAGAGGATCCAAG  
GATCCAGGGTGTACTCTATAACCGGTTCTGAAAGGTTAGTACT  
TATTGAGAGCATGCTTCAATGGACTCTGGTGAAGAAGGAGA  
AAGAAAAGTACATAGCAAAGGTTCTTCTATAGACATGGGTTAGGGCT  
TTAGTGTGTTGAGAGAAGAAGGATGGAATTCTTACGTTGATATA  
TAACGTCACCTAAAGCTGTTAAATTCTGTTATGTTGAGAGCTTGG  
TGAGAGAGATAAGGATATCGTGTAGGGCAGTAGGTAATGATCCAGAA  
CAACAAATTATGTTGATAACCTGGAGATGCCAGCGATATAACTACCA  
ACAAGGCTCTGTAATTCTAGGTTAGCTGTAGCTCTGGCAGGCC  
GAGAATACAGACTTAAAGGGCAGAATATGTAATTGATAACAACTGTT  
CCTCACATGGGTCTTCTGGAGGATAGAATAAAGAAGGCTTACTAC  
GGGAATGATGGCTTAAAGGTCTGAGCTCTCCCTCGGAAGAACCTG  
AAAGATGACAAAGGATATTGACAAATAGGGCTAAGGCTGGGGAG  
CTTCTAAAGGACTTGTGCTGTTGACTCTGTTAAAGACAT  
ACAATATCAATGACCAAGACATACCAAAAGGAGAACAAATATACTT  
TTGGAAACATCATGATGTTTATTGAGGACTCTCATAAAGGCCAGCGTCTC  
ACCGAGAGAAATTGACATGCTTGTGCACTGGCTTGGCTGGGGAG  
AACGGGTTGCTGAGCTGAGTGGCTAGATAAGACTAACATGTCACATT  
CCACCTTAAAGAGCTGCTTCACTAGACGAGAACGCCCAACTC  
GAGGCTAGAGACCTTACCGGAACCTACTGGGGAGAATTGTCACCGGA  
GACTCTGAGGCTTAAAGCTGTTGCTGAGAAGATTAGGATTGATGG  
CTCAAAATAACACGGCTTCTGGTCAAGGAGGATTAGGGAGTACTTA  
TCTTGGGAATAGTCCAAATTGAGTTAGAAGGACGATCCAACATTG  
GAGAGTAACTCTTAACTGGTGTCTTCTCATGGAAACAAATAGGGATGG  
CTTATGTTCTGGAGAATTAGGCAAGACAGAAGAAGTGGAAAGATCAG  
GTTATCAATGTCGATACTATGAGGAGCTTAAAGAAGTCTACGTTAACAG  
CGACGACTGTTGAGGTTAGGGCTTAACTGGTGTGAAAGATTAGGG  
CTTAAACTAACAGACGATGTTCAAGGAGTAAAGAAGGAAAGCTCAAG  
TGAGTGACTTAAATTGAGGAAATTGCAATTGAGGCTGG  
GGGAGGAAACGGCATATGGCAATGTCGCTGGGAAGTACTGAGAG  
ACACTCACTGAGAATTATGCTGAGGAGCTTGGCTTGG  
CTTGTCCCTATCTGAAACACAACGCACTGCAAGAACACGATGTTGAG  
AGGATGGCCTGAGGAGCTGGACTCTGGATGGCAAGAACATT  
GGTAGATGAGGACTTCACTTATTACACAGATTGGGAGTACTTA  
AACTCAAGAAATATGAAAGCAGTAGGATATGAGGATAGGCTGCTG  
GAACCTGGCTGTTGCTGAGCTTACCGGTTAACAGGTTATAACATG  
CTATCATATAAAACACGGCTTCACTAGAGAGAGGACTCGCCAGATCA  
TCTTCTGGAGAAGTGAAGCTGAGGAGAAAGAACATCTGG  
TGACAAAGTTGAAATTCCGGATCTGTAAGGTTACAGAGGAGAAA  
AGTACTATGAGAACCTGATGAGAAGTGGCTCATATTCCAGAAC  
GTCGAGGGAGAAGGGTCTGGAAACTGTTGCTCTGGAGAGAAGAG  
TGAAGAGAGAGGTTCTGGAAACTGTTGCTCTGGAGAGAAGAG  
GTGTTGCTGTAACGGCAGTGGAGAAGGAGTTGTTGATAAGA  
ACCGACGAGGGAGATGGAACACTGAGTTAGGTAAGGTTACTG  
GAGAATCCCGAACACTGGAGACAATTCGCTCAAGACATGGCAG  
GTGTCATTGGACTTATGTCGCCCCAGGAGGACATGCTTGG  
ACCGAGAGT

GGAATTGTAACCGATCTCATAGTTAATCCACACGGTATACCATCAAGAAT  
GACCGTGGTCAGCTAATTGAGCTATAGGGTAAAGGTTGCTTACTTA  
AGGGAAAGGAGAGGTTGATGGAACAGCATTATCGGAGAACCGAAGAG  
CTAAGAAAAGAGCTAGAAGAGCTGGTAAAGCTGATTAAC  
AATGATGATGGTATAACAGGGAAAGGATTAGGGCTGATATATT  
GTGTGATCTACTATCAGAGGCTCACCACATGGTGCAGATAAG  
GCACGTTCAAGAGGCTCTGCTCAAGTCTTAACTAAGCAACGG  
ACGGAGAGTAGAGGGTGTCTAAGATTGAGAAATGGGAGAG  
TTGTTGCCATGGAGCTGCCATGTTACTAATAGGGTTACT  
AGGCTGTTGAGGAGAGAGAGAG  
AGCATAAGACTGAAGTATGGGTATGAGAGCTGACACATAG  
AGGGACAAGGAAGAAGAAGAGTCTACTGTCCTTTGAGAGAAG  
AAAGAATTAGCAAAGTGTGAGGATGAGCTACCGGTTCAAA  
GAGTGTGAGGAGGGTATTAGACCTAAGTTAAACCTCTCAGAG  
GTGA

# Protein overexpression: pitfalls and limitations

## 2) Protein product is not functional or hard to purify

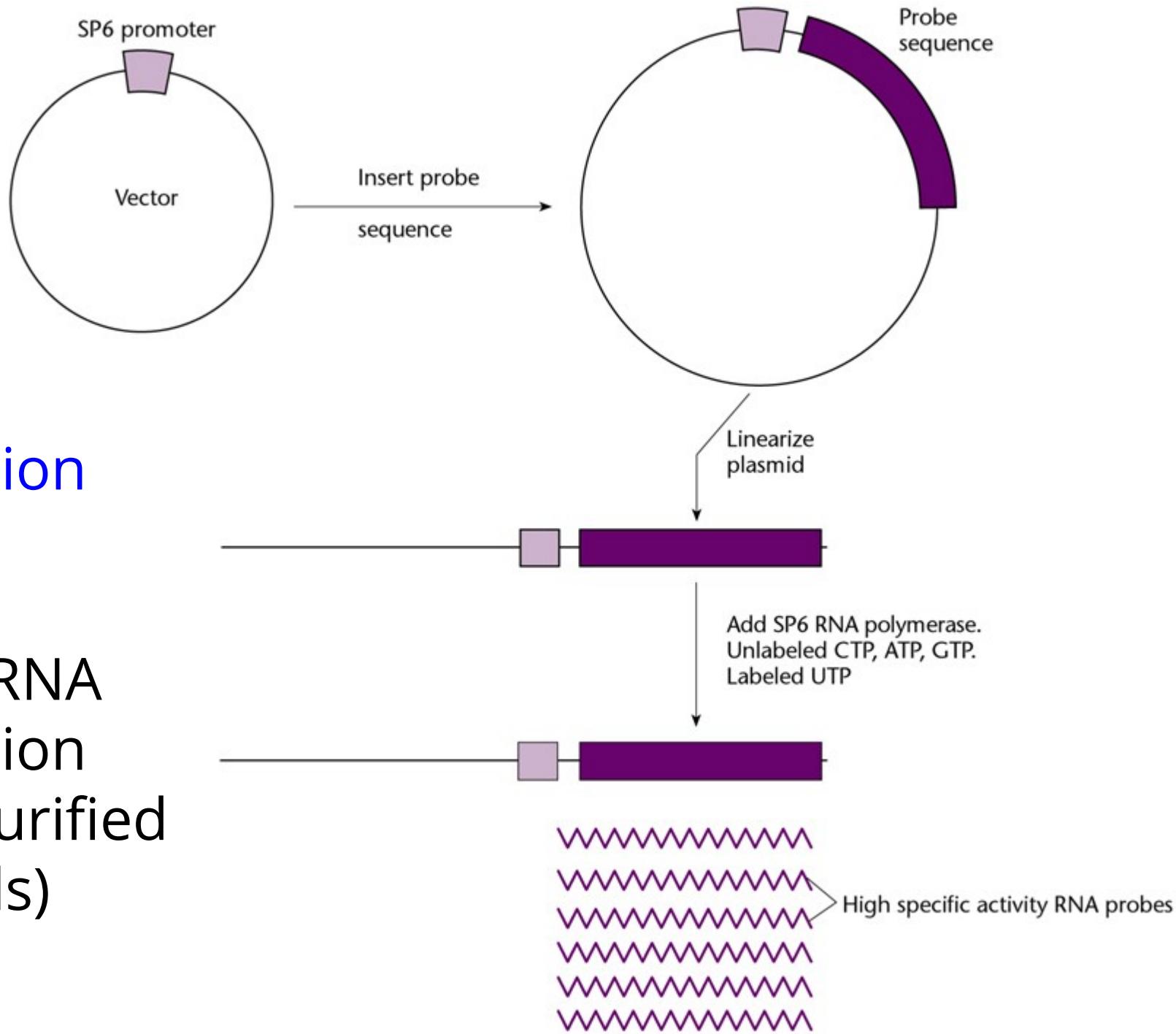
### a) *Inactive protein*

- i. Lacking necessary cofactor
- ii. Does not fold properly
- iii. Needs posttranslational modification
  - Glycosylation
  - Acetylation
  - Phosphorylation
- iv. Functions as part of a multi-protein complex

### b) *Protein purification or detection difficulties:* hard to separate or distinguish from host proteins

- Add epitope tag

RNA  
expression  
vectors:  
*in vitro* RNA  
expression  
(from purified  
plasmids)



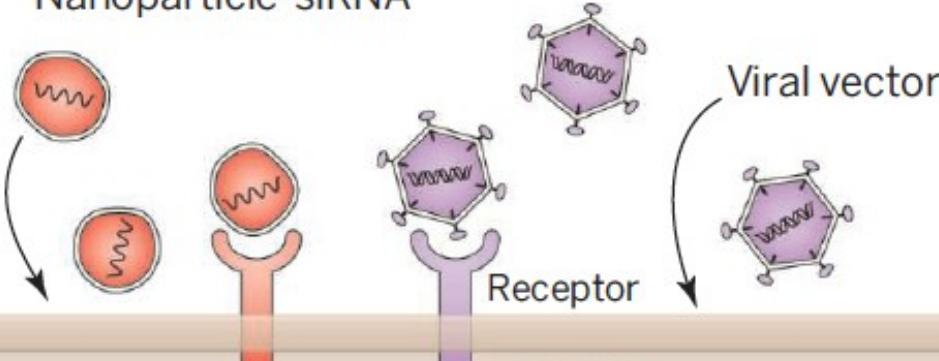
# Therapeutic RNAs: RNA interference (RNAi)

- Micro RNAs and small interfering RNAs can be used to knock down expression of disease-causing genes.
- The current best RNAi therapeutic target is the liver
  - because of the liver's physiological role in removing particles from circulation, the RNAi therapeutic agent is very likely to reach it
- Some diseases that can be targeted at the liver:
  - Transthyretin, causing amyloid neuropathy
  - Hepatitis B
  - Liver cancer
  - Hepatic porphyria (heme biosynthesis disorder)

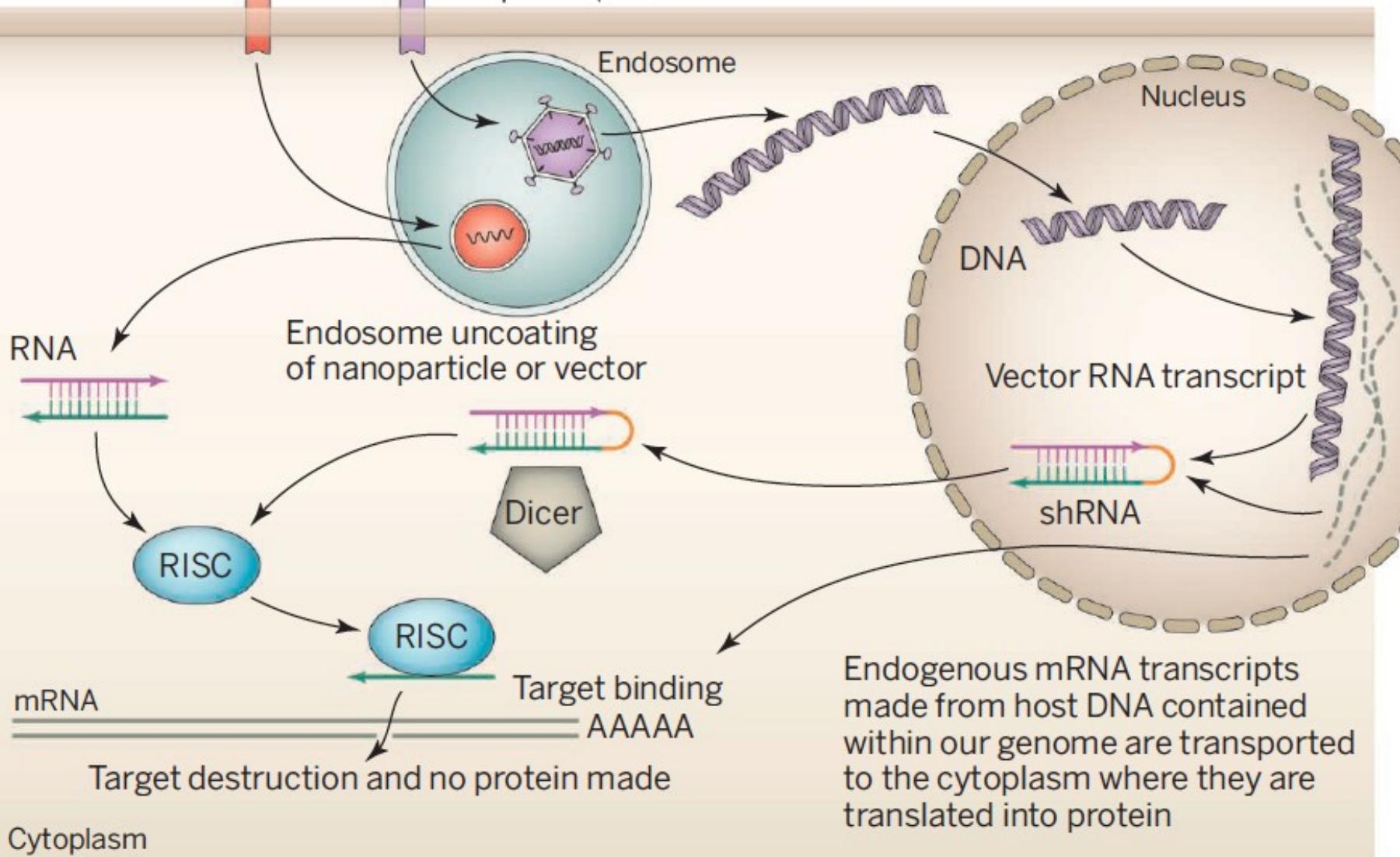
## Two modes of delivery of therapeutic RNAi

- 1) Premade RNAi precursors produced by *in vitro* transcription or chemical synthesis
  - Lipid nanoparticles (LNPs)
  - GalNac -RNA conjugates, which induce receptor mediated endocytosis
- 2) RNA expression vectors (recombinant virus) that make RNAi precursors upon entry into cells
  - Theoretically better for life-long treatments or hard to reach tissues
  - Precise control of expression is hard to achieve

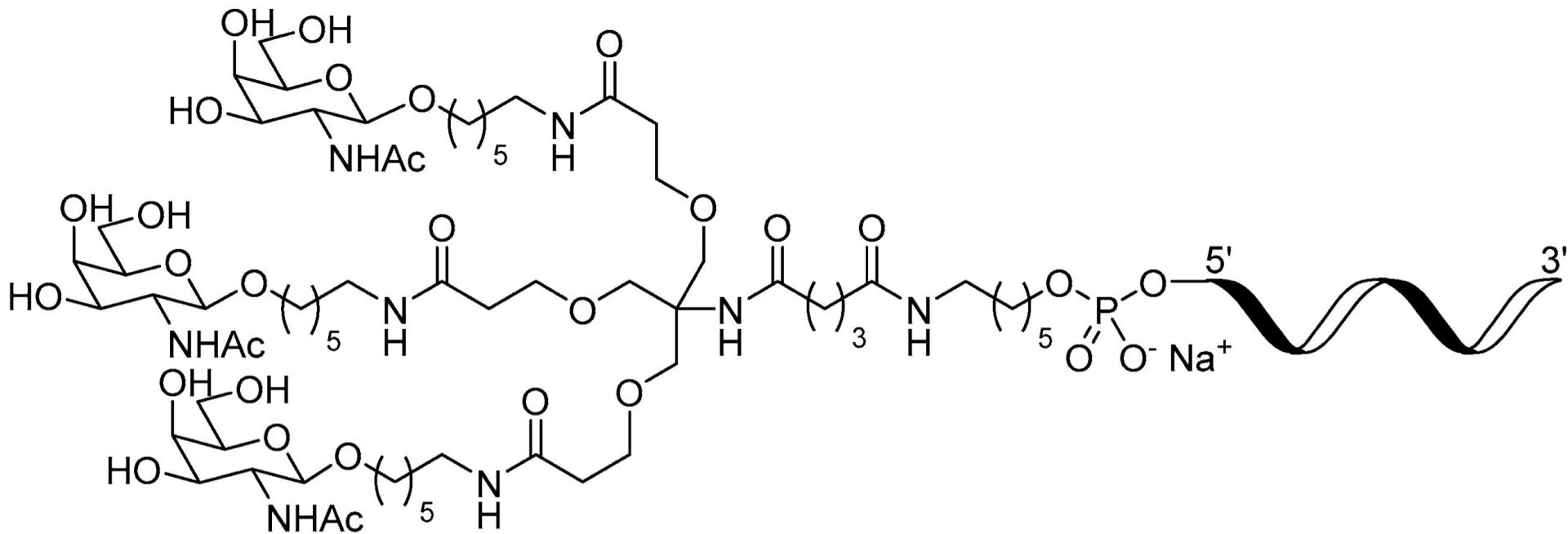
## Nanoparticle-siRNA



Two delivery modes  
for RNAi therapy:  
nanoparticle versus  
viral vector

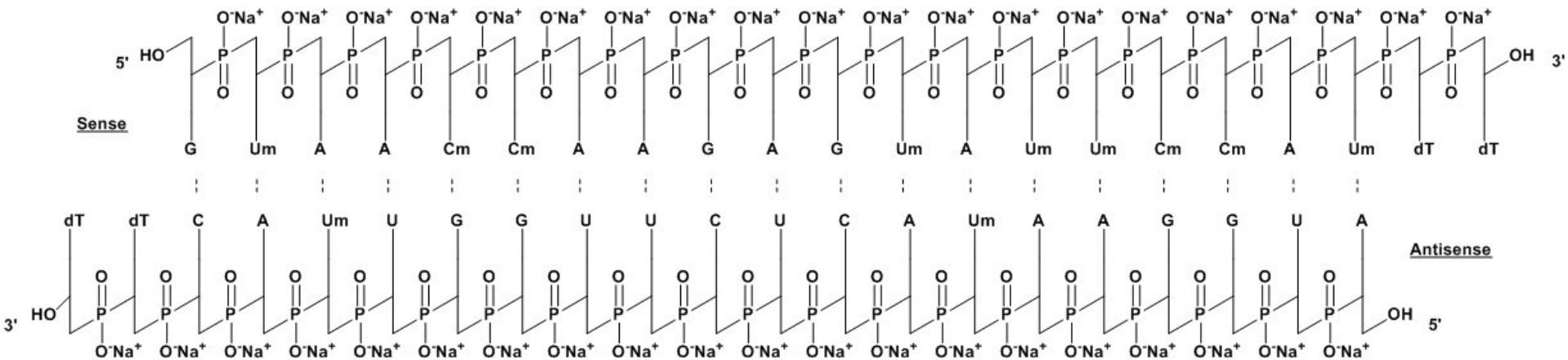


# Nanoparticle (NP) example: N-acetylgalactosamine (GalNac)/RNA conjugate



- Binds to the asialoglycoprotein receptor that is highly expressed on hepatocytes
- Results in rapid endocytosis

# Patisiran: first RNAi drug approved by FDA (2018)

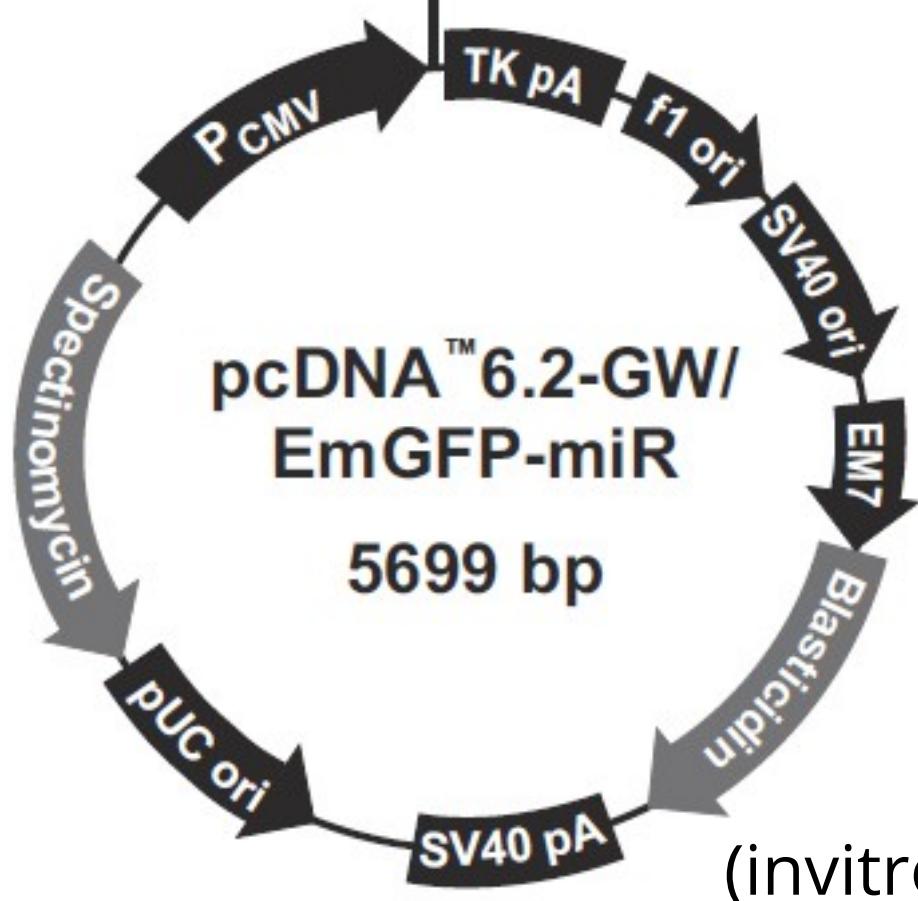


A, adenosine; C, cytidine; G, guanosine; U, uridine; Cm, 2'-O-methylcytidine; Um, 2'-O-methyluridine; dT, thymidine (modified nucleotides help tame dendritic cell immune response)

- Formulated as lipid NPs for delivery to hepatocytes
- RNAi specifically binds to a genetically conserved sequence in the 3'-untranslated region (3'-UTR) of mutant and wild-type transthyretin (TTR) messenger RNA (mRNA)
- Treats transthyretin amyloidosis (peripheral neuropathy)

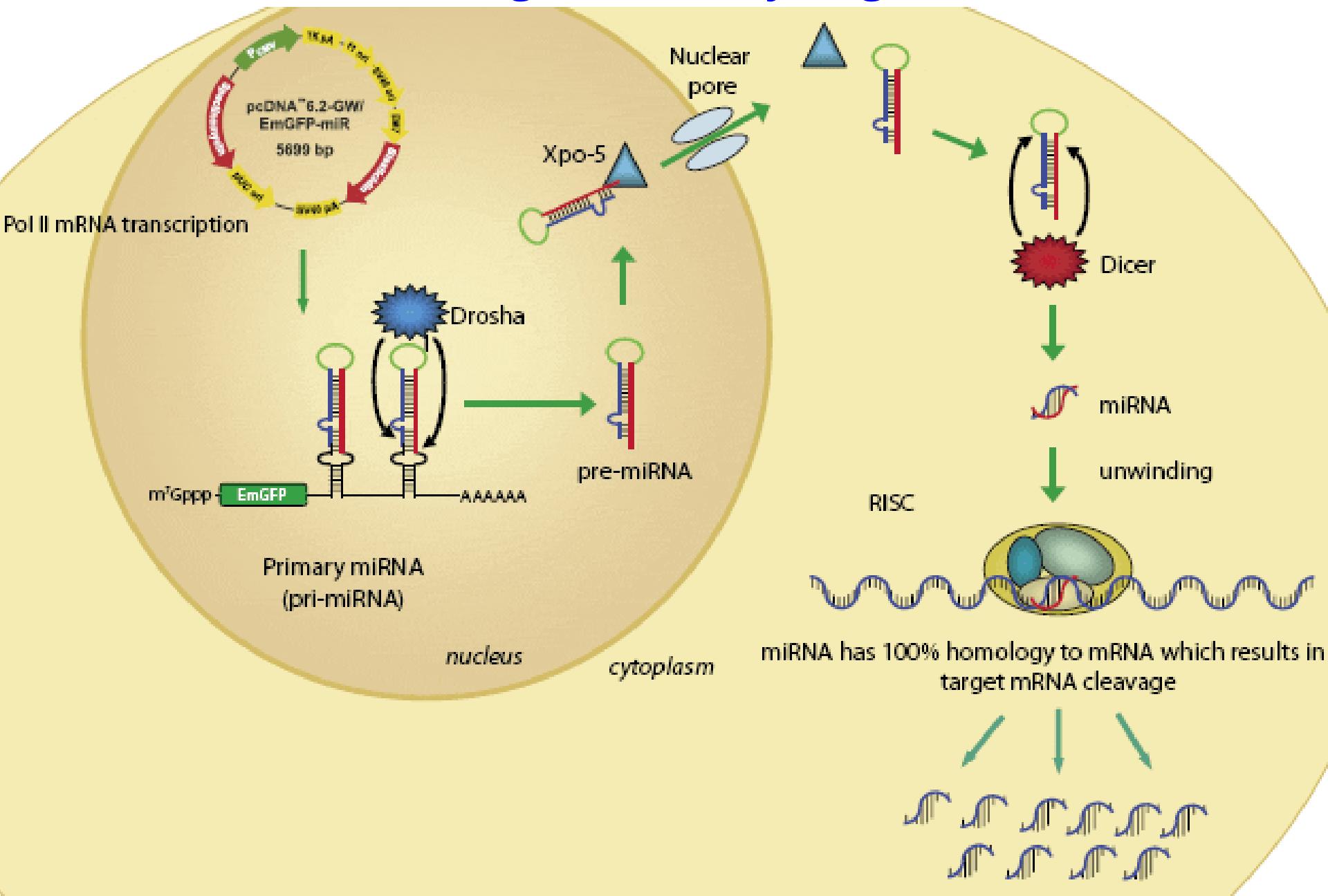


Eukaryotic  
vector for  
micro RNAs: *in*  
*vivo* RNAi



(invitrogen.com)

# Directed mRNA degradation by engineered miRNA



# Gene cloning: Expression of protein and RNA

## Controlled gene expression

### I. proteins

- o Various tags and their functions
- o A protocol for purification of 6xHis tagged proteins expressed in *E. coli*

### II. RNA

# Study and engineering of gene function: mutagenesis

- I. Random mutagenesis, mutant selection schemes
- II. Site-directed mutagenesis, assembly of new DNA fragments
- III. Rational engineering of proteins
- IV. Genetic code expansion

## Readings

- 1) *34 MC4 Mutagenesis*. Very short summary of types of mutagenesis techniques, with lists of specific techniques.
- 2) *35 MC4 Mutagenesis techniques*. Some specific mutagenesis protocols.
- 3) *GFP mut 1994*. A mutant of Green Fluorescent Protein with different spectral properties.
- 4) *Gene shuffling for glyphosate resistance 2004*.
- 5) *Unnatural amino acids 2005*. A review concerning systems encoding new amino acids.

# Mutations change the genetic program – does this influence biomolecular function?

- What is the biological importance of a protein or non-coding RNA – what happens to the organism when the gene is mutated or deleted?
- Which parts of a protein or RNA influence function of that biomolecule?
- How can we intentionally alter or improve protein or RNA function, using random or directed mutation?

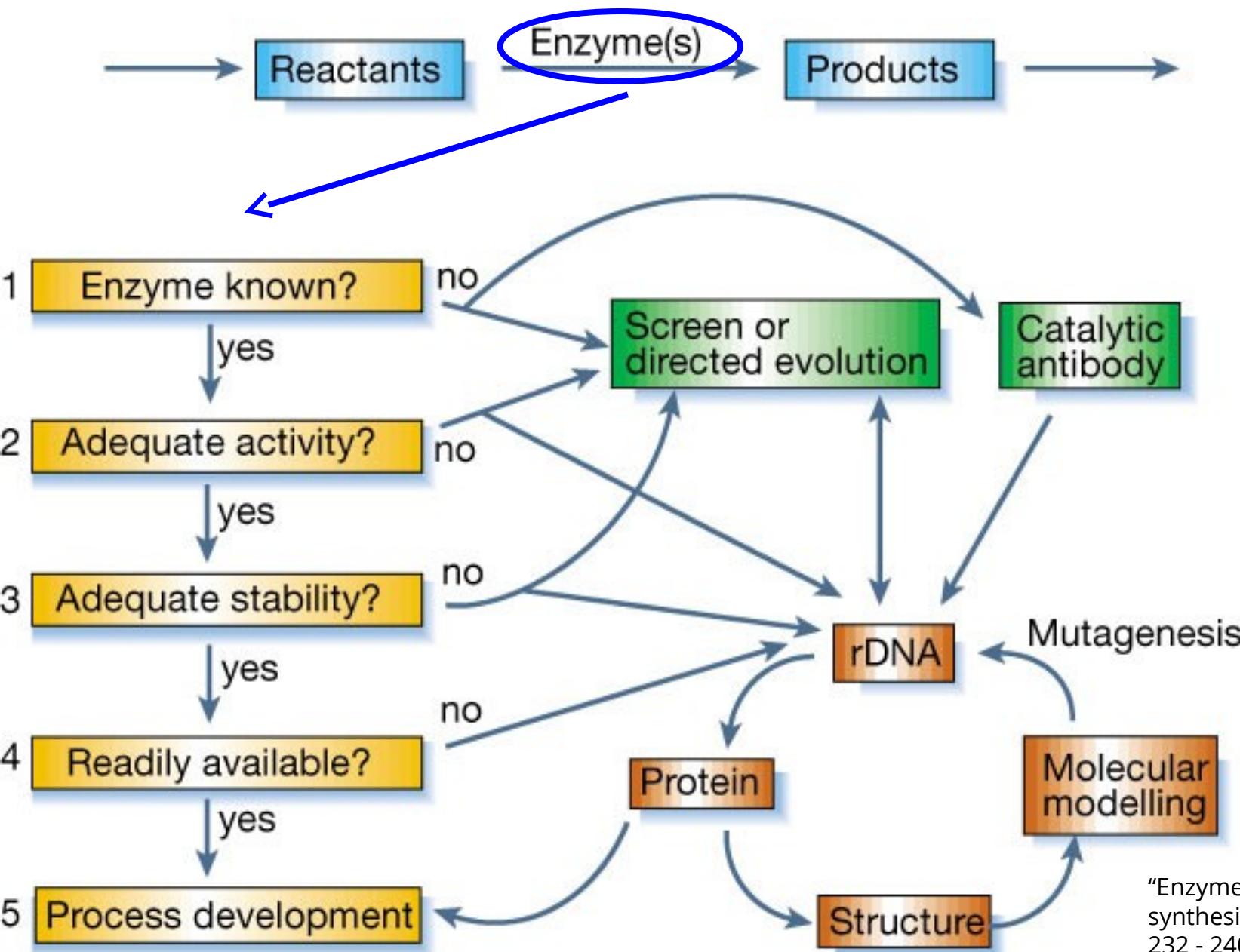
# Many enzymes are commercially important, e.g.:

Enzyme	Industrial use(s)
$\alpha$ -Amylase	Beer making, alcohol production
Aminoacylase	Preparation of L-amino acids
Bromelain	Meat tenderizer, juice clarification
Catalase	Antioxidant in prepared foods
Cellulase	Alcohol and glucose production
Ficin	Meat tenderizer, juice clarification
Glucoamylase	Beer making, alcohol production
Glucose isomerase	Manufacture of high-fructose syrups
Glucose oxidase	Antioxidant in prepared foods
Invertase	Sucrose inversion
Lactase	Whey utilization, lactose hydrolysis
Lipase	Cheese making, preparation of flavorings
Papain	Meat tenderizer, juice clarification
Pectinase	Clarifying fruit juices, alcohol production
Protease	Detergent, alcohol production
Rennet	Cheese making

# Protein engineering through mutation

- Improve stability/function under new conditions
  - temperature, pH, organic/aqueous solvent, salt, etc.
- Change enzyme substrate specificity
- Enhance enzymatic rate
- Change molecular binding properties
- Create brand new enzymes

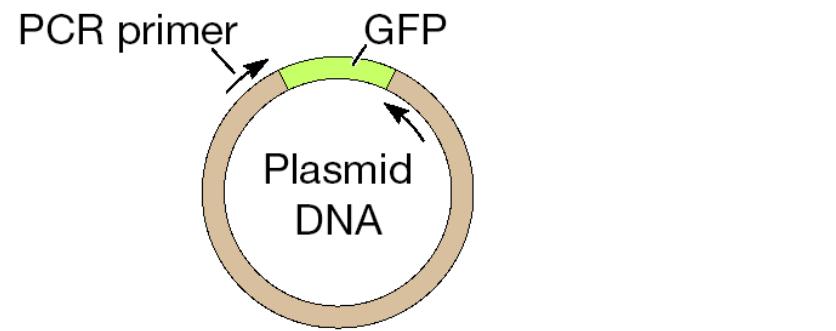
# Obtaining and improving useful enzymes



## Introduce random mutations followed by some kind of screen or selection: simulated evolution

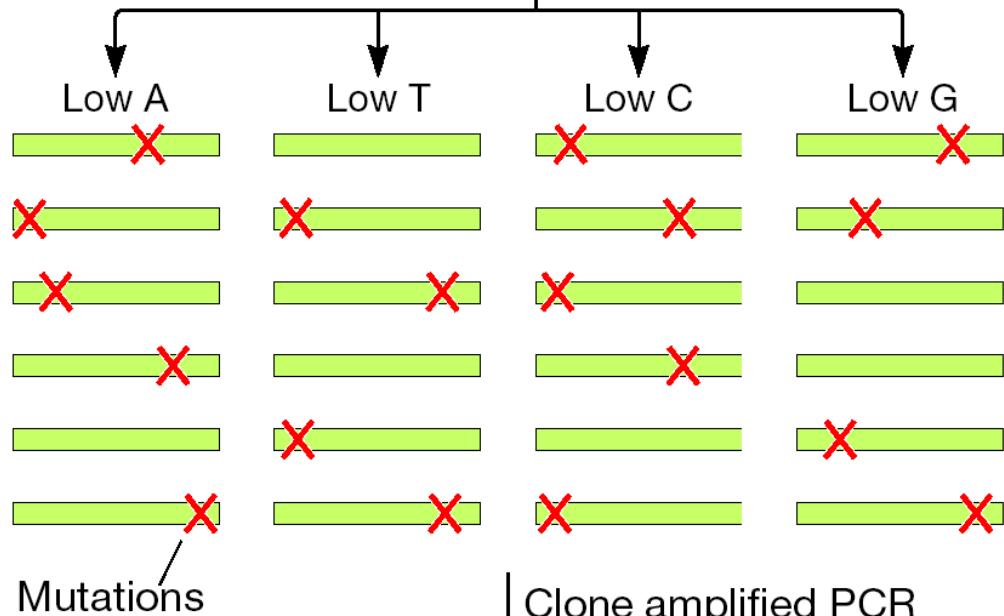
- a) General mutagenesis: expose the organism or its DNA to chemical, UV, or radiation mutagen
- b) PCR mutagenesis: base misincorporation
  - Include Mn<sup>2+</sup> in reaction
  - Reduce concentration of one dNTP
- c) 'Cassette' mutagenesis
  - Partly randomized oligonucleotides used in cloning
- Many possible mutations in a LARGE library of clones
- There has to be some kind of phenotypic screen or selection for the valuable sub-population

# Random mutagenesis by PCR: Green Fluorescent Protein (GFP)



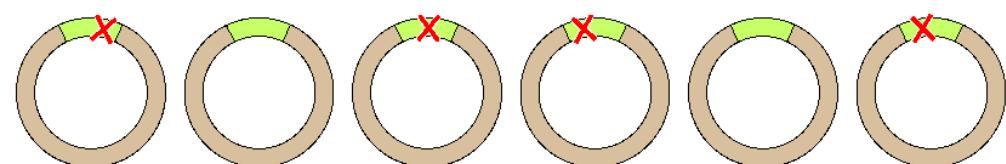
Four PCR reactions, with each nucleotide deficient

Mutations in  
PCR products

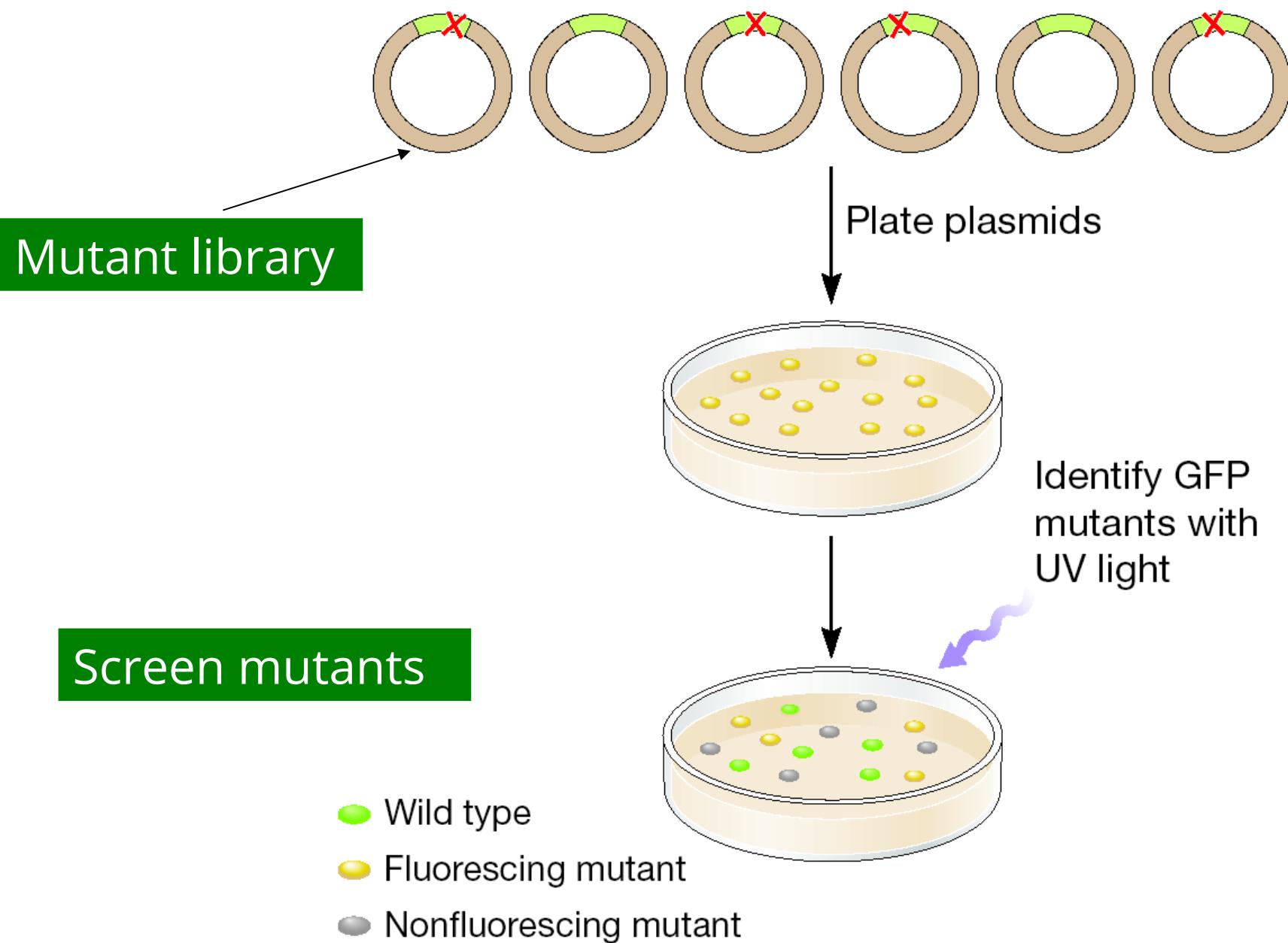


Clone amplified PCR  
products containing  
mutations into plasmids

Mutant library

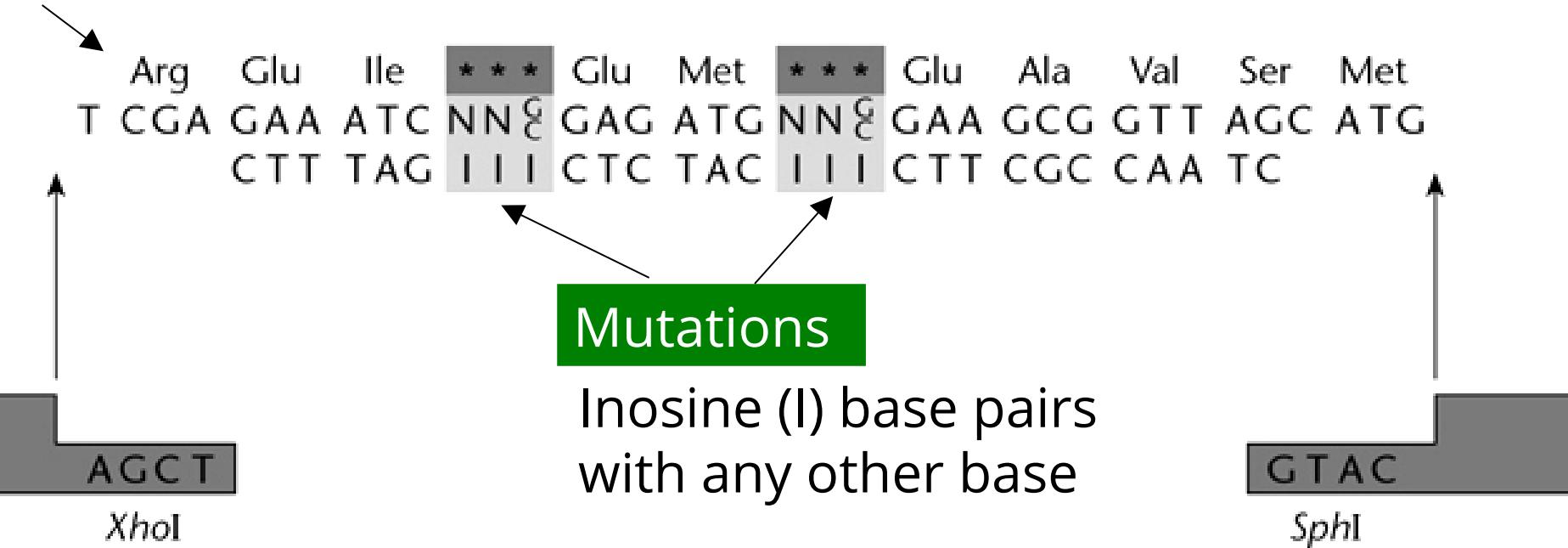


# Random mutagenesis by PCR: the Green Fluorescent Protein



# Cassette mutagenesis (semi-random)

## Translation of sequence



Synthetic cassette DNA strands are synthesized, and annealed

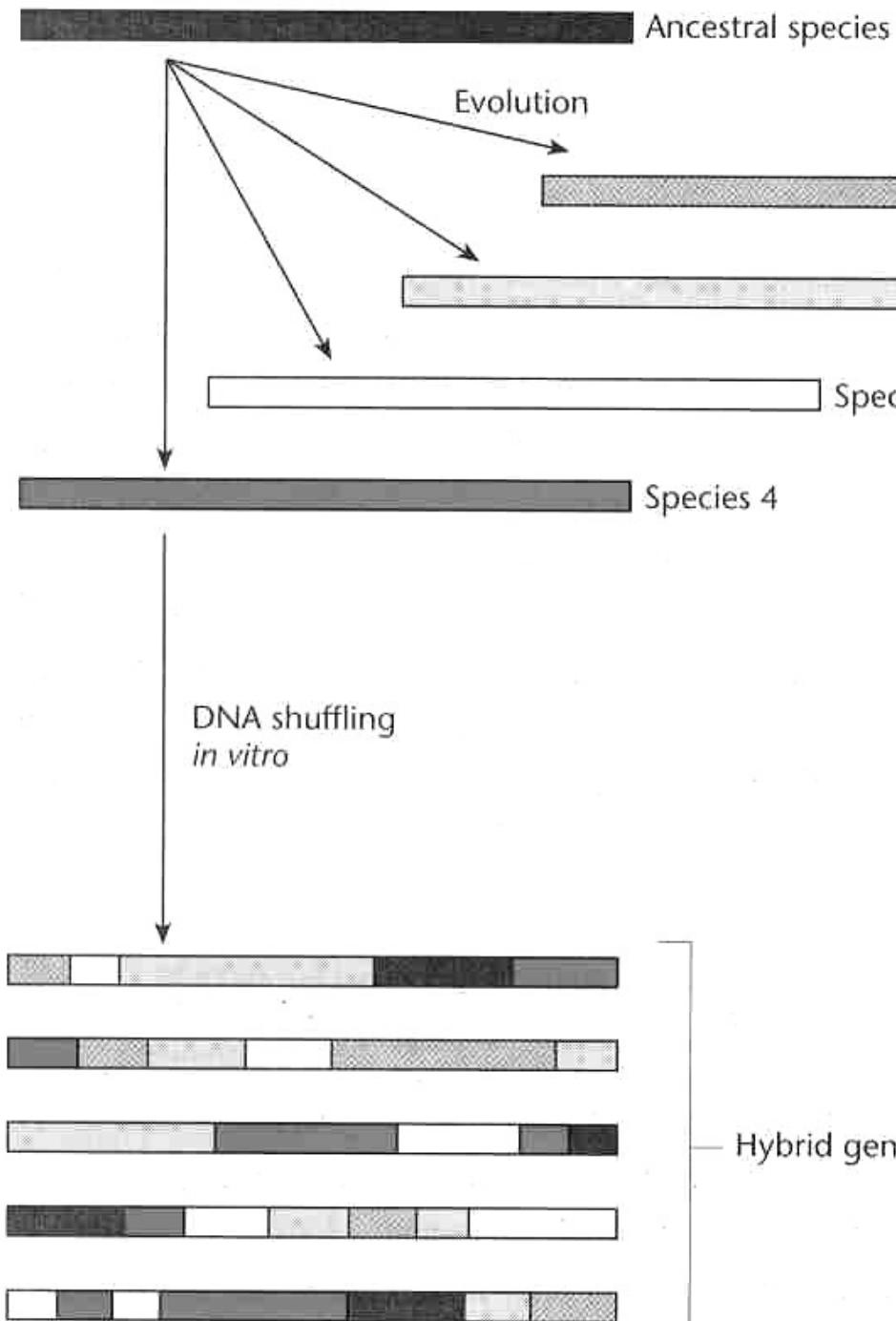
The cassette is ligated to make library, then screened

Allows random insertion of any amino acid at defined positions

# Random and semi-random mutagenesis: directed evolution

- Mutagenize existing protein, eg. error-prone PCR, randomized oligo cassette mutagenesis
  - and/or --
- Do “ gene shuffling”
- Create library
- Screen library of mutations for proteins with altered properties
  - Standard plate screen: 10,000-100,000 mutants
  - (Phage display:  $10^9$  mutants)

# PCR shuffling – like meiotic recombination



For gene shuffling protocols, have related genes in original pool:

- 1) evolutionary variants, or
- 2) variants mutated *in vitro*

Shuffling allows rapid scanning through sequence space:  
faster than doing multiple rounds of random mutagenesis and screening

# GAT: an example for gene shuffling

- GAT gene encodes glyphosate N-acetyl transferase
- Background: glyphosate resistance (in “roundup ready” crops) can be conferred to plants with a microbial EPSPS gene, which is indifferent to glyphosate while the plant version is susceptible.
- EPSPS doesn’t modify glyphosate, allowing the glyphosate to accumulate in the plant, potentially causing problems with crop yields
- GAT acetylates glyphosate, making N-acetyl glyphosate which is not herbicidal
- Several GAT genes were isolated from bacteria by a biotech firm – they provide an alternative to EPSPS
- *The genes weren’t very active, so shuffling was done*

# Three GAT genes from *Bacillus licheniformis*

*	20	*	40	*
401	MIEVKPIN <del>A</del> EDTYEIRHRILRPNQPLEACMYETDLLGGAFHLGGYYRGKL			
B6	MIEVKPIN <del>A</del> EDTYEIRHRILRPNQPLEACKYETDLLGGFHLGGYYRDRL			
DS3	MIEVKPIN <del>A</del> EDTYEIRHRILRPNQPLEACMYETDLLGGFHLGGYYRGKL			
60	*	80	*	100
401	ISIASFH <del>A</del> EHSELEGEEQYQLRG <del>M</del> ATLEGYREQKAGSTLIRHAEELLRK			
B6	ISIASFH <del>O</del> AEHSELEGOKQYQLRG <del>M</del> ATLEGYREQKAGSTLIRHAEELLRK			
DS3	ISIASFH <del>N</del> AEHSELEGOKQYQLRG <del>M</del> ATLEGYREQKAGSTLIRHAEELLRK			
*	120	*	140	
401	KGADLLWCNARTSVSGYYEKLGFSEQGEVYDIPPIGPHILMYKKLT			
B6	KGADLLWCNARTSVSGYYKKLG <del>F</del> SEQGGVYDIPPIGPHILMYKKLT			
DS3	KGADLLWCNARTSVSGYYEKLGFSEQGGVYDIPPIGPHILMYKKLA			

Sites of natural variation are shaded

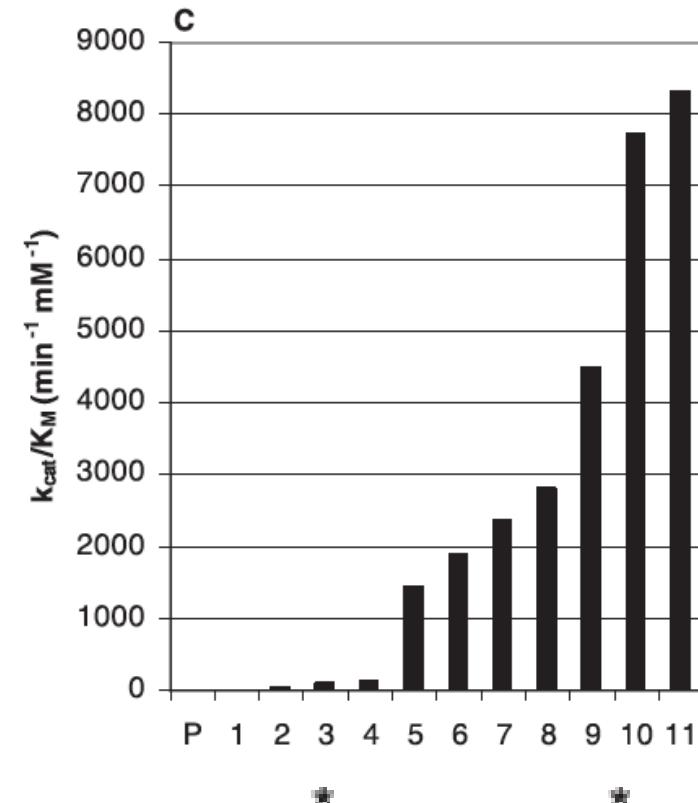
## Discovery and Directed Evolution of a Glyphosate Tolerance Gene

Linda A. Castle,<sup>1\*</sup> Daniel L. Siehl,<sup>1</sup> Rebecca Gorton,<sup>1</sup>  
Phillip A. Patten,<sup>2</sup> Yong Hong Chen,<sup>2</sup> Sean Bertain,<sup>1</sup>  
Hyeon-Je Cho,<sup>1</sup> Nicholas Duck,<sup>3†</sup> James Wong,<sup>3</sup> Donglong Liu,<sup>3</sup>  
Michael W. Lassner<sup>1</sup>

## Mutant creation, screening, and assays

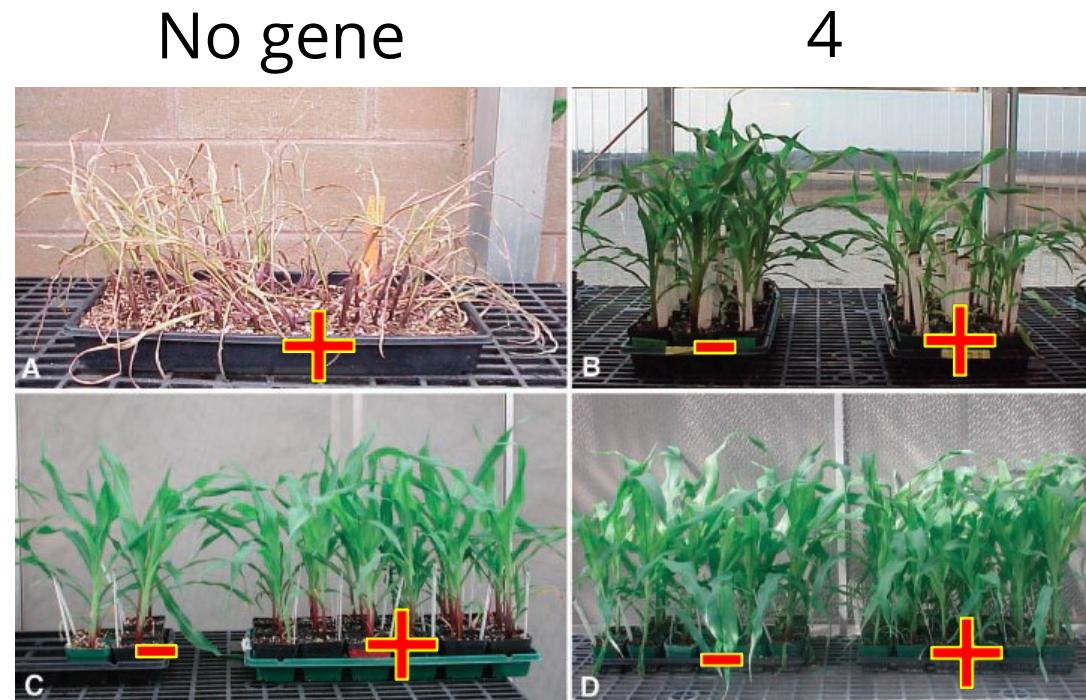
- The three GAT genes were fragmented and shuffled by PCR over 11 rounds
- Additional mutations were added at rounds 4 and 8, based on sequences of related genes in other *Bacillus* species
- After each shuffle, libraries were created
- 5000 gene variants were screened for GAT activity
- Top performing mutants were selected for further shuffling in each round

# Top performer ( $K_{cat}/K_m$ ) for each round



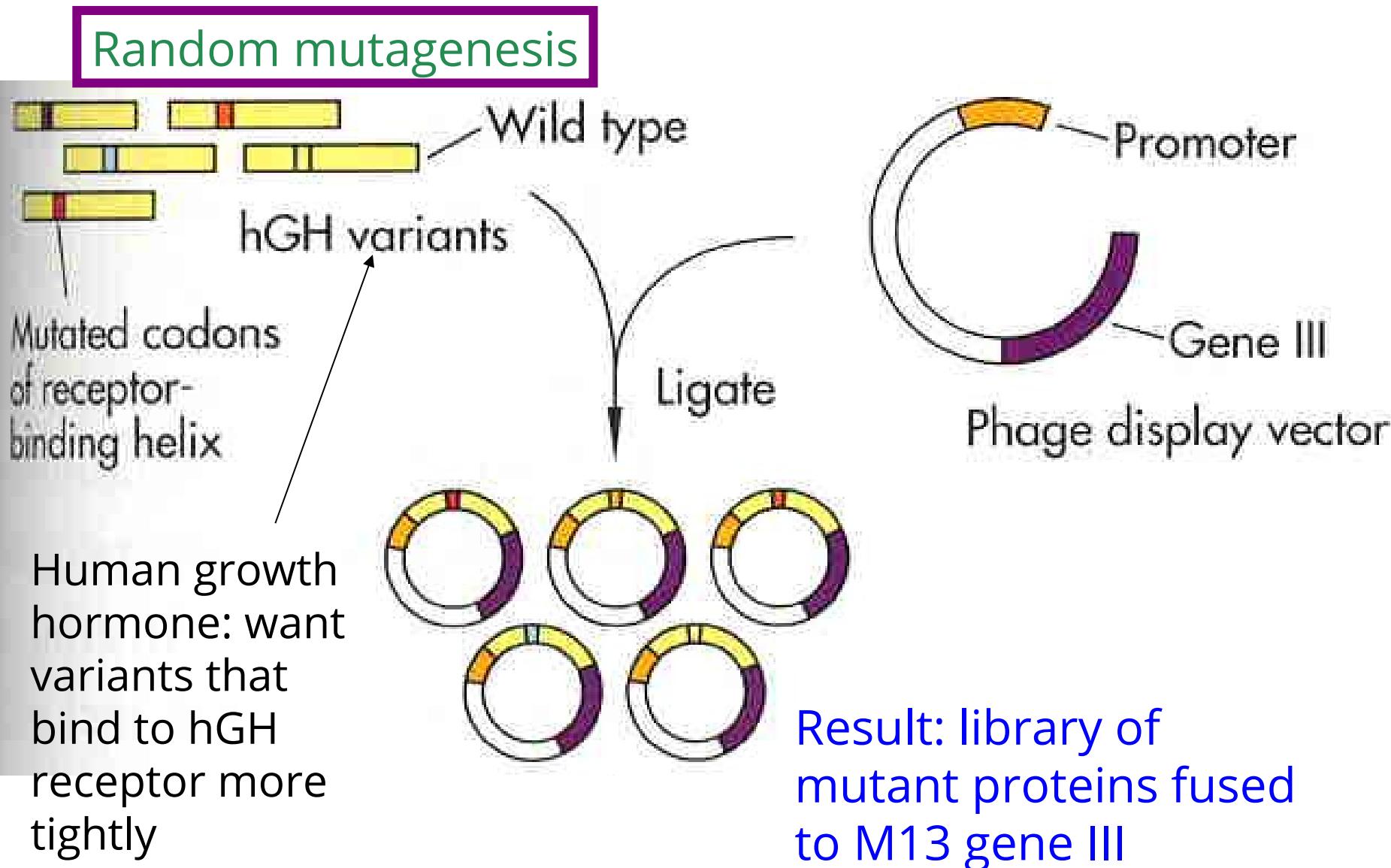
MIEVKPINNAEDTYD**L**RHRV**L**RPNQPIEACMF**E**SDLTRSAFHLGGF**Y**GGKLISVASFHQAE  
 HSEL**O**G**K**K**Q**YQLRG**V**ATLEGY**R**EQKAGSS**L**V**K**HAEE**I**LRKRG**A**DMI**W**CNART**S**ASGYY**R**K  
 LGFSE**Q**GE**V**F**D**T**P**PG**G**PHIL**M**YKRIT

\* 146

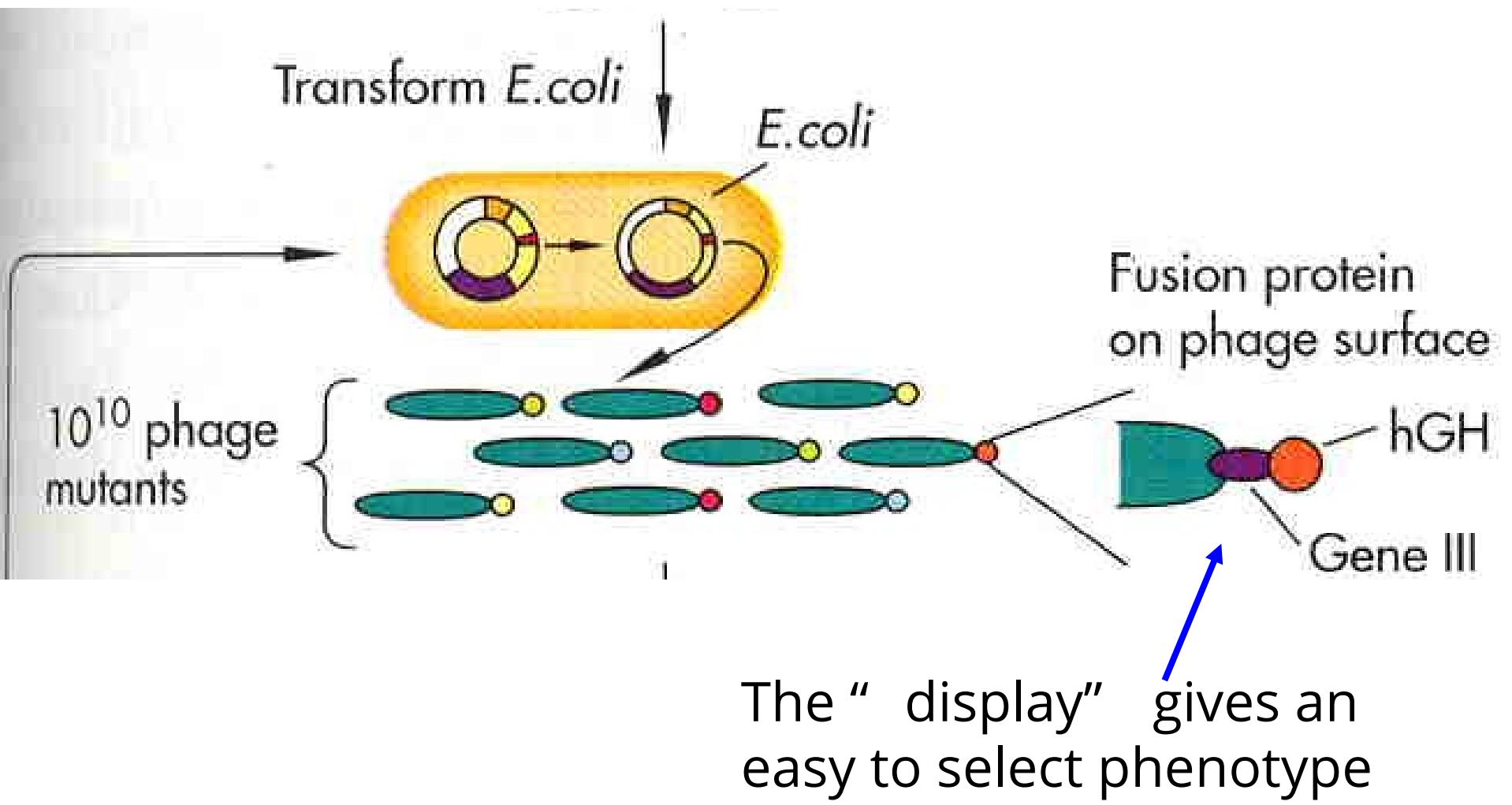


Blue: original diversity, Green: added rd.4, Pink: added rd. 8, Orange: random occurrence

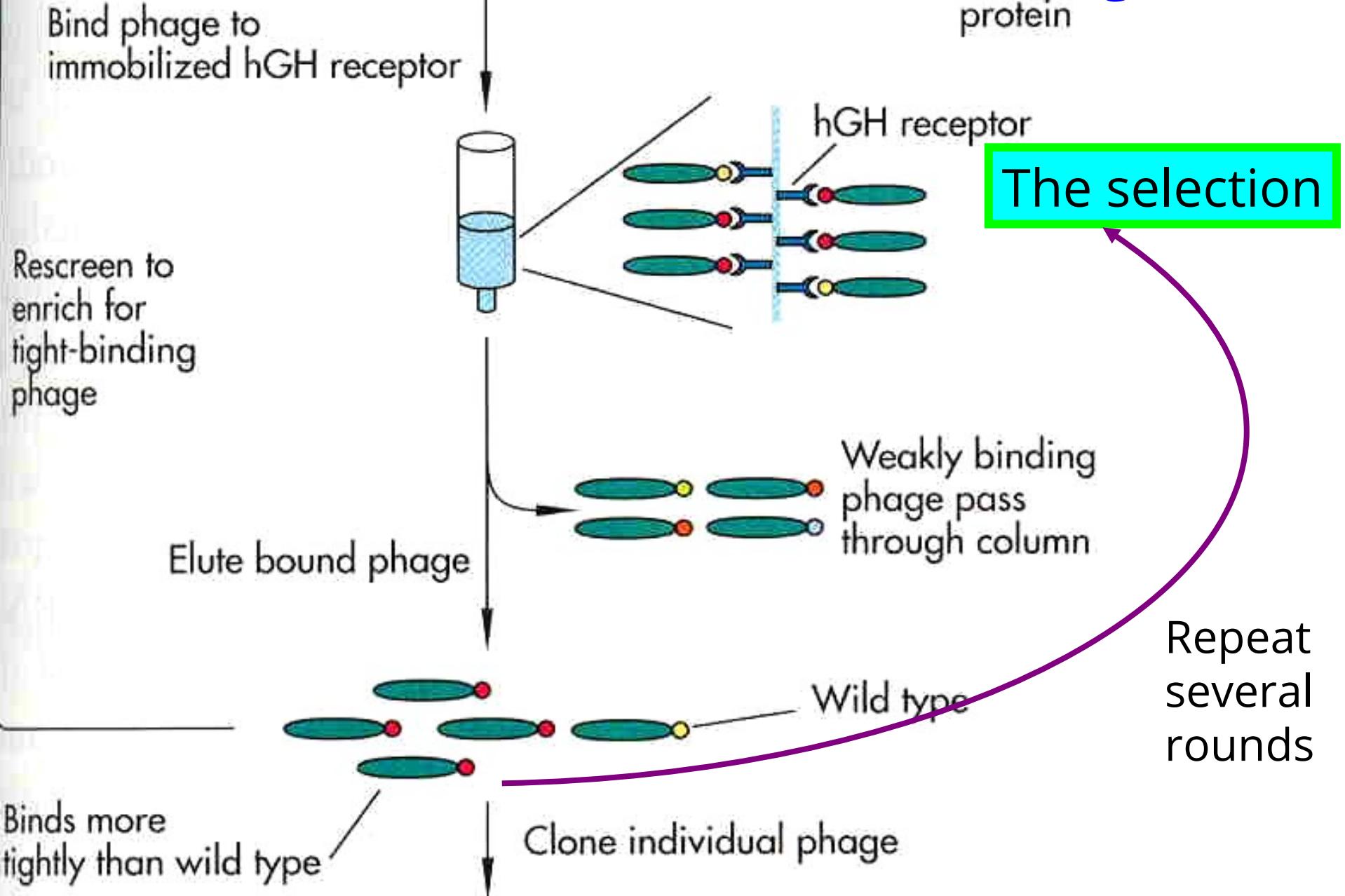
# Screening huge ( $10^9$ ) libraries: phage display



# Phage display: production of recombinant phage



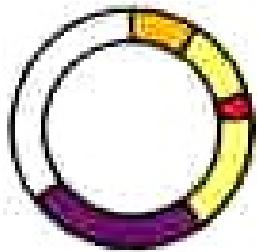
# Phage display: collect tight-binding phage



# Which sequences were selected?



Isolate phage DNA

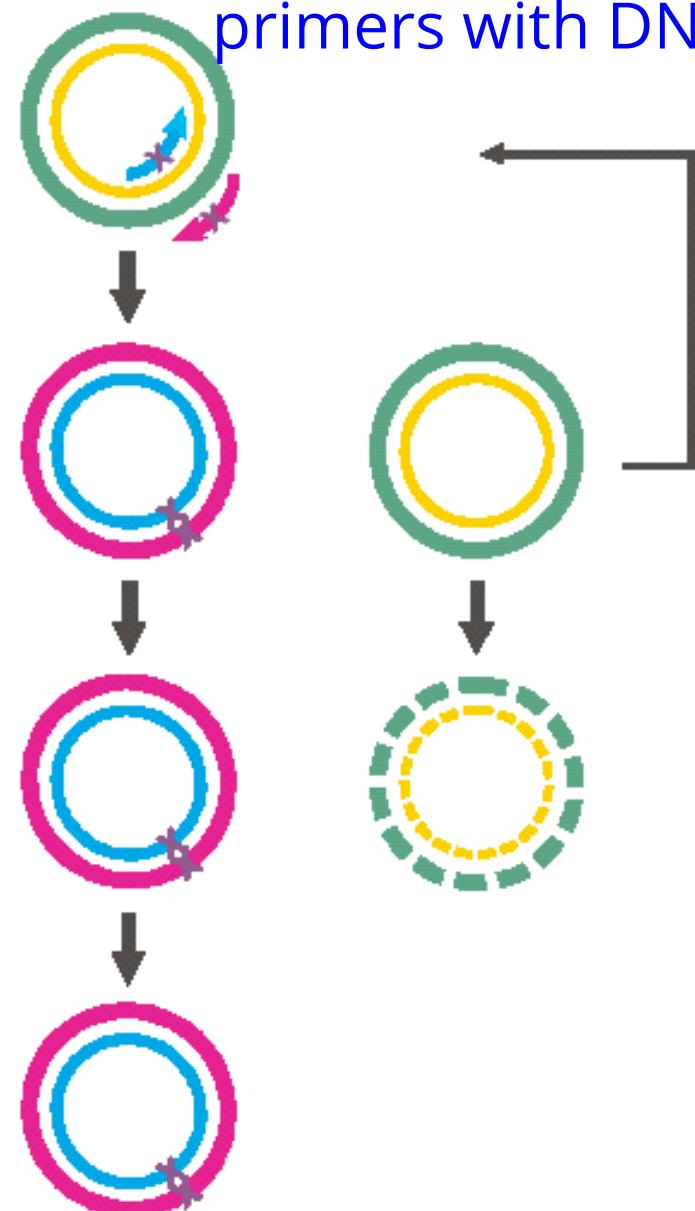


Determine sequence of  
tight-binding mutant

## Site directed mutagenesis: when you know exactly what you want to make

- Changes to amino acids that are likely to be important for function
- New combinations of protein domains
- Rational re-design of protein structure/function
- Addition of new amino acids to the genetic code

# Site-directed mutagenesis: extension of 2 mutagenic primers with DNA polymerase: the PCR/*Dpn* I protocol



## Mutant Strand Synthesis

- Perform thermal cycling to:
- 1) Denature DNA template
  - 2) Anneal mutagenic primers containing desired mutation
  - 3) Extend primers with *PfuUltra* DNA polymerase

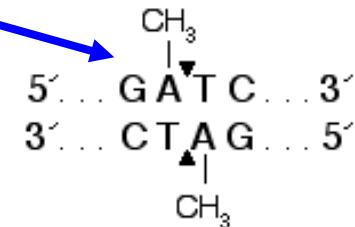
Template DNA is destroyed, only mutated DNA remains

## *Dpn* I Digestion of Template

Digest parental methylated and hemimethylated DNA with *Dpn* I

## Transformation

Transform mutated molecule into competent cells for nick repair



Uses double-stranded plasmid DNA

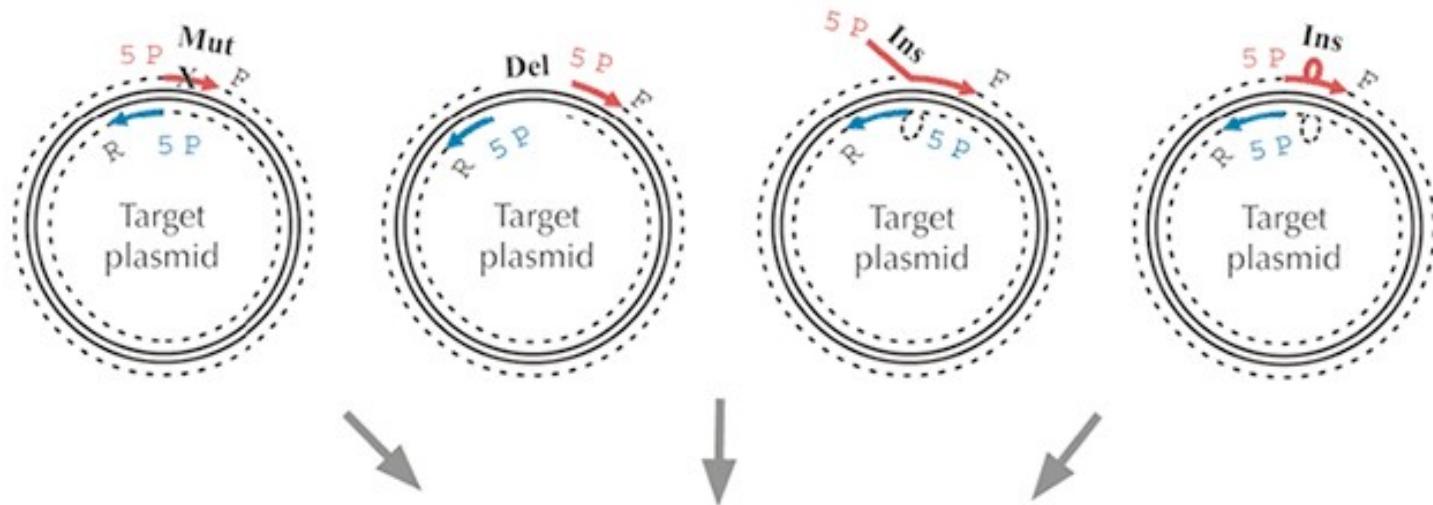
# Site directed mutagenesis: plasmid PCR/ligation/transformation

Point mutation

Deletion

Insertion option 1

Insertion option 2



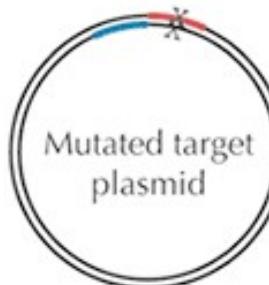
Linear amplified target plasmid  
with desired mutation



**Step 1.**  
Amplification of target  
plasmid with two  
phosphorylated primers.

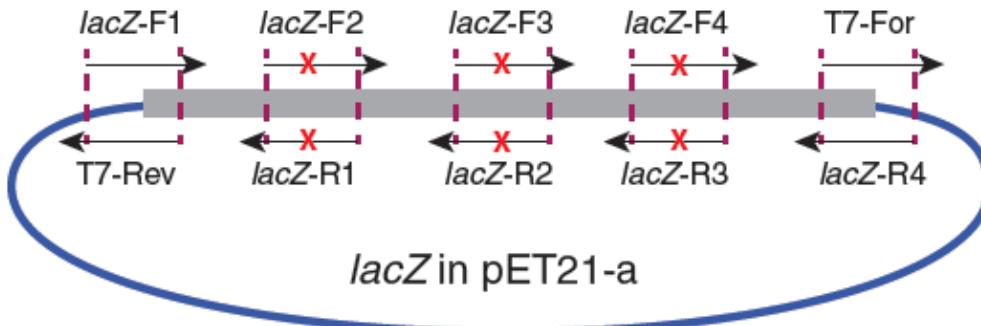
<http://www.thermoscientificbio.com/mutagenesis/phusion-site-directed-mutation-kit/>

**Step 3.**  
Transformation  
into *E. coli*.

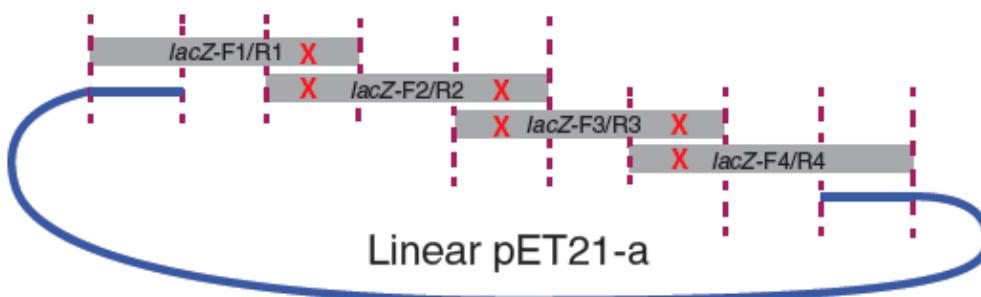


**Step 2.**  
Plasmid circularization  
by ligation.

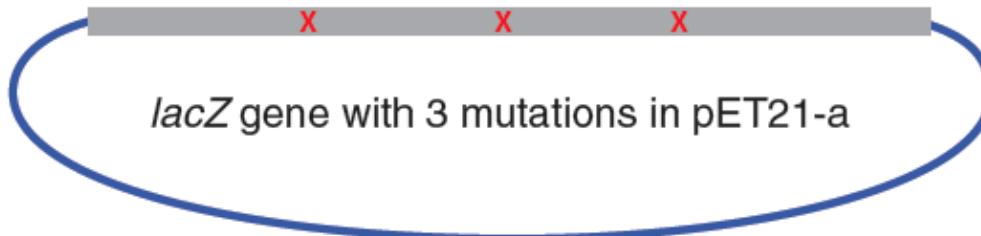
# Site directed mutagenesis: Gibson assembly



PCR to generate fragments with designed mutations for assembly.



Gibson Assembly Master Mix to join fragments at 50°C.



Multiple, overlapping DNA fragments, each with a mutation

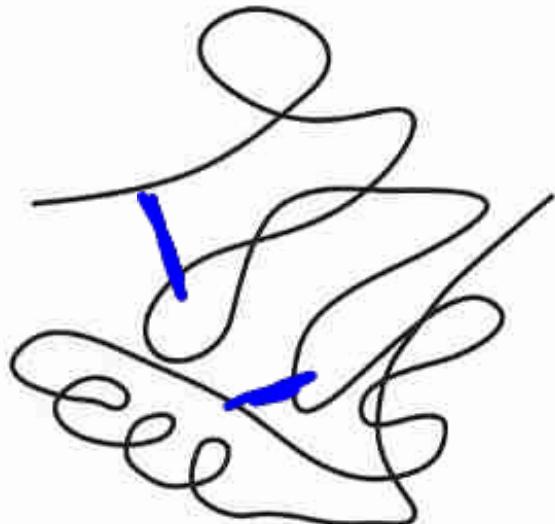
Gibson assembly stitches the fragments together

# Rational site-directed mutagenesis

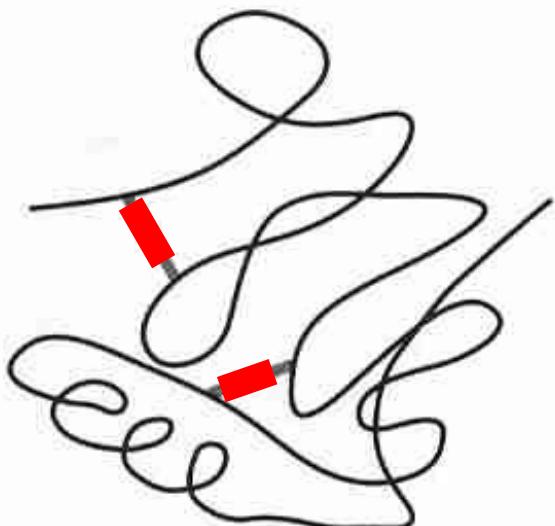
Re-program a protein by making deliberate changes in amino acid sequence, based on structures or other prior knowledge

- If structure is known, mutate amino acids in any part of protein thought to influence activity, interactions, stability, etc.
- For a protein with multiple family members: mutate desired protein in positions that bring it closer to another family member with desired properties

# Site-directed mutagenesis: T4 lysozyme



Native protein



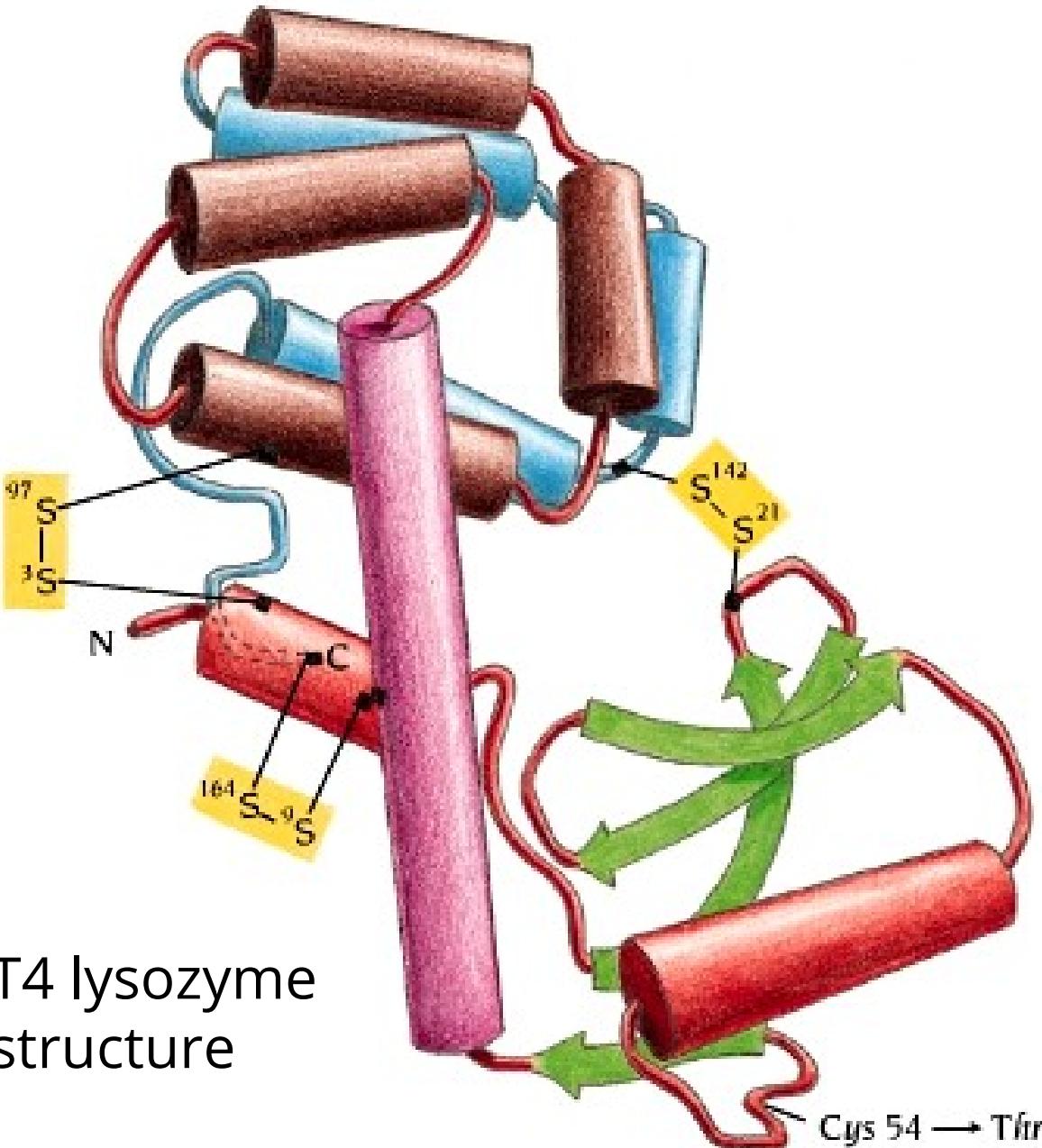
Engineered protein

T4 lysozyme: structure known

Can it be made more stable by the addition of pairs of cysteine residues (allowing disulfide bridges to form)

Does this affect the activity of the protein?

# T4 lysozyme: a model for protein stability



T4 lysozyme  
structure

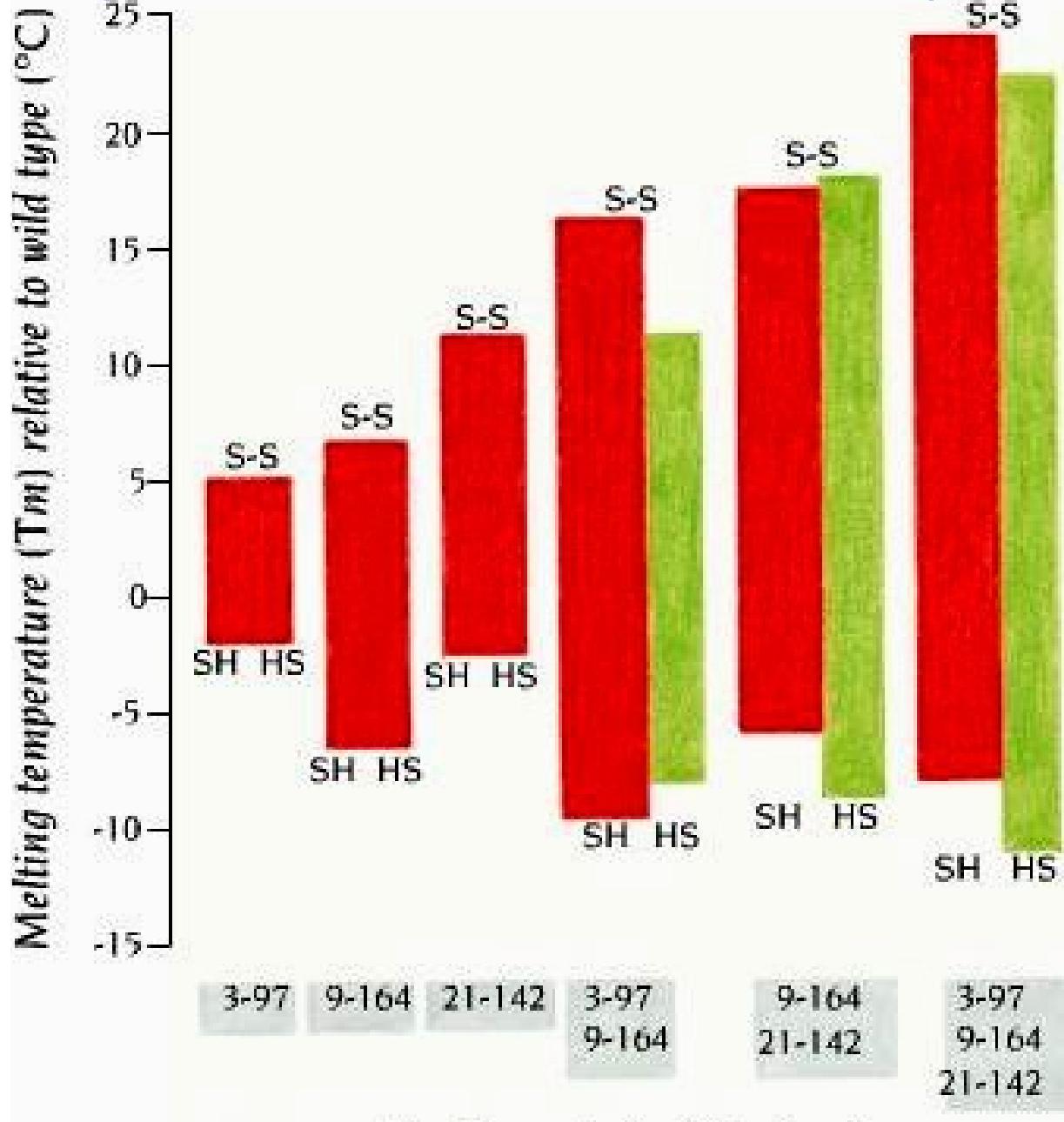
Gene was  
mutagenized to  
add cysteine  
residues in close  
proximity

Can disulfide  
bridges be  
deliberately  
added?

Does this enhance  
protein stability?

Does this affect  
enzyme activity?

# Engineered disulfides stabilize the protein to high temps



Bottom of bar:  
melting temp in  
reducing conditions

Top of bar:  
Melting temp in  
oxidizing conditions

Green bars: if the  
individual mutation  
effects were added  
together

# Phenotypic trade-off: Increased stability can reduce enzyme activity

Enzyme	Amino acid at position:							No. of -S-S-	% Activity	$T_m$ (°C)
	3	9	21	54	97	142	164			
wt	Ile	Ile	Thr	Cys	Cys	Thr	Leu	0	100	41.9
pwt	Ile	Ile	Thr	Thr	Ala	Thr	Leu	0	100	41.9
A	Cys	Ile	Thr	Thr	Cys	Thr	Leu	1	96	46.7
B	Ile	Cys	Thr	Thr	Ala	Thr	Cys	1	106	48.3
C	Ile	Ile	Cys	Thr	Ala	Cys	Leu	1	0	52.9
D	Cys	Cys	Thr	Thr	Cys	Thr	Cys	2	95	57.6
E	Ile	Cys	Cys	Thr	Ala	Cys	Cys	2	0	58.9
F	Cys	Cys	Cys	Thr	Cys	Cys	Cys	3	0	65.5

Adapted from Matsumura et al., *Nature* 342:291–293, 1989.

wt, wild-type T4 lysozyme; pwt, pseudo-wild-type enzyme; A through F, six engineered cysteine variants; -S-S-, disulfide bonds;  $T_m$ , “melting” temperature (a measure of thermostability).

**F:** 3 S-S bonds, highest stability, but no activity

**D:** 2 S-S bonds, increased stability and nearly 100% active

# New amino acids in proteins?

- New amino acid = new functional group
- Alter or enhance protein function (rational design)
- Chemically modify protein following synthesis (chemical derivitization)
  - Probe protein structure, function
  - Modify protein *in vivo*, add labels and monitor protein localization, movement, dynamics in living cells

# Mutation by altering the genetic code

- 61 sense codons, 3 non-sense (stop) codons
- 20 amino acids (plus selenocysteine & pyrrolysine)
- Other amino acids exist, some in the cell (as precursors to the 20 amino acids), but have not been added to the genetic code in a living system (as far as we know)
- Can other amino acids be added to biological systems?

Xie and Schultz (2005) “ A chemical toolkit for proteins -- an expanded genetic code” *Nat Rev Mol Cell Biol* 7, p. 775.

# How to modify genetic code?

Adding new amino acids to the code--must bypass the fidelity mechanisms that have evolved to prevent this from occurring

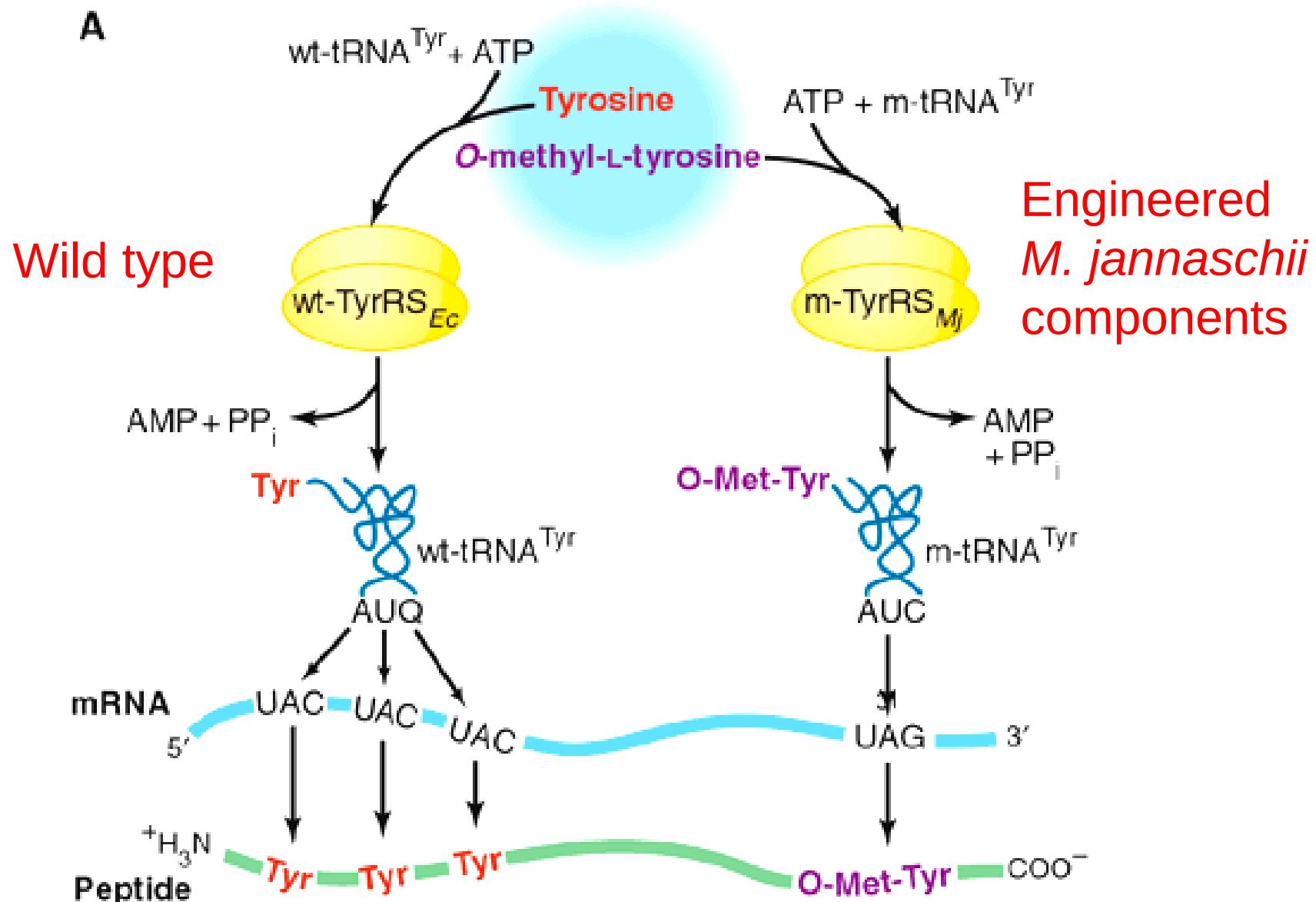
## 2 key mechanisms of fidelity

- Correct amino acid inserted by ribosome through interactions between tRNA anti-codon and mRNA codon of the mRNA in the ribosome
- Specific tRNA charged with correct amino acid because of high specificity of tRNA synthetase interaction
- Add **new tRNA**, add **new tRNA synthetase** to make a new amino acid available

# Altering the genetic code

	Second letter					
	U	C	A	G		
First letter	UUU } Phe UUC } UUA } Leu UUG }	UCU } Ser UCC } UCA }	UAU } Tyr UAC } UAA } Stop UAG }	UGU } Cys UGC } UGA }	UGG } Trp UGA }	U C A G
C	CUU } CUC } CUA } Leu CUG }	CCU } CCC } CCA }	CAU } His CAC } CAA }	CGU } CGC } CGA }	CGG }	U C A G
A	AUU } AUC } Ile AUA }	ACU } ACC }	AAU } Asn AAC }	AGU } AGC }	AGA }	U C A G
G	GUU } GUC } GUA }	GCU } GCC }	GAU } Asp GAC }	GGU } GGC }	GGG }	U C A G
	Val GUG }	Ala GCG }	GAA } Glu GAG }	Gly GGA }		

# Site-specific insertion of unnatural amino acids



Wang et al. (2001) *Science* **292**, p. 498.  
Böck (2001) *Science* **292**, p. 453.

What can you do with a tRNA that recognizes stop codons, and that gets charged with a new amino acid?

You can program an mRNA that directs specific insertion of new amino acid

- Design protein to have UAG stop codon where you'd like the new amino acid to go
- Transform engineered *E. coli* with plasmid containing the engineered gene
- Feed cells the unnatural amino acid to get synthesis of full length gene

# Some questions:

- What are the consequences for the cell with an expanded code?
- Do new amino acids confer any kind of evolutionary advantage to organisms that have them? (assuming they get a ready supply of the new amino acid...)
- Why do cells have/need 3 stop codons????

# A system for site-specific insertion of new (unnatural) amino acids into proteins: summary

Three components were engineered:

- 1) new “ codon” (use amber, the rarest stop codon: UAG)
- 2) tRNA recognizing UAG codon, and charged with unnatural amino acid
- 3) aminoacyl tRNA synthetase (aaRS) to add new amino acid to the UAG tRNA

Unique proteins can be created

## What's next: Quadruplet-encoding ribosomes

Mutant ribosomes have been isolated that can decode a quadruplet base sequence (as opposed to the standard triplet)

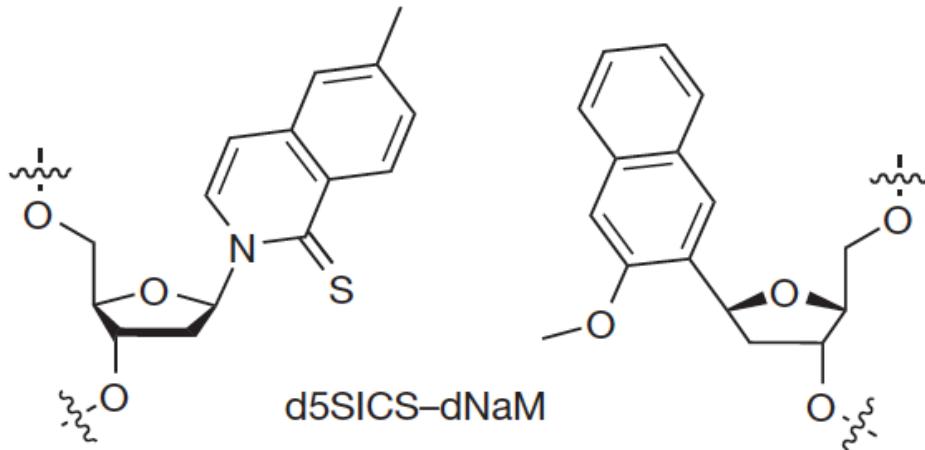
Quadruplet tRNAs charged with unnatural amino acids cause insertion of the unnatural amino acid at the quadruplet

These "ribo-Q1" ribosomes are specific to only specific (orthogonal) mRNA, and won't translate "normal" mRNAs

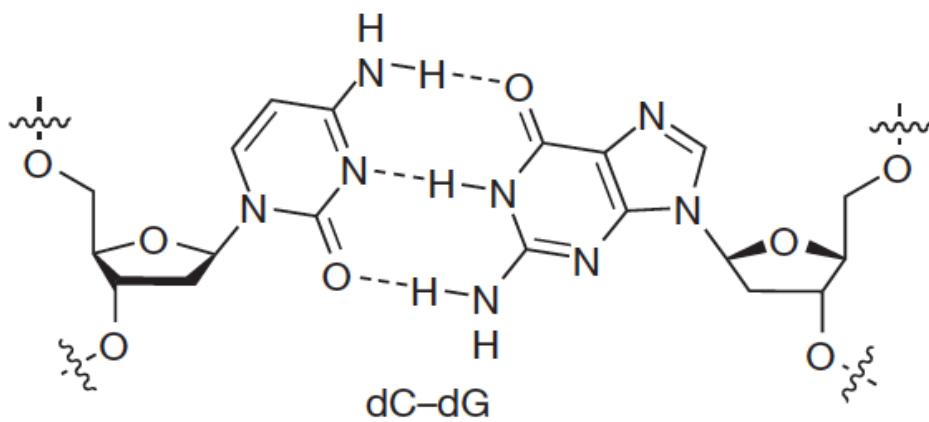
256 blank codons available (as opposed to 3 stop codons), so it could be possible to encode many different unnatural amino acids into proteins

"...foundational technologies for the encoded synthesis and synthetic evolution of unnatural polymers in cells."

# What's next: new base pairs in DNA



Hydrophobic nucleobase:  
d5SICS-dNaM



Number of codons  
with four bases:

$$4^3 = 64$$

Number of codons  
with six bases:

$$6^3 = 216$$

A semi-synthetic organism with an expanded genetic alphabet

Denis A. Malyshev<sup>1</sup>, Kirandeep Dhami<sup>1</sup>, Thomas Lavergne<sup>1</sup>, Tingjian Chen<sup>1</sup>, Nan Dai<sup>2</sup>, Jeremy M. Foster<sup>2</sup>, Ivan R. Corrêa Jr<sup>2</sup>  
& Floyd E. Romesberg<sup>1</sup>

(see also  
hachimoji DNA)

# **Study and engineering of gene function: mutagenesis**

- I. Random mutagenesis, mutant selection schemes
- II. Site-directed mutagenesis, assembly of new DNA fragments
- III. Rational engineering of proteins
- IV. Alterations in the genetic code

# Applied mutagenesis: pathway engineering and synthetic biology

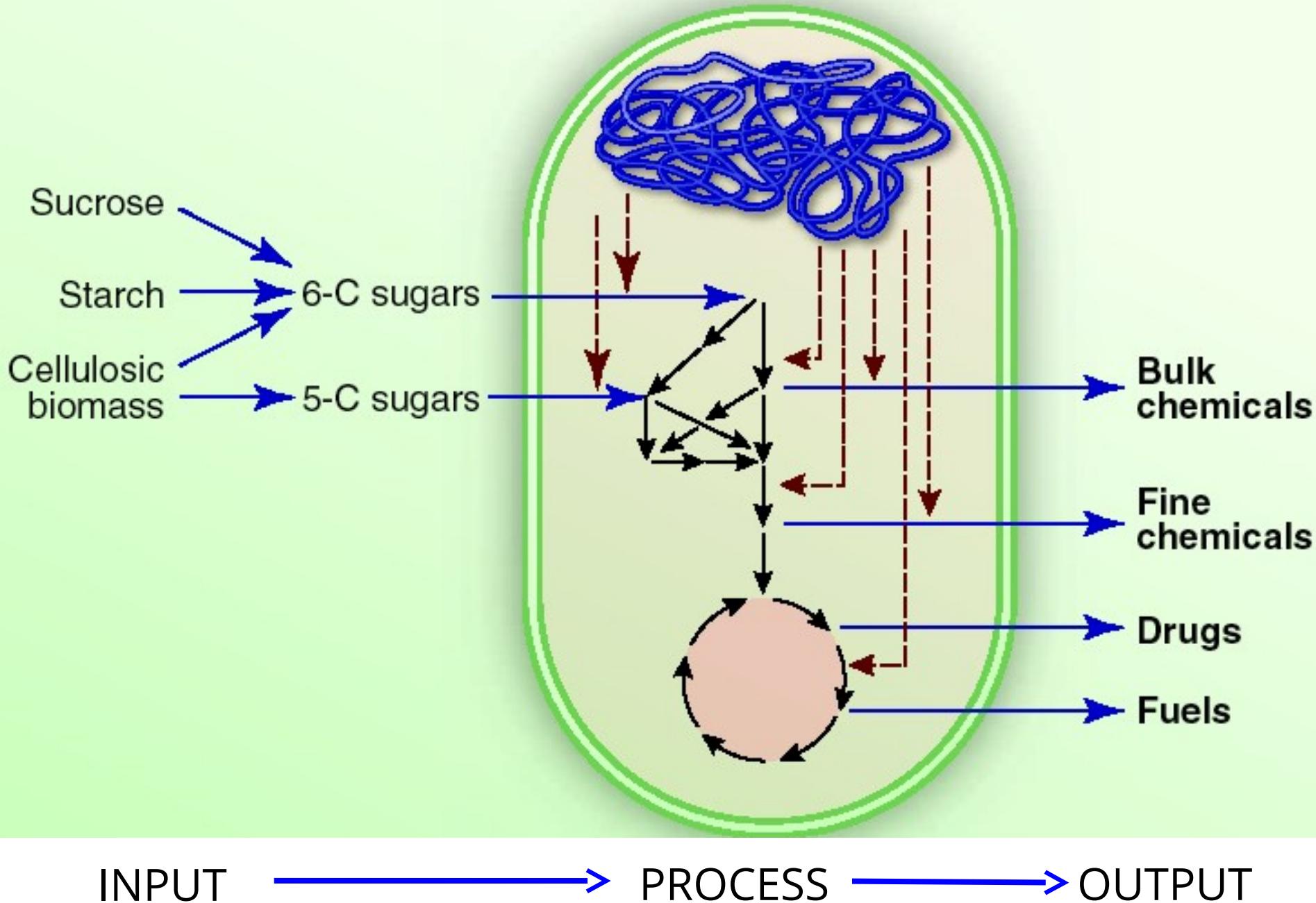
Increase biological production of useful molecules

- Random screening for overproducing strains (genome shuffling)
- Rational engineering of pathways and organisms

## Guide to readings

- 1) Metabolic engineering short review 2010.
- 2) Original genome shuffle. A paper reporting a recombination approach for improvement of antibiotic production
- 3) MAGE 2009. A paper describing a method to rapidly introduce genetic variation to targeted regions of the genome
- 4) Synthetic artemisinin. A paper describing metabolic engineering of *S. cerevisiae* to produce the antimalarial compound artemisinin.
- 5) Implications of synthetic genomes (2010). Perspective on the first completely synthesized bacterial genome.
- 6) Biocontainment of genetically modified bacteria (2015)

# The cell as a bioreactor



# Biologically derived molecules for sale:

- Antibiotics
- Vitamins
- Metabolic by-products (ethanol, lactic acid)
- Amino acids and derivatives (indigo, aspartame)
- “secondary metabolites” from plants – e.g. alkaloids (caffeine, theobromine, etc.)
- Hydrocarbons for fuel
- Synthesis often requires multiple steps and enzymes, making these molecules difficult to synthesize chemically

# Increased production of antibiotics: Classical Strain Improvement

- 1) Obtain organism that produces the compound of interest  
-- the original strain of *Penicillium* mold made penicillin at micrograms per liter of culture
- 2) Random mutagenesis followed by screen for increased production.
- 3) With top producer, repeat mutagenesis and screen
- 4) Outcome: milligrams of penicillin per liter of culture (1000-fold increase in production)
  - Time consuming and expensive process!

# Genome shuffling: an alternative to Classical Strain Improvement

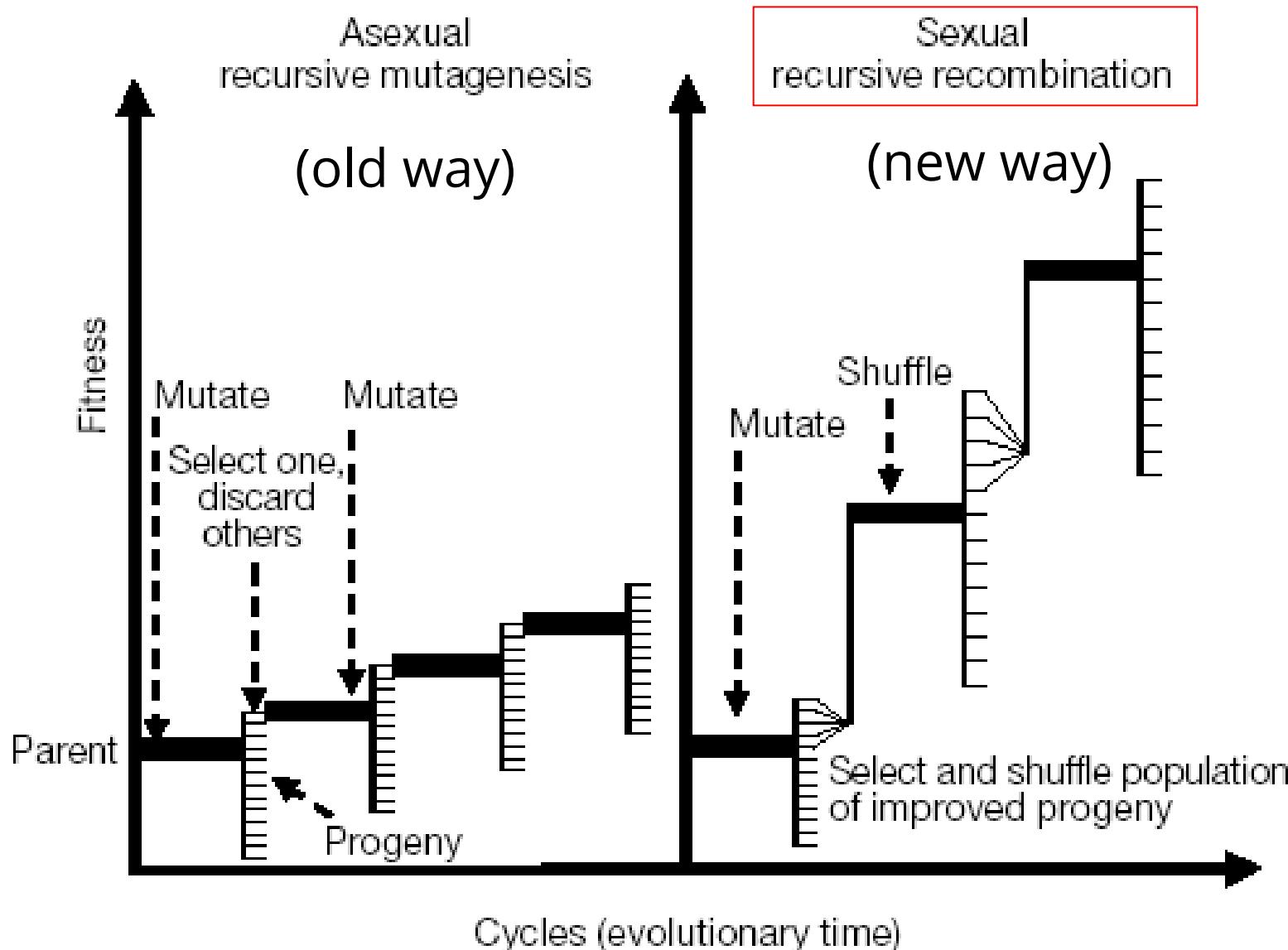
## Genome shuffling leads to rapid phenotypic improvement in bacteria

**Ying-Xin Zhang\*, Kim Perry\*, Victor A. Vinci†, Keith Powell\*,  
Willem P. C. Stemmer\* & Stephen B. del Cardayré\***

\* Maxygen, 515 Galveston Drive, Redwood City, California 94063, USA

† Eli Lilly and Company, Lilly Corporate Center, Indianapolis, Indiana 46285,  
USA

# The shuffling advantage: simultaneous recombination of entire genomes (breeding) with multiple parents



# Testing recursive shuffling

- Compare classical strain improvement (CSI) to genome shuffling
- *Streptomyces sp.*: produce polyketide antibiotics
- Induce recombination by recursive protoplast fusion:
  - Fuse protoplasts
  - Regenerate cell walls, grow as a population (F1)
  - Make protoplasts with F1, repeat until F4
- Test with 4 auxotrophy markers (next page)
- Test for increased antibiotic production

# Test of recursive shuffling

4 parental strains

Supplements required:

Strain	Genotype
<i>S. coelicolor</i> 2684	<i>proA1 argA1 cys<sup>wt</sup> uraA1</i>
<i>S. coelicolor</i> 2685	<i>proA1 arg<sup>wt</sup> cysD18 uraA1</i>
<i>S. coelicolor</i> 2686	<i>pro<sup>wt</sup> argA1 cysD18 uraA1</i>
<i>S. coelicolor</i> M124	<i>proA1 argA1 cysD18 ura<sup>wt</sup></i>

pro, arg, ura (not *cys*)

pro, cys, ura (not *arg*)

*arg*, cys, ura (not *pro*)

pro, arg, cys (not *ura*)

“ Shuffle” (recombine) all 4 strains

Can *progeny* be isolated that can *grow without pro, arg, ura, and cys supplementation* (indicating progeny with all 4 genes wild type)?

Shuffling: increased efficiency of recombination



**Table 1 Distribution of phenotypes in a four-strain cross of *S. coelicolor***

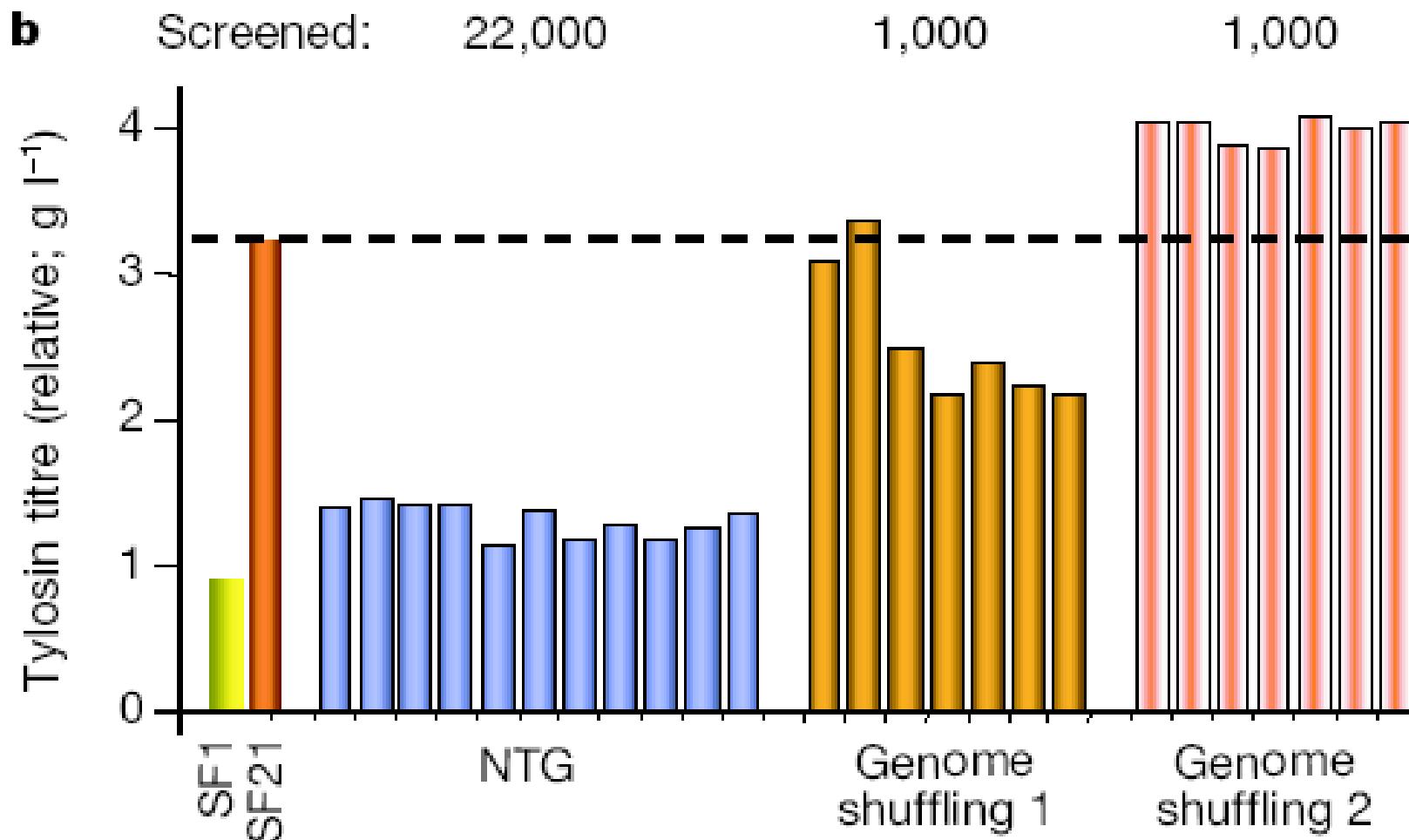
Phenotype	Single fusion*	Recursive fusion†
Two markers	8.4%	60%
Three markers	0.73%	17%
Four markers	0.000045%	2.5%

The distribution of phenotypes from each fusion is reported in Supplementary Information.

\* Phenotypes were determined from colony counts on defined medium containing 16 combinations of the four supplements (See Methods). Each phenotype is corrected for dilution and the presence of prototrophic markers, and divided by the total colonies growing on completely supplemented medium. The value shown represents the sum of the frequencies from each phenotypic class.

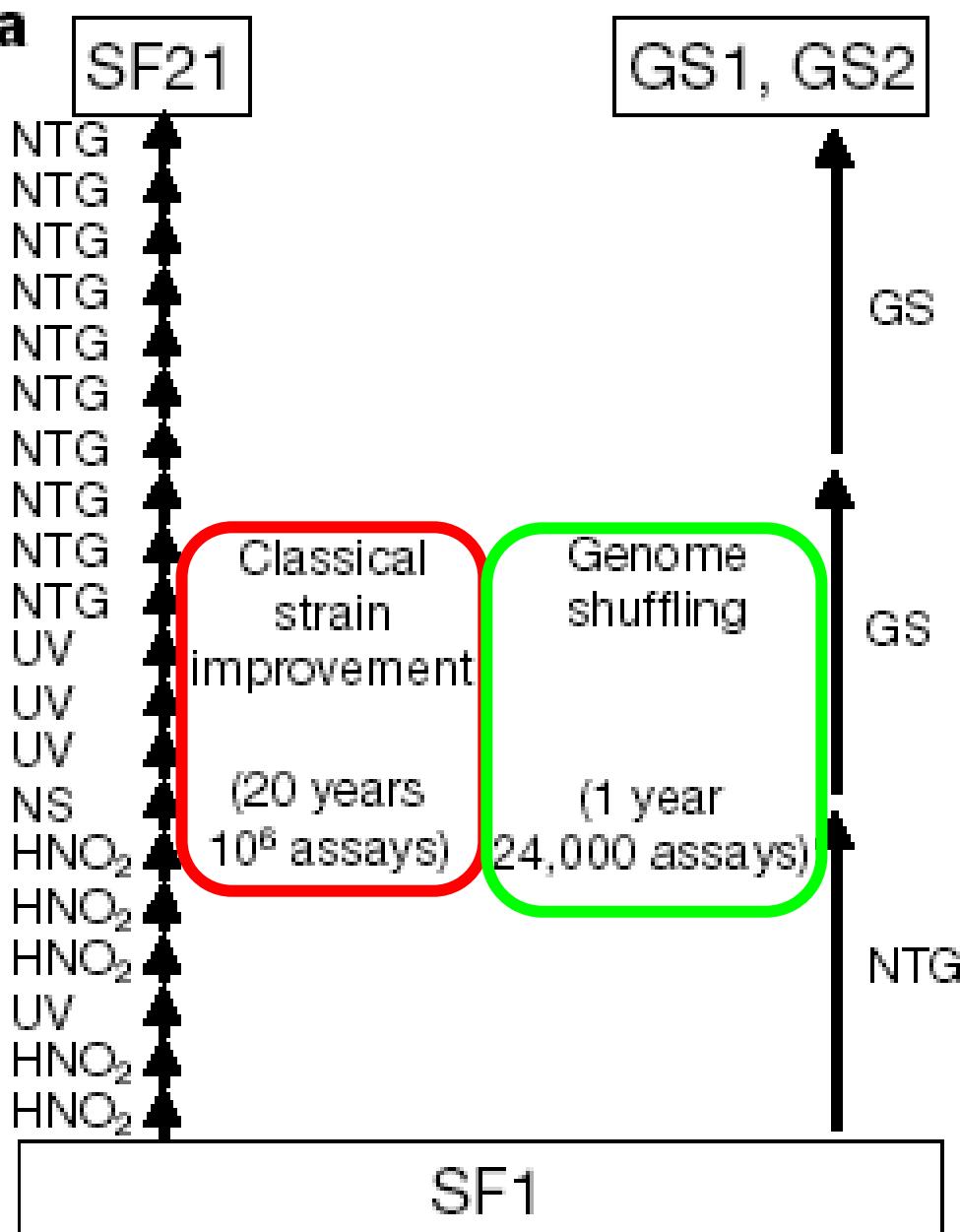
† The distribution of 483 individual colonies characterized for marker phenotype.

# Test case: increase tylosin production by *S. fradiae*



SF1 was treated with NTG, 11 strains selected (22,000 screened), those 11 strains were shuffled once (GS1) and then again (GS2)

# Comparison: CSI versus genome shuffling

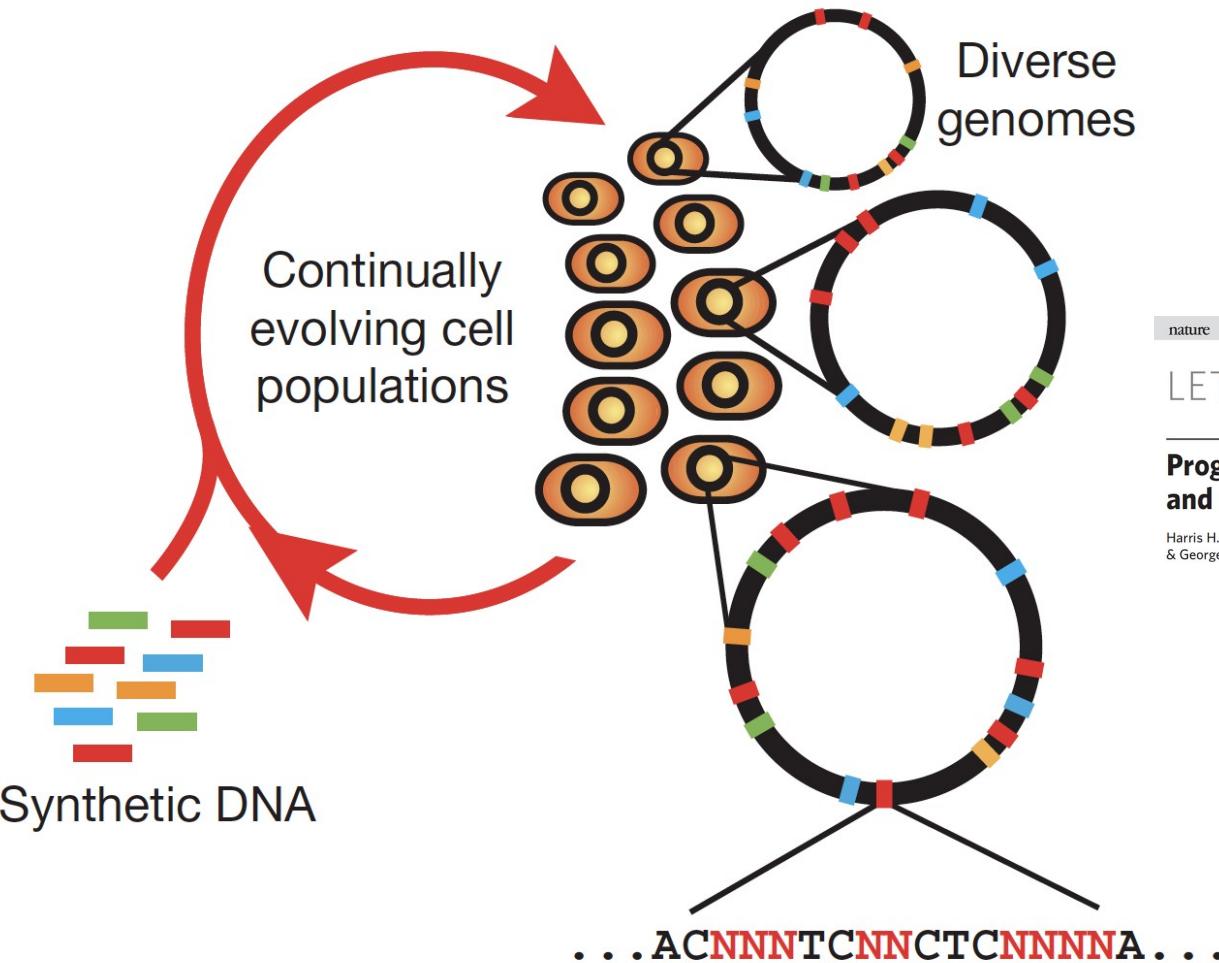


**c**

Strain	Titre (rel. g l <sup>-1</sup> )
SF1	1.0 ± 0.1
SF21	6.2 ± 2.4
GS1	8.1 ± 1.2
GS2	6.2 ± 1.2

Similar results &  
much faster with GS

# Targeted genome evolution by MAGE (Multiplex Automated Genome Engineering)



Cells are repeatedly transformed with synthetic oligonucleotides that recombine with genome and increase sequence diversity

nature

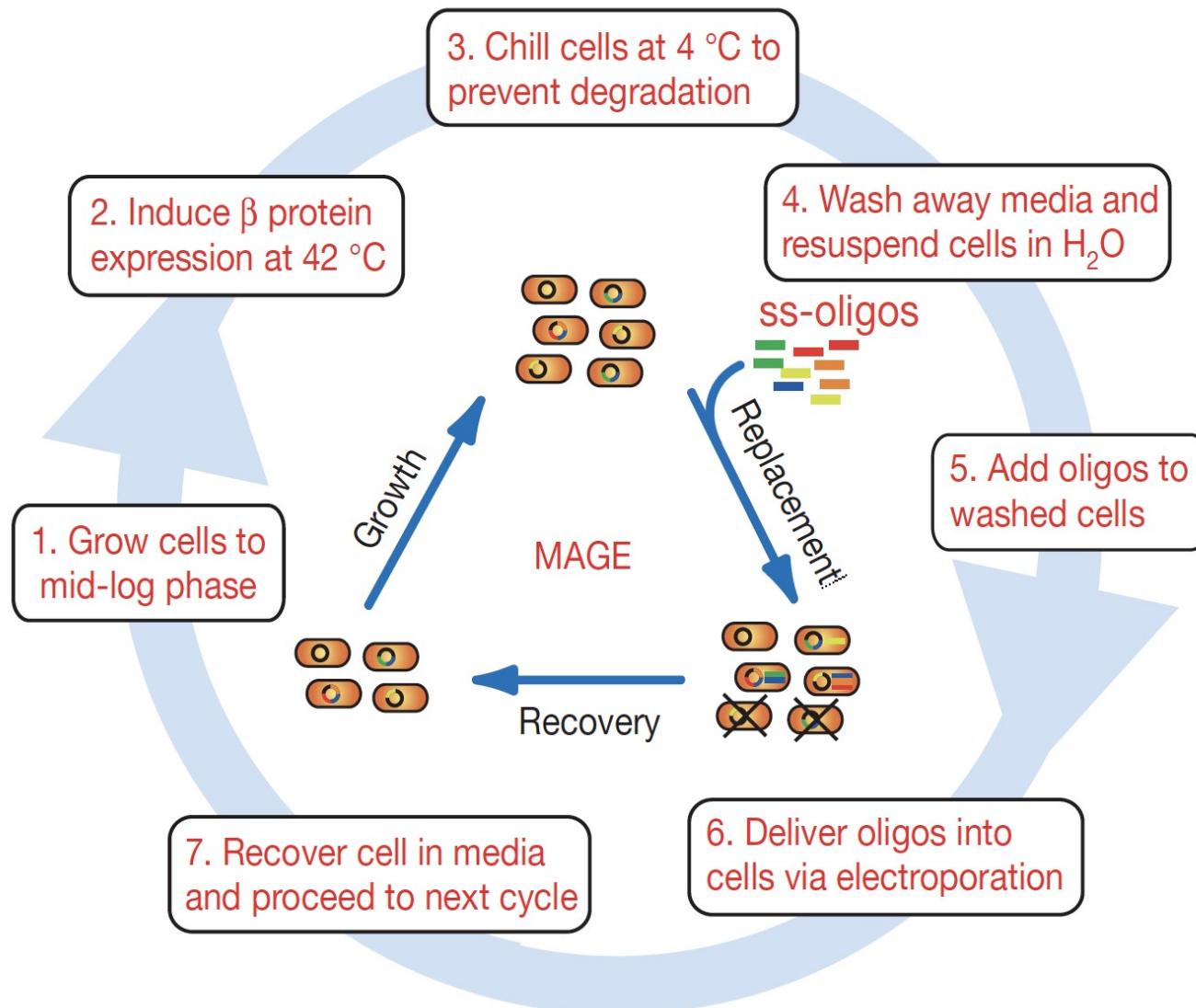
Vol 460 | 13 August 2009 | doi:10.1038/nature08187

## LETTERS

### Programming cells by multiplex genome engineering and accelerated evolution

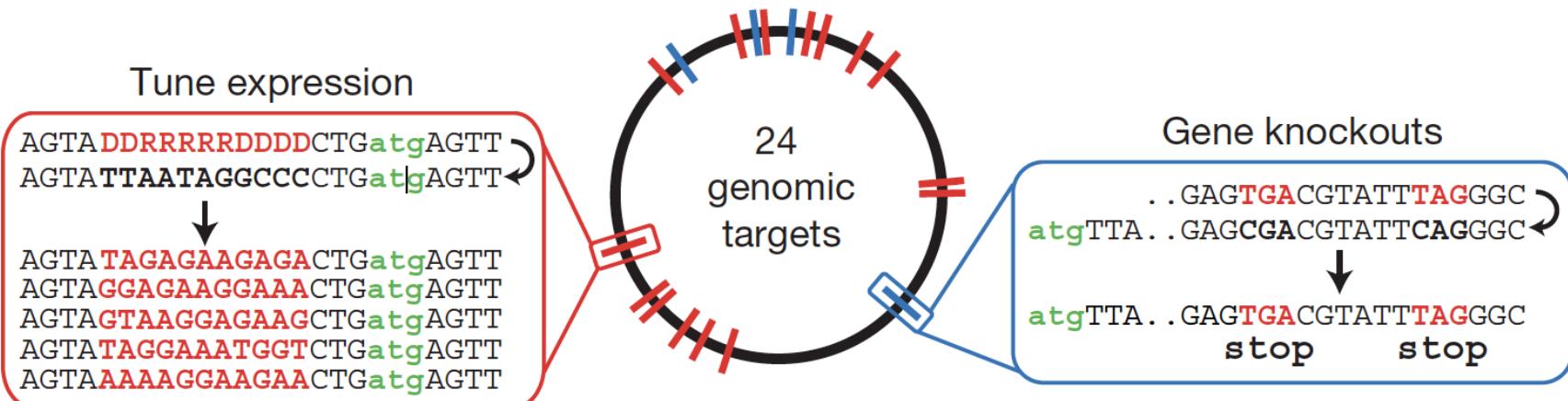
Harris H. Wang<sup>1,2,3\*</sup>, Farren J. Isaacs<sup>1\*</sup>, Peter A. Carr<sup>4,5</sup>, Zachary Z. Sun<sup>6</sup>, George Xu<sup>6</sup>, Craig R. Forest<sup>7</sup> & George M. Church<sup>1</sup>

# One automated cycle (takes ~2.5 hours )



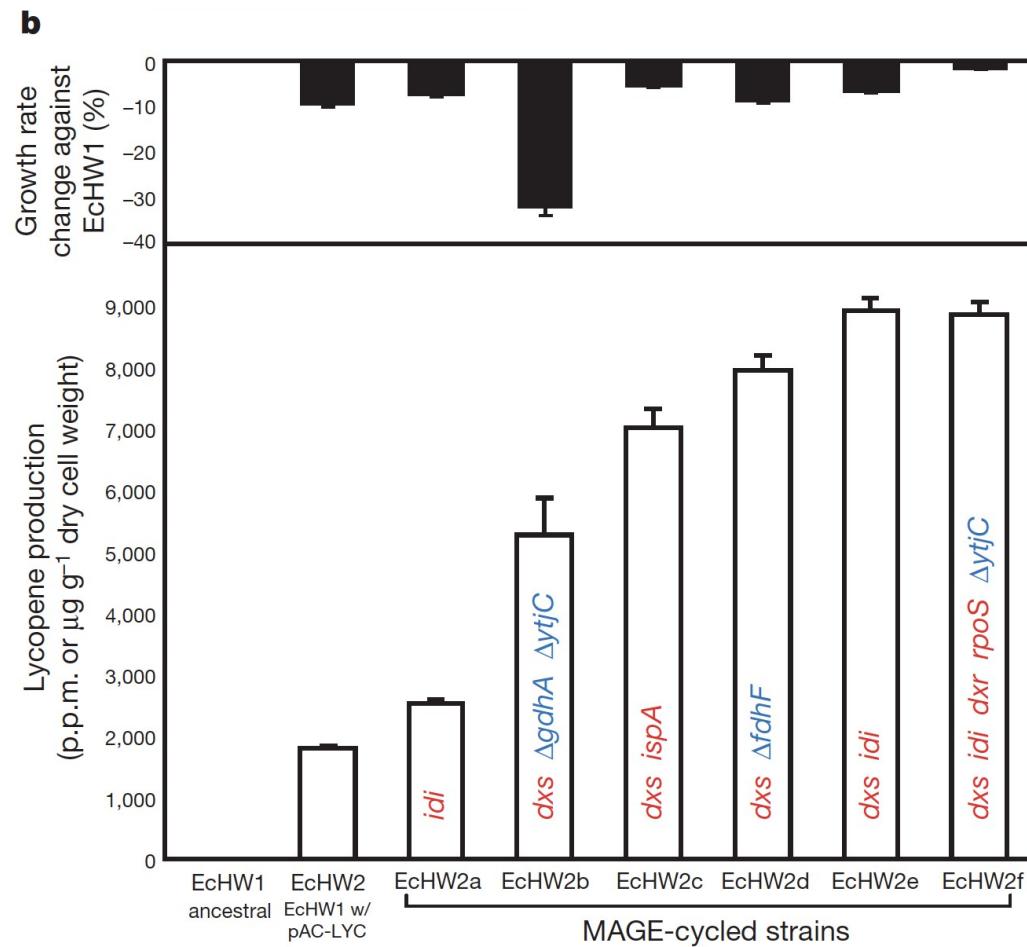
After 5 cycles, measured average of 3.1 base changes per cell

# Test of method: lycopene production



- 20 genes in *E. coli* known to enhance lycopene production, and 4 genes known to divert resources away from the lycopene pathway
- DNA was added to randomize the 20 genes' ribosome binding sites, as well as to knock out the diverting genes
- MAGE cycling was done and colonies were selected for analysis on the basis of increased red color (lycopene)

Optimized translation of a subset of the genes in pathway, and knock out of 3 of the 4 diverting genes, provided up to 5X increases in lycopene production



# Increasing production of a biological compound: rational design

- 1) Increase production of a naturally produced commercial compound
  - Modify existing genes
- 2) Obtain a new organism that can convert an existing compound into a commercial compound
  - Introduce new genes
  - Modify existing genes

natural source of indigo: woad [*Isatis tinctoria*]



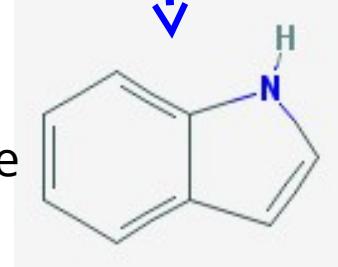
woad

tryptophan



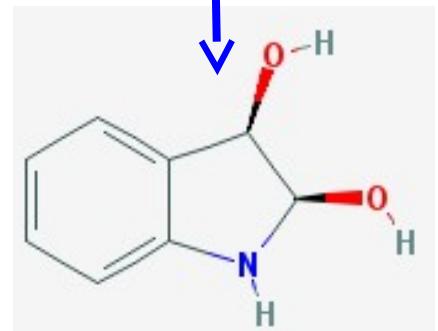
Tryptophanase (*E. coli*)

indole



Naphthalene dioxygenase (cloned)

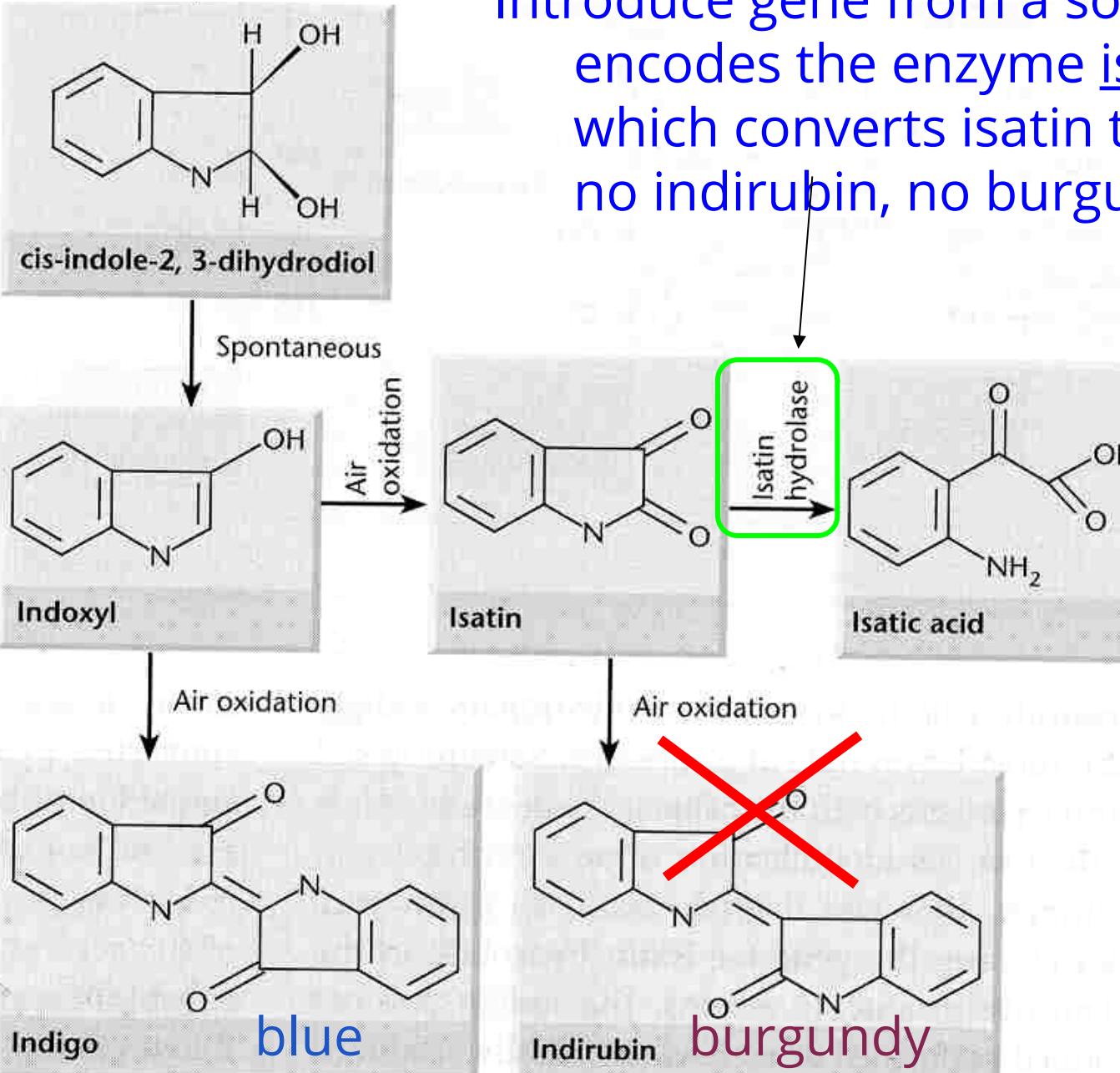
cis-indole-2,3-dihydrodiol



### Engineering *E. coli* to produce indigo

- Mutate tryptophan synthase complex to release indole
- Introduce naphthalene dioxygenase (from *Pseudomonas putida*)

Introduce gene from a soil microbe that encodes the enzyme isatin hydrolase which converts isatin to isatic acid (so no indirubin, no burgundy color)

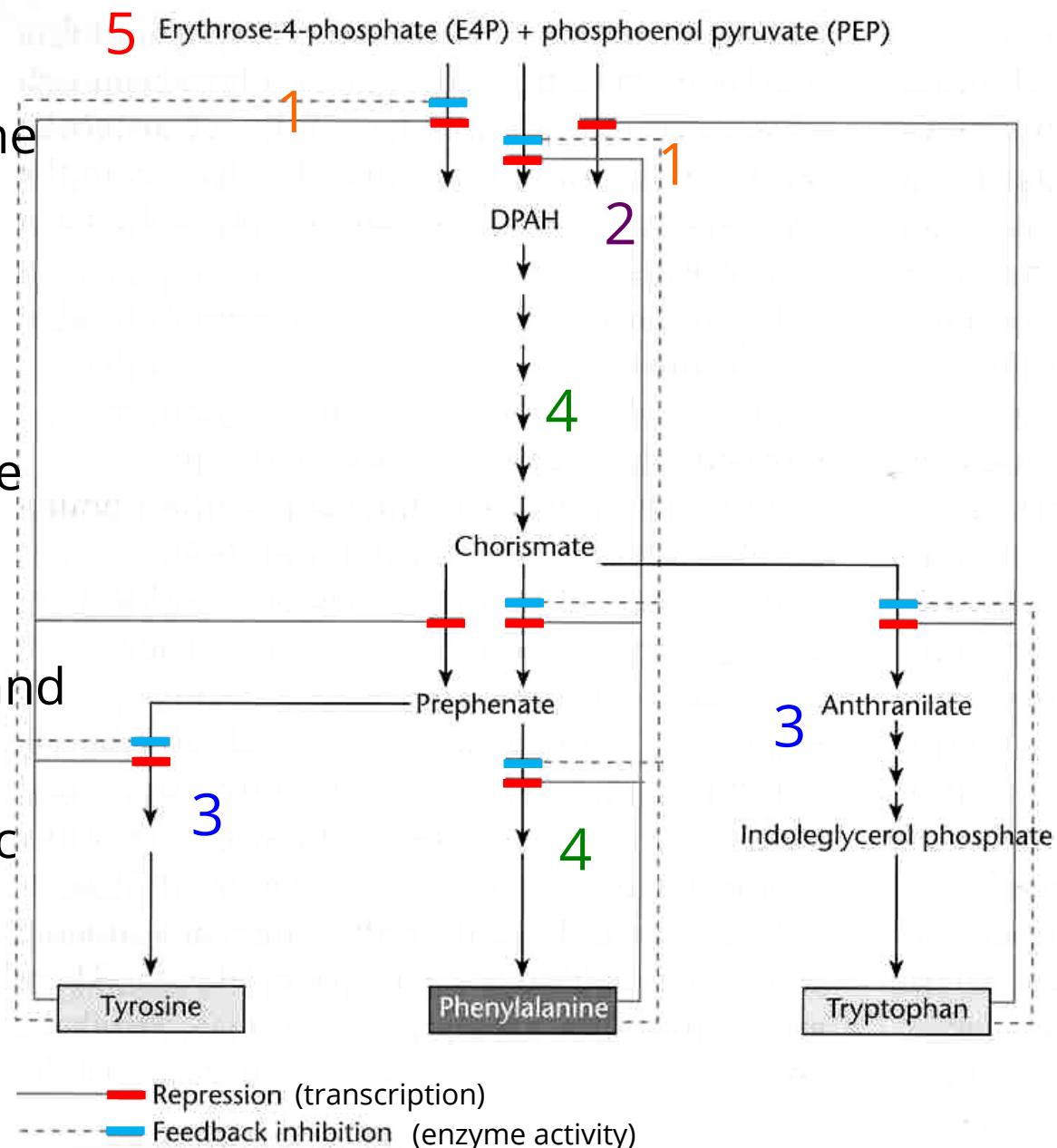


# Metabolic engineering: genetic manipulations to maximize synthesis of biological molecules

- Remove transcription, translation repressors
- Defeat enzyme feedback control
- Speed up rate-limiting enzymes
- Block competing pathways
- Funnel carbon to the pathway of interest
- Increase transport of compound out of cell

# How to overproduce phenylalanine in *E. coli*

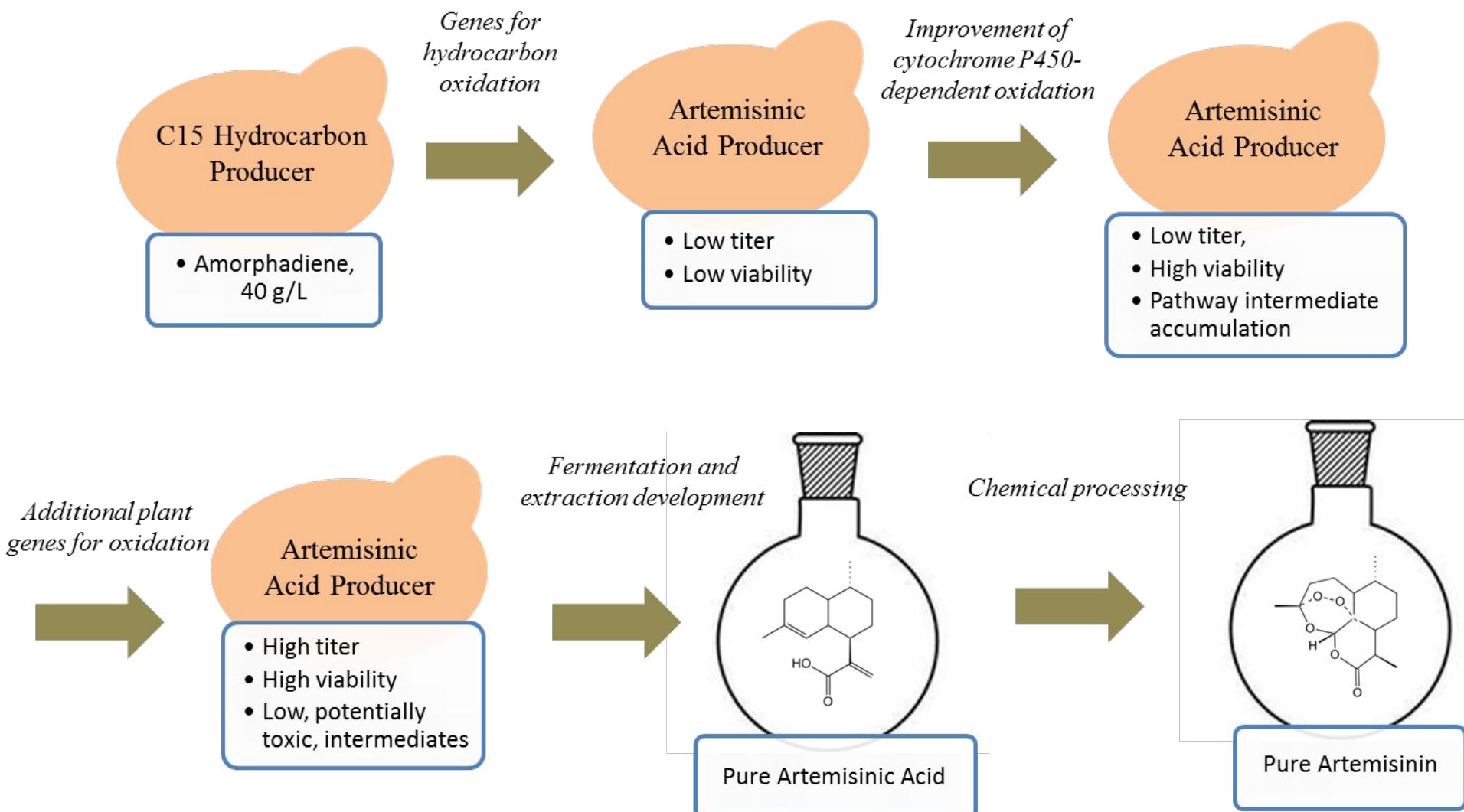
- 1) Remove feedback inhibition (select strains resistant to phenylalanine analogue feedback inhibitors)
- 2) Avoid transcriptional repression (place genes under control of non-phe controlled promoters)
- 3) Remove pathway competition (delete tyr and trp specific genes)
- 4) Overexpress phe-specific genes
- 5) Increase E4P and PEP synthesis



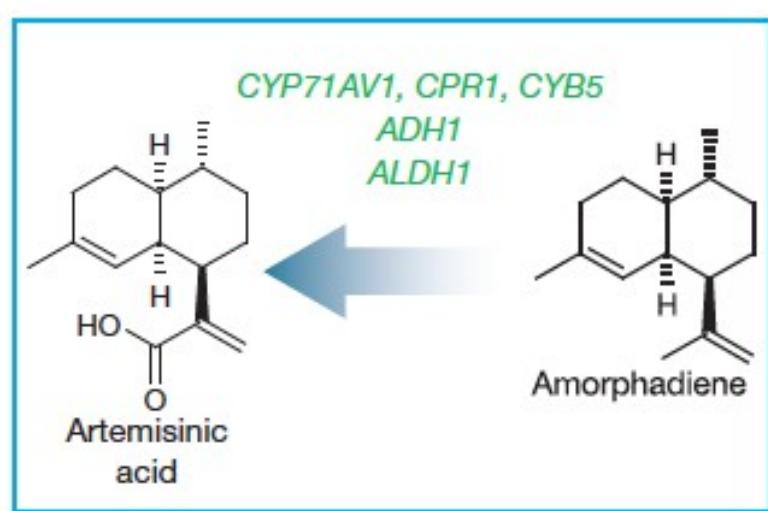
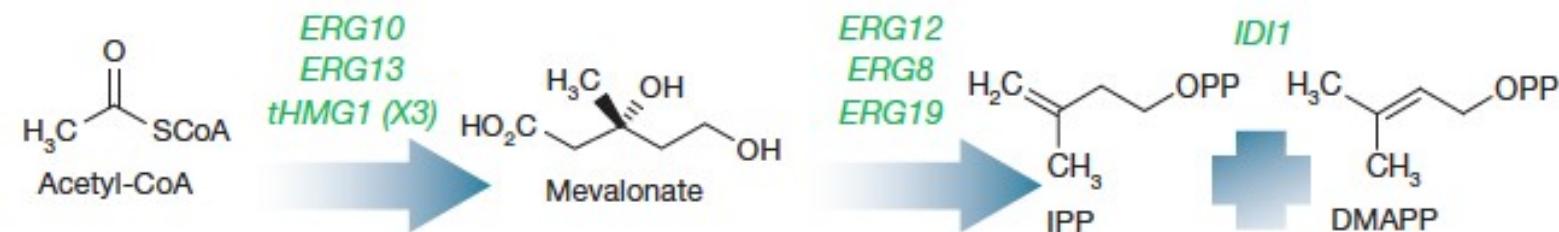
# *S. cerevisiae* engineered to produce artemisinic acid

- Artemisinin is a primary medicine for treating malaria, which caused 660,000 deaths worldwide in 2010
- Production: the plant *Artemisia annua*
- The supply line depends on robust crop yields, so off- years could cause drug shortages
- The “semi-synthetic” production of artemisinin through engineering of yeast was recently reported
- Expected production of artemisinin by this approach: 50-60 tons/year, or 80-150 million doses, produced relatively cheaply

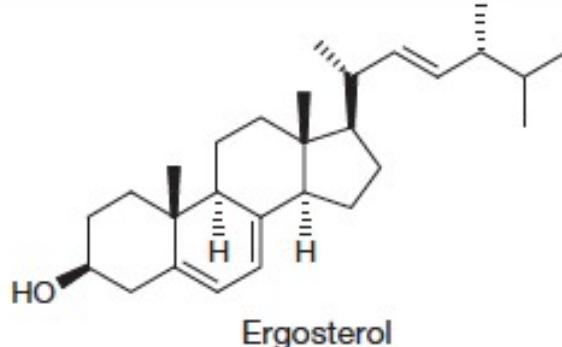
# Synthesis of the antimalarial drug artemisinin (originally produced only by the wormwood plant, *Artemisia annua*)



# Pathway for synthesis of artemisinic acid

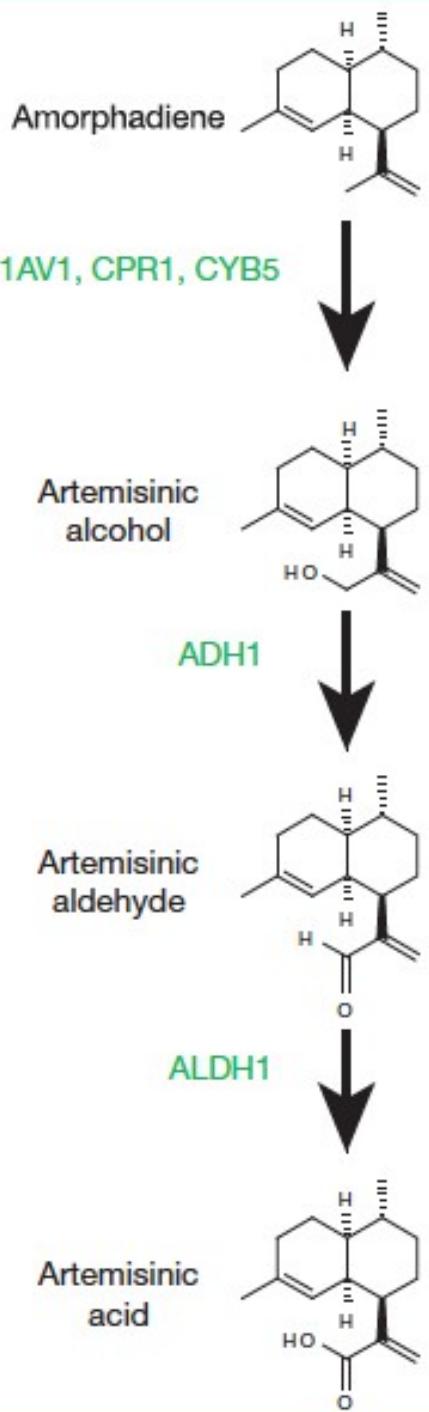


**Green genes:**  
controlled by Gal induction



*ERG1*  
*ERG7*  
*ERG11*  
*ERG24*  
*ERG2*  
*ERG25*  
*ERG6*  
*ERG2*  
*ERG3*  
*ERG5*  
*ERG4*

**Red genes:** engineered to  
be repressible by addition  
of Cu<sup>++</sup> or methionine



Genes isolated and cloned from *Artemisia annua* were the key

However: farmers who grow *Artemisia* (in Asia and Africa) could lose buyers – can they adapt?

Other potentially disruptive synthetic biology projects on the horizon:

- Vanilla
- Vetiver
- Patchouli
- Rubber
- Coconut
- Saffron
- Opioids

<http://www.guardian.co.uk/global-development/poverty-matters/2013/apr/12/synthetic-malaria-compound-artemisia-farmers>

# Rational metabolic engineering

- Requires at least some knowledge of the biochemical pathway required for compound synthesis
- High degree of control
- Trial and error approach can be time consuming
  - try something
  - see if it works
  - find out where the new block to production is
  - change it, too
  - and so on...

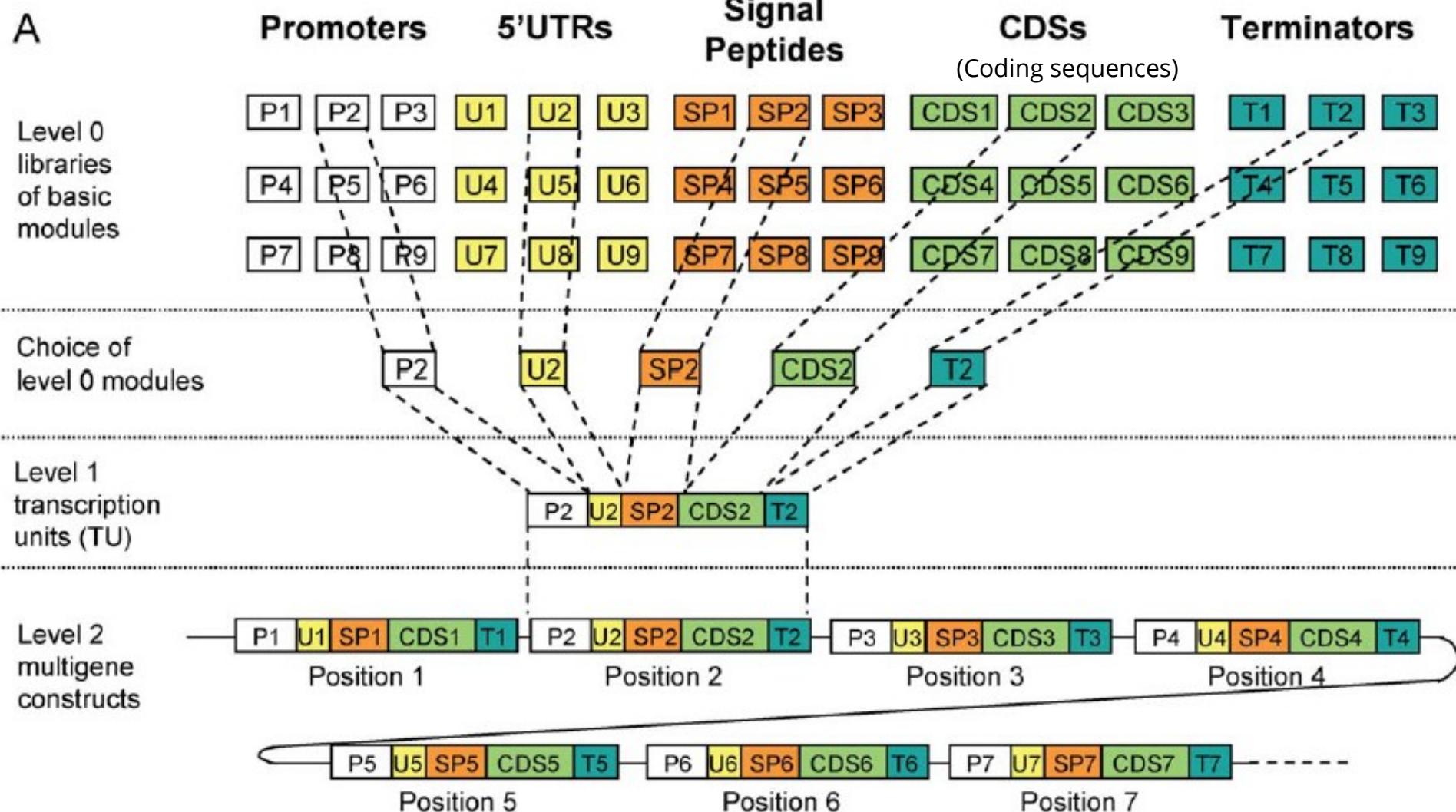
## Synthetic biology:

Construction of cellular systems from component parts to reprogram an organism, or to create a new organism

- engineering principles are followed
- synthesis of modules, followed by ordered assembly
- complete, synthetic genome construction is possible

However: it can be difficult to predict how the modules will interact with each other or with the environment in a living system

# The biobrick approach: mix and match modules



## A Modular Cloning System for Standardized Assembly of Multigene Constructs

PLoS ONE | www.plosone.org

Ernst Weber<sup>1</sup>, Carola Engler<sup>1</sup>, Ramona Gruetzner, Stefan Werner, Sylvestre Marillonnet<sup>\*</sup>

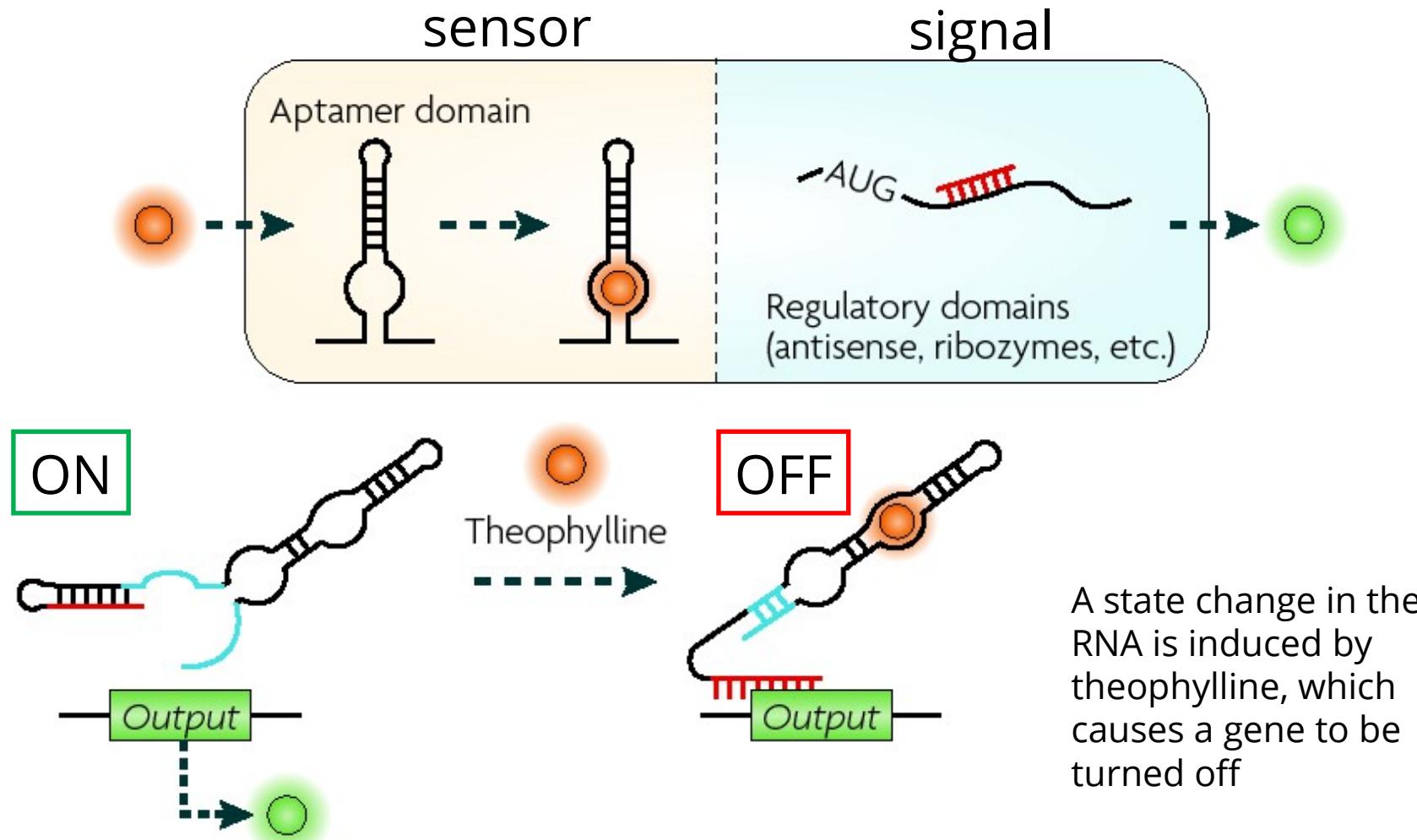
Icon Genetics GmbH, Halle/Saale, Germany

February 2011 | Volume 6 | Issue 2 | e16765

# Synthetic biology module:

An engineered genetic element that performs a specific task

BIOSENSOR: a genetic control module. Presence of a small molecule shuts off signal



# Registry of biological parts:

[http://parts.igem.org/Main\\_Page](http://parts.igem.org/Main_Page)

## Registry of Standard Biological Parts



tools catalog repository assembly protocols help search

BBA\_



### iGEM 2020: An exceptional year

This is not a normal year. Together, we are facing the biggest pandemic in the past 100 years. Based on your feedback, we are adapting the iGEM experience - we have examined each part of the competition to make it even better this year. This is iGEM in the time of a pandemic. It will be different but it will be worth it.

[See the full announcement here](#)

Take a look at some of the changes on our [New for 2020](#) hub.

### Add and Document Parts

Start [adding and documenting](#) your parts now! Your parts should be well characterized and measured, and follow the Registry's requirements.

### Sample Submissions

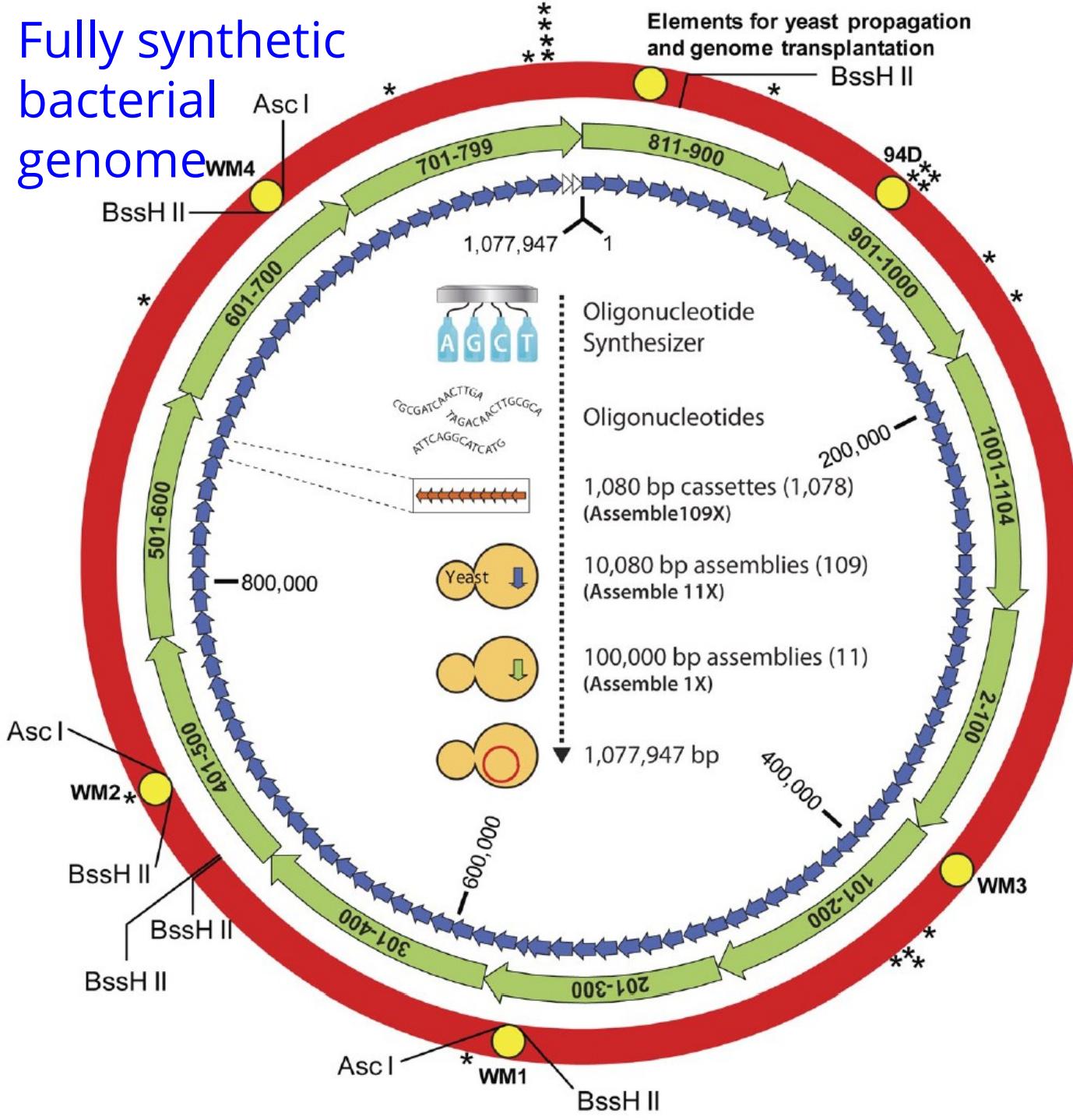
iGEM teams do not need to send samples of their parts for 2020. We want teams to focus on the documentation of their parts! Teams must follow 2020 requirements for parts, including [BioBrick RFC10](#) or [Type IIS compatibility](#).

### Update: Distribution Kit

As a result of the current pandemic, we will not be able to manufacture and ship our [2020 DNA Distribution Kits](#) to teams and labs this year. We sincerely apologize for any inconvenience this may cause.

iGEM: International Genetically Engineered Machine

# Fully synthetic bacterial genome



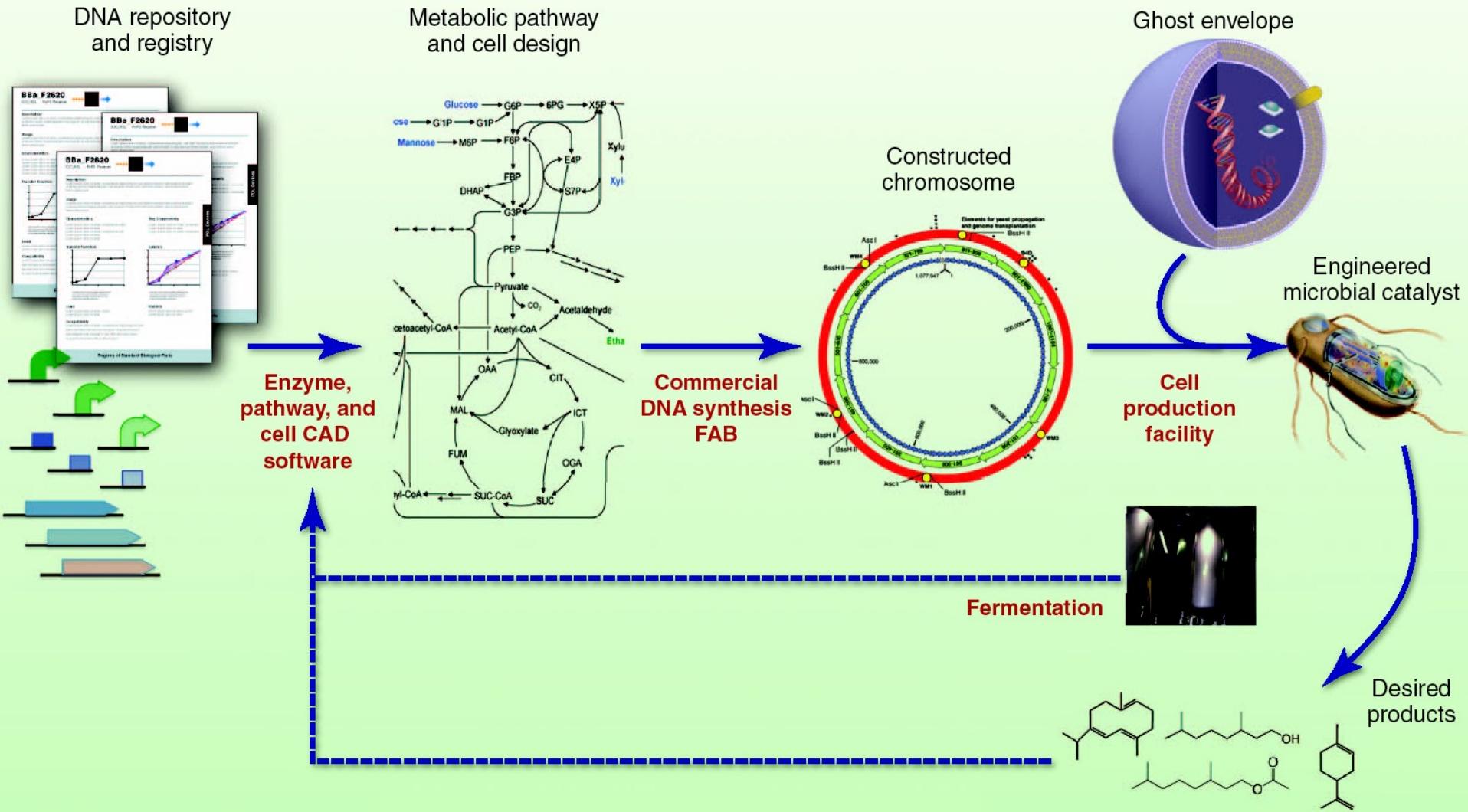
1080 bp synthetic cassettes assembled into sets (10 per set, 109 sets total) by recombination in yeast

10,800 bp fragments assembled into sets (10 per set, 11 sets total)

11 ~100 kb fragments assembled into complete genome

~ \$40M  
20 people  
>10 years

# The future of bioengineering?



## Risks? How to manage them?

How do synthetic life forms interact with various potential environments?

How long do synthetic organisms persist, and how well do they compete with non-synthetic organisms?

How quickly can the synthetic organism evolve?

Can synthetic organisms transfer their DNA to natural organisms and vice versa?

“Once released, synthetic organisms cannot be retrieved”

# Genetic engineering for bio-containment

- Make the organism dependent on one or more non-canonical amino acids
  - Genetic code is re engineered
  - Requires feeding with non-canonical amino acid
  - Genetic code differs from other organisms: genetic exchange can't readily occur
- Make the organism dependent on added small molecules not found in nature
  - 'riboregulators' based on riboswitches controlling many essential genes in an organism
  - Addiction modules encode nucleases that kill the cell if the antidote isn't made, and the antidote is controlled by a small molecule

# Applied mutagenesis: pathway engineering and synthetic biology

Increase biological production of useful molecules

- Random screening for overproducing strains (genome shuffling)
- Rational engineering of pathways and organisms

# Cloning in eukaryotes (part 1)

*Saccharomyces cerevisiae*

- 1) Types of vectors
- 2) Integrative gene deletion and whole genome screens
- 3) TAP cloning

## Guide to readings:

- 1) *Yeast as a model organism* 2011
- 2) *Glaever et al. 2002.* Creation of an ordered library of barcoded yeast deletion strains.
- 3) *TAR cloning* 2008. A protocol describing how to use the transformation-associated recombination cloning protocol in yeast
- 4) *Use of TAR to reconstruct SARS CoV2.* A recent description of construction of synthetic virus from sequence info

# *Saccharomyces cerevisiae*: a model eukaryote

- 1) Biochemistry and cell biology similar between yeast and “higher” eukaryotes
  - many gene homologs between yeast and humans, eg. Cell cycle (cancer) genes
- 2) It is easy to convert it from diploid to haploid (by sporulation) and back to diploid (by mating), so you can look at mutations haploid and diploid
- 3) Excellent genetic tools are available
- 4) Very easy to work with in the laboratory

# Yeast and genomics

- The yeast genome was sequenced in 1996
- It became a model system for inventing and testing tools for studying entire eukaryotic genomes/transcriptomes/proteomes etc.
- The functions of many eukaryotic gene families were first discovered in yeast
- <http://www.yeastgenome.org/>

# Transformation using yeast

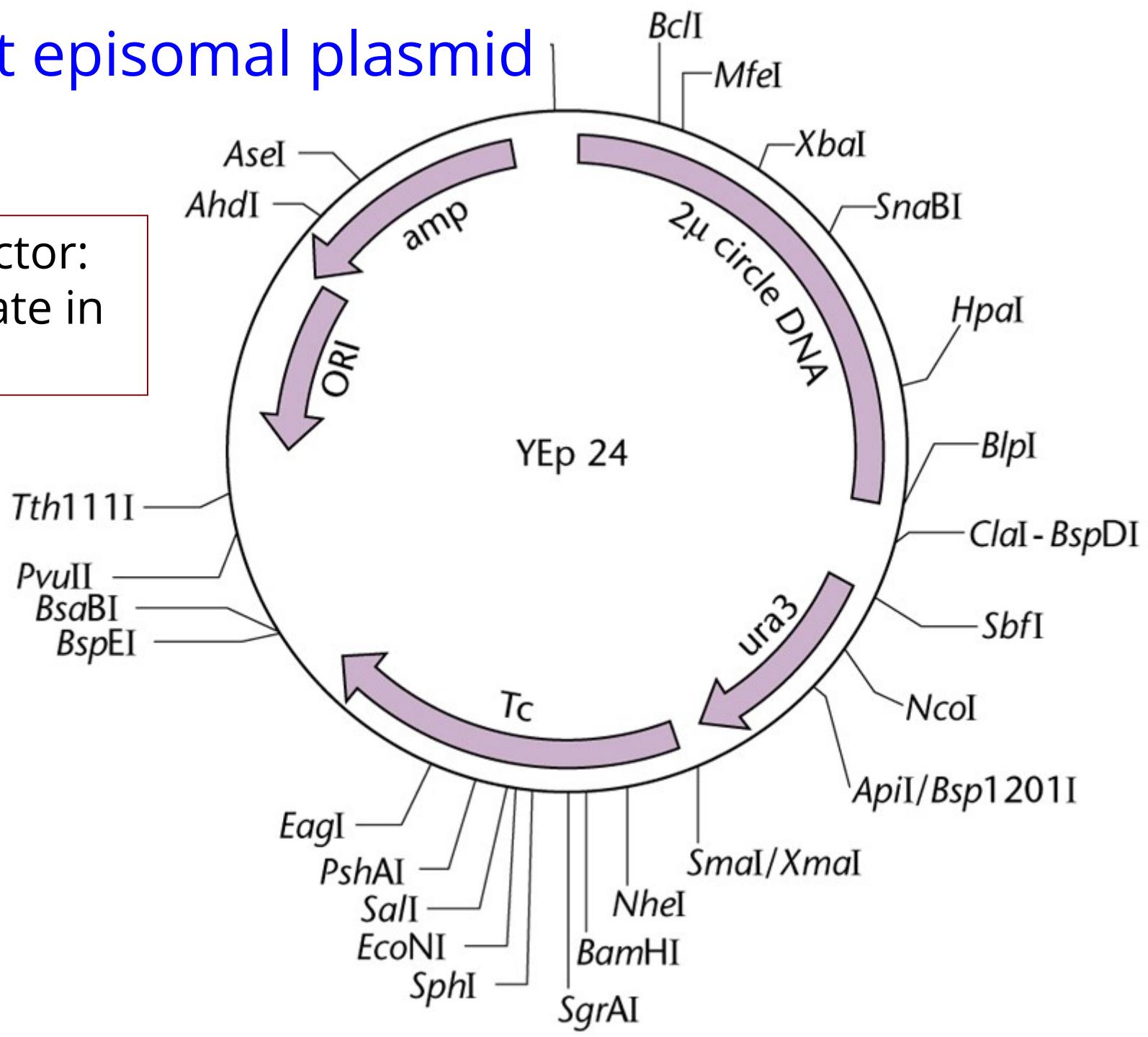
- Electroporation, or chemical competence (Lithium chloride/PEG treatment)
- Selection:
  - nutritional markers
    - His3, Leu2, Trp1: amino acid biosynthetic genes
    - Ura3 – nucleotide biosynthetic gene
    - **these require auxotrophic yeast strains**
  - Aminoglycoside (ribosome inactivating) antibiotic resistance (kanamycin)

# **Yeast Episomal plasmid: high copy number plasmid**

- Contains naturally occurring “2 micron circle” origin of replication
- High copy number: 50-100/cell

# A yeast episomal plasmid

Shuttle vector:  
can replicate in  
*E. coli*, too

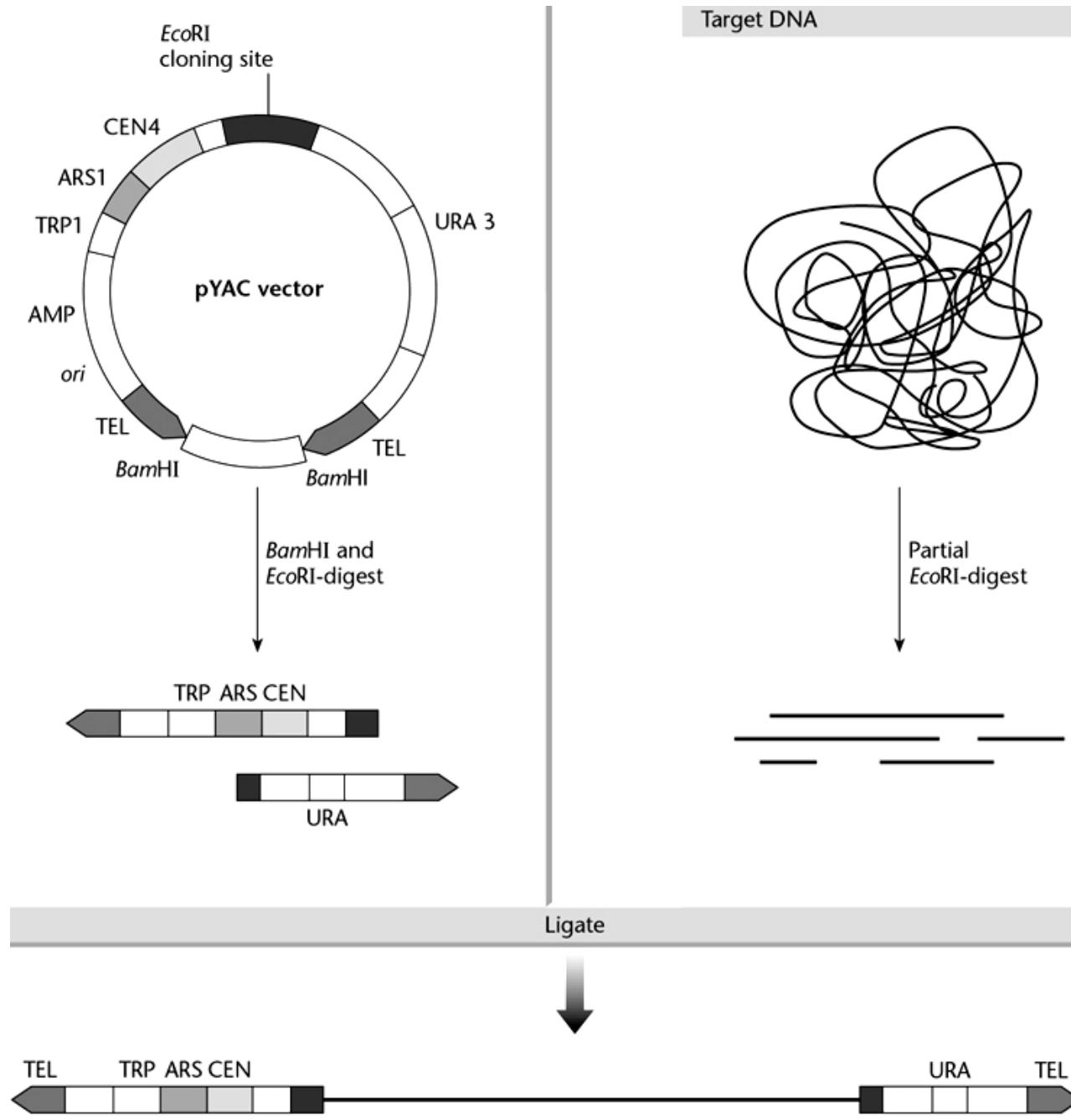


# Yeast Centromeric plasmid: low copy number plasmid

- Yeast Centromeric plasmid
- Contains yeast *ars* (autonomously replicating sequence) for replication
- Contains yeast centromere for proper segregation to daughter cells
- Very low copy number, ~1 per cell (good for cloning genes that are toxic or otherwise affect cell physiology)

# YAC: yeast artificial chromosome

- Replicates as chromosome: centromere and telomeres
- Clone **very large pieces** of DNA



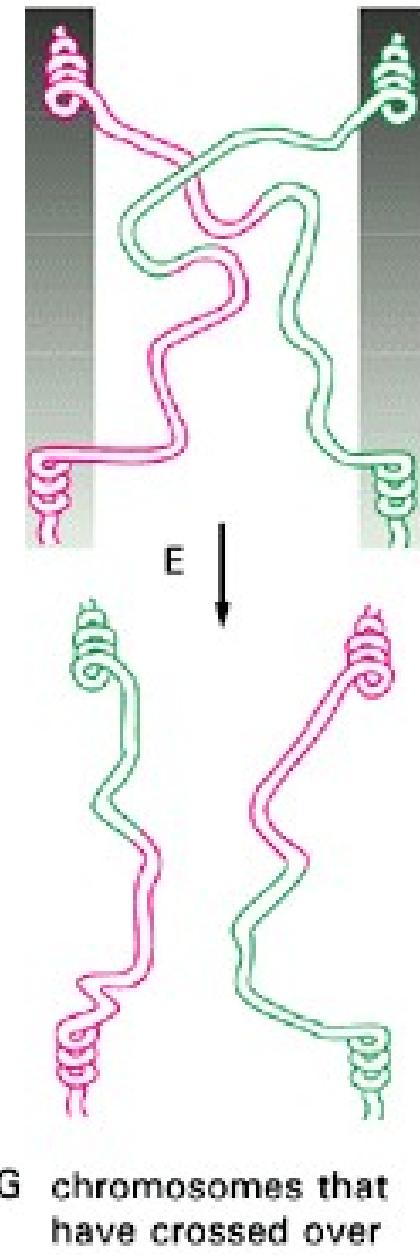
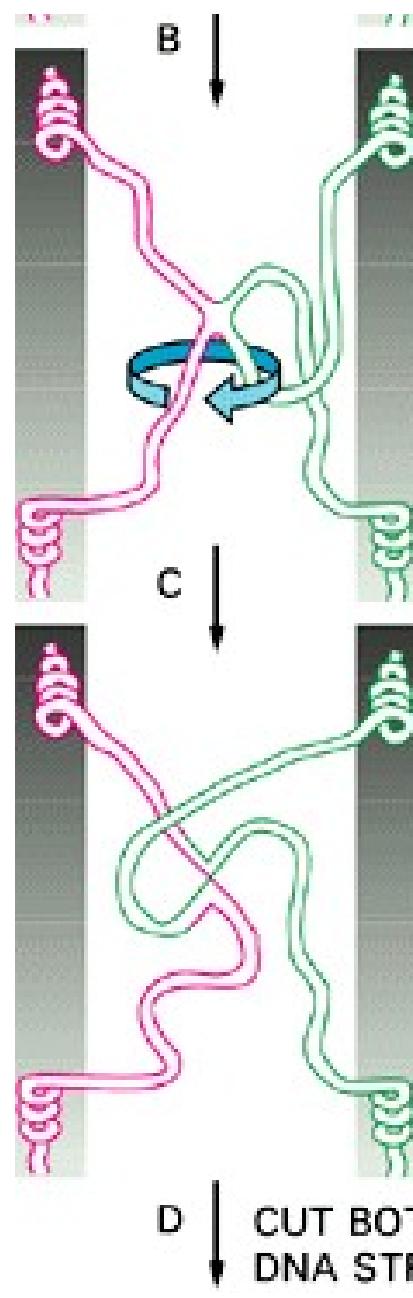
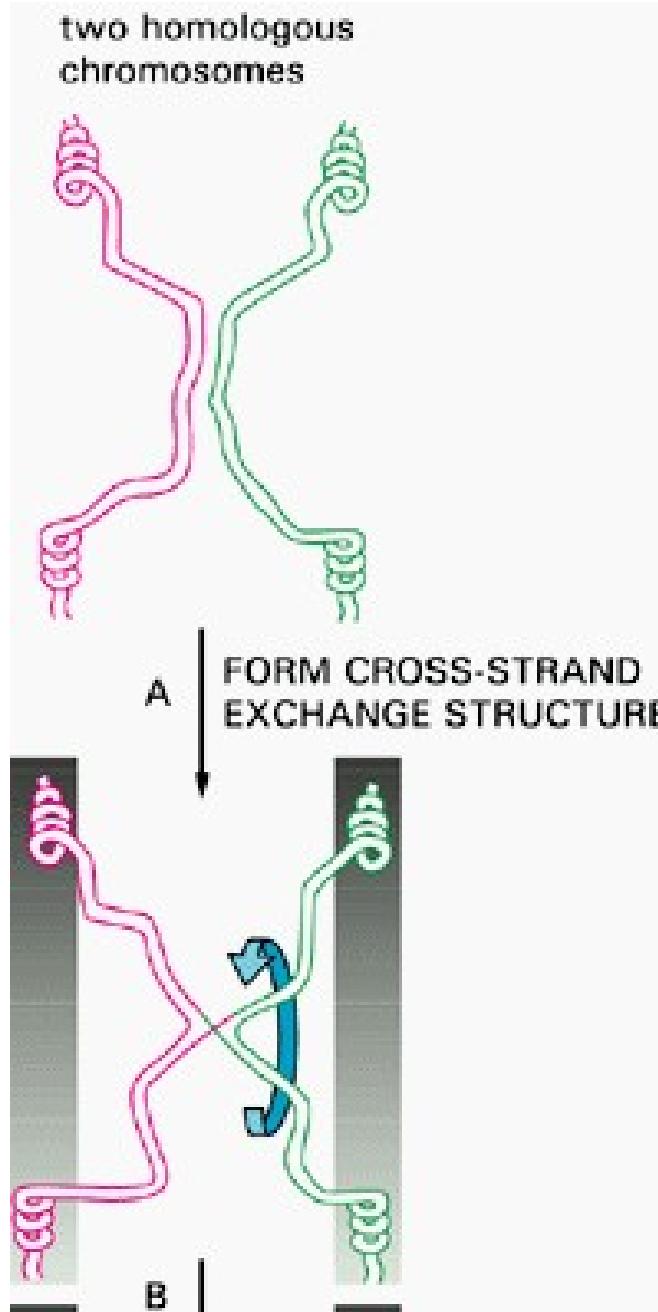
# Yeast integrative plasmid: homologous recombination

- No yeast replicon in the plasmid
- Can transform but cannot replicate
- Requires integration into chromosome for propagation
- Genes on the chromosome can be manipulated/deleted

# DNA can be integrated by recombination

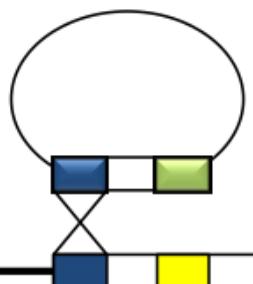
- If transformed DNA has **no way to replicate** and has **homology to chromosome**, it can be integrated by homologous recombination
- Two pieces of DNA with the same (or similar) sequence: RecA protein (bacteria) or Rad51 (eukaryotes) causes strand exchange between homologous sequences
- Homologous recombination occurs with a highly predictable frequency: in a transformation experiment with a reasonable length of homologous DNA, ~1/1000 transformed cells will recombine the transformed DNA

# Homologous recombination: portrait of a single cross-over



# Looping in of a circular DNA by recombination can be followed by looping out

1)



Plasmid (can't replicate)

2)



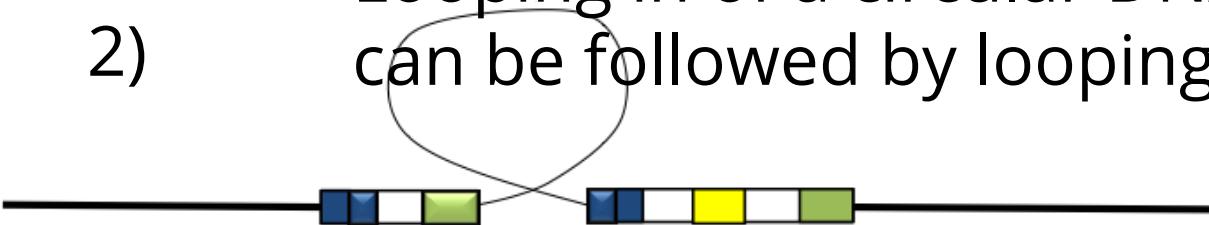
Homologous recombination inserts the plasmid into the chromosomal DNA at one of the two sites flanking the target gene.



	Plasmid homology site 1
	Plasmid homology site 2
	Chromosomal homology site 1
	Chromosomal homology site 2
	Target gene
—	Chromosomal DNA
—	Plasmid DNA

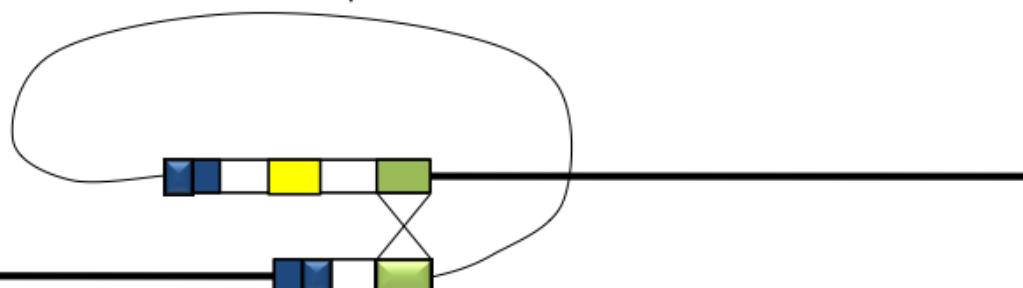
## Looping in of a circular DNA by recombination can be followed by looping out

2)



The DNA realigns itself to undergo recombination at the other homology locus.

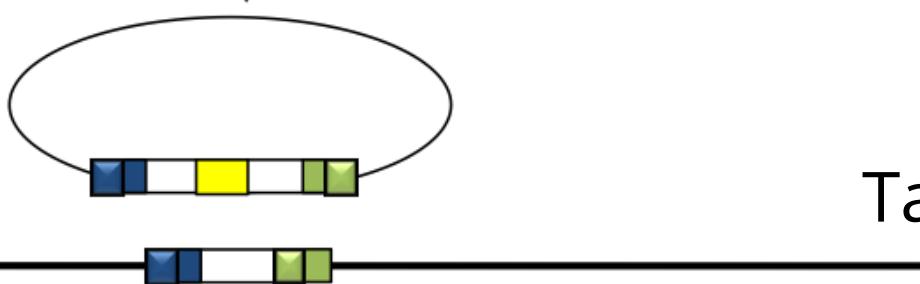
3)



Homologous recombination could occur at the other locus in either of step one and step three, leading to four total possible paths, two of which produce the desired removal. Screening is required in order to determine in which cells the knockout was successful.

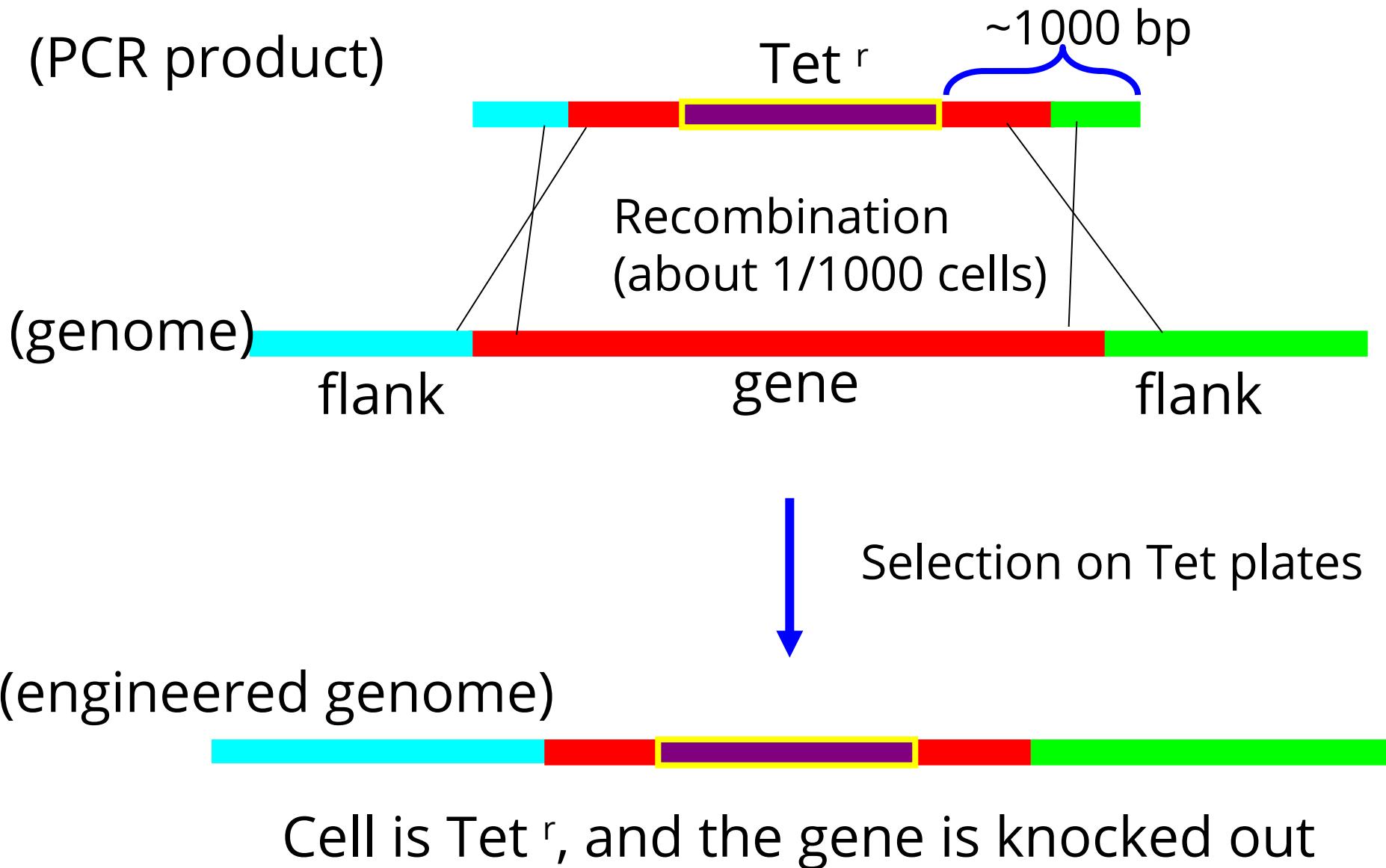
Homologous recombination removes the plasmid from the chromosomal DNA along with the target gene.

4)



Target gene deleted

# Homologous recombination (double crossover) to knock out a gene

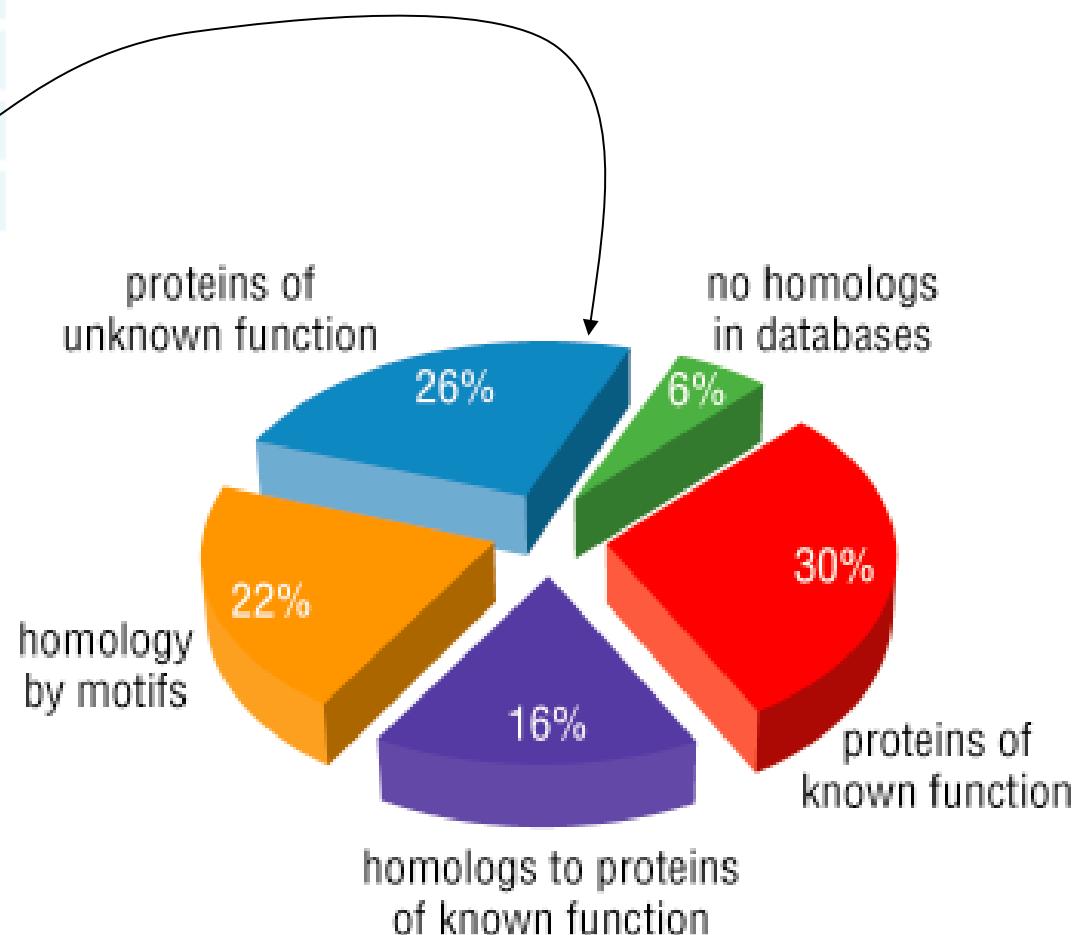


## Genome Sizes of Representative Organisms

Organism	Genome size (base pairs)	Number of genes
<i>Mycoplasma genitalium</i>	$45.8 \times 10^5$	483
<i>Methanococcus jannaschii</i>	$1.6 \times 10^6$	1,783
<i>Escherichia coli</i>	$4.6 \times 10^6$	4,377
<i>Pseudomonas aeruginosa</i>	$6.3 \times 10^6$	5,570
<i>Saccharomyces cerevisiae</i>	$1.2 \times 10^7$	6,282
<i>Caenorhabditis elegans</i>	$1.0 \times 10^8$	19,820
<i>Drosophila melanogaster</i>	$1.8 \times 10^8$	13,601
<i>Arabidopsis thaliana</i>	$1.2 \times 10^8$	25,498
<i>Homo sapiens</i>	$3.3 \times 10^9$	~30,000 (?)

Large percentages of coding proteins cannot be assigned function based on homology

## Homology: the function is only inferred



# **Gene function: phenotypes of knockouts**

All yeast genes were knocked out -- what can this tell us about their functions?

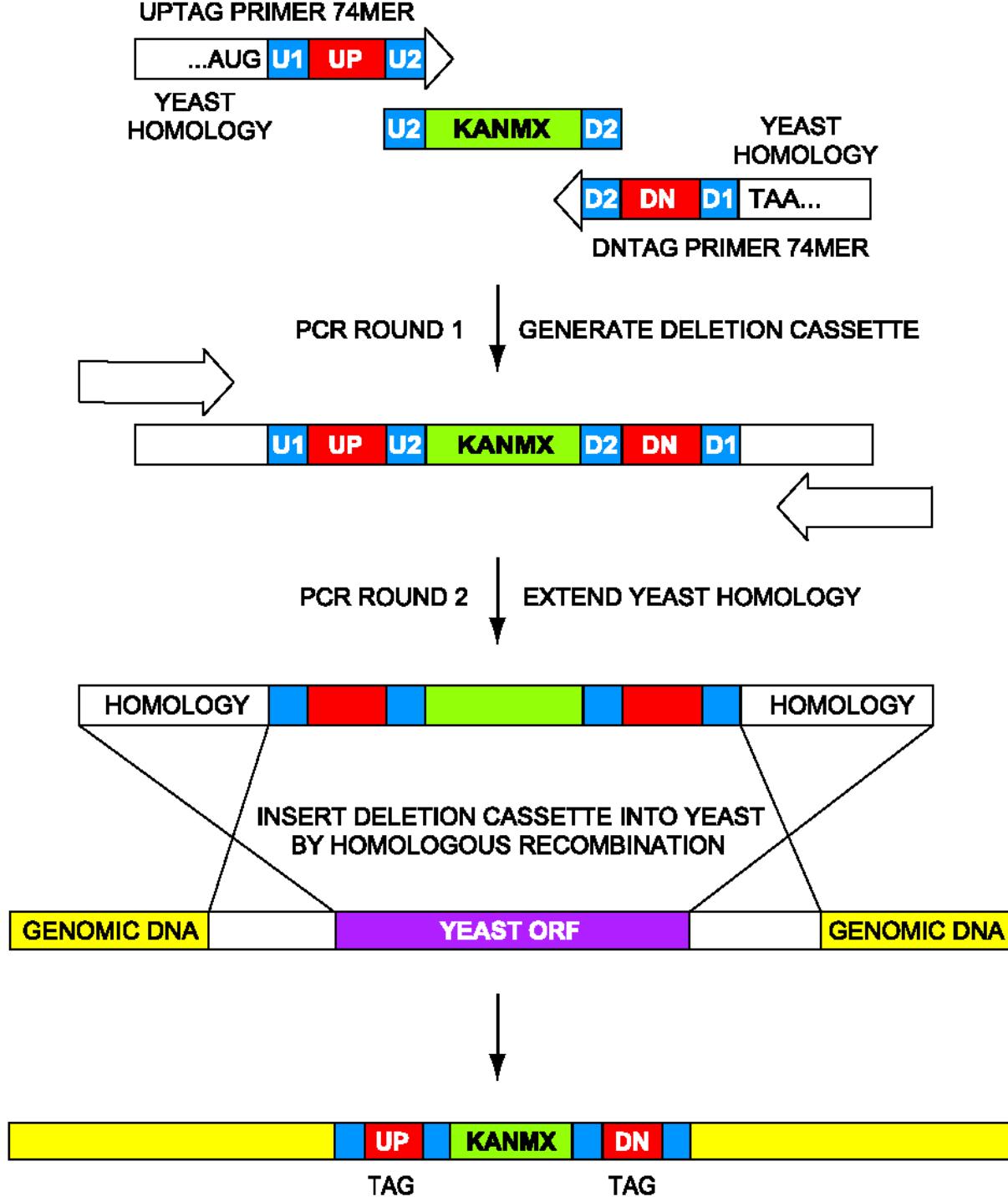
Use “bar codes” for tracing yeast knockout strains...

## **“Functional profiling of the *Saccharomyces cerevisiae* genome.” Glaever et al. *Nature* 2002**

“We systematically constructed a nearly complete collection of gene-deletion mutants (96% of annotated open reading frames, or ORFs) of the yeast *Saccharomyces cerevisiae*. DNA sequences dubbed ‘molecular bar codes’ uniquely identify each strain, enabling their growth to be analysed in parallel and the fitness contribution of each gene to be quantitatively assessed by hybridization to high-density oligonucleotide arrays.”

“We failed to delete 215 genes for unknown reasons; about 62% of these are questionable ORFs that have no known biological function.”

# Knockouts with tags



# Example of gene knock-out screen

Which genes are required for growth with galactose as the sole carbon source?

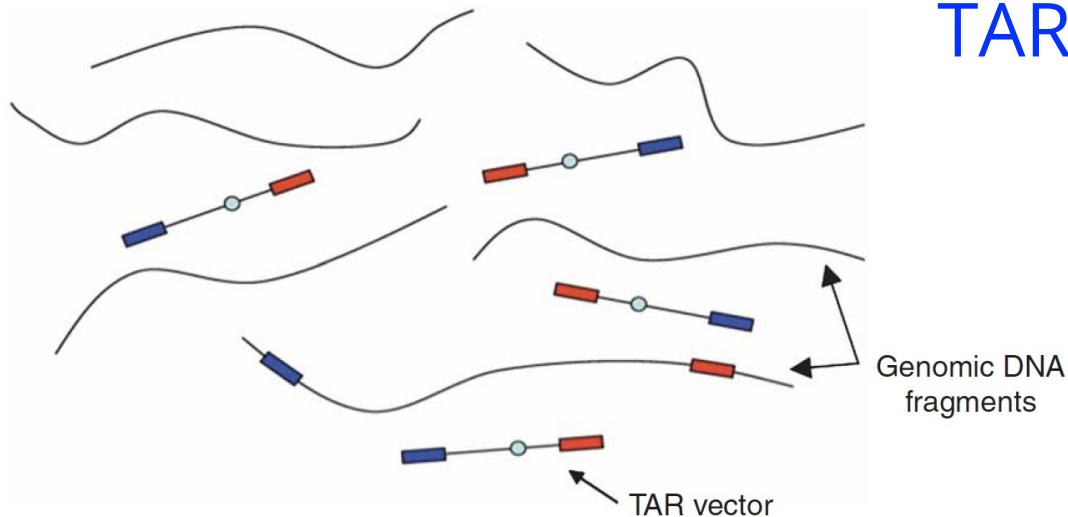
1. Pool collection of viable gene knock-outs
2. Grow on galactose minimal medium
3. Strains with knock-outs of genes essential for growth on galactose will die, removed from population
4. Amplify and label “bar codes”, probe a microarray containing all bar codes--which ones disappear when galactose is sole carbon source?
  - 10 new genes identified in galactose metabolism

# Cloning in yeast: TAR (Transformation-associated recombination)

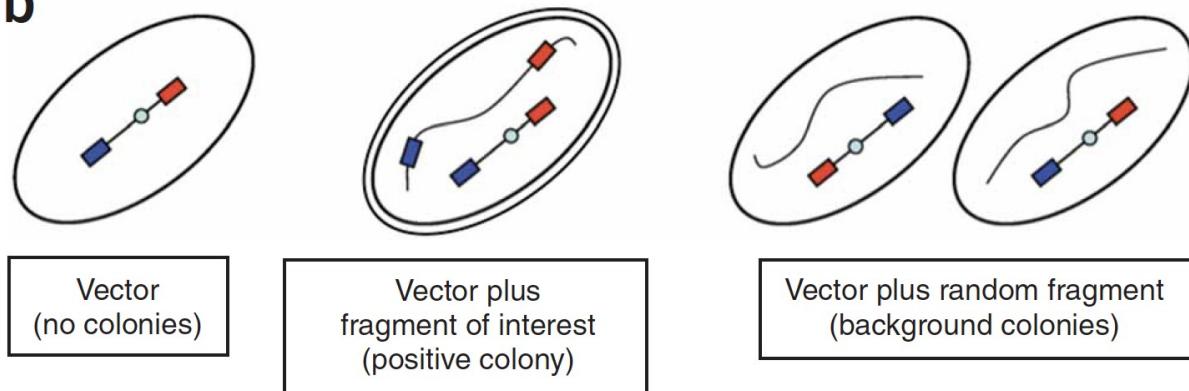
- Free DNA ends are good substrates for homologous recombination in yeast
- Any chromosomal fragment up to 300 kb in length can be isolated in yACs within weeks and with high efficiency
- This means you can clone entire eukaryotic genes, with introns, exons, and control regions, with relative ease

# TAR cloning

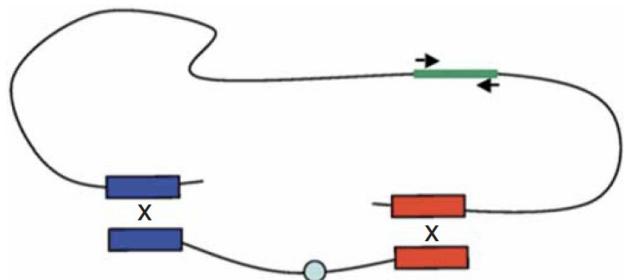
a



b



c

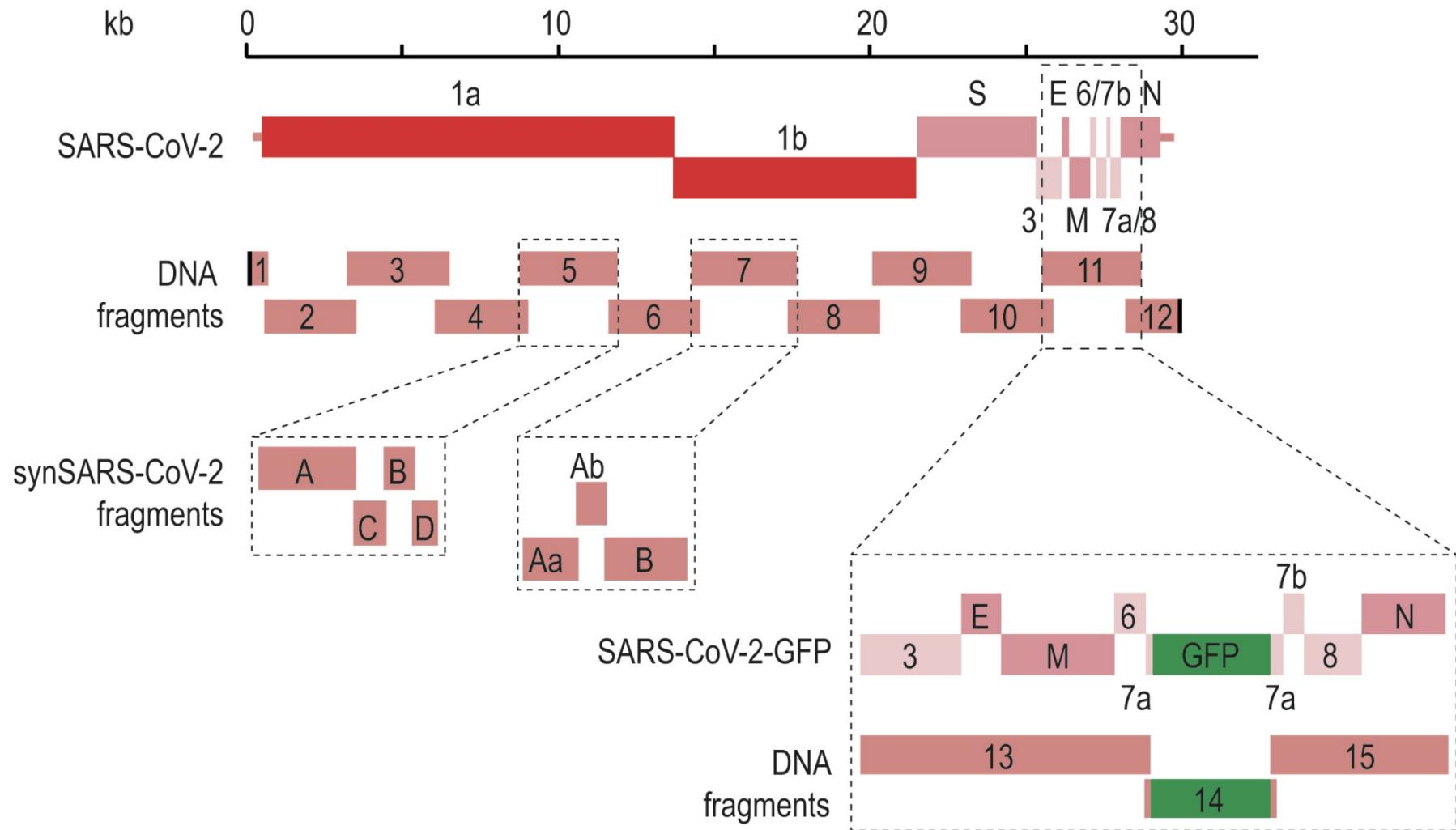


- Fragment genome
- Transform yeast with fragments along with TAR circular yAC vector containing 'hooks' for homologous recombination
- Check transformants with PCR primers in middle of desired clone fragment

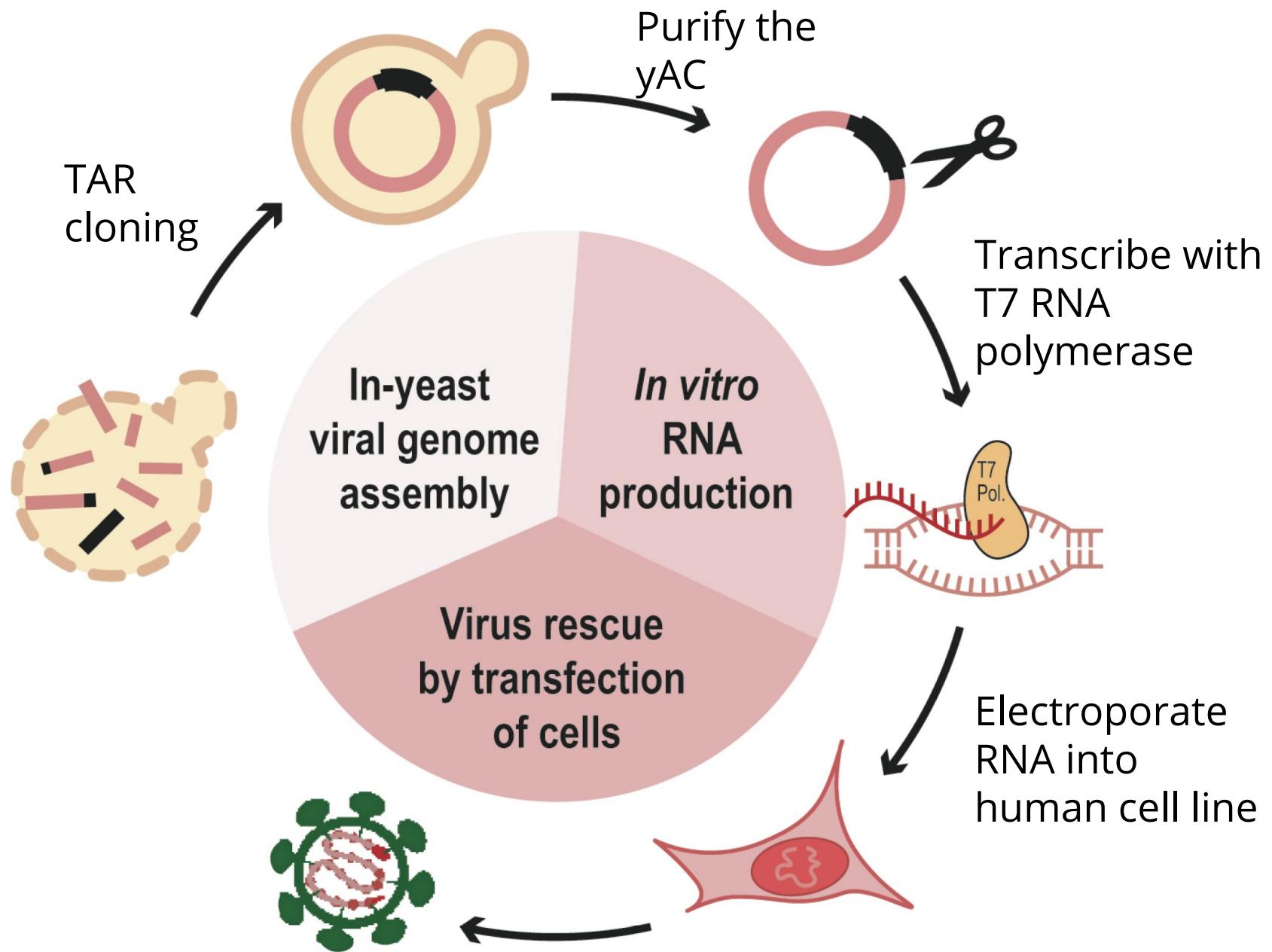
# Reconstruction of the SARS CoV 2 genome using TAR cloning

- The size of the genome (~30 kilobases) makes it difficult to clone and maintain in E. coli
- TAR cloning permits one step transformation/assembly from numerous smaller fragments
- A GFP expressing version of the virus was created to assist in screening of antivirals–easy to track presence/absence of virus

# DNA fragments for assembly of SARS CoV 2 and variants



# Reviving virus from DNA fragments



# Cloning in *S. cerevisiae* (cloning in eukaryotes, part 1)

- 1) Types of vectors
- 2) Integrative gene deletion and whole genome screens

# Cloning in eukaryotes: cell culture and stem cells

- 1) Animal cell culture
- 2) Stem cells: easy genetic manipulation, and capable of differentiation into other cell types
- 3) ES cells derived from Somatic Cell Nuclear Transfer (SCNT)
- 4) Induced pluripotent stem (iPS) cells, reprogrammed by directed gene expression

## **Guide to readings:**

- 1) *ATCC Animal Cell Culture Guide*. Basic guide for animal cell culture from the American Type Culture Collection
- 2) *Stem Cell Basics 2020*. Guide to stem cell biology from the National Institutes of Health
- 3) *Review: Pluripotent stem cells 2010*. Methods for getting pluripotent stem cells, and the clinical implications.
- 4) *ES cells from SCNT 2013*. Perspective on pioneering human cloned stem cell work, done at OHSU.
- 5) *Stem cell comparison 2014*. Comparison between methods of stem cell reprogramming.
- 6) *iPS and organoids in disease modeling 2019*.

# Gene transfer to animal cells

- Studying complex eukaryotes
  - Work with isolated cells to simplify cell biology studies
  - Human cell lines allow convenient studies of basic human cell & molecular biology
- Cells in culture can be used to establish conditions for gene therapy – treatment of genetic disorders by restoration of gene function
- Large scale animal cell culture can be useful in production of recombinant animal proteins: accurate post-translational modifications

# Growth of mammalian cells: tissue culture

Culture media contains

- Glucose, amino acids, vitamins
- Salts/buffers
- antibiotics (anti-bacterial, anti-fungal)
- whole serum (undefined, provides necessary growth factors and other macromolecules)
- Or, specific growth factors, plus transferrin, and insulin (if no serum is added)

Grow at 37°C, in 5% CO<sub>2</sub>

- Some cell types require a “feeder layer” of other, non-dividing cells (DNA-damaged to prevent cell division). Feeder cells provide a surface for growth, as well as some nutrients

# Mammalian cell culture: 3 Types

1. Primary cell culture/finite cell line/cell strain, not derived from tumors
  - a. Released from tissues by enzymatic digestion (which breaks up the extracellular connections)
  - b. Limited number of cell divisions: 40 to 60 divisions, the “Hayflick Limit”, followed by senescence
2. Continuous cell line, ‘transformed’, may be from a tumor
  - a. Earliest example: HeLa (also, lots of others exist)
  - b. Unlimited number of cell divisions: “immortalized”
3. Embryonic/pluripotent stem cells
  - a. Derived from early embryo (blastocyst) or de-programmed adult cell
  - b. Undefined cell type, but changes can be induced
  - c. Potential for unlimited number of cell divisions

# Biology of cells in culture

1. Primary cultures, non-tumor derived
  - a. Essentially normal biology
  - b. New cultures must constantly be made
2. Tumor cell lines
  - a. Unlimited proliferation (not normal biology, in that respect)
  - b. Mutations accumulate over time – how does the biology compare to non-tumor cells?
3. Embryonic stem cells
  - a. Normal biology -- differentiation into various cell/tissue types can be induced
  - b. Cells may be engineered, eg. to correct genetic defect

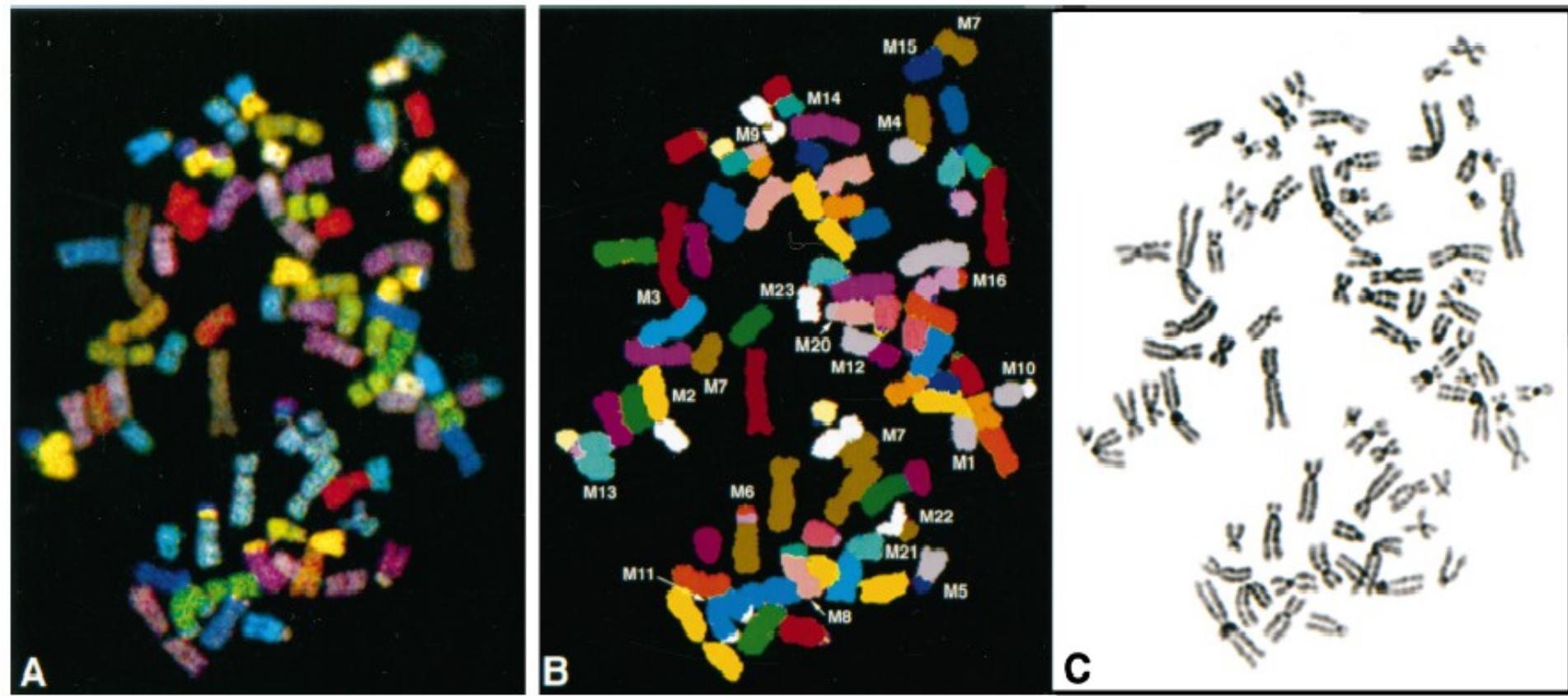
# HeLa cells: the first human tumor cell line

- Henrietta Lacks was an African American woman, and a cancer patient at the Johns Hopkins Hospital in Baltimore. She died from the disease in 1951, at the age of 31
- A tissue sample from her cervical tumor was given to cell biologist George Gey, *without patient or family consent*
- Gey was able to produce the first immortalized tumor cell line from this sample, and shared the HeLa cell culture with labs worldwide
- HeLa cells became one of the central model systems for human cell and molecular biology
- Henrietta Lacks' surviving family members were unaware of the fate of her cells until more than 20 years later
- Her family has only recently received acknowledgment of ethical lapses surrounding the use of her cells for so long, and in many cases for profit

# Proposed research & consent policy changes inspired by the case of Henrietta Lacks and her cells

- Acknowledge whom HeLa cells came from, that she was an African American woman, with a family and her own story
- Require consent for use of cells in research, even if those cells are 'de-identified'
- Acknowledge and undo disparities that exist in basic research, particularly as affected by past and current systemic racism
- Offer financial compensation to help make amends for past injustices
- See editorial "  
[Henrietta Lacks: science must right a historical wrong](#)"

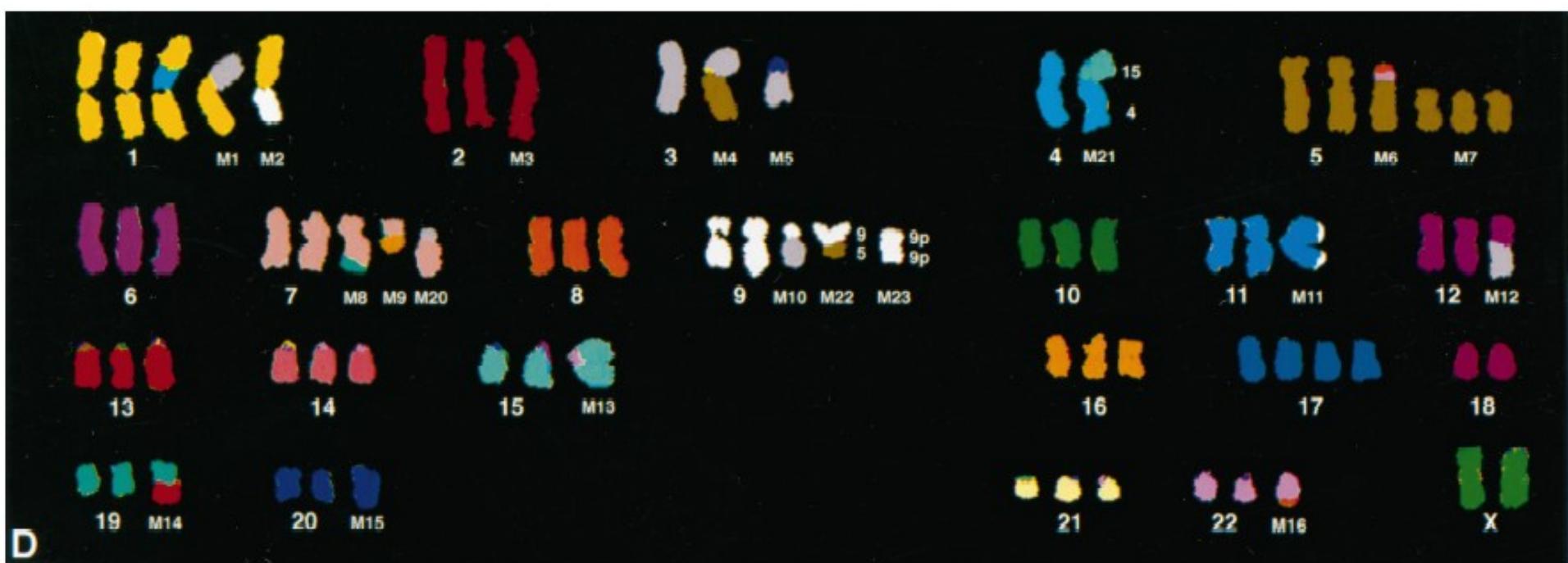
# Tumor cell line biology: HeLa chromosomes are very different from those in normal human cells



“spectral karyotype” (SKY)  
Helps to identify chromosomes

DAPI banding  
(general  
fluorescent stain)

# HeLa: chromosome number and forms are altered



Many chromosomes duplicated or rearranged, & numerous “derivative” chromosomes present

[CANCER RESEARCH 59, 141–150, January 1, 1999]

## Comprehensive and Definitive Molecular Cytogenetic Characterization of HeLa Cells by Spectral Karyotyping

Merryn Macville,<sup>1</sup> Evelin Schröck, Hesed Padilla-Nash, Catherine Keck, B. Michael Ghadimi,<sup>2</sup> Drazen Zimonjic,  
Nicholas Popescu, and Thomas Ried<sup>3</sup>

Genome was sequenced in 2013:

- confirms gene duplications, rearrangement, & mutations
- Altered gene regulation leads to questions re. general relevance of biology to normal human cells

# Stem cells: what they are

- Unspecialized, undifferentiated cells
- “Renewable” through cell divisions, capable of dividing many times
- Can be induced to differentiate into specialized cell types, e.g. cardiac, neural, skin, etc.
- <http://stemcells.nih.gov/>

Three types:

- Embryonic stem (ES) cells: from embryos, pluripotent (giving rise to any cell type), also totipotent? (able to develop into a new individual organism?)
- Adult stem (AS) cells: from adult tissues, multipotent (giving rise to specific cell types)
- Induced pluripotent stem (iPS) cells: somatic cells reverted to an undifferentiated, pluripotent state

# Kinds of stem cells

**Totipotent:** Able to give rise to all the cell types of the body, plus all of the cell types that make up the extra-embryonic tissues such as the placenta. Only zygote → ~4 cell stage cells are totipotent

**Pluripotent:** Able to give rise to all of the various cell types of the body. Pluripotent cells cannot make extra-embryonic tissues such as the amnion, chorion, and other components of the placenta.  
<blastocyst inner cell mass>

**Multipotent:** Able to develop into more than one cell type of the body. <eg: hematopoietic stem cells>

# Why stem cells are important

## Medicine

- ES and iPS cells are pluripotent, and could be used to produce new tissues for “regenerative” medicine
- Cloned ES and iPS cells could be used to generate engineered cells and tissues that could be modified by gene therapy and not rejected by the recipient
- ES- and iPS-derived cells can be induced to grow ‘organoids’

## Basic science

- How do stem cells remain unspecialized in culture?
- What are the signals that cause specialization in stem cells, and how do these signals function?
- Stem cell development could provide models for human tissue development (and developmental disease)

# Where are stem cells found?

- ES cells: from inner cell mass of early (3 to 5 day old) embryo
  - human ES cells first cultured in 1998, using donated embryos (with consent) created for fertility purposes
- Reprogrammed somatic cells
  - ES cells from cloned somatic cells (SCNT)
  - Cell fusion: somatic cell with ES cell
  - induced pluripotent stem (iPS) cells
- AS cells: from adult tissues

# How do you know if you have ES/iPS cells?

- 1) Growth capacity: ES cells are capable of lots of cell divisions in culture without differentiation
- 2) Cell-type “ markers” tell you what kind of a cell you have:  
**Oct-4 protein** expression is high in ES cells but not in differentiated cells
- 3) Chromosomes should be normal: Check the karyotype (many immortalized cell lines are cancer-derived, and often have abnormal karyotypes)
- 4) The cells must be differentiable
  - A) Allow natural differentiation
  - B) Induce differentiation
  - C) Check for teratoma formation in SCID mice  
(Teratoma: tumor containing all three germ layers)  
(SCID: Severe combined immunodeficiency)

# Federal regulations involving human ES cells

1998: The first human ES cell lines were derived.

[www.sciencemag.org/content/282/5391/1145](http://www.sciencemag.org/content/282/5391/1145)

2001: US President (George W. Bush) restricts federally funded research on ES cell lines. Only cell lines made prior to 2001 could be used in federally funded research. Non-federally funded research is exempt from this regulation.

2004: California sets up California Institute for Regenerative Medicine <http://www.cirm.ca.gov/about-cirm/our-history>

2009: US President (Barack Obama) revokes the 2001 Bush policy, allowing federal funding for creation and study of ethically derived, new human ES cell lines.

<https://stemcells.nih.gov/policy/2009-guidelines.htm>

<http://www.nature.com/news/2009/090309/full/458130a.html>

# Alternatives to embryos for “ES-like” cells?

Some adult stem cells are multipotent (& possibly pluripotent?)

- 1) hematopoietic: → develop into all types of blood cells
- 2) bone marrow stromal cells: → bone, cartilage, connective tissue, fat cells
- 3) neural: → brain and nerve cells
- 4) epithelial: → cell types lining digestive tract
- 5) skin: → epidermis, follicles
- 6) germ-line: → sperm, eggs

Some of these stem cell types can do more: brain stem cells have been seen to differentiate into some types of blood or skeletal muscle cells

# AS versus ES cell: important differences

- AS cells are hard to find, generally difficult to isolate, cannot yet be cultured efficiently
- AS cells are generally limited to producing cells from the tissue type that they are found in (multipotent)

-----

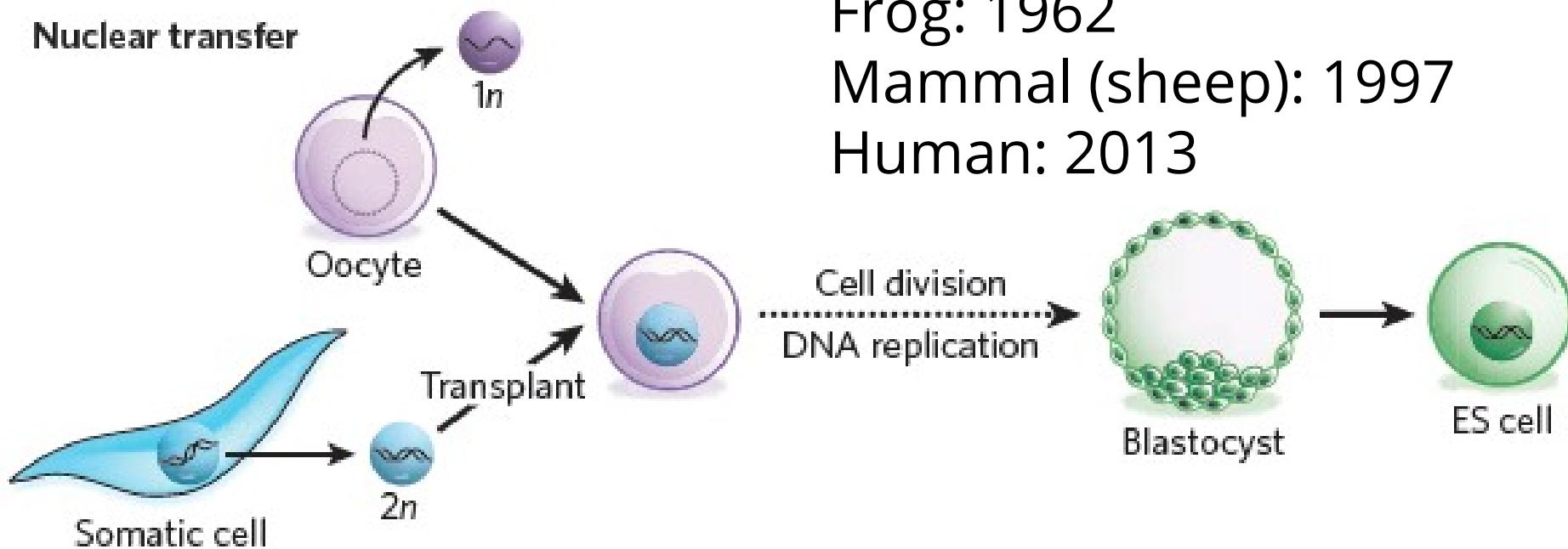
- ES cells divide a lot in culture (easy to manipulate and propagate)
- ES cells are pluripotent

# **Somatic cell nuclear transfer (SCNT): a route to patient-derived ES cells**

- Human egg (donated, ethical issue) has its haploid genome removed, and provides cytoplasm with reprogramming factors
- Adult, differentiated cell provides diploid nucleus
- Diploid egg begins to divide, forming an embryo
- The embryo develops to blastocyst stage. ES cells are taken from the inner cell mass, destroying the clone embryo (ethical issue)
- Could embryo develop to become viable human? (ethical issue)

# Reprogramming for pluripotency: Somatic cell nuclear transfer (SCNT)

## a Nuclear transfer



Frog: 1962

Mammal (sheep): 1997

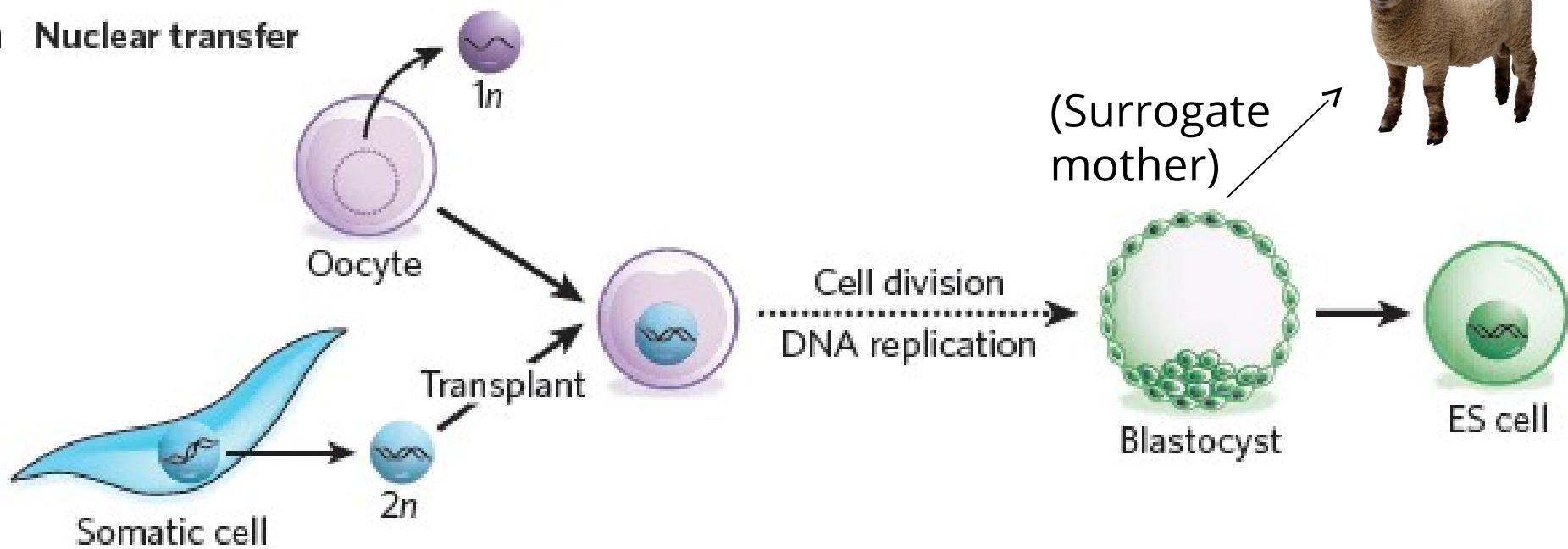
Human: 2013

## Recipient cell:

- can be oocyte ( $n$ ), or fertilized egg ( $2n$ )
- Recipient nucleus contains reprogramming factors
- condensed chromosomes and spindle are removed from recipient cells at a stage of the cell cycle when nuclear membrane is gone (leaves the normally nuclear reprogramming factors for the somatic cell nucleus)

SCNT blastocyst can develop to become cloned organism in some species

a Nuclear transfer



Cloned animals may show abnormalities:

- altered gene expression in embryo
- elongated telomeres
- impaired immune system
- tendency to obesity
- increased cancer rates

Presumably because of failure to erase “epigenetic memory”

2013

# Human Embryonic Stem Cells Derived by Somatic Cell Nuclear Transfer

Masahito Tachibana,<sup>1</sup> Paula Amato,<sup>2</sup> Michelle Sparman,<sup>1</sup> Nuria Marti Gutierrez,<sup>1</sup> Rebecca Tippner-Hedges,<sup>1</sup> Hong Ma,<sup>1</sup> Eunju Kang,<sup>1</sup> Alimujiang Fulati,<sup>1</sup> Hyo-Sang Lee,<sup>1,6</sup> Hathaitip Sritanaudomchai,<sup>3</sup> Keith Masterson,<sup>2</sup> Janine Larson,<sup>2</sup> Deborah Eaton,<sup>2</sup> Karen Sadler-Fredd,<sup>2</sup> David Battaglia,<sup>2</sup> David Lee,<sup>2</sup> Diana Wu,<sup>2</sup> Jeffrey Jensen,<sup>1,4</sup> Phillip Patton,<sup>2</sup> Sumita Gokhale,<sup>5</sup> Richard L. Stouffer,<sup>1,2</sup> Don Wolf,<sup>1</sup> and Shoukhrat Mitalipov<sup>1,2,\*</sup>

<sup>1</sup>Division of Reproductive & Developmental Sciences, Oregon National Primate Research Center, Oregon Health & Science University, 505 NW 185th Avenue, Beaverton, OR 97006, USA

<sup>2</sup>Division of Reproductive Endocrinology, Department of Obstetrics and Gynecology, Oregon Health & Science University, 3181 SW Sam Jackson Park Road, Portland, OR 97239, USA

<sup>3</sup>Department of Oral Biology, Faculty of Dentistry, Mahidol University, Bangkok 10400, Thailand

<sup>4</sup>Women's Health Research Unit, Oregon Health & Science University, 3303 SW Bond Avenue, Portland, OR 79239, USA

<sup>5</sup>Boston University School of Medicine, 72 East Concord Street, Boston, MA 02118, USA

<sup>6</sup>Present address: Laboratory Animal Center, Osong Medical Innovation Foundation, Chungbuk 363-951, Republic of Korea

\*Correspondence: mitalipo@ohsu.edu

<http://dx.doi.org/10.1016/j.cell.2013.05.006>

# Somatic cell nuclear transfer (SCNT) 2013

- Somatic cells (fetal fibroblasts) provided the diploid nuclei
- Donated oocytes had spindle/condensed chromosomes removed in the presence of **caffeine** (helps maintain meiotic arrest)
- Fibroblasts were fused with oocyte
- Cells were activated with electroporation, and then grown to early embryo stage on feeder layer cells
- Inner cell mass removed and cells cultured
- Colonies resembling ES cells were collected for analysis
- Efficiency of ES cell production high (about 1 in 10 SCNT attempts gives an ES cell line)

## Conclusions:

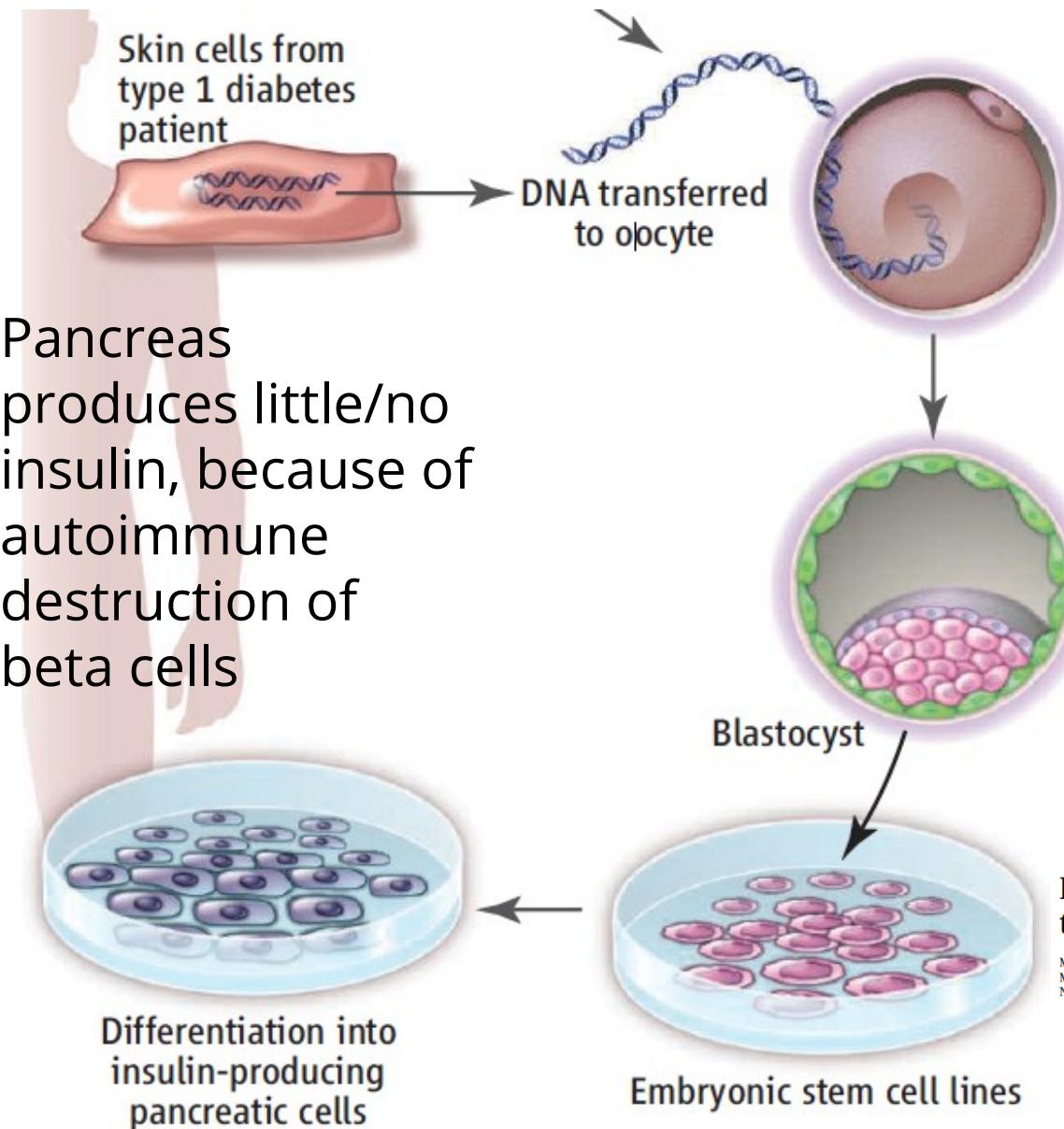
Human ES cells can be derived by SCNT (cloning)  
cells display ES cell markers  
cells can form teratomas  
normal karyotype  
genotypes as expected from the fusion

The process is fairly efficient

\* *How do these cells compare with iPS cells?*

# Applications for human SCNT-derived ES cells

Example: patient-derived therapy for type I diabetes



Adult cells provide the nucleus, allows to make more beta cells

In mice: new beta cells make insulin, respond to glucose levels

How well might it work in humans?

Human oocytes reprogram adult somatic nuclei of a type 1 diabetic to diploid pluripotent stem cells

Mitsutoshi Yamada<sup>1\*</sup>, Bjarki Johannesson<sup>1\*</sup>, Ido Sagi<sup>2</sup>, Lisa Cole Burnett<sup>1</sup>, Daniel H. Kort<sup>4,5</sup>, Robert W. Prosser<sup>4,5</sup>, Daniel Paull<sup>1</sup>, Michael W. Nestor<sup>1</sup>, Matthew Freiby<sup>3</sup>, Ellen Greenberg<sup>3</sup>, Robin S. Goland<sup>3</sup>, Rudolph L. Leibel<sup>3</sup>, Susan L. Solomon<sup>1</sup>, Nissim Benvenisty<sup>2</sup>, Mark V. Sauer<sup>4,5</sup> & Dieter Egli<sup>1</sup>

doi:10.1038/nature13287

# **Ethical issue: what about the oocytes/eggs required for somatic cell nuclear transfer?**

- Availability of human eggs is limited
- Egg donation is a medical procedure
- Should egg donors be paid? How much?
- Can human eggs be produced by animal chimeras?

## Other things to consider:

- SCNT ES cell derivation requires destruction of an embryo (a clone embryo, but still, a theoretically viable human embryo)
- Are cloned ES cells totipotent (giving rise to a whole person)?
- What if someone attempts to clone humans? Lots of ethical problems arise
  - What's the success rate of the process?
  - Who raises the cloned child?
  - Would a cloned person be able live a normal life, or would there be health complications derived from the reprogramming/embryogenesis process?
  - What's the relationship between a clone and its original?
  - ...

# **Another method for reprogramming somatic cells to ES cell-like state**

**“ Induced pluripotent stem cells” (iPSCs)**

In SCNT, mammalian differentiated cells can be reprogrammed to an undifferentiated state by factors present in oocyte, but what are those factors?

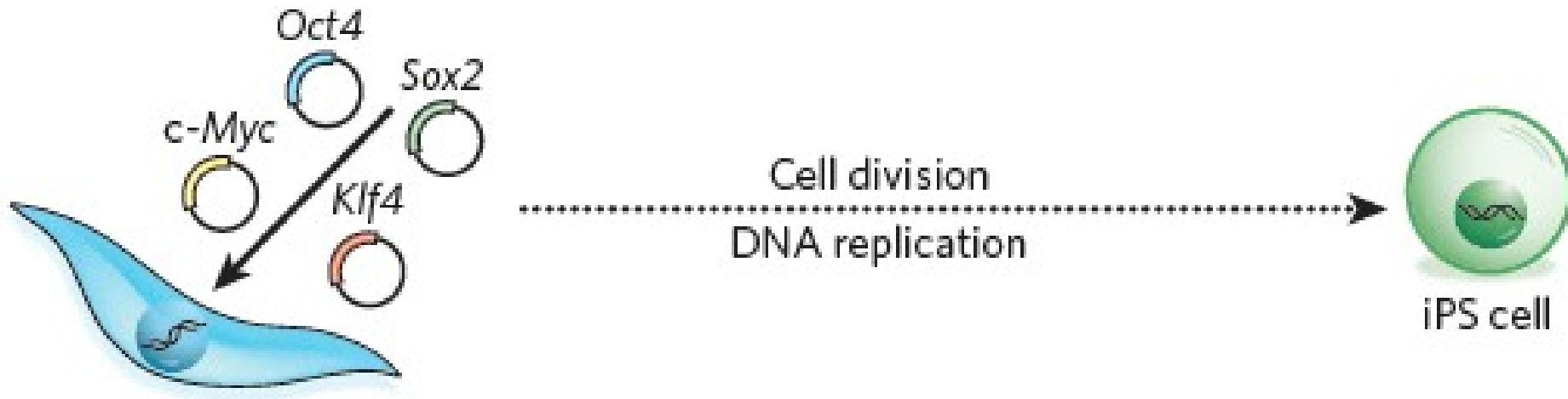
Mouse cells can be converted to iPSCs by the action of four transcription factors:

Oct4  
Sox2  
c-myc  
Klf4

(Takahashi and Yamanaka, Cell. 2006 Aug 25;126(4):663-76.)

# Transcription factor induction of iPS cells from adult somatic cells

## c Transcription-factor transduction



- Four transcription factors: Oct4, Sox2, c-Myc, & Klf4 are sufficient to reprogram to an ES-cell like state
- Other factors can substitute
- Efficiency of reprogramming is low
- There may be a stochastic (random) element to the reprogramming process (in addition to the factors)

## Some issues with iPSCs

- Reprogramming can be inefficient
- Reprogrammed cells can accumulate mutations during the reprogramming process, show signs of premature aging
- Reduced differentiation into some cell types, eg. neural or heart tissue (probably from' epigenetic memory' )
- iPS cells can cause immune response when returned to same mouse from which cells originally came from

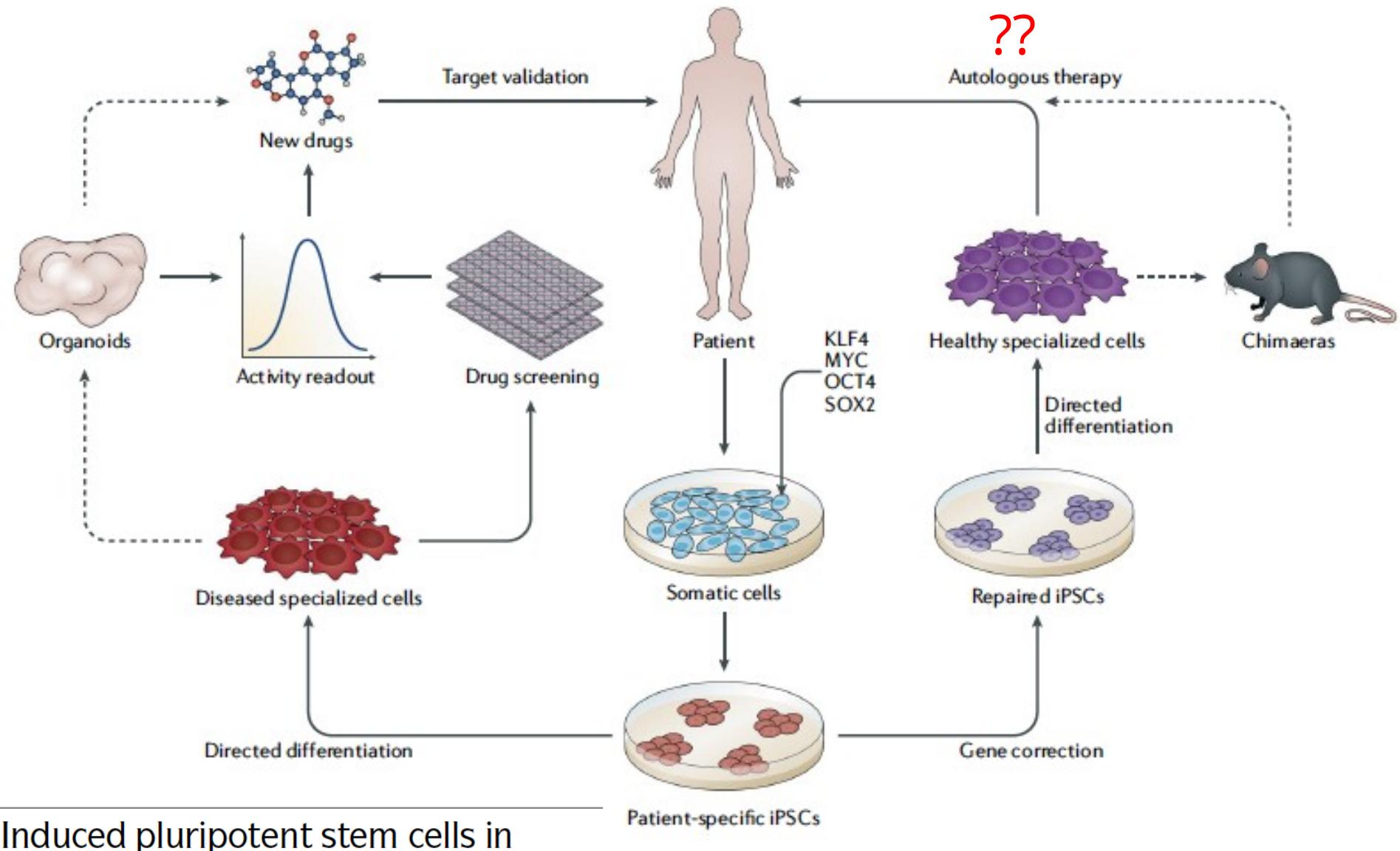
# How do ESCs, NT ESCs, and iPSCs compare?

- Studies have directly compared:
  - Gene expression patterns
  - Epigenetic modifications (methylation)
  - Mutations
- Does it matter if a nucleus is reprogrammed by an egg cell, or whether it is reprogrammed by individual genes?
- Which method is more promising for production of patient-derived stem cells?

## How do ESCs, NT ESCs, and iPSCs compare?

- Johannesson et al. (Egli, also Mitalipov) 2014
  - Cell type used for conversion: adult and neonatal human fibroblasts
  - iPSC conversion method: transient RNA transfection
- The number of methylation, mutation, and gene expression changes is very similar between these iPSCs and NT-ESCs, with both having more than ESCs
- The methylation and gene expression changes vary from one reprogramming to another: therefore the process is naturally error prone
- Neither reprogramming method works better than the other

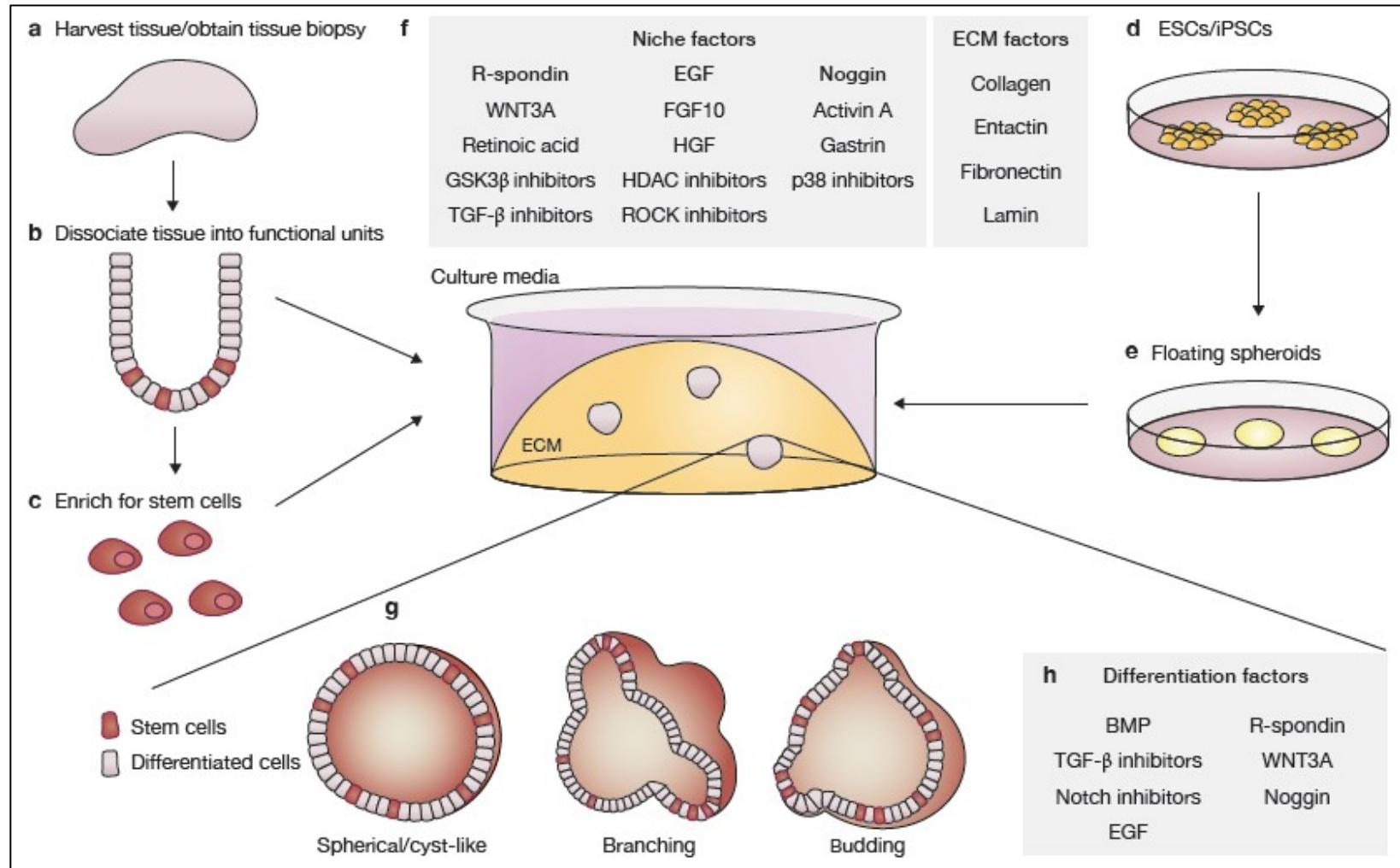
# Development and application of stem cells



Induced pluripotent stem cells in  
disease modelling and drug discovery

# Development and application of stem cells

**Organoids:** *in vitro*-grown tissue clusters derived from primary tissues, or stem cells (ES or iPS)



Organoids as an *in vitro* model of human development and disease

## Stem cells: a cure all?

You have to be cautious...

[https://sciencebasedmedicine.org/?  
s=stem+cell+therapy&category\\_name=&submit=Search](https://sciencebasedmedicine.org/?s=stem+cell+therapy&category_name=&submit=Search)

Stories linked above provide a discussion of stem cell therapies, hype, and unethical or outright fraudulent claims

More good stuff at <https://sciencebasedmedicine.org>

# Summary

- 1) Animal cell culture
- 1) Stem cells: culturable, easy genetic manipulation, and capable of differentiation into other cell types
- 2) ES cells and their derivation by SCNT
- 3) Induced pluripotent stem (iPS) cells and their derivation

# Cloning in eukaryotes: transformation and viral transduction

Transformation and transduction of higher eukaryotes

- 1) DNA transfer methods
- 2) Non-replicative transformation (transient transfection) vs. Stable transformation (recombination)
- 3) Utilizing viral machinery for DNA transduction

## **Guide to readings:**

- 1) 37 *MC4 Mammalian cell transfection*. Short summary of transfection methods.
- 2) 38 *MC4 Transfection selection*. Selective agents used in stable transformation. Also, some info on lipofection and calcium phosphate transfection methods.
- 3) 39 *MC4 Viral transduction*. The use of viral vectors.
- 4) 40 *MC4 Virus vectors*. Specific viral vectors

# Strategies for gene transfer: mammalian cells

- Transfection
  - Biochemical:
    - Cells take up DNA from medium following some kind of chemical treatment
  - Physical:
    - Electroporation
    - Microinjection into nucleus
    - “ gene gun” : particles coated with DNA bombarding cells
- Transduction
  - Virus infection is used to transfer nucleic acids into cells

# Transfection does not require a vector

- Any kind of DNA can be used (linear DNA, plasmids, etc.)
- The DNA may contain selectable markers, eukaryotic replicons, eukaryote-specific gene expression signals
- Bacterial shuttle vectors are often used during the engineering phase

# Transfection of mammalian cells

Chemical:

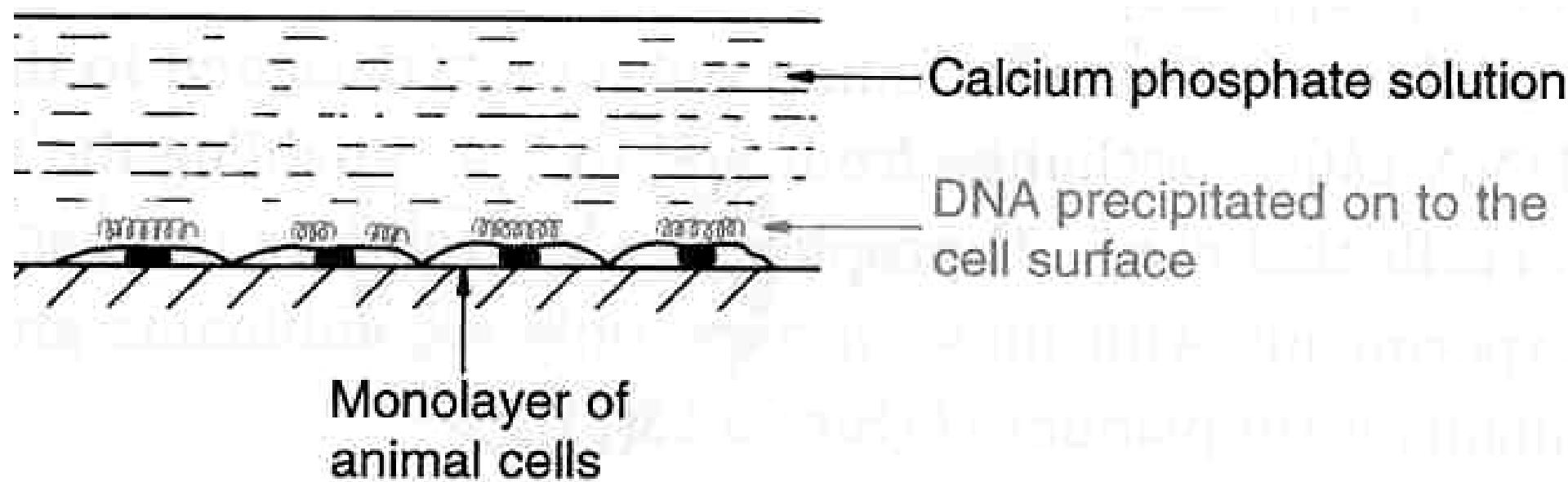
- Calcium phosphate/DNA co-precipitate
- Liposomes containing DNA
- DEAE-dextran/DNA co-precipitate
- Polybrene (a polycation)

Physical:

- electroporation
- biolistics
- microinjection

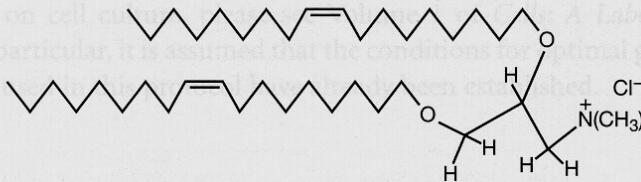
# Transfection of mammalian cells: DNA/Calcium phosphate coprecipitate

- In cell monolayers, up to 20% of cells take up DNA
- endocytosis of the precipitate?

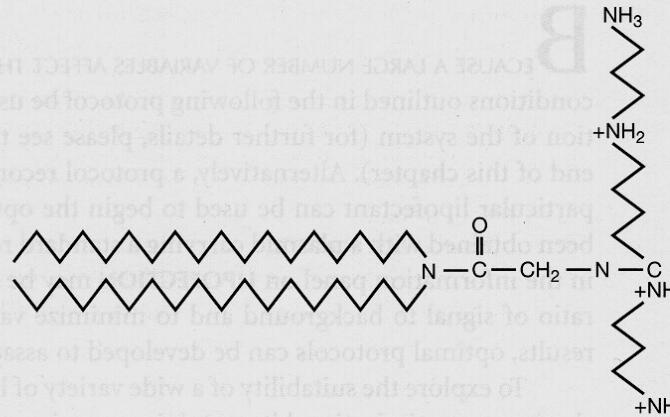


# Lipofection: transfection using liposomes

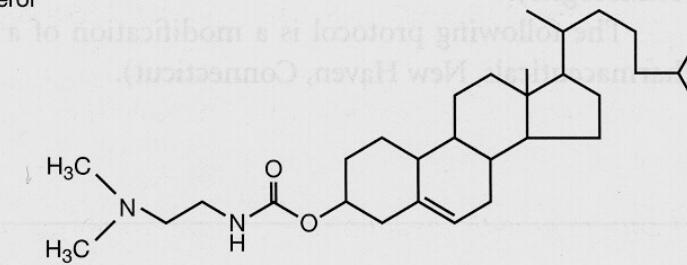
- Cationic/neutral lipid mixtures spontaneously form complexes with DNA
- The lipid vesicles that form are called liposomes
- Liposomes interact with negatively charged cell membranes, and the DNA enters the cell through membrane fusion
- The efficiency is very high: up to 90% of cells in the culture can be transfected



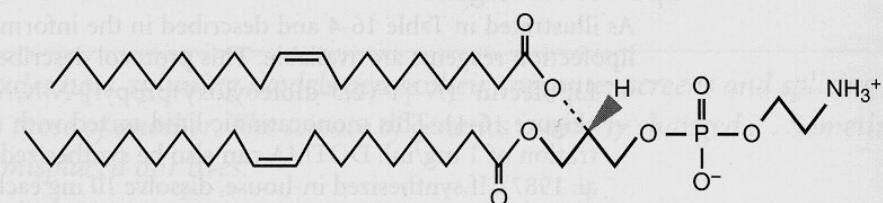
B. DOGS



C. DC-cholesterol



D. DOPE



# Cationic and neutral lipids create artificial membranes that bind to DNA

The lipids interact with cell membranes, & fusion  
delivers the DNA

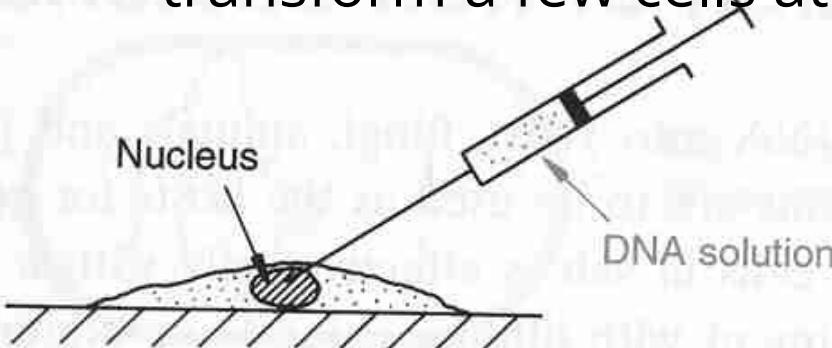
(common example:  
lipofectamine)

# Direct DNA transfer

-- when other, simpler methods are unavailable

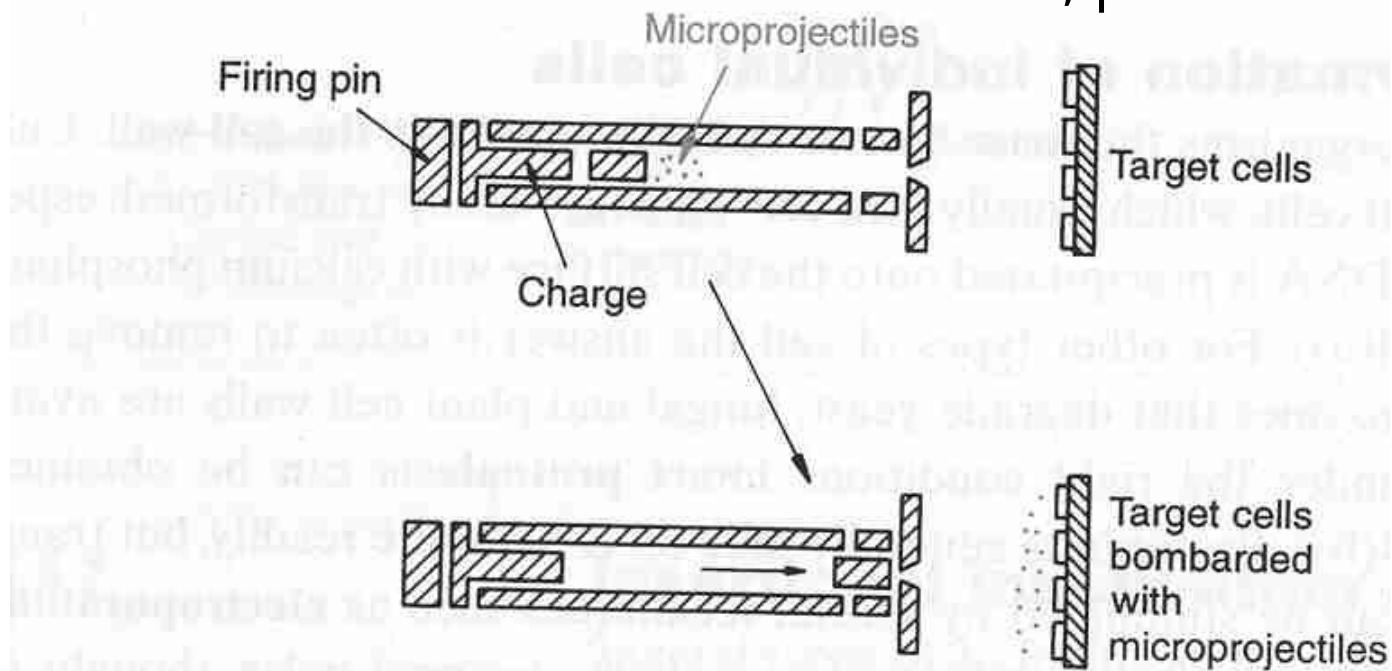
## (a) Microinjection

Need large cells, & can only transform a few cells at a time



## (b) Transformation with microprojectiles

Works well for tissues, plant cells



# Transient transfection

- DNA maintained in nucleus for short time
- Extra-chromosomal, no replicon, no integration
- No selection is required

# How is transient transfection used?

- One-time genetic experiments
  - For example:
    - Transfect with a gene and measure phenotypic effect
    - Examine gene regulation of transfected genetic element
- Testing recombinant DNA prior to time-consuming and difficult cell-line construction -- are genes expressed properly?

# Stable transfection

- DNA integrates into the genome, thus giving stable, heritable transformation
- Homologous recombination is one possible mechanism, but DNA can also randomly integrate
- Results in formation of a “cell line” that carries and expresses the transgene indefinitely
- **Selectable markers** assist in capturing these rare events

# Stable transfection/transformation

- Mechanism of transport of DNA is not known: “ Some DNA” is transported to the nucleus
- Linear DNA works better than circular plasmids, because non-homologous end-joining (NHEJ) pathways may be activated
- Large, concatameric rDNA structure may eventually integrate, usually by non-homologous insertion
- There is no way to control where in the genome the DNA is integrated
- Up to 1 in 1000 transfected cells may carry the transfected gene in a stable fashion

# Viral transduction

- Elements of the viral life cycle can be borrowed
  - attachment to cells, introduction of genetic material
  - Integration into host genome
  - Expression of genes (especially strong promoters)
- Transfer genes to cultured cells or living animals
- Efficiency of transduction makes this an important method for gene therapy

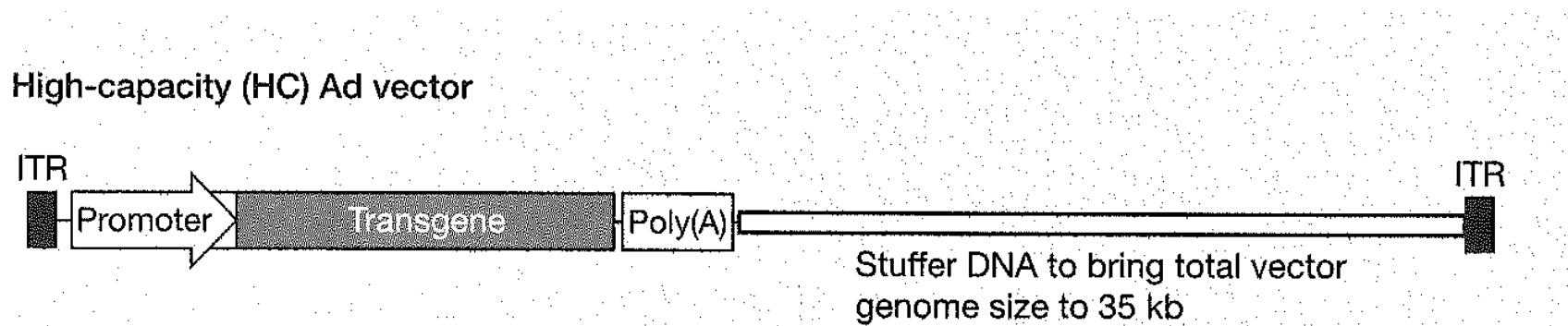
# Issues with a viral approach

- Size of DNA that can be cloned is limited
- Some viruses have limited host range (e.g. retroviruses target only dividing cells)
- Pathogenic virus backbones must have harmful aspects removed/ inactivated
- For gene therapy, patient may have prior immunity to that virus (e.g. adenovirus)

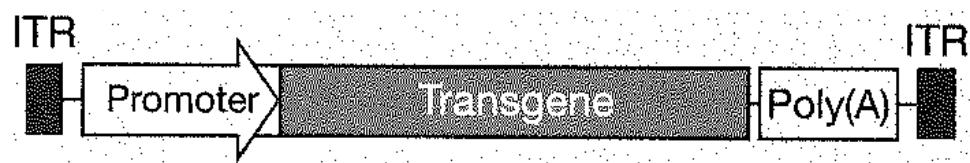
# Viral vectors

- Adenovirus: double stranded DNA, 37 kb
  - Easy to handle, high titer ( $>10^9$  viral particles/ml)
  - Infects both dividing and non-dividing cells
  - “gutless” adenovirus vectors can carry up to 36 kb DNA
- Adeno-associated virus: single stranded DNA, 4.7 kb
  - Small, single stranded DNA genome (4.7 kb)
  - Integrates in host genome, where it's stable for years
  - Infects both dividing and non-dividing cells
  - Non-pathogenic

# Adenovirus vectors



# AAV vectors

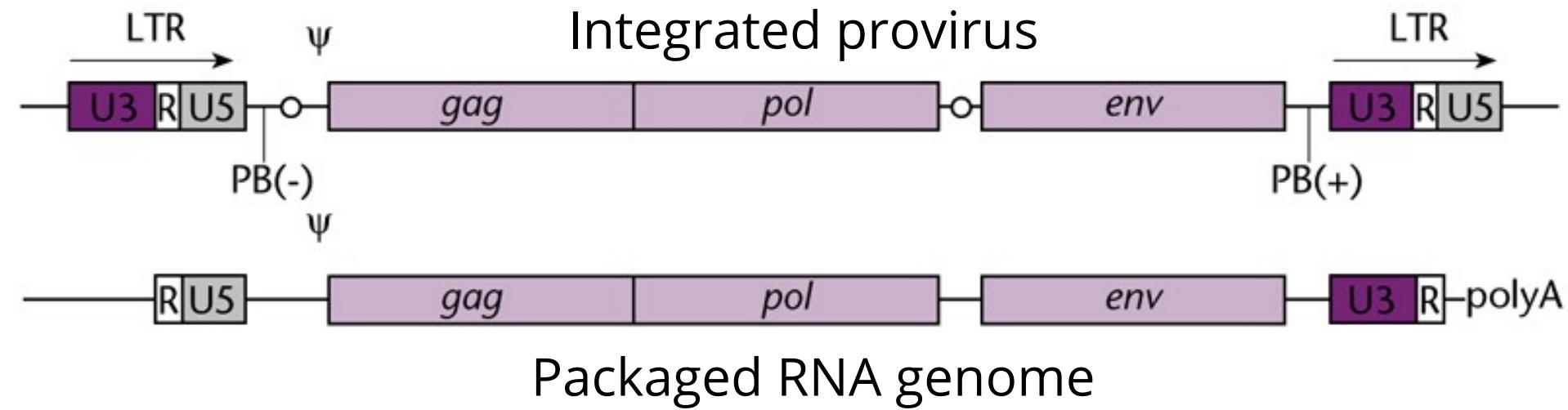


ITR = inverted terminal repeats

# Viral vectors: Retroviridae

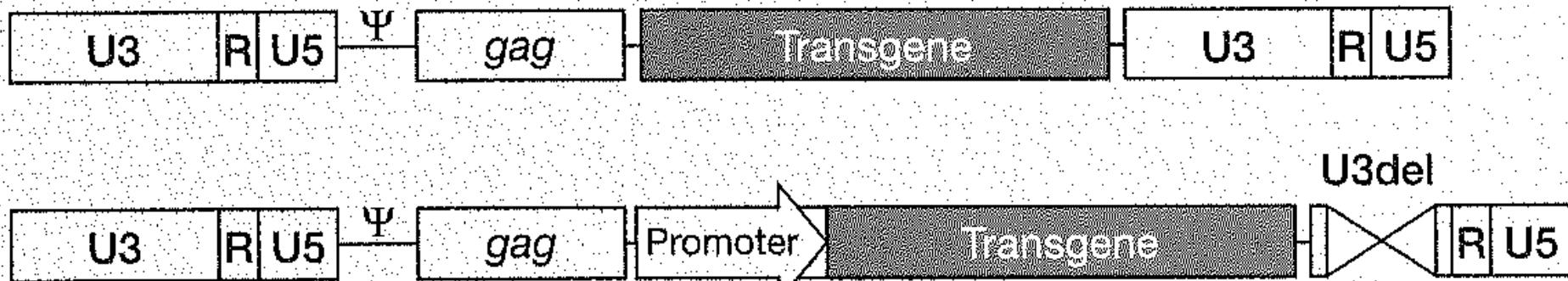
- Single stranded, RNA genome, 8-10 kb
- Replication intermediate is double stranded DNA
- DNA version of viral genome is integrated into the host genome following infection
- Retrovirus:
  - Gag, pol, and env genes
  - Vectors derived from Moloney murine leukemia virus
  - Up to 6.5 kb DNA can be cloned
  - Only targets dividing cells
- Lentivirus:
  - Additional genes: tat, rev, vpr, vpu, nef, vif
  - Vectors derived from HIV-1 and related viruses. Pathogenicity has been engineered out
  - Infects dividing and non-dividing cells

# Moloney Murine Leukemia Virus (MMLV) an oncoretroviral vector



U5: unique 5' region (poly adenylation signal)  
U3: unique 3' region (strong promoter)

# Engineered retroviral vectors



Typical retroviral vector: LTR (promoter), packaging signal (psi), promoter, transgene, LTR

U3 deletion causes 'self-inactivation'

8 kb max transgene size

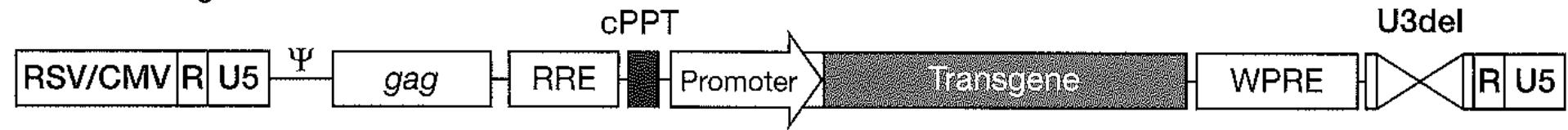
LTRs are retained for:

- transcription activity
- packaging
- insertion into genome

# Essential lentiviral vector features

## Lentivirus vectors

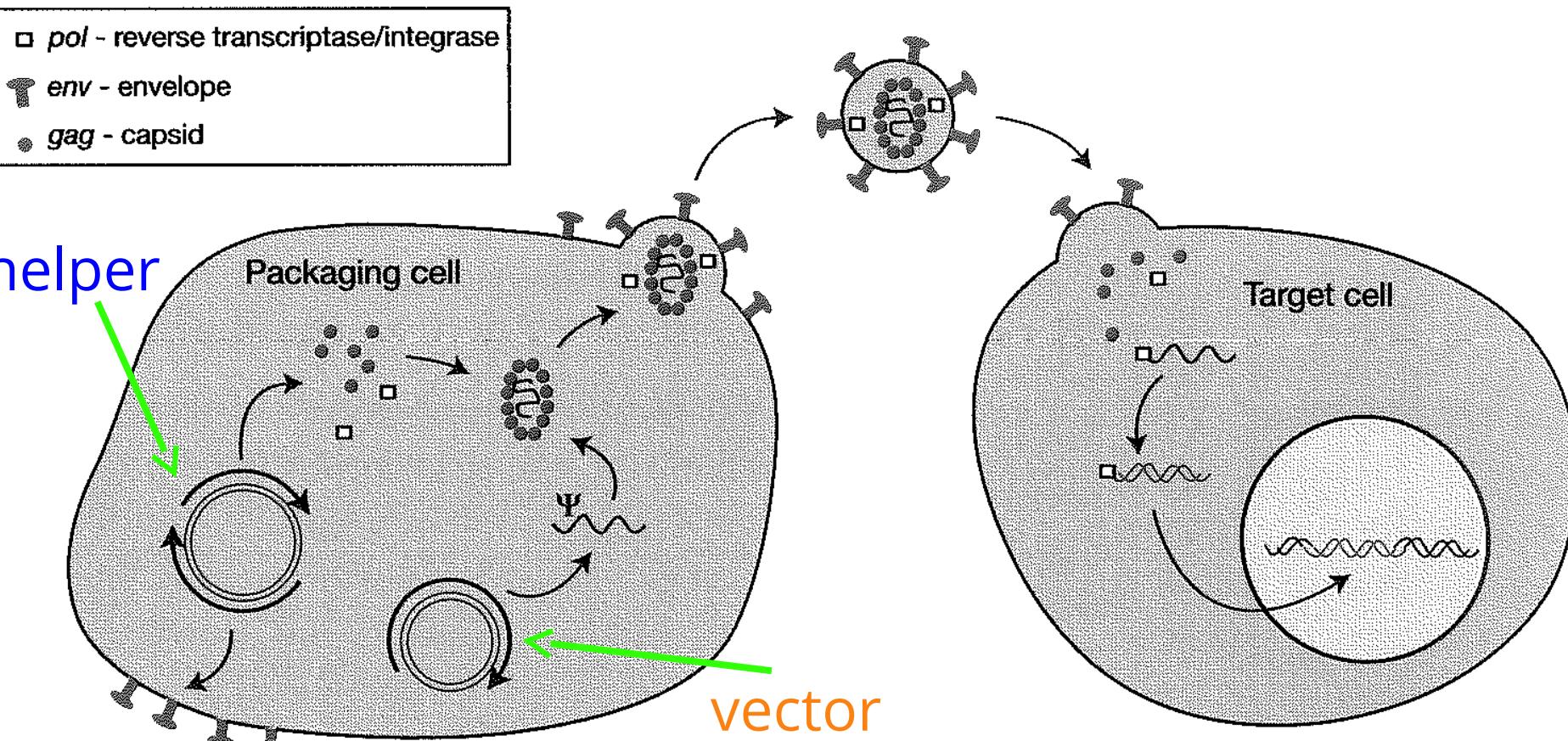
### C Current generation



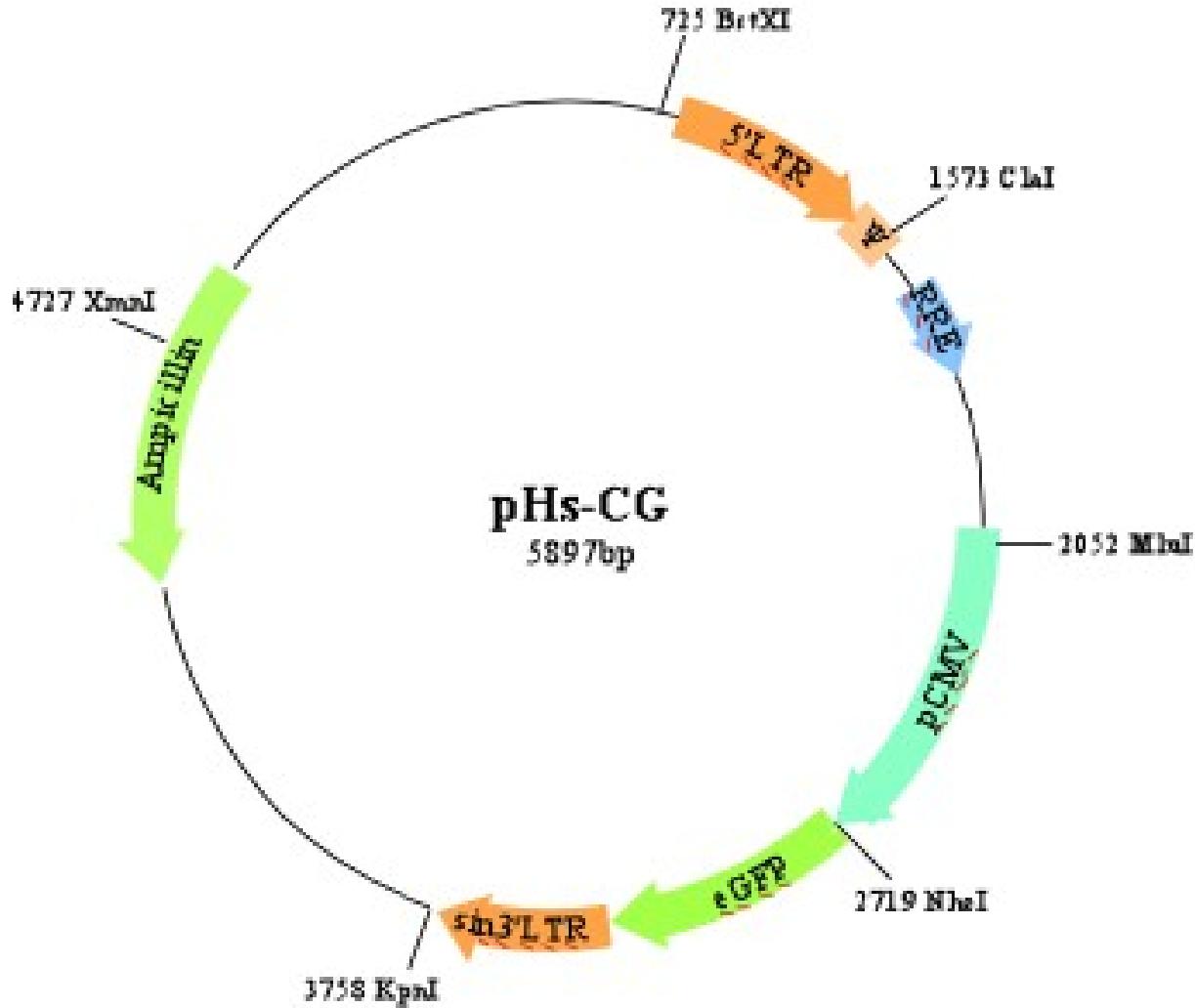
- RSV/CMV: enhancer sequences
- U5, R: Reverse transcriptase sites
- psi: packaging signal
- RRE: rev responsive element, gets RNA to the cytoplasm
- cPPT: poly purine tract for cDNA synthesis
- WPRE: woodchuck post transcriptional regulatory element (increases nuclear export)

# Safety features of lentiviral vectors

- Most viral genes are deleted
- Gag, pol, and env genes are expressed in the production cell, not the vector itself
- Self-inactivating (sin): no functional U3 promoter, prevents formation of replication competent virus



# An example of a lentiviral vector



Gfp marker allows identification of infected cells

Basic HIV-1 based sin vector with the GFP marker driven by pCMV.

[https://web.stanford.edu/group/nolan/\\_OldWebsite/retroviral\\_systems/helix.html](https://web.stanford.edu/group/nolan/_OldWebsite/retroviral_systems/helix.html)

[https://web.stanford.edu/group/nolan/\\_OldWebsite/retroviral\\_systems/retsyst.html](https://web.stanford.edu/group/nolan/_OldWebsite/retroviral_systems/retsyst.html)

# Cloning in eukaryotes

## Transformation and transduction of higher eukaryotes

- 1) DNA transfer methods: biochemical vs. physical treatments
- 2) Non-replicative transformation: (transient transfection) vs. stable transformation (recombination)
- 3) Utilizing viral machinery for DNA transduction: adenovirus, AAV, and retroviruses

# Gene transfer to higher eukaryotes: selection and control

- 1) Positive / negative selection, reporter genes
- 2) Gene targeting by recombination
- 3) Use of cre/lox site-specific recombination
- 4) Controlled gene expression (examples: tetracycline and light)

## Readings guide:

- 1) *Capecchi 2005*. Retrospective by Mario Capecchi, who made key discoveries in mammalian genome engineering
- 2) *Brainbow 2007*. A cre/lox based system for giving color to neurons
- 3) *41 MC4 Reporter genes*. Review of reporters, also overview of the TetR system
- 4) *CRY2 optogenetics 2010*. Control of gene expression w/ light

# Positive selection for transformation (part I)

## “Dominant” selectable markers

- Neo: aminoglycoside phosphotransferase confers survival in presence of aminoglycoside antibiotics, e.g. G418 (similar to neomycin, kanamycin)
- Pac: puromycin N-acetyltransferase, confers resistance to puromycin
- Ble: glycopeptide binding protein, confers resistance to bleomycin, Zeocin
- These selectable markers do not require a specific genotype in the transfected cell-line

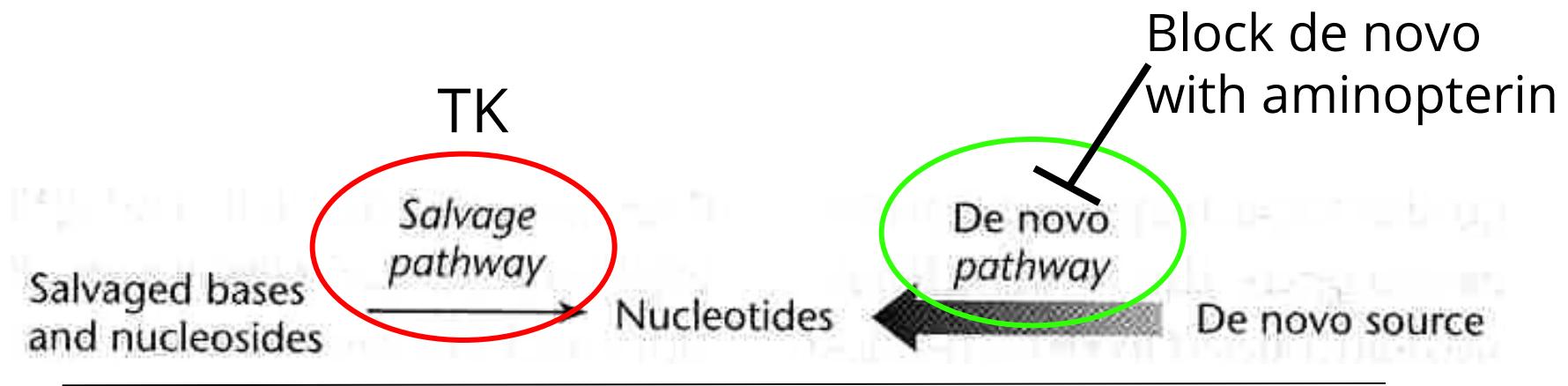
# Positive selection for transformation (Part II)

## Endogenous markers

- Confer a property that is normally present in cells
  - thymidine kinase (TK) (required for salvage pathway of nucleotide biosynthesis)
- These markers may only be used with cell lines that already contain mutations in the marker genes

# Positive selection for transformation:

The Thymidine Kinase (TK) gene functions in the salvage pathway in production of nucleotides



TK is only essential if the cell is forced to use the salvage pathway (the de novo pathway is shut off)

# Counter-selectable markers

You can select AGAINST thymidine kinase (TK):

Add a nucleotide analogue that is TOXIC if the TK protein is present

examples: 5-bromo-deoxyuridine, ganciclovir, AZT

Cells that express TK die in the presence of these compounds, Cells without the Tk gene survive

This allows SELECTION for the loss of a specific piece of DNA (the term is 'counter selection')

Other toxic genes may be used: e.g. diphtheria toxin gene [dipA] is sometimes used in counter-selection)

# Eukaryotic cell transformation

- 1) Getting DNA in: several possible methods
- 2) Transient transformation: no selection
- 3) Stable transformation: selection is required  
(also, counter-selection can be useful)

# Gene targeting (genome engineering)

- Homozygous, null mutants (“ knock-out” ): what is the effect on the organism?
- Exchange of one gene for another (gene “ knock-in” )
- Correction of mutated genes: gene therapy for diseases with a genetic origin

# Gene replacement vectors: homologous recombination

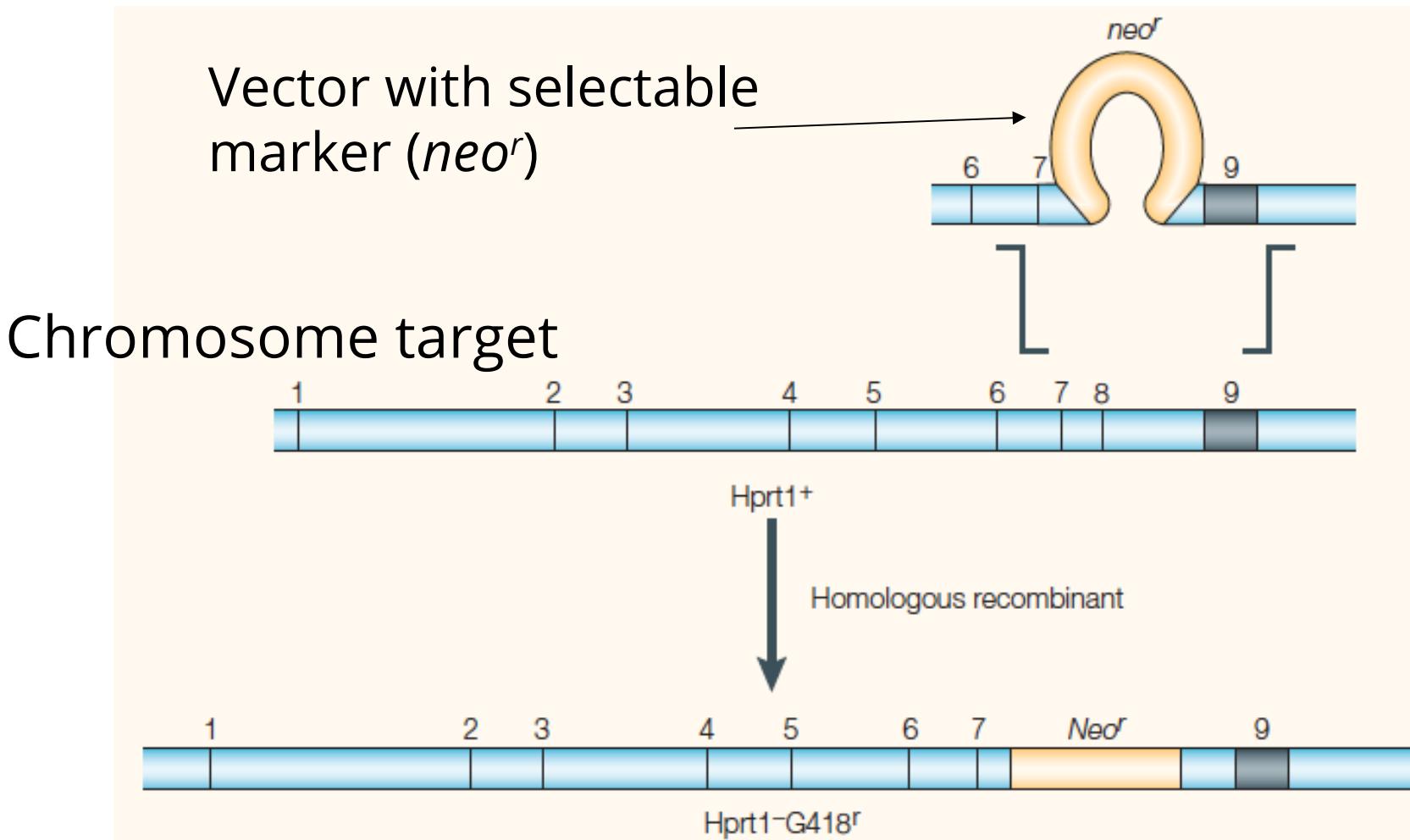


Figure 3 | Disruption of the endogenous hypoxanthine phosphoribosyl transferase gene by gene targeting in embryonic stem cells. The targeting vector contains genomic hypoxanthine phosphoribosyl

HPRT gene disruption strategy

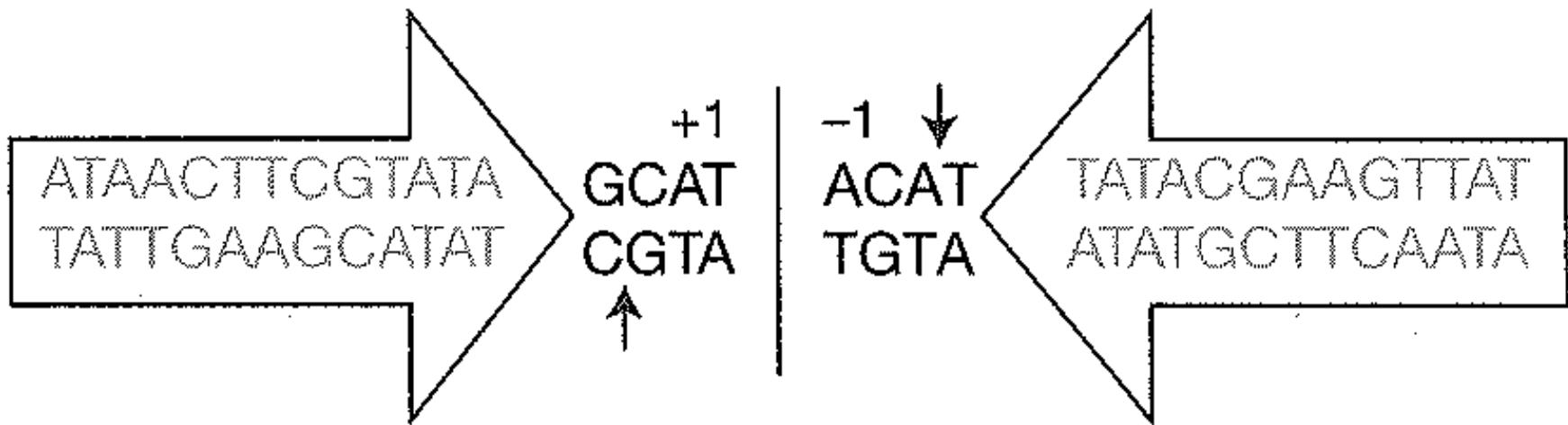
# Homologous recombination

- Ubiquitous process
- Requires regions of homology between recombining DNAs

# Site-specific recombination

- Specialized machinery governs process
- Recombination occurs at short, specific recognition sites

# Cre and Lox: a site-specific recombination system



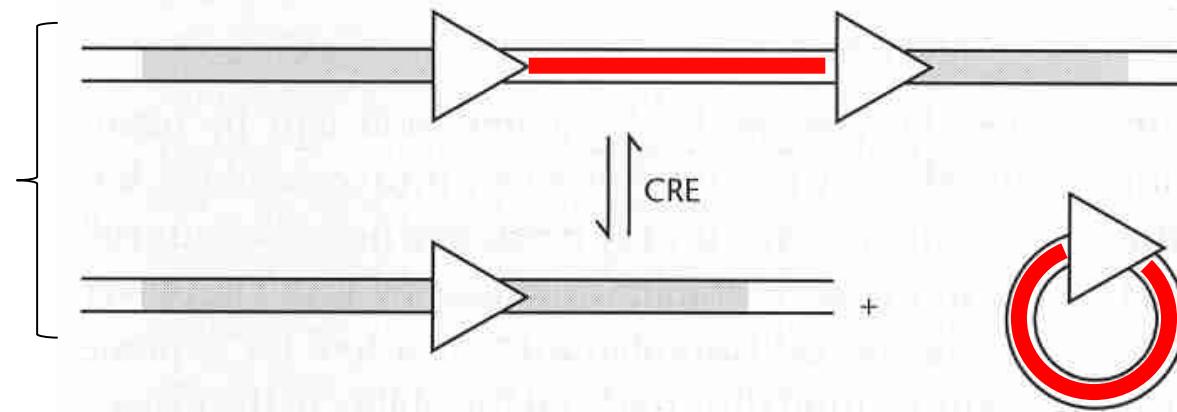
A single loxP site includes two 13 bp inverted repeats separated by an asymmetric linker

# Cre-Lox (site-specific) recombination

- Cre is a protein that catalyzes the recombination process (recombinase)
- LoxP sites: DNA sequences recognized by the Cre recombinase

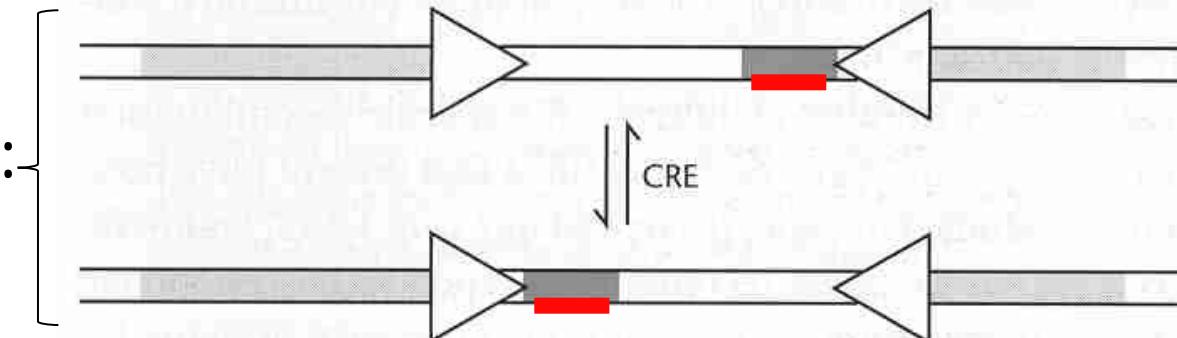
## Direct repeats:

Deletion of intervening sequences



## Inverted repeats:

inversion



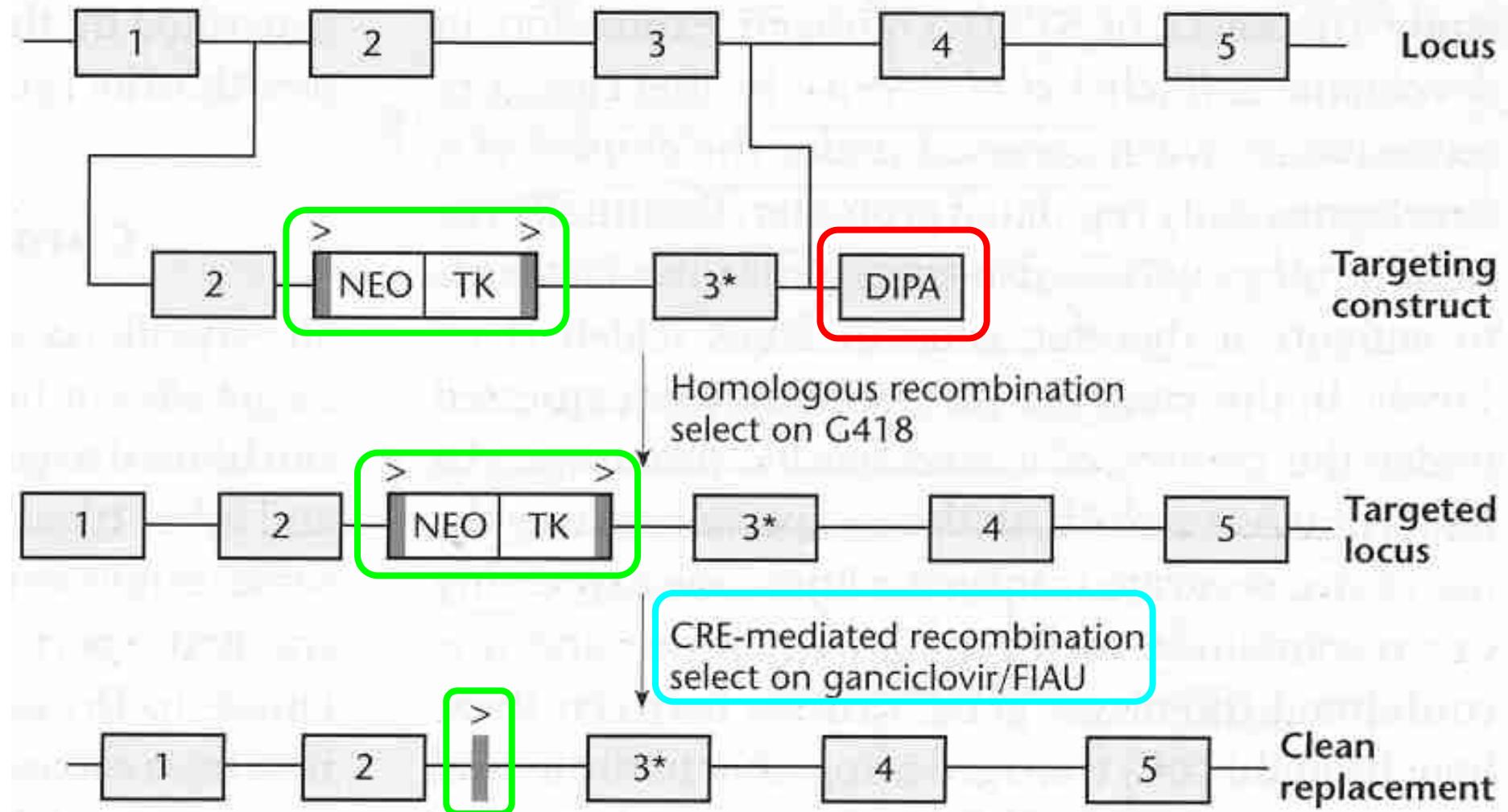
# Cre-mediated conditional deletion

- Surround gene of interest with lox sites (gene is “ floxed” , flanked by lox sites)
- Place Cre gene under inducible control
- Induce Cre at appropriate time or in appropriate place
- Gene of interest can be deleted whenever necessary, e.g. deletions that are lethal in embryo stage can be made once animal has made it to adulthood

# Considerations in homologous recombination strategies

Random insertion of DNA often occurs--how to get around this problem?

- 1) Add a negative selection gene to the DNA outside of the region of homology (ensure that the cells containing this gene via non-specific integration will die)
  
- 2) Screen transformants by PCR for correct position of recombinant DNA insertion

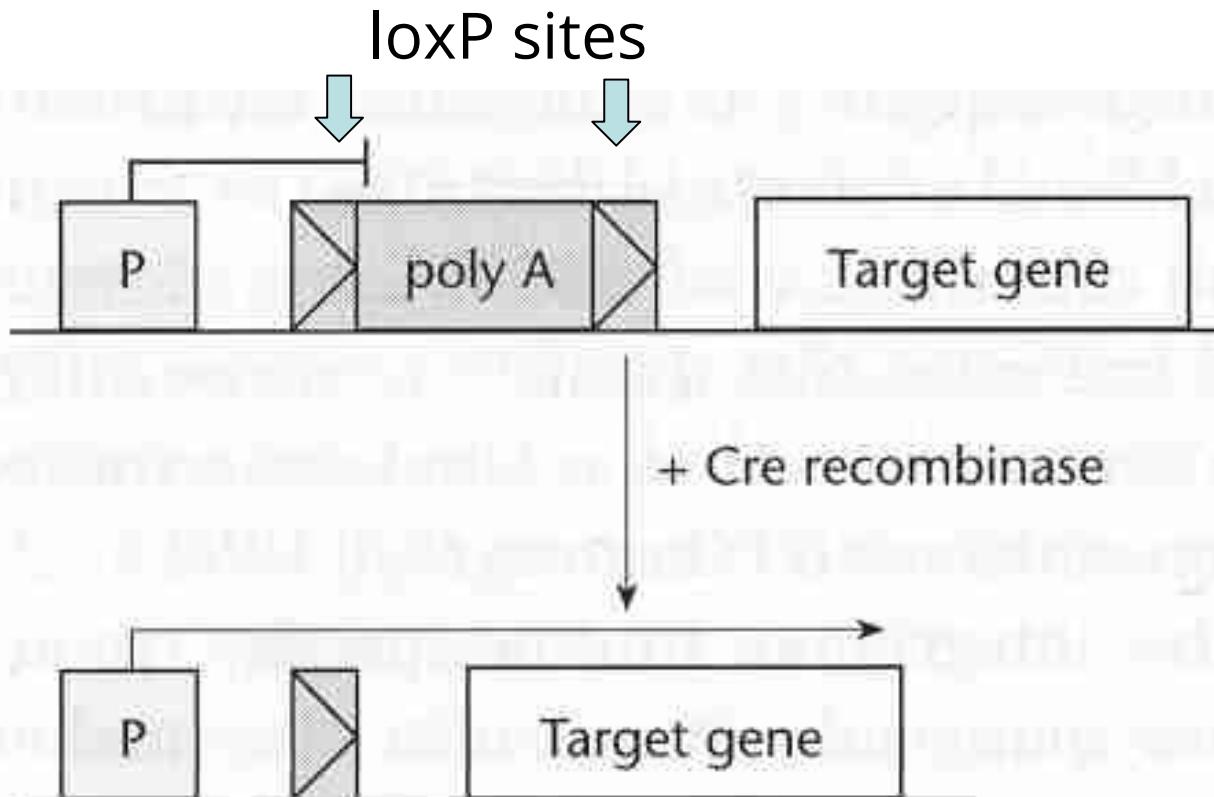


Selection and counter-selection markers flanked by loxP sites

Diphtheria toxin:  
Prevents non-homologous recombination

Cre expression induced by transient transfection

# Cre recombinase-dependent activation of gene expression



Transcription blocked by terminator sequence

Terminator sequence removed, gene is expressed

Conditional expression of Cre defines gene activation

## Another use for cre-lox recombination: Mapping neurons in brains

It can be difficult to track the paths and connection of neurons in brain tissue

This makes it difficult to understand neuronal network architecture

The solution? Give give neurons different colors with various fluorescent proteins

How can different neurons be given different coloration?

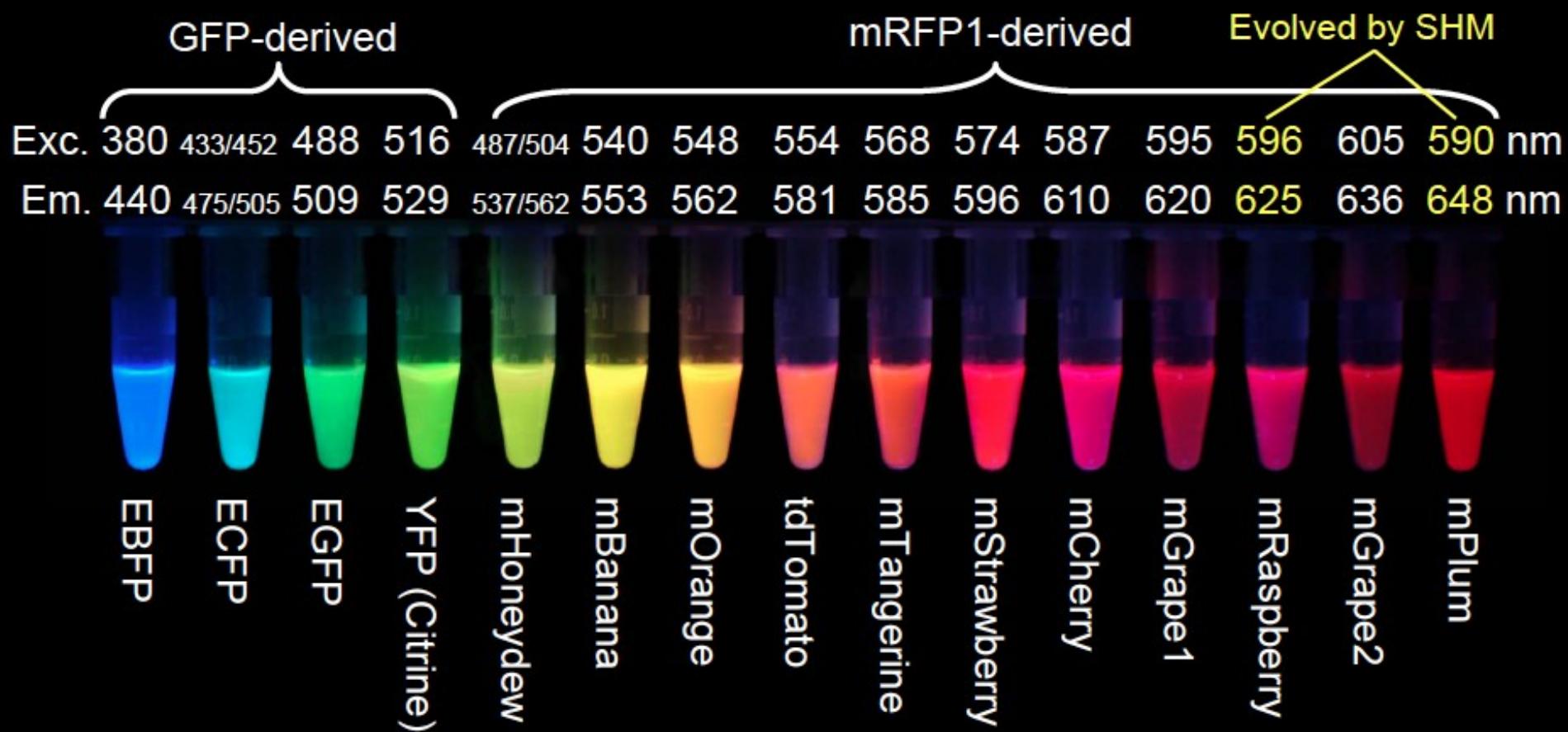
# Reporter genes

- Is the DNA construct present in the transformed cell?
- Is the gene being expressed?
- Where and when is the gene expressed?

Fusion proteins: track position, expression of genes by adding a reporter tag to it

- Common reporters:
  - Fluorescent proteins: detect with UV illumination
  - Beta galactosidase: detect with chromophore
  - Luciferase: emits light

# Wide variety of fluorescent proteins (2004)



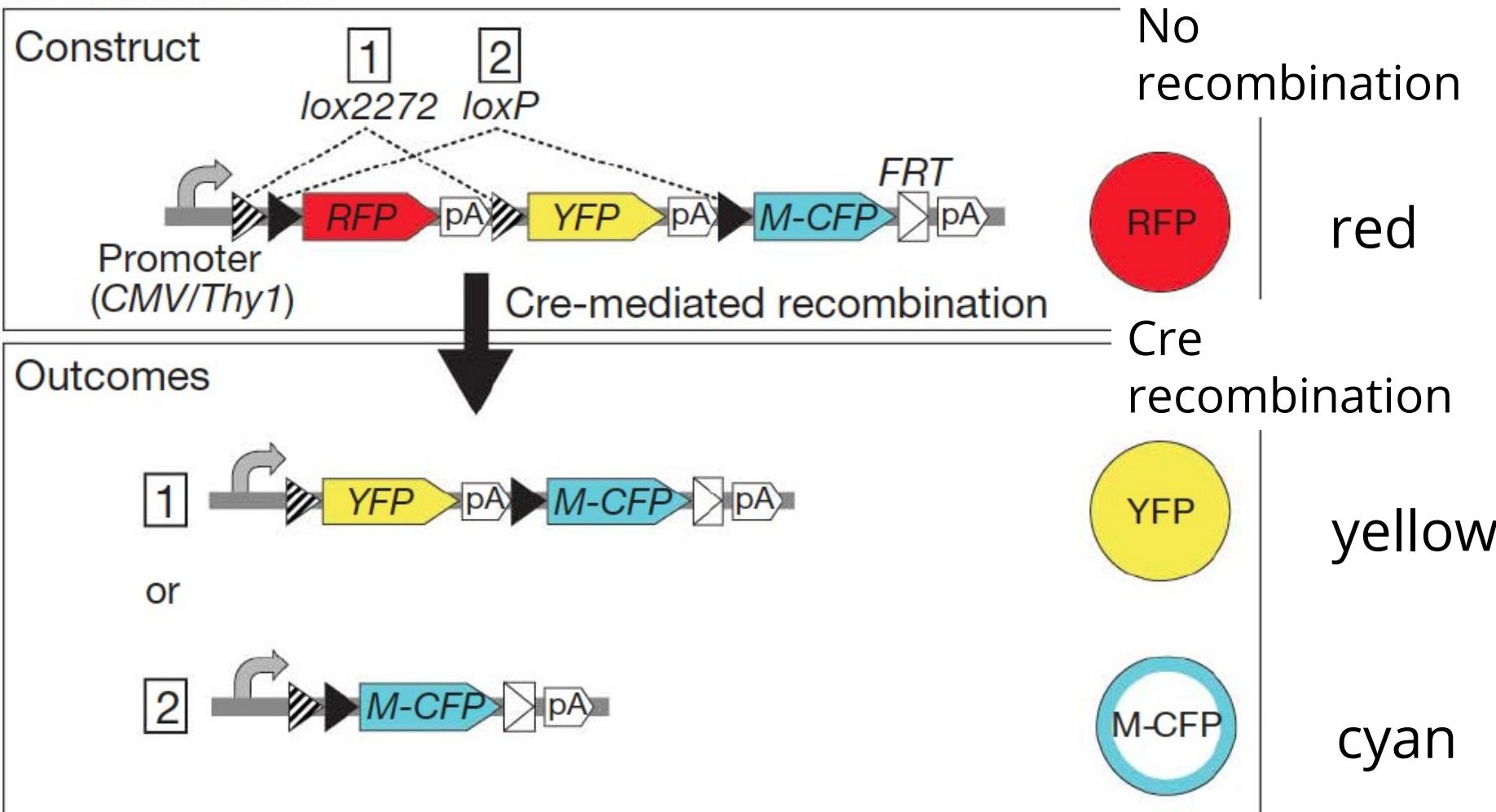
Nathan Shaner et al (2004) *Nature Biotech.* **22**: 1567-1572

Lei Wang et al (2004) *Proc. Natl. Acad. Sci. USA* **101**: 16745-16749

# Cre-Lox is at the heart of the 'brainbow' technique

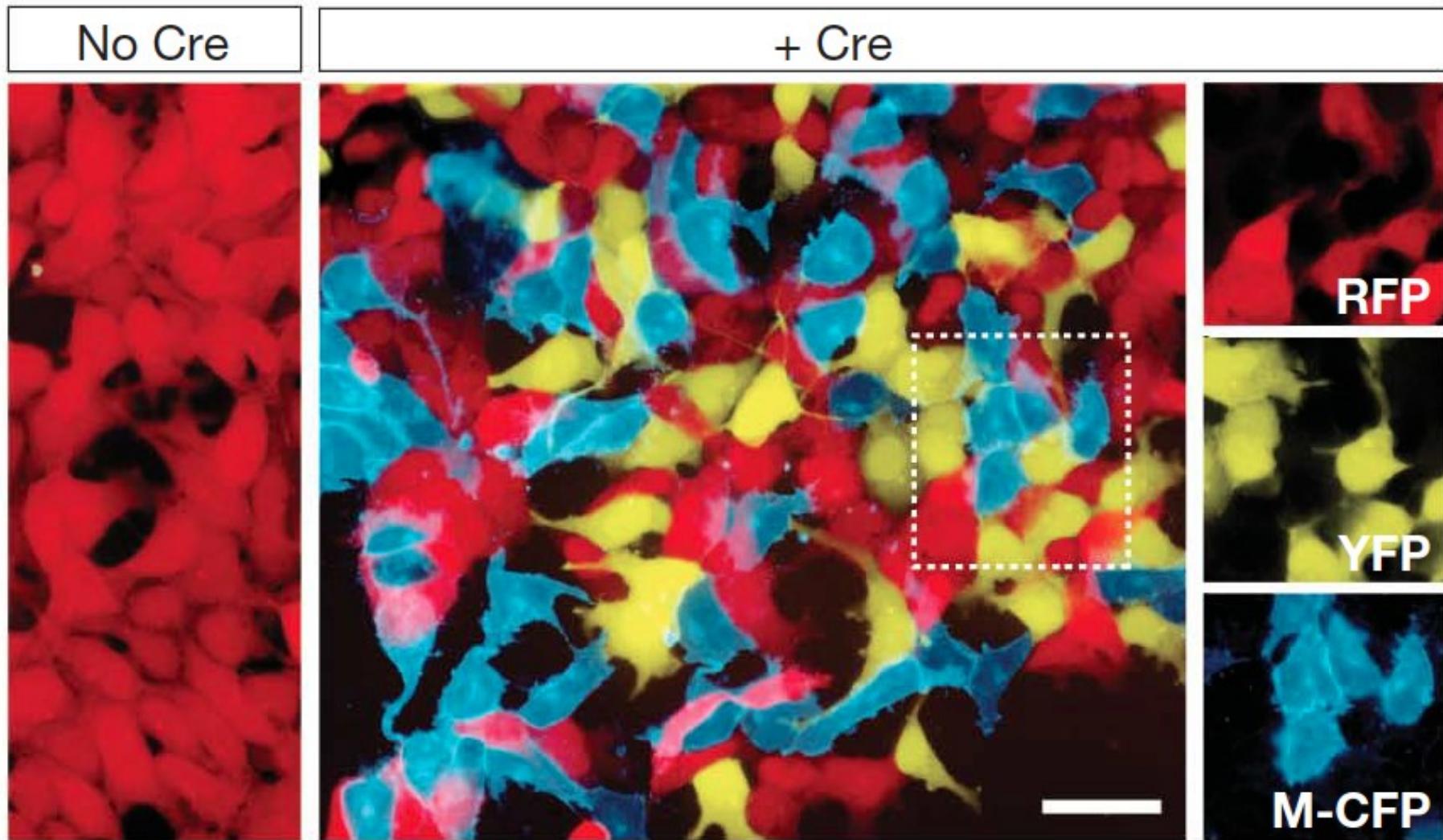
- Two incompatible lox sites (P and 2272)
- Induce cre expression in a cell, one lox site is chosen
- Three possible colors

## a Brainbow-1.0



HEK cells, transformed with DNA construct, with and without Cre induction

**b** Test *in vitro*



# What if there are three constructs in the same cell?

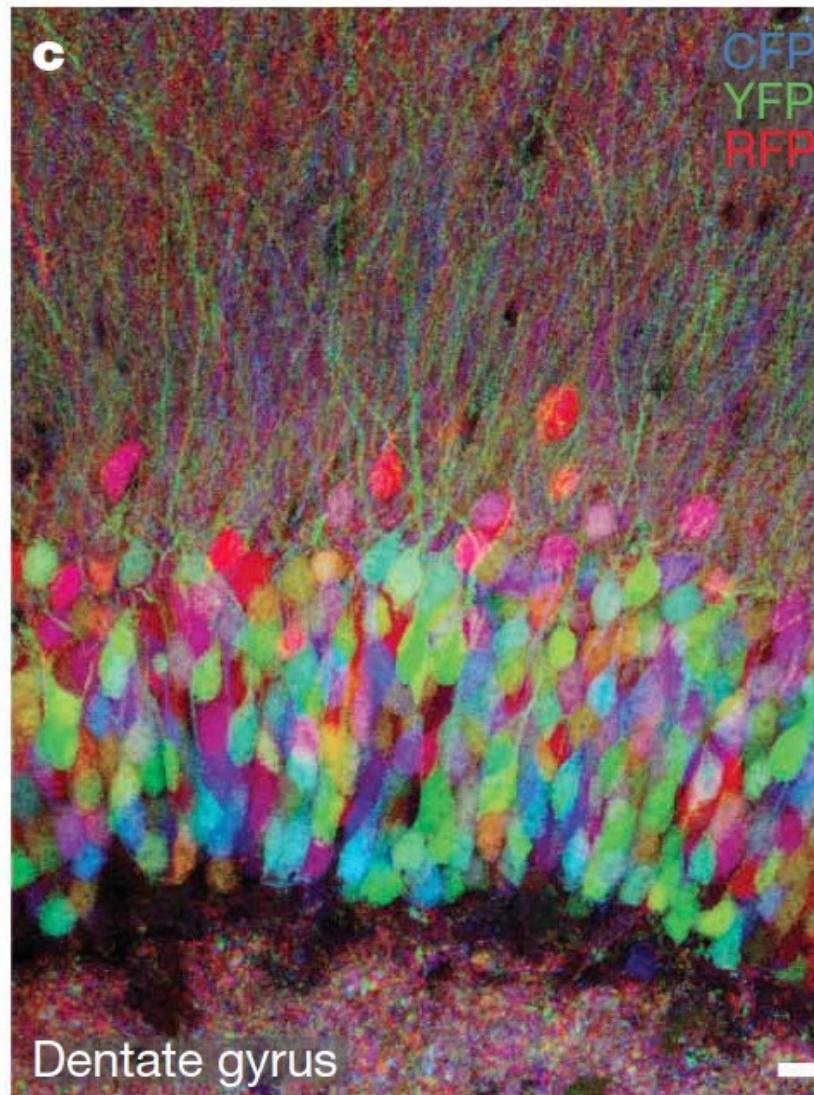
## a XFP combinations

Outcome for Resulting  
each copy colour

1	2	3	
C	C	C	Blue
C	C	Y	Light blue
C	Y	Y	Blue-green
Y	Y	Y	Green
Y	Y	R	Light green
Y	R	R	Orange
R	R	R	Red
R	R	C	Magenta
R	C	C	Purple
R	C	Y	Grey

10 colors?

Brainbow-1 under control of the Thy1 gene (expression in neurons)

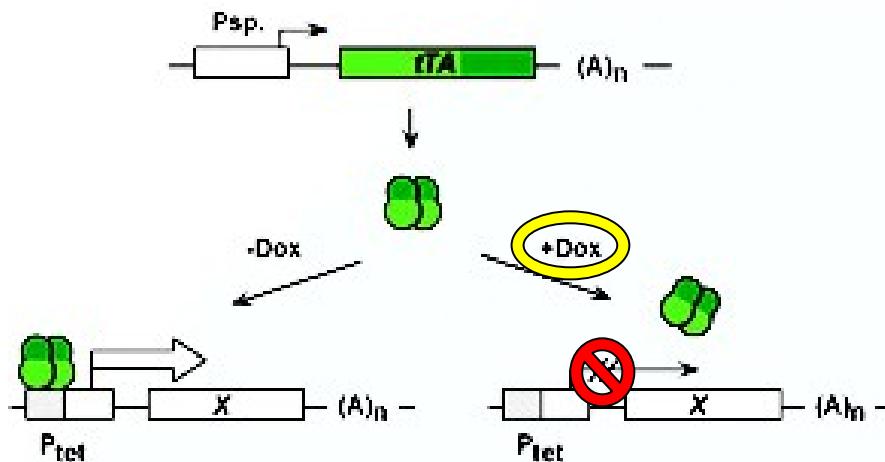


In transgenic mouse, Cre induced by tamoxifen

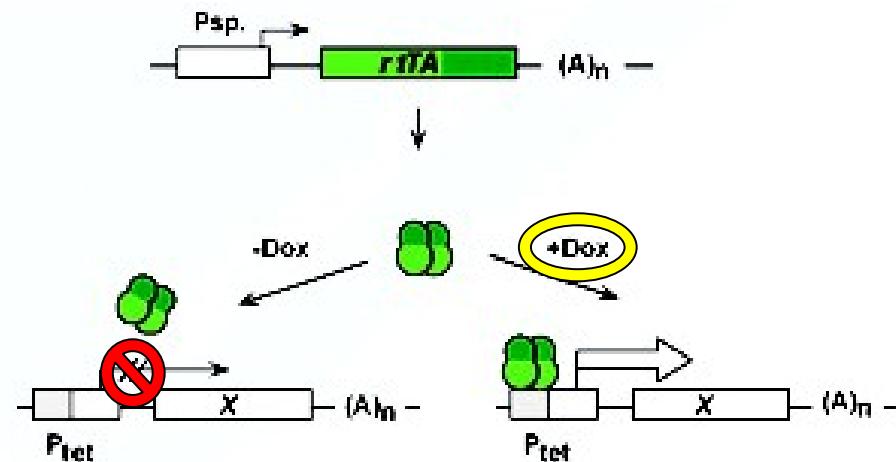
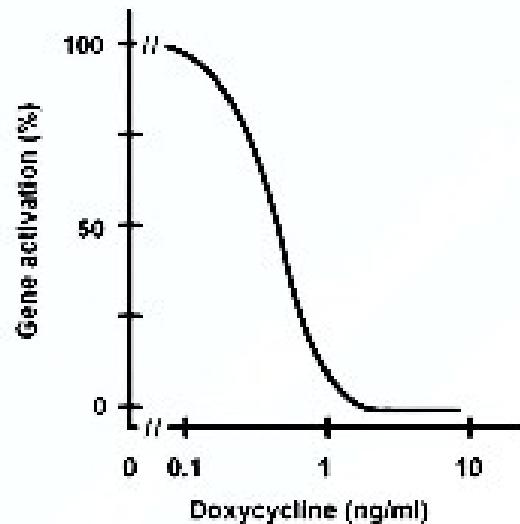
Recombination leads to numerous combinations of fluorescent proteins; about 100 distinct hues detectable

# Control of gene expression: the Tet system

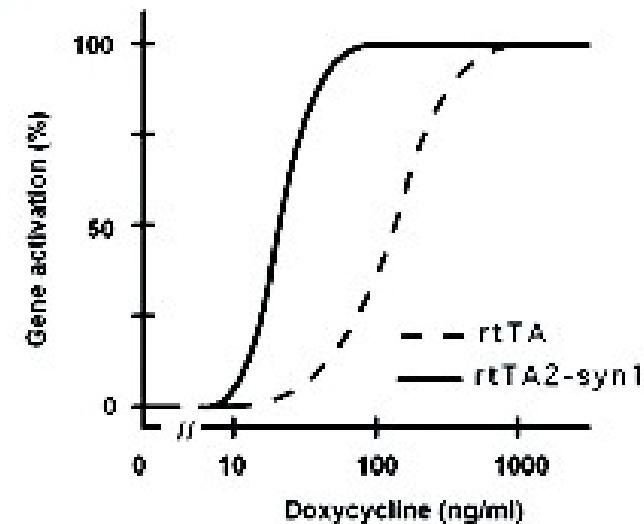
tTA = tet TransActivator: tet repressor fused to the VP16 transactivator. Doxycycline either prevents or allows tTA binding



Induce shut-down



Induce activation



# Inhibiting a gene (**without knock out**)

- Antisense RNA transgenes: synthesize complement to mRNA, prevent expression of that gene
- RNA interference (RNAi): short double-stranded RNAs (siRNAs) silence gene of interest--can be made by transgenes or injected, or by soaking in a solution of dsRNA (*C. elegans*)
- Intracellular antibody inhibition: transgene expresses antibody protein, antibody binds protein of interest, inhibits expression

# Optogenetics: controlling protein activity with light

A number of light-responsive proteins are known, and have been used for controlling cellular processes

Two examples:

- Prokaryotic opsins: in response to illumination, protons or other ions are pumped across a membrane

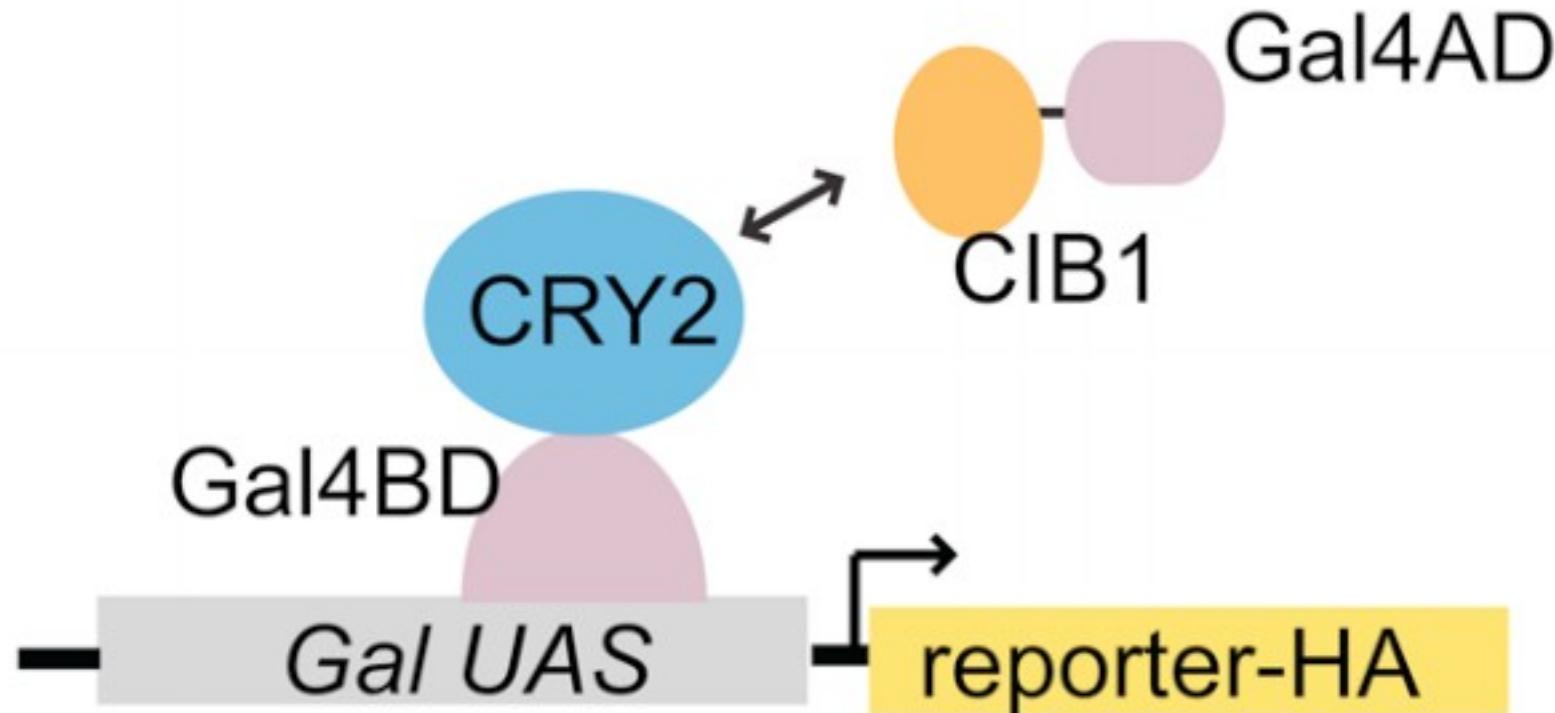
- Allows neurons to be turned on and off with millisecond precision
- <https://www.youtube.com/watch?v=I64X7vHSHOE>

-

- Plant cryptochrome 2 (CRY2) interacts with a partner protein (CIB1) following blue light illumination

# Optogenetics: controlling gene expression with light

Illuminate with blue light, CRY2 interacts with CIB1, and the reporter is expressed



*Nat Methods*. 2010 December ; 7(12): 973–975. doi:10.1038/nmeth.1524.

Rapid blue light induction of protein interactions in living cells

Matthew J. Kennedy<sup>1,4</sup>, Robert M. Hughes<sup>2,4</sup>, Leslie A. Petey<sup>2</sup>, Joel W. Schwartz<sup>1</sup>, Michael D. Ehlers<sup>1,3</sup>, and Chandra L. Tucker<sup>2,\*</sup>

## Summary:

- 1) Selection methods
- 2) Homologous recombination can be used to manipulate the genome with positive selection as well as negative selection
- 3) Site specific recombination (cre/lox) allows controlled removal of specific sequences: deletion of a gene, or activation of some other genetic program
- 4) Controlling gene expression

# Gene therapy and 'genetic surgery'

- 1) Pioneering gene therapy with retroviruses:  
successes and failures
  
- 2) Precision genome engineering: site-specific  
genome cleavage and repair

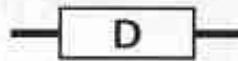
## Guide to readings: Gene therapy

- 1) SCID Gene therapy. The original report of success in SCID treatment (2000)
- 2) Future of gene therapy. Perspectives on early gene therapy attempts (2004)
- 3) Gene therapy state of the art 2011. Review of the field.
- 4) Programmable nucleases 2014

# Gene therapy: treating a disease by adding a new gene

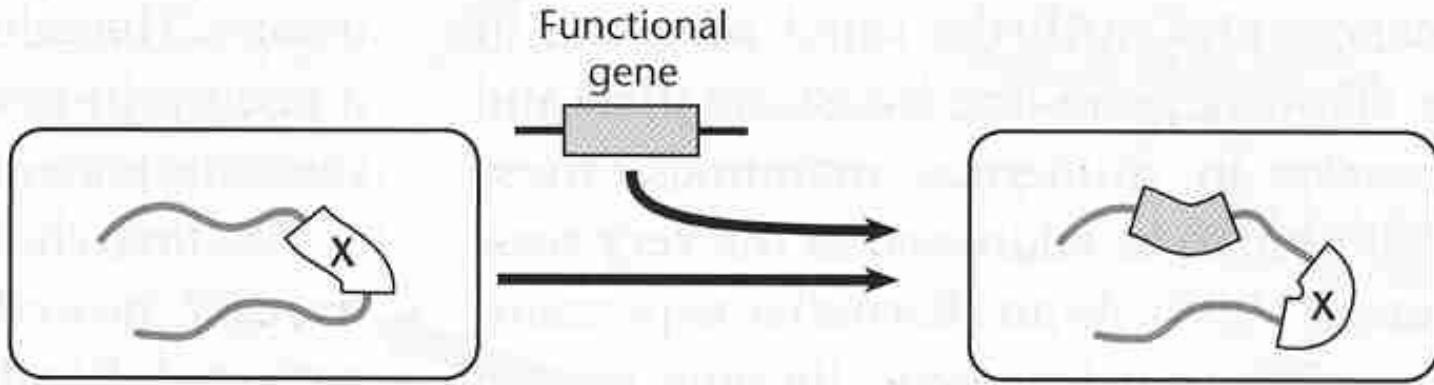


Mutant (disease) gene – loss of function

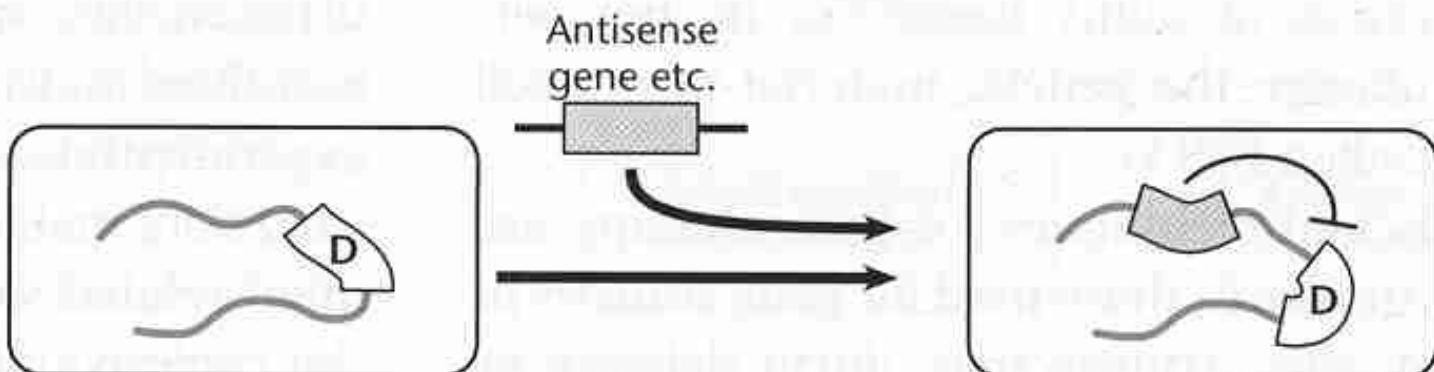


Mutant (disease) gene – dominant gain of function

## 1 Gene augmentation therapy

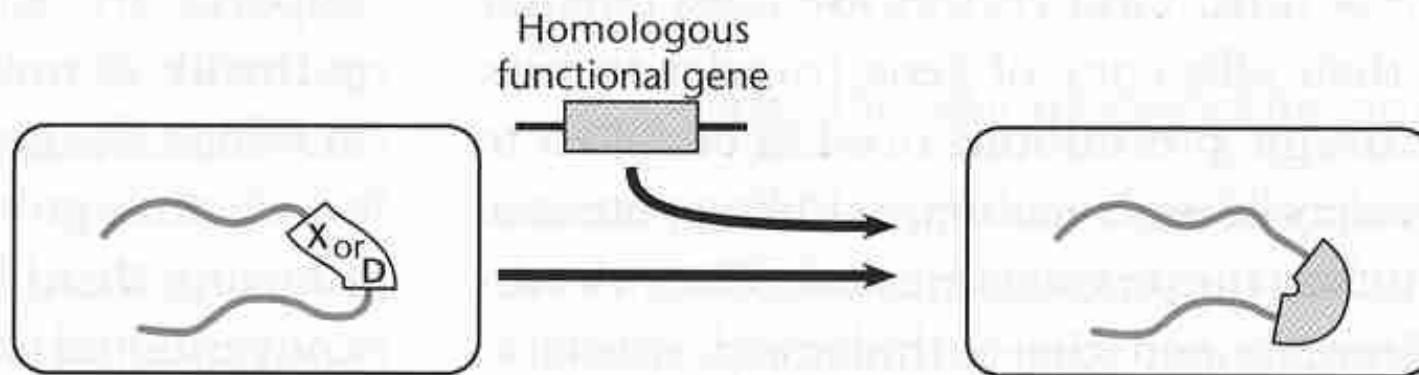


## 2 Gene inhibition therapy

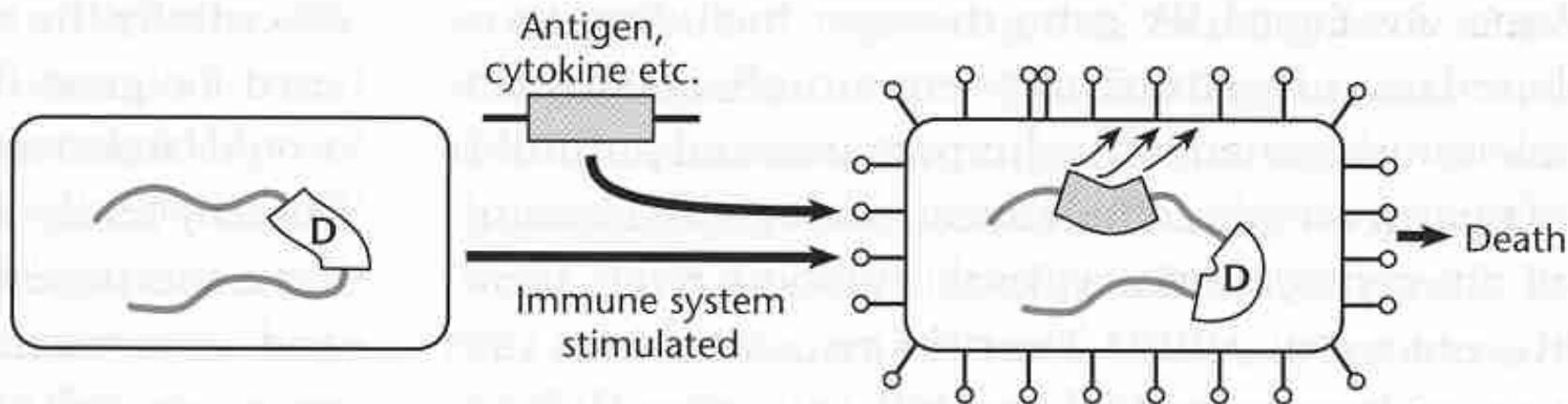


# Gene therapy: treat a disease by adding a new gene

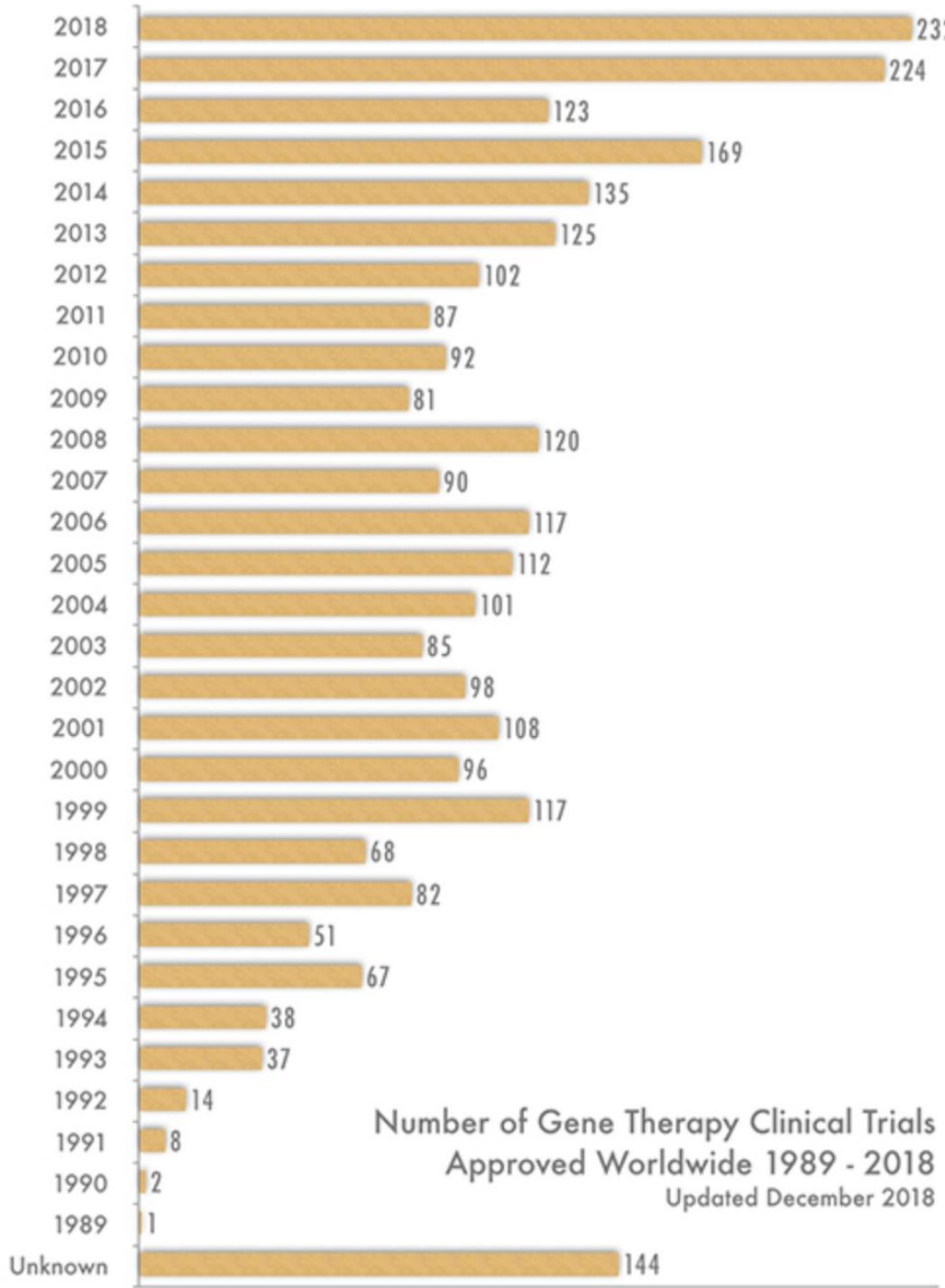
## 3 Gene targeting



## 4 Assisted killing

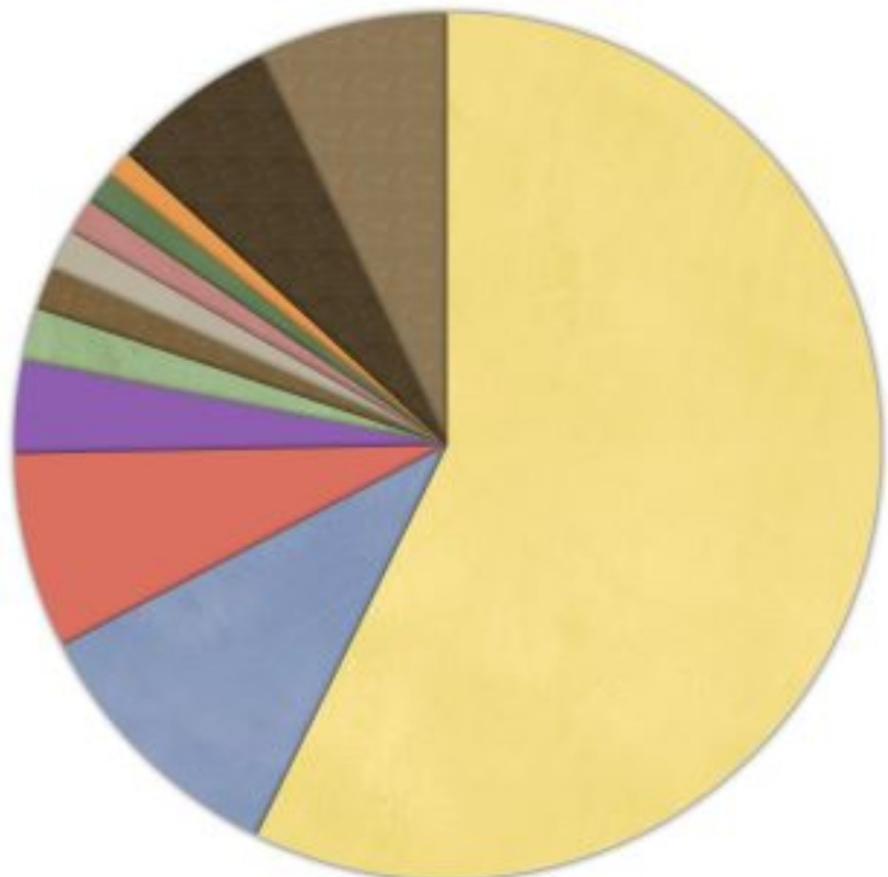


# Gene therapy trials worldwide



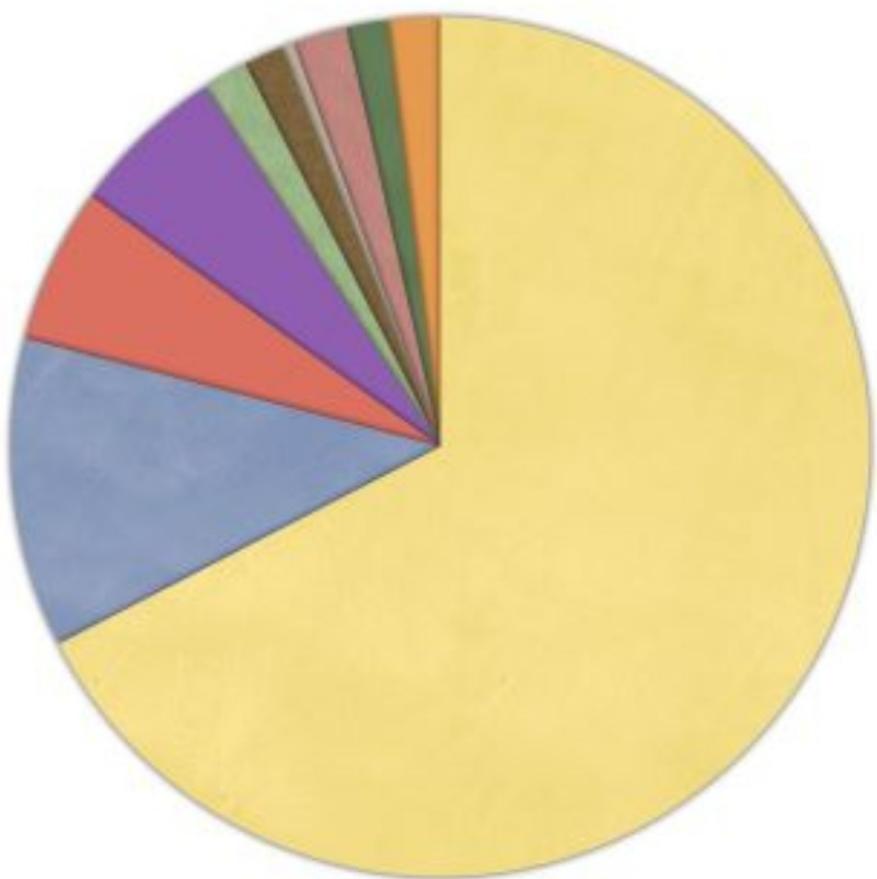
<http://www.abedia.com/wiley/>

## Geographical Distribution of Gene Therapy Clinical Trials By Country



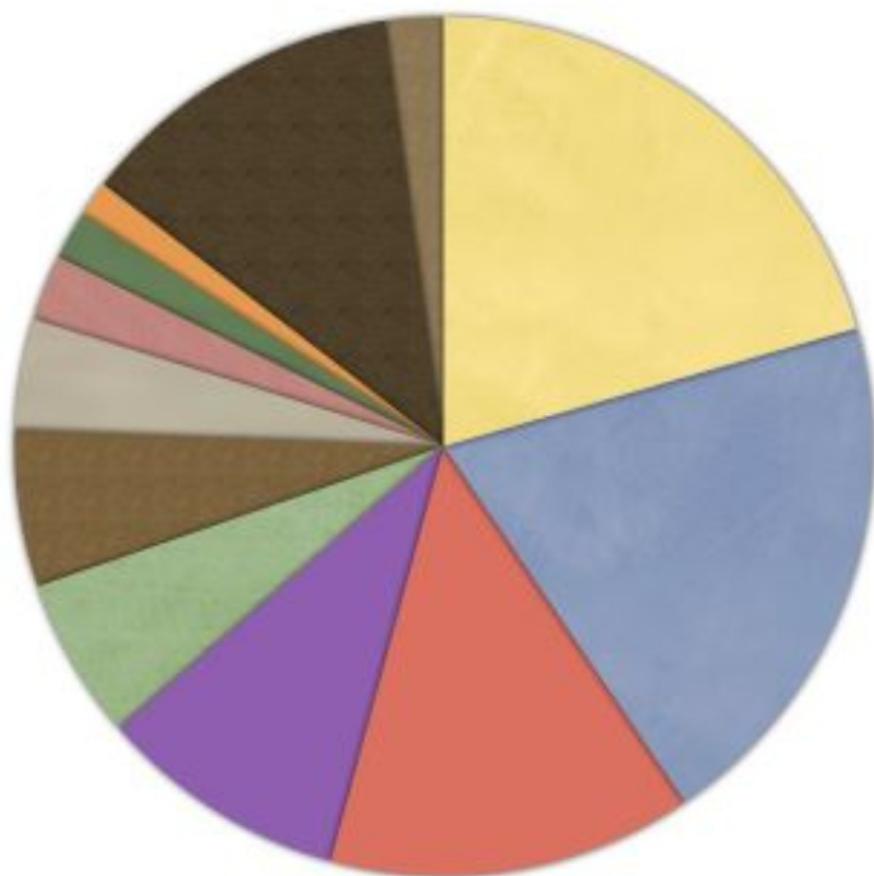
- United States 57.2% (n=1820)
- China 10.1% (n=322)
- United Kingdom 7.3% (n=233)
- Germany 3.5% (n=110)
- France 2.0% (n=63)
- Switzerland 1.6% (n=50)
- Japan 1.5% (n=48)
- Spain 1.3% (n=42)
- Netherlands 1.2% (n=38)
- Australia 1.0% (n=33)
- Other countries 6.1% (n=194)
- Multi-country 7.1% (n=227)

## Indications Addressed by Gene Therapy Clinical Trials



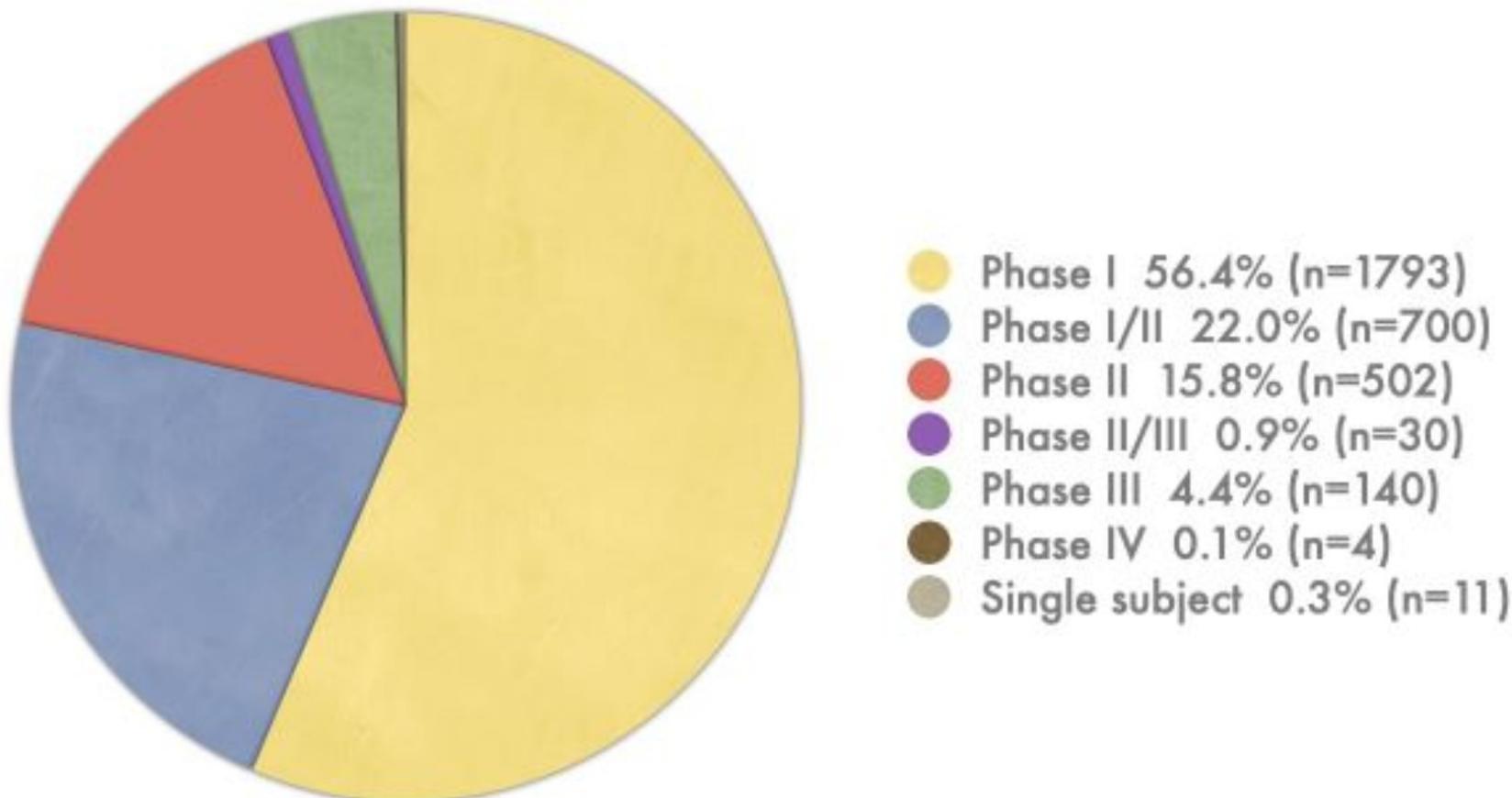
- Cancer diseases 67.4% (n=2144)
- Monogenic diseases 11.6% (n=370)
- Cardiovascular diseases 5.8% (n=186)
- Infectious diseases 5.8% (n=186)
- Neurological diseases 1.7% (n=55)
- Ocular diseases 1.5% (n=47)
- Inflammatory diseases 0.5% (n=15)
- Other diseases 2.0% (n=65)
- Gene marking 1.5% (n=49)
- Healthy volunteers 2.0% (n=63)

## Gene Types Transferred in Gene Therapy Clinical Trials



- Receptor 20.6% (n=654)
- Antigen 19.9% (n=634)
- Cytokine 13.7% (n=437)
- Deficiency 9.4% (n=298)
- Suicide 6.1% (n=194)
- Growth factor 5.9% (n=188)
- Tumor suppressor 4.2% (n=132)
- Oncolytic virus 2.5% (n=79)
- Marker 1.8% (n=58)
- Antisense 1.4% (n=43)
- Other types 12.4% (n=393)
- Not-known 2.2% (n=70)

## Clinical Phases of Gene Therapy Clinical Trials



Phase I: Researchers test a new drug or treatment in a small group of people for the first time to evaluate its safety, determine a safe dosage range, and identify side effects.

Phase II: The drug or treatment is given to a larger group of people to see if it is effective and to further evaluate its safety.

Phase III: The drug or treatment is given to large groups of people to confirm its effectiveness, monitor side effects, compare it to commonly used treatments, and collect information that will allow the drug or treatment to be used safely.

Phase IV: Studies are done after the drug or treatment has been marketed to gather information on the drug's effect in various populations and any side effects associated with long-term use.  
<https://www.nlm.nih.gov/services/ctphases.html>

# Gene therapy vectors

Viral: Most widely used, the viral infection and integration/replication modes bring high efficiency of transformation

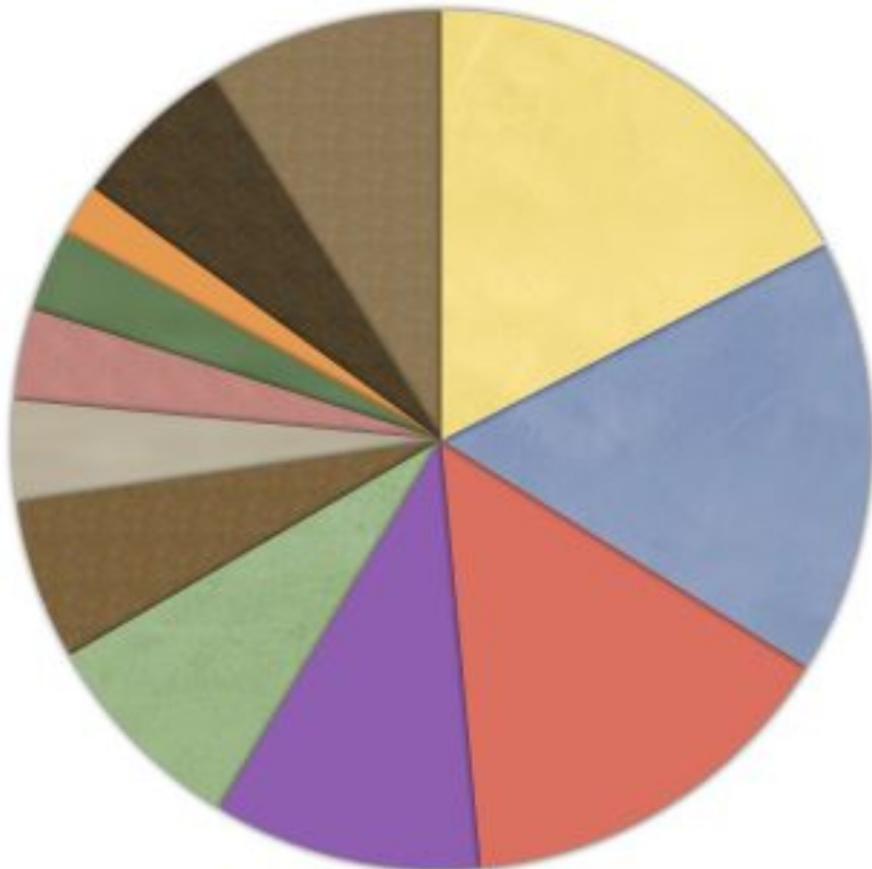
- ★ Retrovirus (actively dividing cells, semi-random integ.)
- ★ Lentivirus (all cells, semi-random integration)
- ★ Adenovirus (relatively toxic, replicates episomally)
- ★ Adeno-associated virus (easy to purify, infects all cells, but DNA insert size limited to 4 kb)

Non-viral: inefficient, because DNA must be delivered somehow (e.g. liposomes), Mainly *ex vivo* because of their low efficiency. Immune responses less of a problem, though

- ★ Plasmids
- ★ Minicircles (lacking unwanted plasmid DNA)

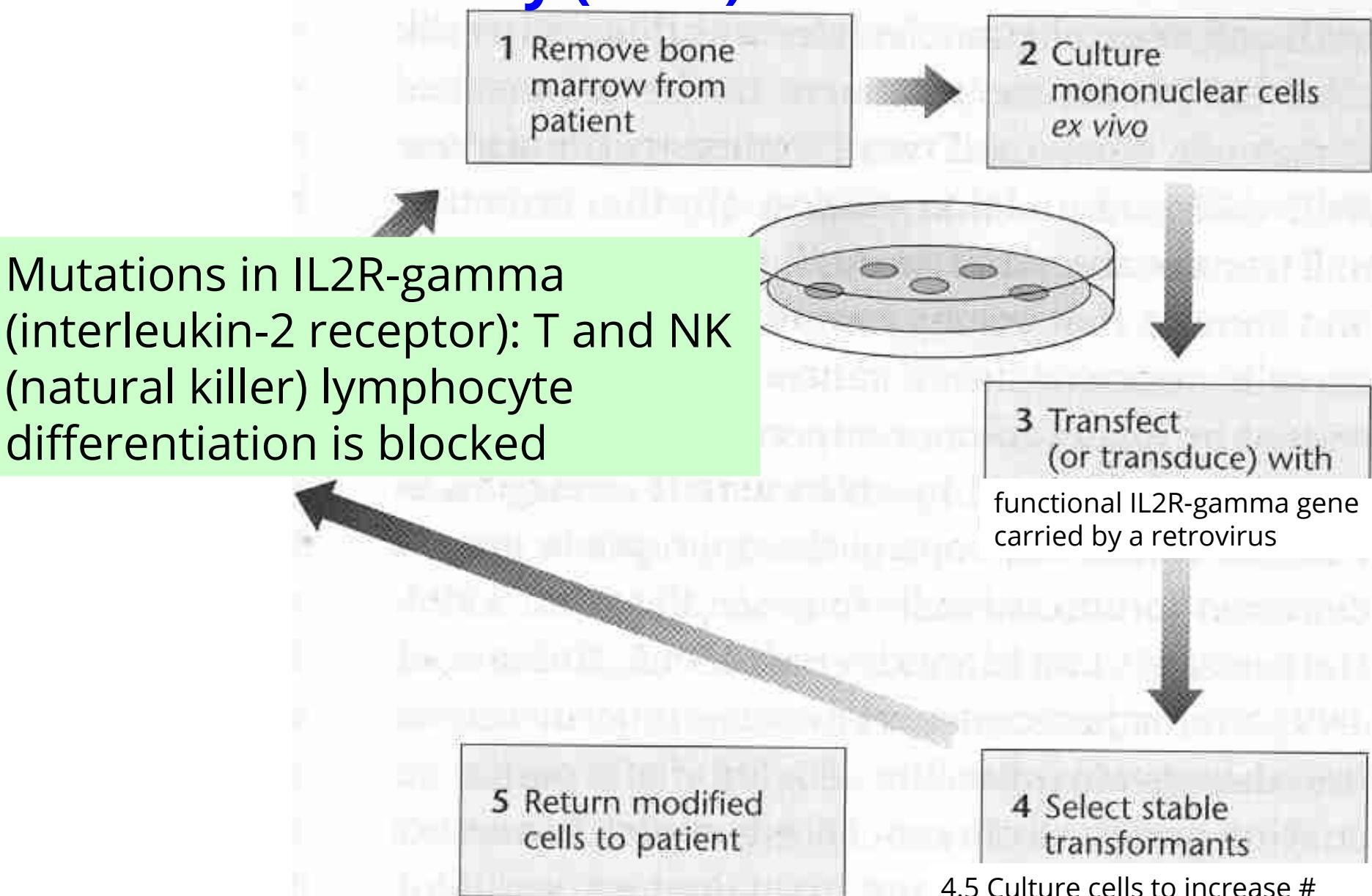
Viral transduction in ~70% of clinical trials

## Vectors Used for Gene Transfer in Gene Therapy Clinical Trials



- Adenovirus 17.5% (n=573)
- Retrovirus 16.4% (n=536)
- Plasmid DNA 14.7% (n=482)
- Lentivirus 10.1% (n=331)
- Adeno-associated virus 8.0% (n=263)
- Vaccinia virus 6.0% (n=197)
- Lipofection 3.8% (n=125)
- Poxvirus 3.4% (n=113)
- Herpes simplex virus 3.1% (n=101)
- RNA transfer 2.0% (n=64)
- Other vectors 6.0% (n=198)
- Not-known 9.0% (n=295)

# *Ex vivo* gene therapy for severe combined immunodeficiency (SCID)

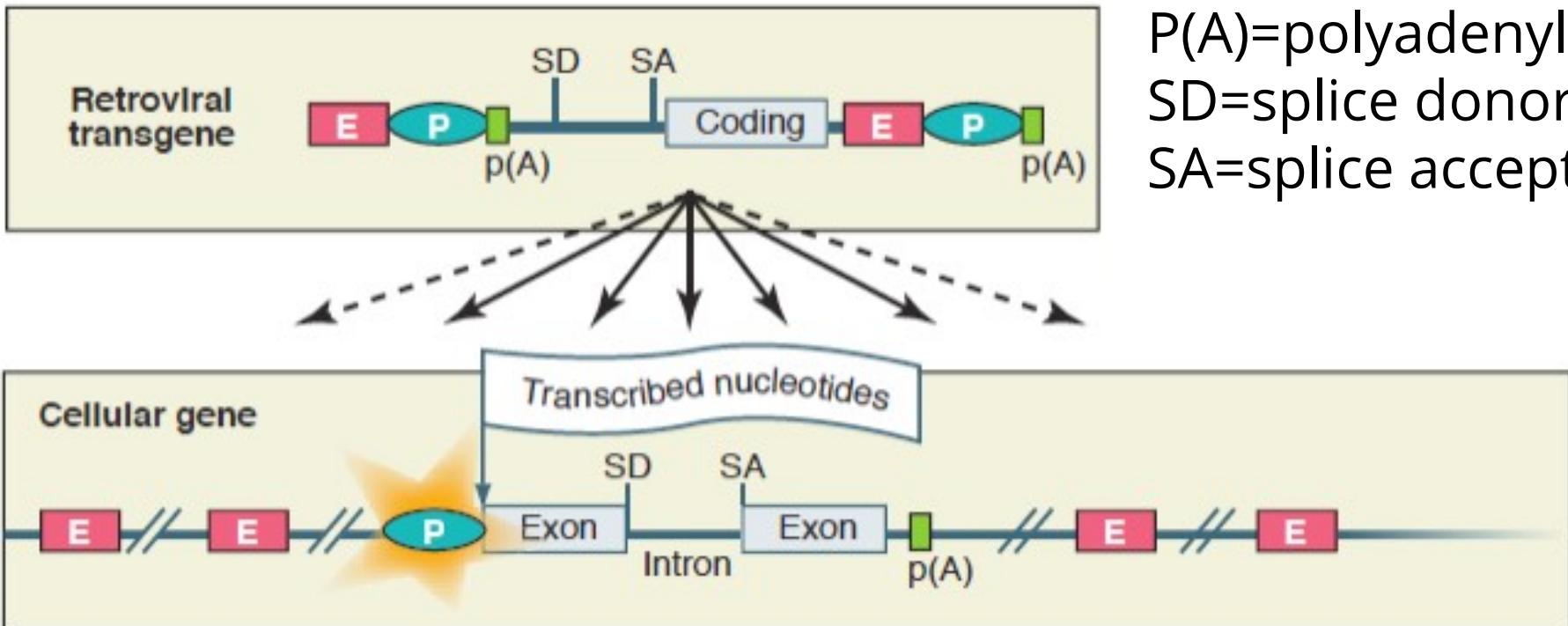


## Problems with gene therapy of SCID patients:

- 10 patients: bone marrow stem cells treated with retrovirus containing IL2R-gamma
- retrovirus inserted randomly into a large population of stem cells (lots of potential insertion sites)
- 9/10 patients survived an otherwise fatal disease
- BUT 2/9 developed leukemia because of insertion of virus near a proto-oncogene. The enhancers in the retroviral vector probably caused overexpression of oncogene (dominant effect)
- Oncogene dysregulation: in about 0.1 to 1% of all retroviral insertion events (or 0.01%, depends on who you ask)
- Oncogene activated cells may grow faster (selection)
- Expression of IL2R- gamma may itself be oncogenic (although this is debatable)

# Retroviral integration: semi-random, but most commonly integrates near expressed genes

E=enhancer  
P=promoter  
P(A)=polyadenyl  
SD=splice donor  
SA=splice acceptor



The proto-oncogene LMO-2 was inadvertently up-regulated in the two patients that developed leukemia

# What to do about the oncogene activation problem?

- retrovirus vectors have their problems, but no better alternatives exist yet
- all medicine has side effects
- benefits still outweigh problems (recent studies: 17/18 SCID patients treated are currently alive, compared to 75% success rates of bone marrow transplants)
- clinical trials on many patients would assist in improving treatment

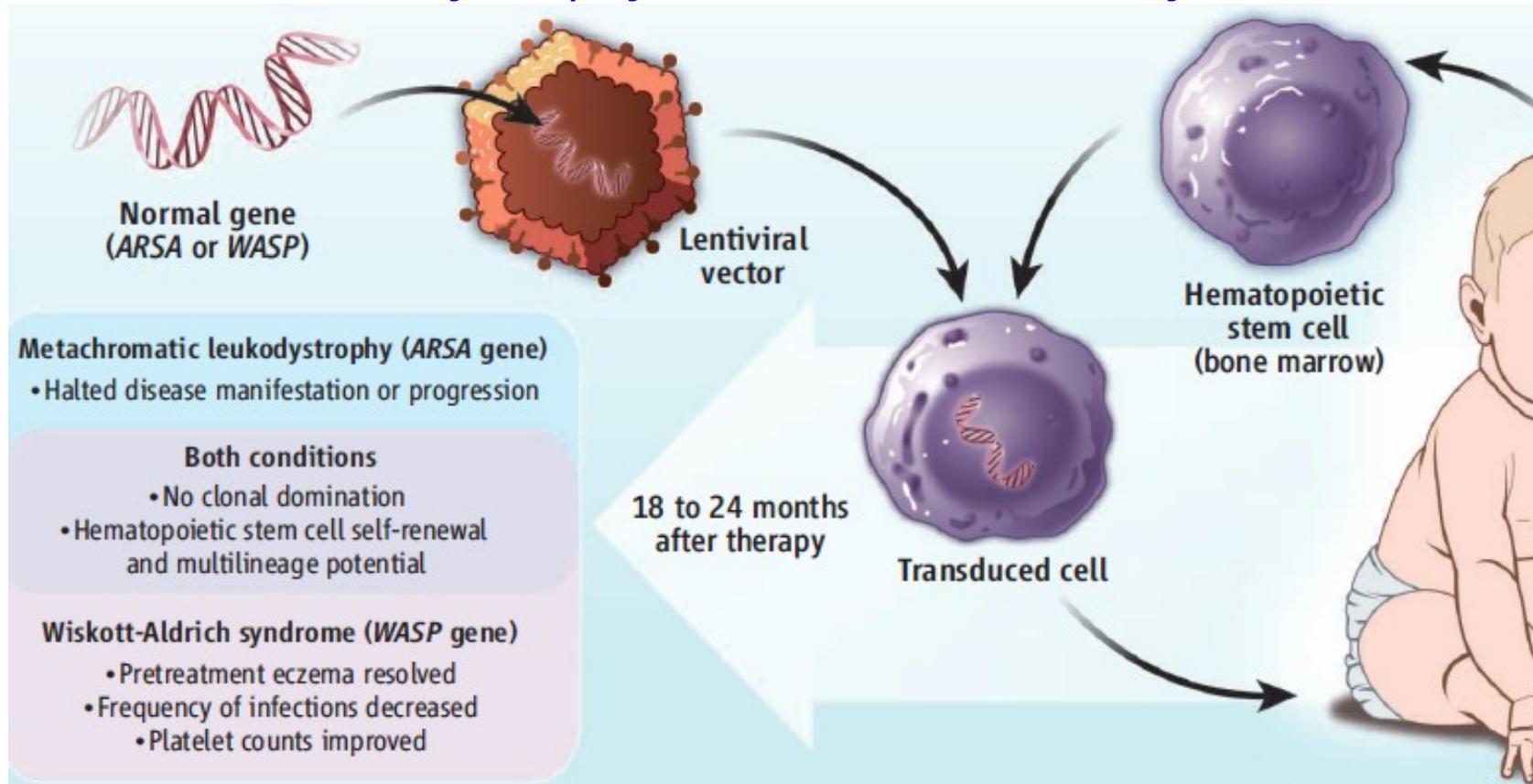
# Following up on gene therapy using retroviruses

- Gamma retroviral insertion alters nearby growth-regulation genes by insertional mutagenesis (activation)
- 'self-inactivating vectors (mutations in U3 region of LTR) are less mutagenic in vitro and in vivo (the insertional pattern is different)
- However, self inactivating vectors do not express transgene at high enough levels
- Cancer development may also depend on other poorly understood factors: transduction protocol, target cell population characteristics
- Still more work...

# New report on IL2R gamma gene therapy (2019)

- Lentiviral, not retroviral vector
- Engineered to have insulator sequences near end to prevent activation of nearby genes
- Patients pre-treated with busulfan (kills off immune cells) to create niche for engineered cells to grow in
- 10 infants were treated who had no sibling matches for bone marrow transplant
- All could leave isolation and go home within 3-4 months of treatment
- Long term results not yet available

# More recent successes in gene therapy: single gene disorders *metachromatic leukodystrophy* and *Wiskott-Aldrich syndrome*

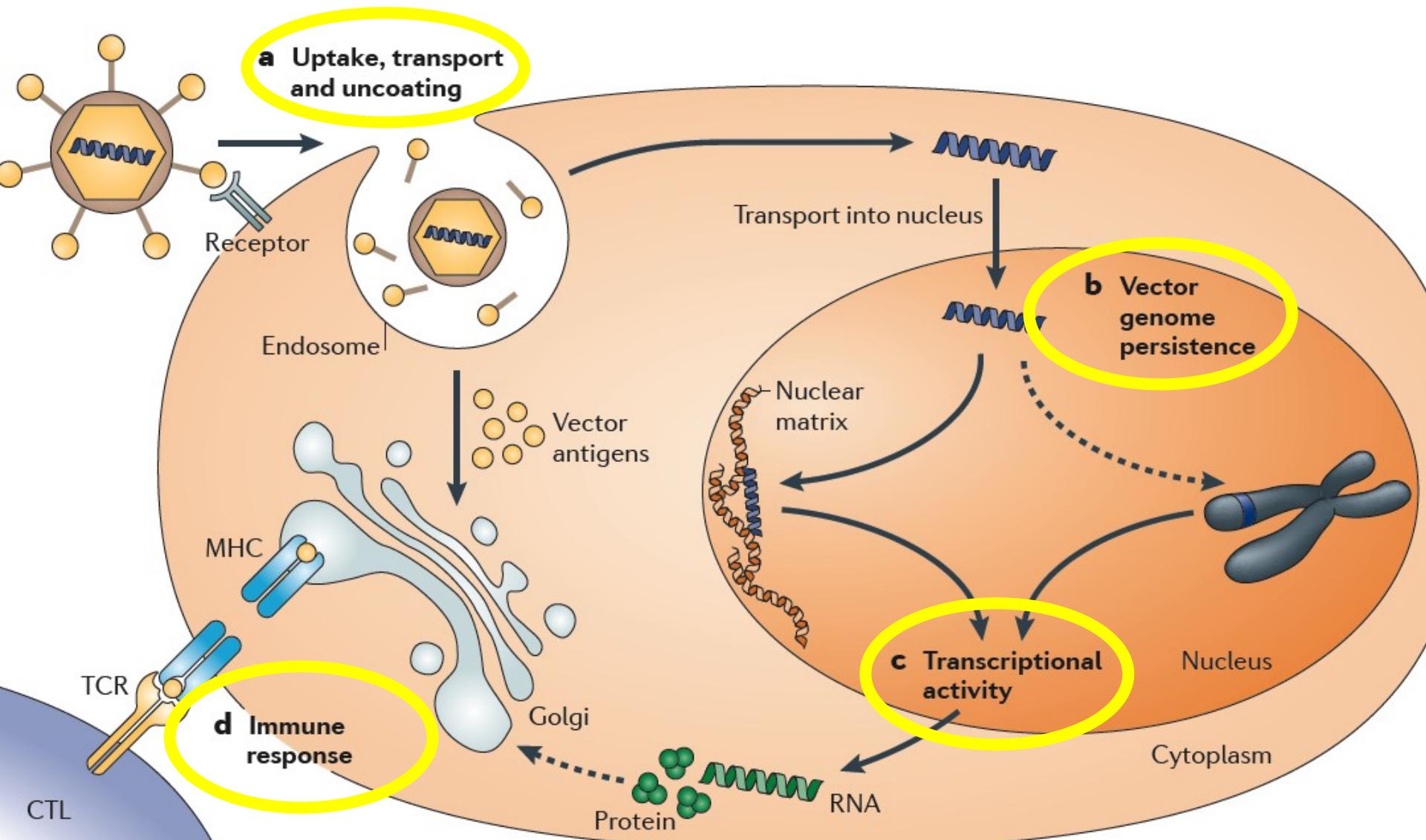


- Lentivirus-mediated, *ex vivo*
- No “clonal domination” was seen
- Random integration of vector (in contrast to retroviral vectors), less likely to cause abnormally proliferating cells

# Important factors for gene therapy in the clinic

- 1) Production of (mostly viral) vectors is costly
- 2) Vector targeting/tissue specificity
  - Target cells must be treated, non-target cells must not be harmed
  - Transcription signals (promoters) must be specific for cell types that are targeted
  - It is difficult to direct vectors only to target cells
  - virus targeting is difficult to control/change

# Obstacles to effective gene delivery in eukaryotic cells

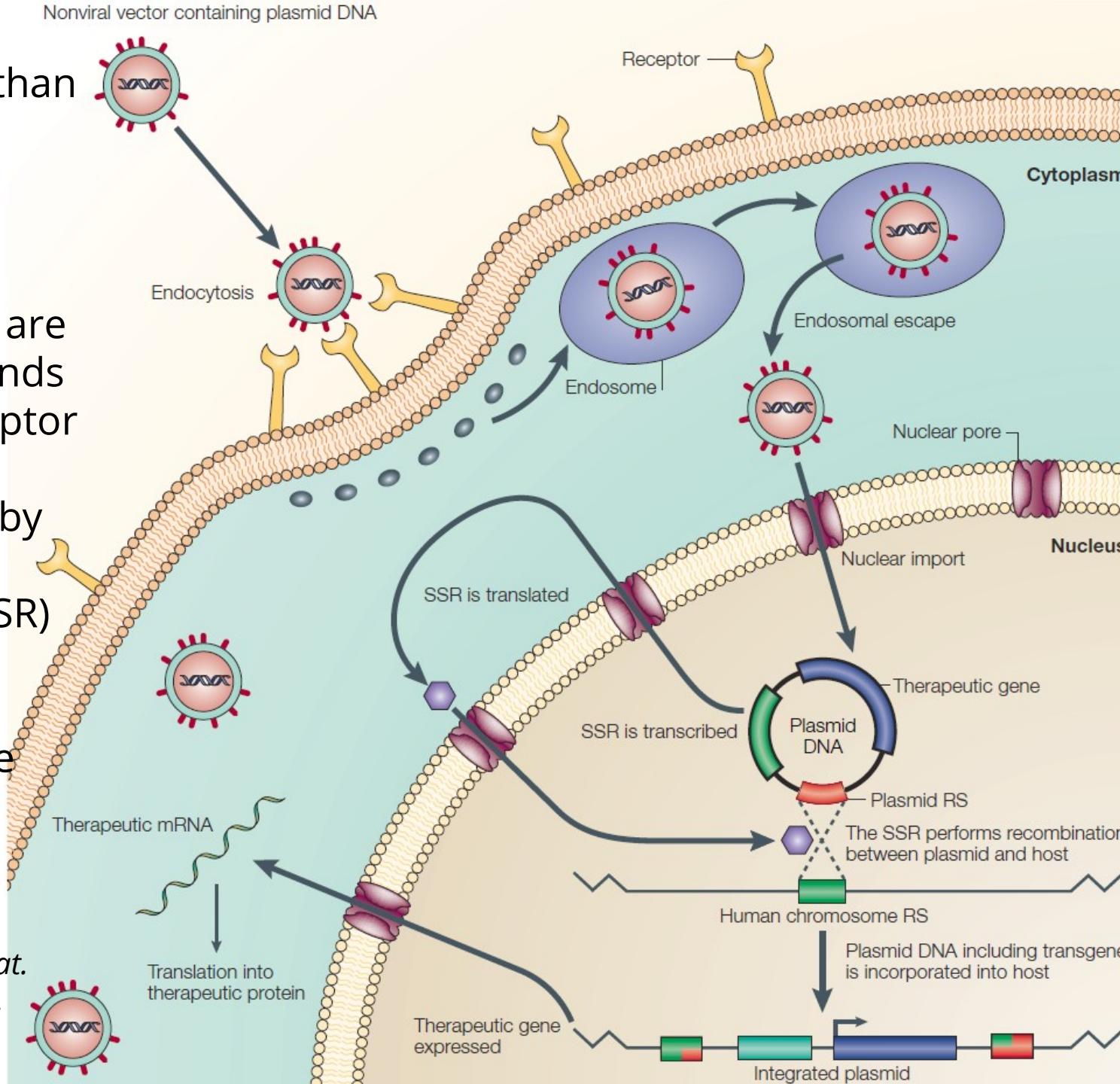


# Non-viral gene delivery: safer than retroviral integration?

Liposomes containing DNA are coated with ligands for specific receptor

DNA integrates by site specific recombinase (SSR) with naturally occurring pseudosites (like loxP)

Glover et al. (2005) *Nat. Rev. Genetics* 6 p. 299.



# Another approach to gene insertion: targeted recombination induced by DNA breaks

DNA can be inserted by homologous recombination (gene targeting)

- Inefficient: won't happen in most cells receiving DNA
- Requires selection in culture (e.g. stem cells)

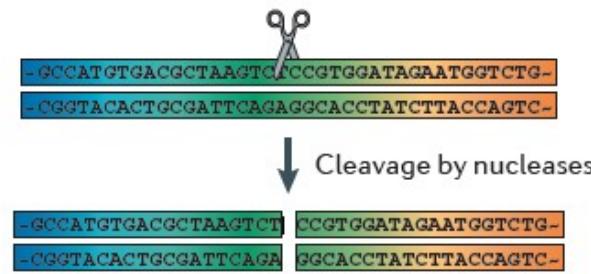
Efficiency of homologous recombination by nicking or breaking the genome at the region you want to engineer ("genome surgery")

1. DNA recognition
  2. DNA cleavage
- Make a machine that can do both, with selectivity for only one site in the genome

# What happens after a double stranded break in DNA?

a

HDR: Homology Dependent Repair



DSB

HDR

NHEJ

-TGTGACGCTAAGTCT **TAT**... CTTA CCGTGGATAGAATGG-  
-ACACTGCGATTCAAG **ATA**... GAAT GGCACCTATCTTACC-

Donor DNA

-TGTGACGCTAAGTC **C**CCGTGGATAGAATGG-  
-ACACTGCGATTCAAG **G**GGCACCTATCTTACC-

Donor DNA

Or

-TGTGACGCTAAGTC **C**CCGTGGATAGAATGG-

ssODN

Gene or tag insertion

Gene correction or point mutagenesis

Small indels

NHEJ: Non Homologous End Joining

-GCCATGTGACGCTAAGTCT **ATG**CCGTGGATAGAATGGTCTG-  
-CGGTACACTGCGATTCAAG **TAC**GGCACCTATCTTACCAAGTC-

-GCCATGTGACGCTAAGTCT **T**CCGTGGATAGAATGGTCTG-  
-CGGTACACTGCGATTCAAG **A**GGCACCTATCTTACCAAGTC-

-GCCATGTGACGCTAAGTCT **GTG**GGATAGAATGGTCTG-  
-CGGTACACTGCGATTCAAG **CAC**CTATCTTACCAAGTC-

-GCCATGTGACGCTAAGTCT **GTG**GGATAGAATGGTCTG-  
-CGGTACACTGCGATTCAAG **CAC**CTATCTTACCAAGTC-

-GCCATGTGACGCT **GG**GGATAGAATGGTCTG-  
-CGGTACACTGCGATTCAAG **CCT**ATCTTACCAAGTC-

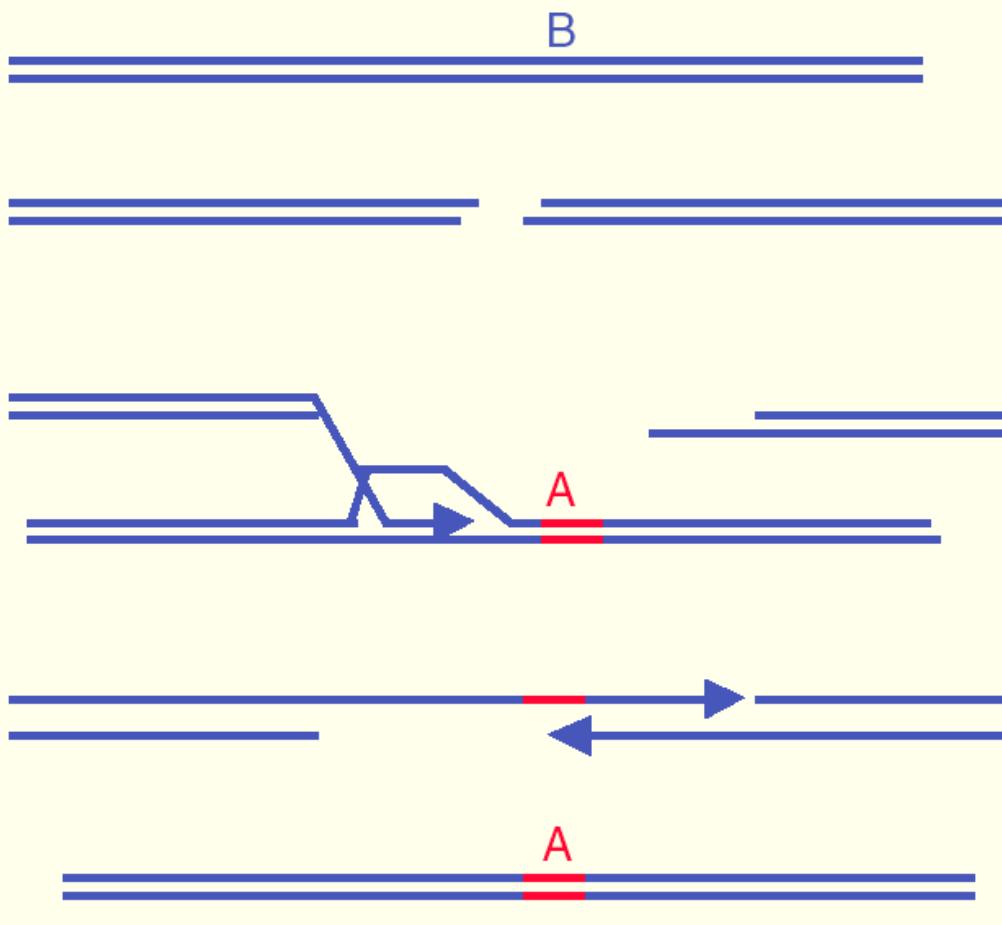
A guide to genome engineering  
with programmable nucleases

Hyongbum Kim<sup>1</sup> and Jin-Soo Kim<sup>2,3</sup>

NATURE REVIEWS | GENETICS

VOLUME 15 | MAY 2014 | 321

# Double stranded breaks and homologous recombination



Undamaged DNA (allele B)

DSB created (spontaneous or induced, e.g., by ZFN)

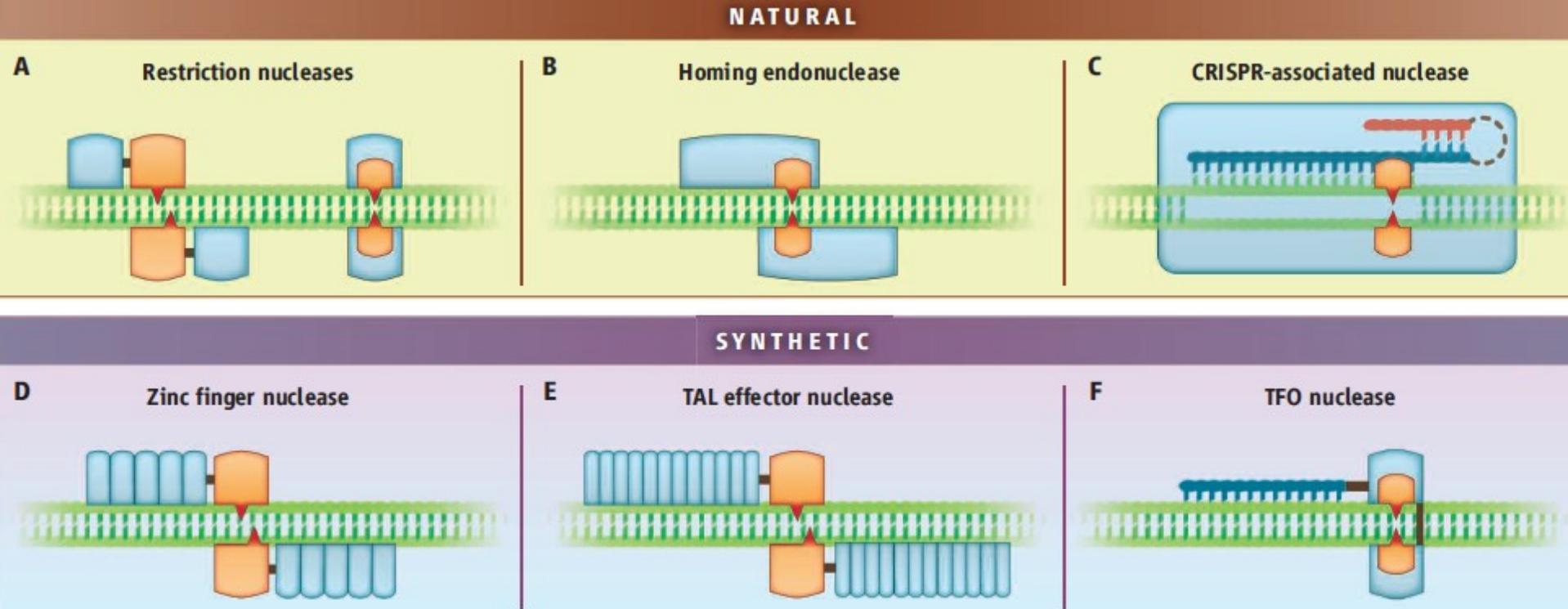
Strand invasion into undamaged homologous DNA (allele A)

In gene targeting, exogenous DNA serves as homologous DNA donor

Repairing of original strands of DNA  
Gaps filled by DNA polymerase and nicks sealed by DNA ligase

Conversion of blue allele (B) into red allele (A) in region of DSB

# Several strategies for defining specific DNA cut sites



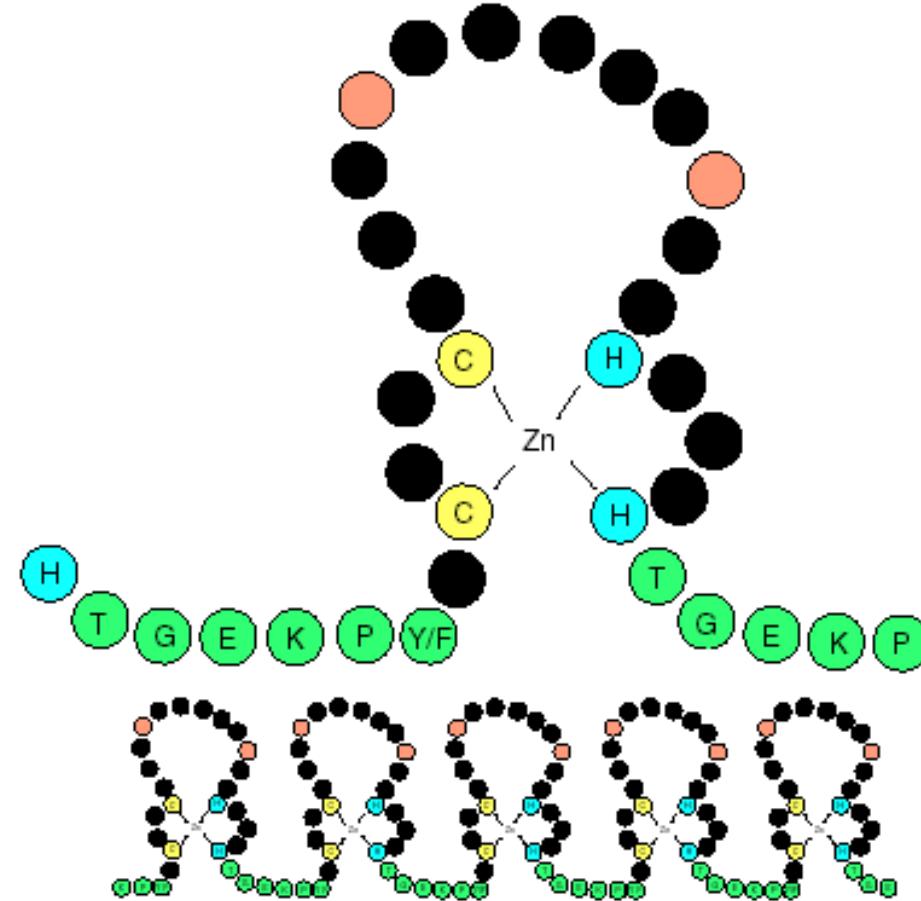
Blue: DNA recognition, Orange: DNA nuclease

**New Tool for Genome Surgery**  
John van der Oost  
*Science* **339**, 768 (2013);  
DOI: 10.1126/science.1234726

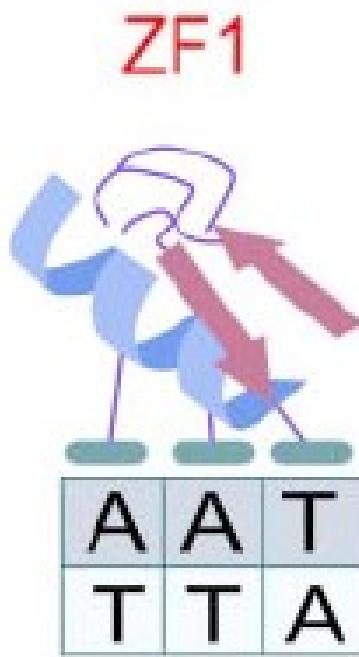
To introduce mutations, co transfect with

- A) programmable nuclease
- B) Targeting vector (containing desired DNA)

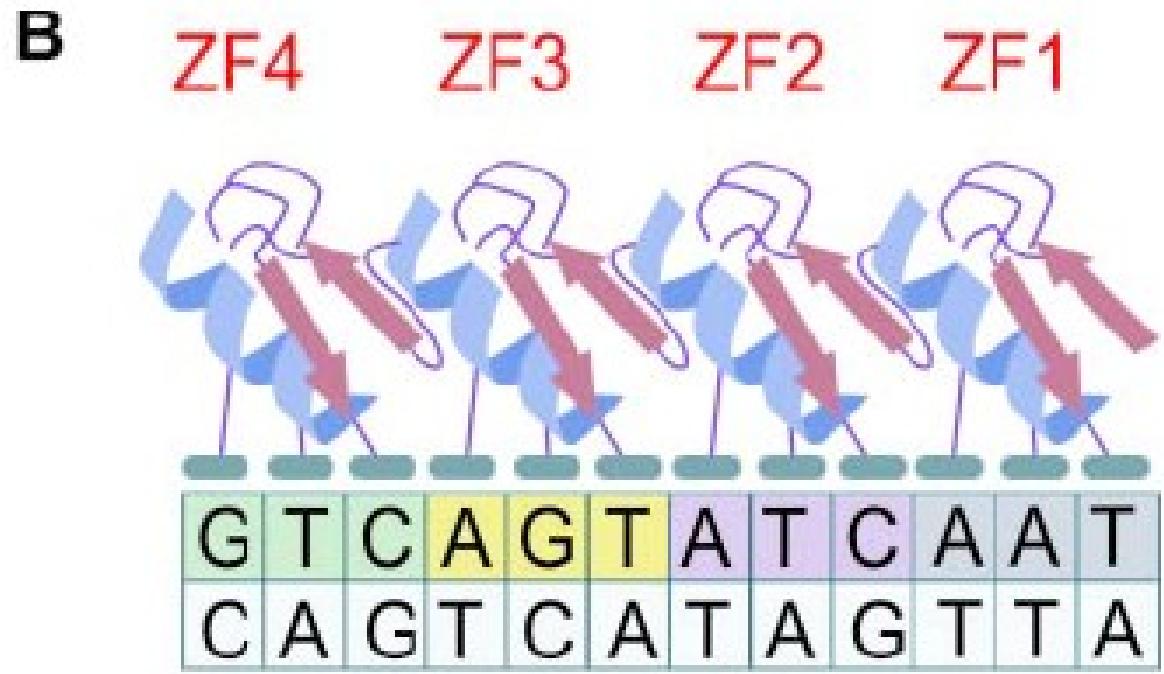
# Zinc finger transcription factors



- DNA binding domains stabilized by Zn<sup>+2</sup> binding
- Zn-ribbons fused together can recognize longer DNA sequences



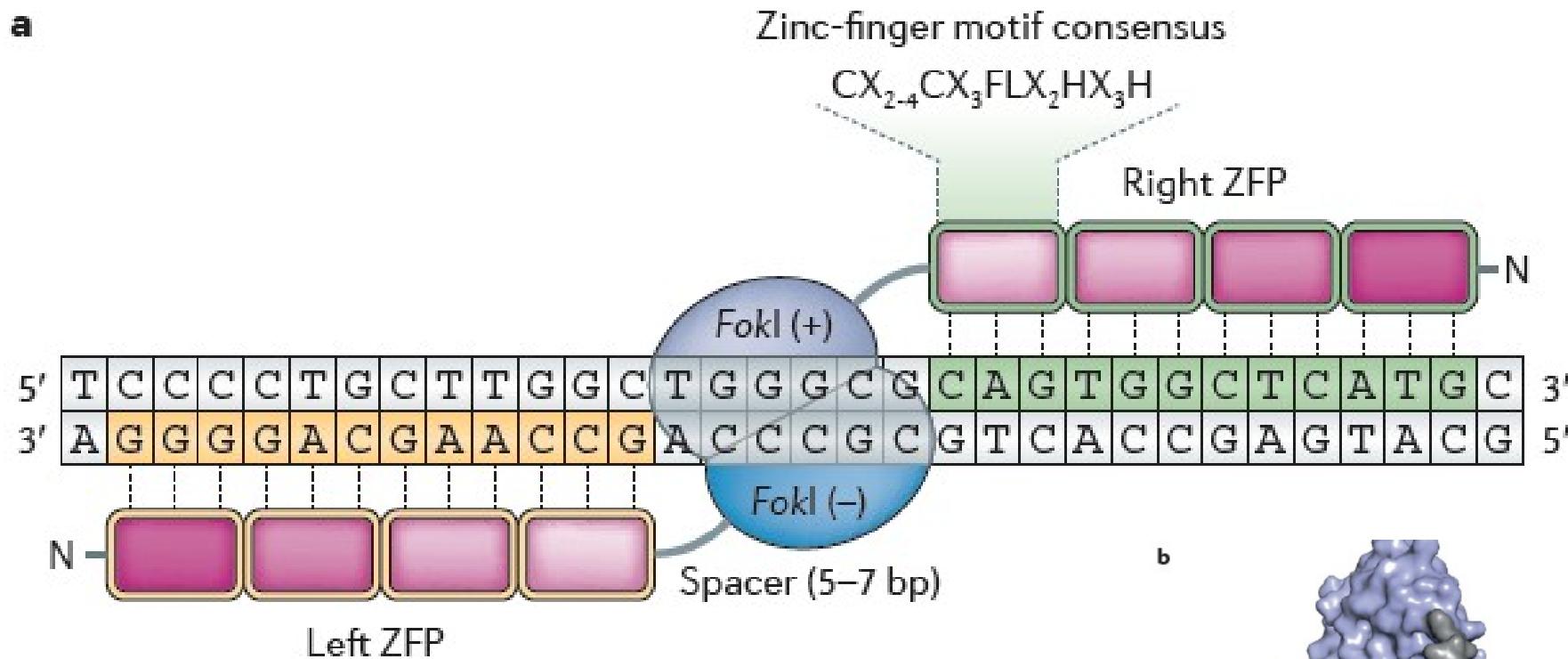
Single Zn finger



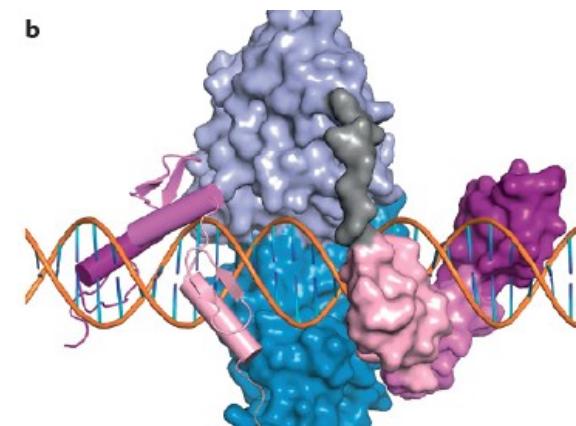
four Zn fingers in tandem

# Zinc finger nuclease (ZFN) mechanism

a

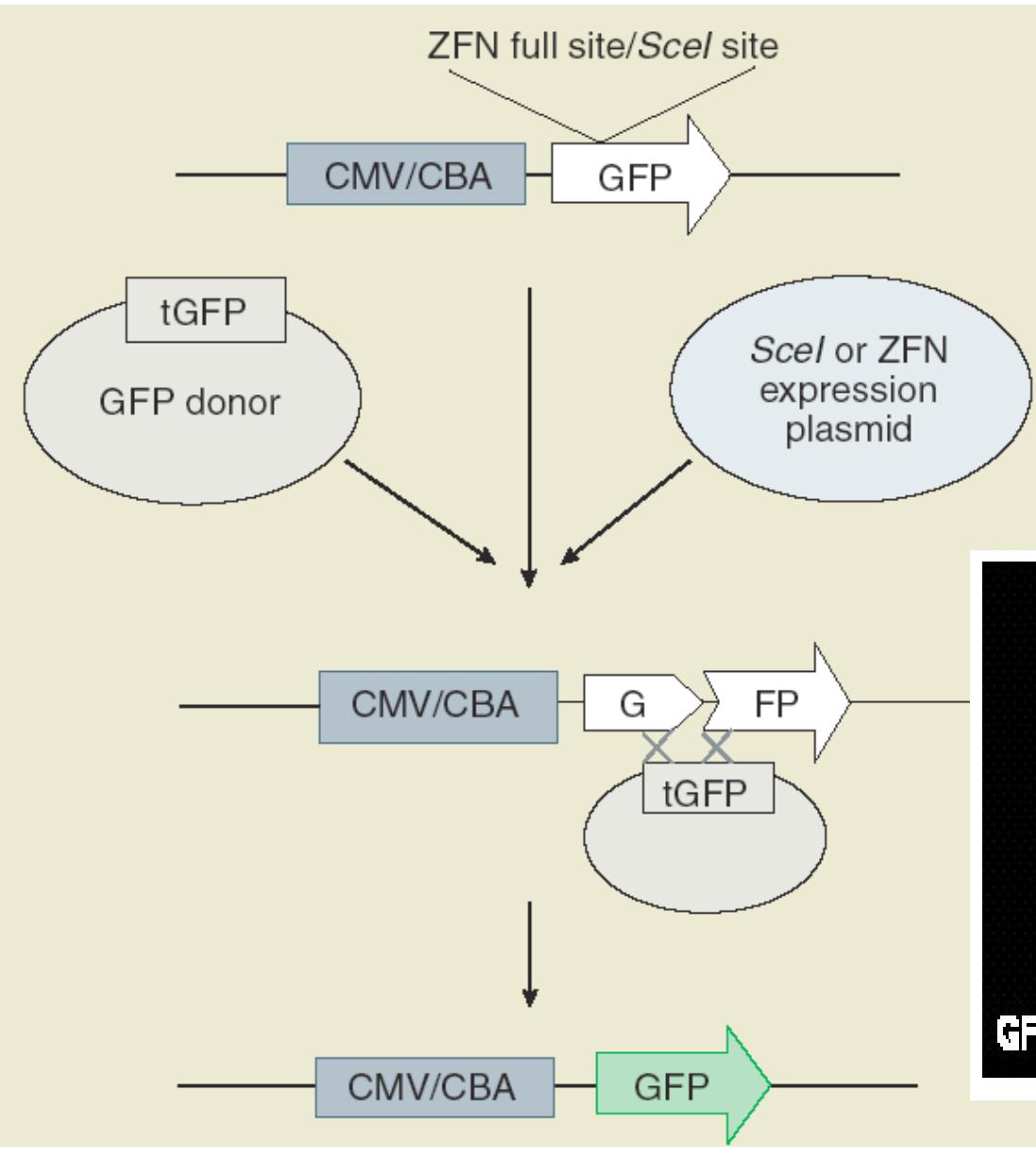


b



Zinc fingers can be hypothetically tailored to any DNA site; sequence must be long enough to ensure one site/genome

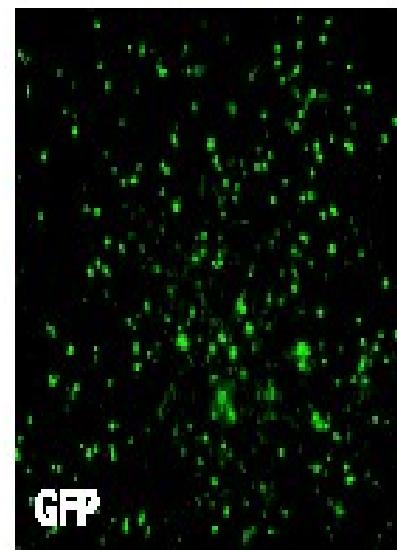
# Test of method: make GFP functional



Mammalian cell  
culture transfection  
GFP plasmid



-ZFN



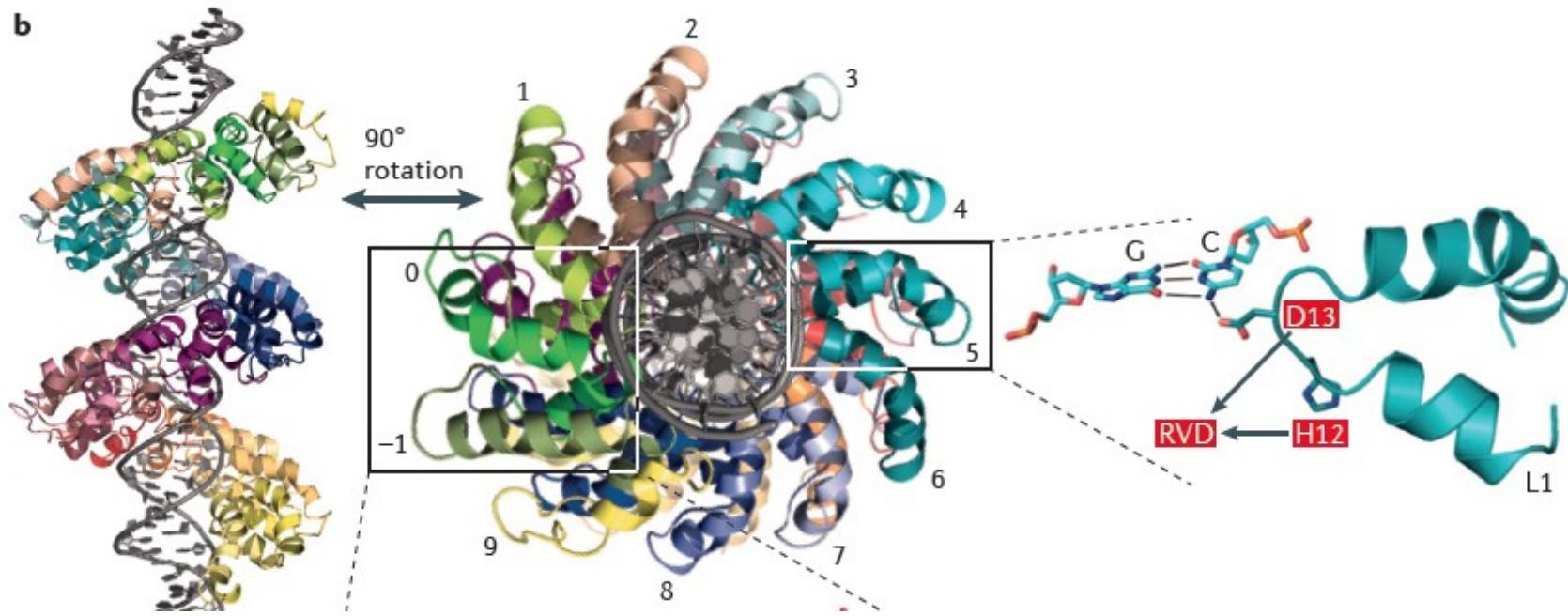
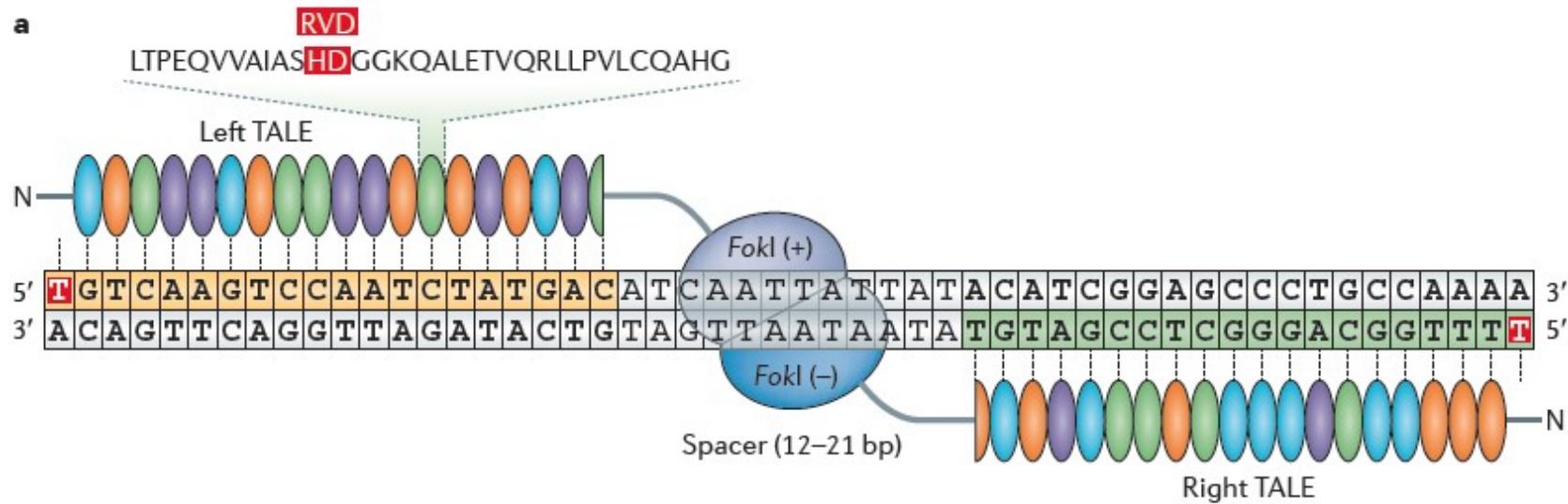
+ZFN

# Practical uses of ZFN-mediated recombination

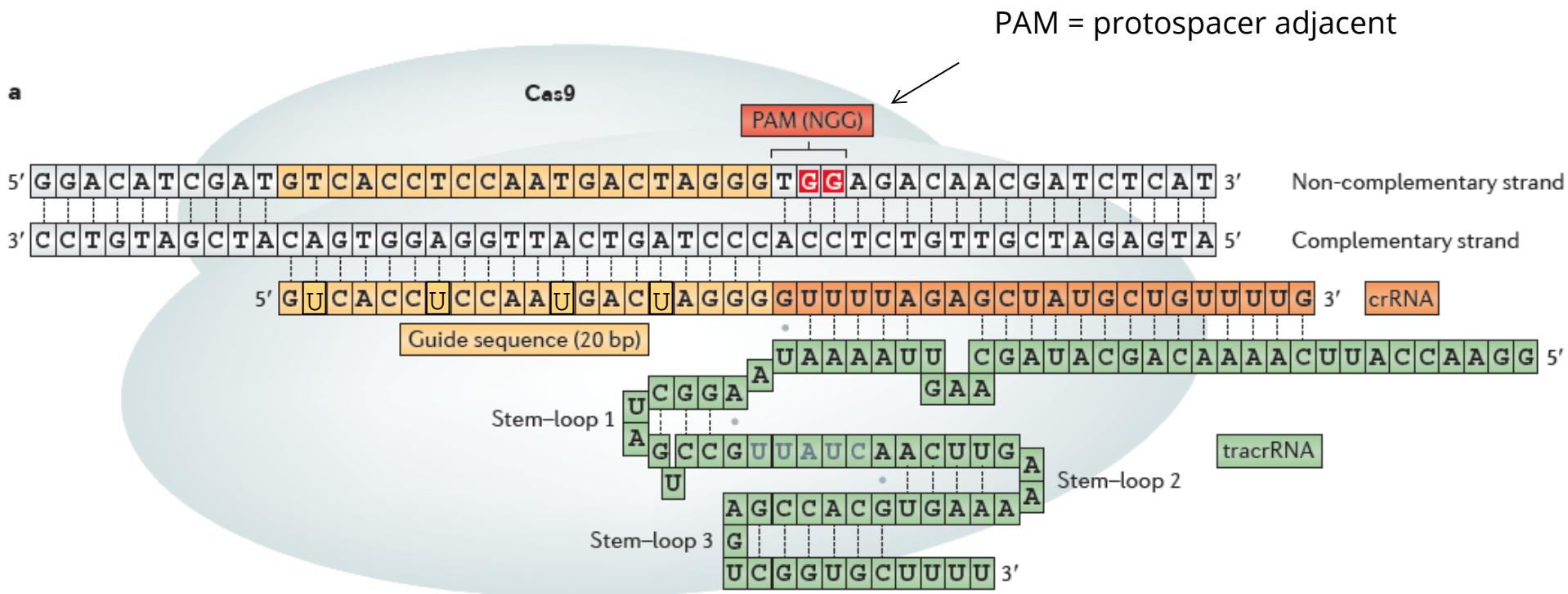
- ZFN has been used to alter IL2R-gamma to wild type in cultured cells, with high efficiency
- However, ZFNs can be toxic because of “off-target” binding and genome cleavage
- Also, for use in gene therapy (precision genome surgery) it will be necessary to optimize systems for delivery of the ZFN and the DNA to be recombined
  - Transfection?
  - Viral delivery?
  - Direct injection?

# TALENs: transcription factor-like effector nucleases

Single base recognized by each repeat-variable di-residue (RVD)

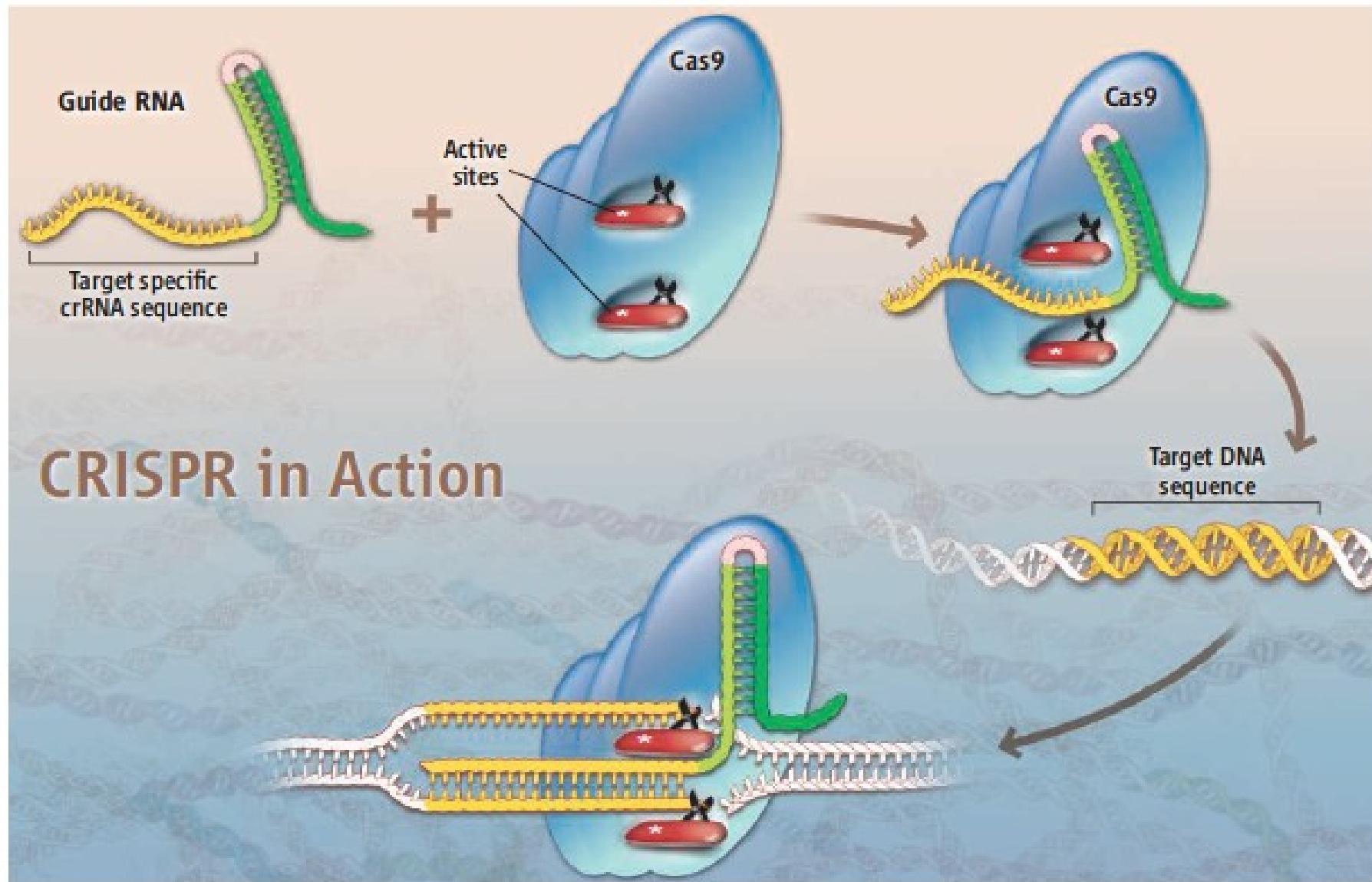


# Cas9: an RNA-guided nuclease, used in microbial self-defense systems



The guide sequence-targeted interactions are very stable, and can also provide a tethering platform for proteins or RNAs, if the Cas9 nuclease activity is shut down

# CRISPR: clustered regularly interspaced short palindromic repeats

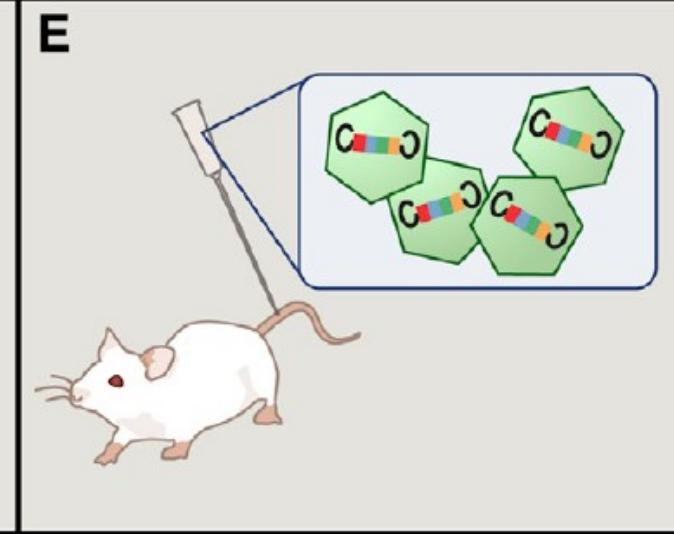
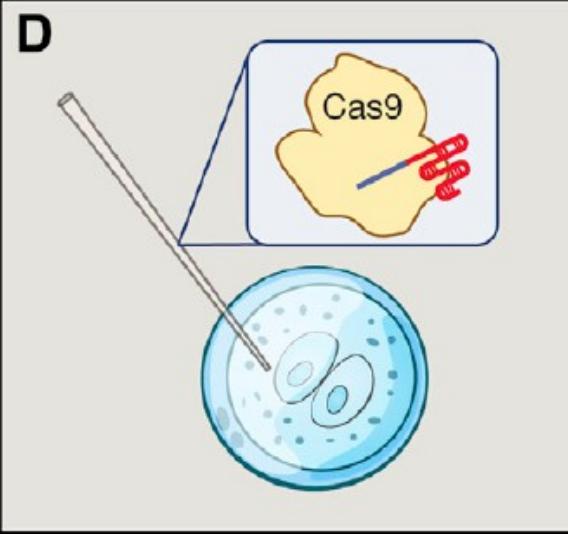
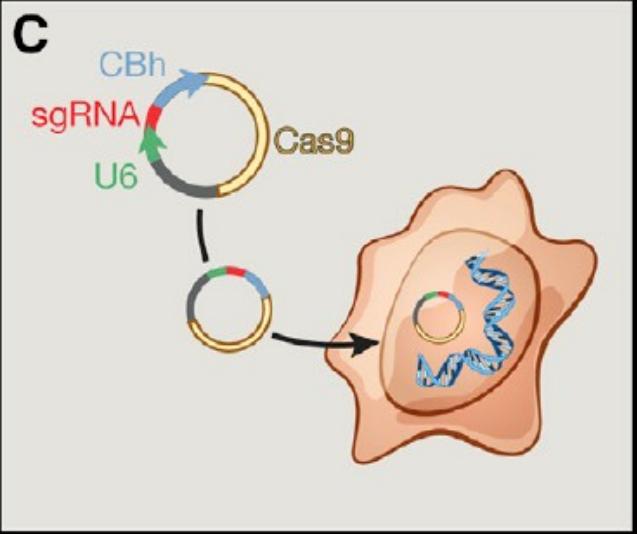


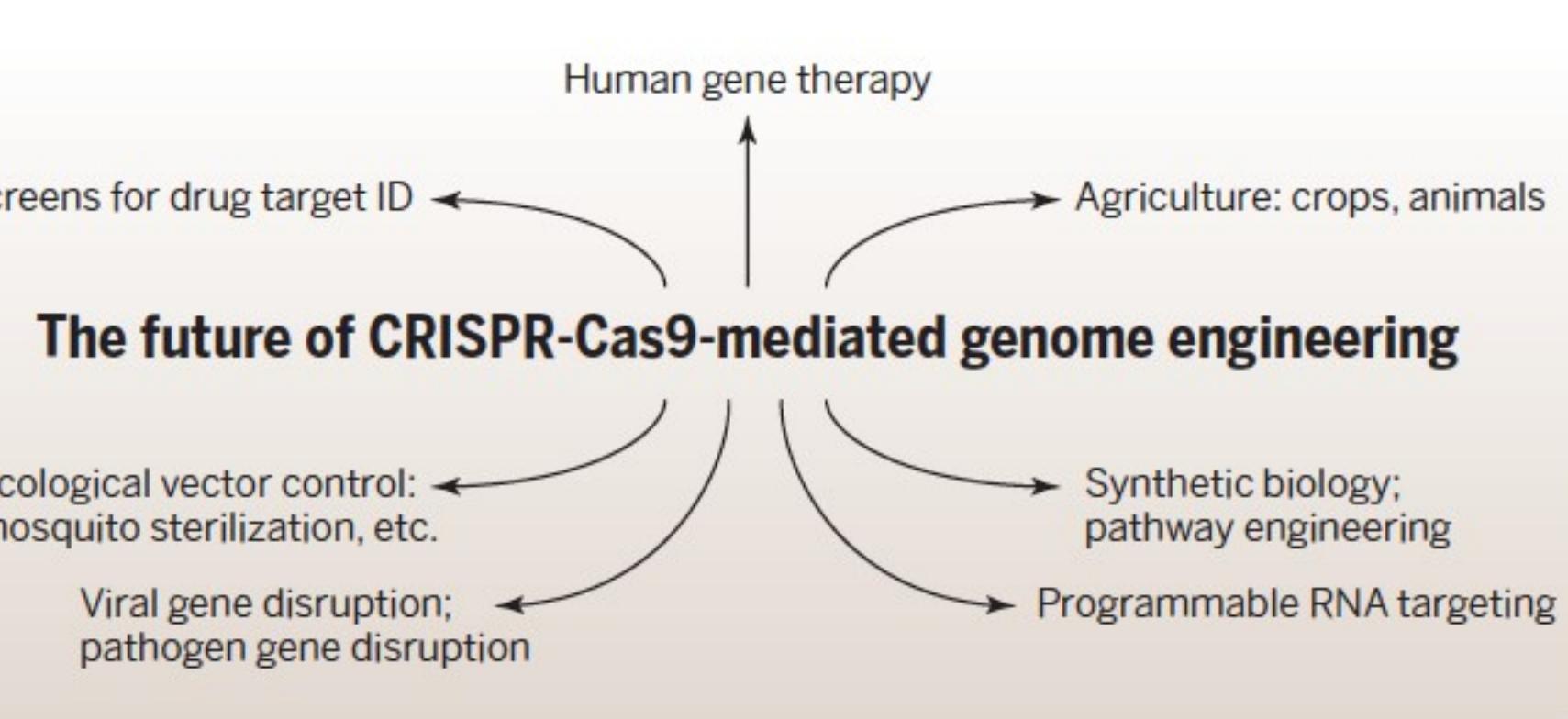
# CRISPR Cas 9 is very easy to use

Transfection

Direct injection

Viral transduction





2015: Human embryos were engineered with CRISPR-Cas9

2015: Calls for ban on using CRISPR-Cas9 with human embryos, supported by the US NIH

# Summary of the main nucleases for genetic surgery

	ZFNs	TALENs	RGENs
<b>DNA targeting specificity determinant</b>	Zinc-finger proteins	Transcription activator-like effectors	crRNA or sgRNA
<b>Nuclease</b>	FokI	FokI	Cas9
<b>Success rate<sup>‡</sup></b>	Low (~24%)	High (>99%)	High (~90%)
<b>Average mutation rate<sup>§</sup></b>	Low or variable (~10%)	High (~20%)	High (~20%)
<b>Specificity-determining length of target site</b>	18–36 bp	30–40 bp	22 bp (total length 23 bp)
<b>Restriction in target site</b>	G-rich	Start with T and end with A (owing to the heterodimer structure)	End with an NGG or NAG (lower activity) sequence (that is, PAM)
<b>Design density</b>	One per ~100 bp	At least one per base pair	One per 8 bp (NGG PAM) or 4 bp (NGG and NAG PAM)
<b>Off-target effects</b>	High	Low	Variable
<b>Cytotoxicity</b>	Variable to high	Low	Low
<b>Size</b>	~1 kb × 2	~3 kb × 2	4.2 kb (Cas9 from <i>Streptococcus pyogenes</i> ) + 0.1 kb (sgRNA)

A guide to genome engineering  
with programmable nucleases

ZFN: zinc finger nuclease  
TALEN: TAL effector nuclease  
RGEN: RNA guided endonuclease

## Gene therapy summary

- 1) Gene therapy targets, types of vectors
- 2) Retroviral gene therapy, advantages and drawbacks
- 3) Precision genome engineering site-specific genome cleavage and recombinational repair