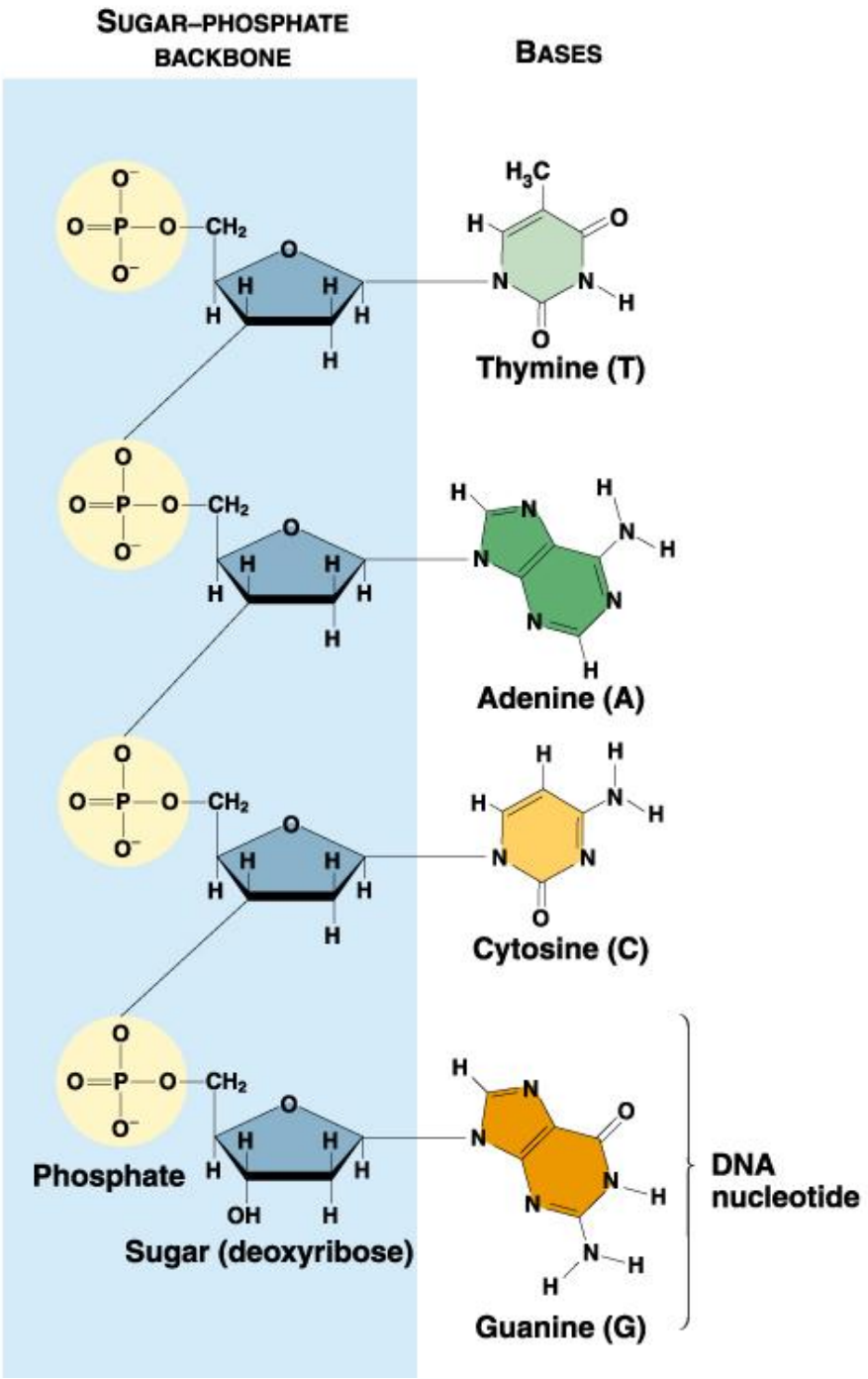
Gene Technologies

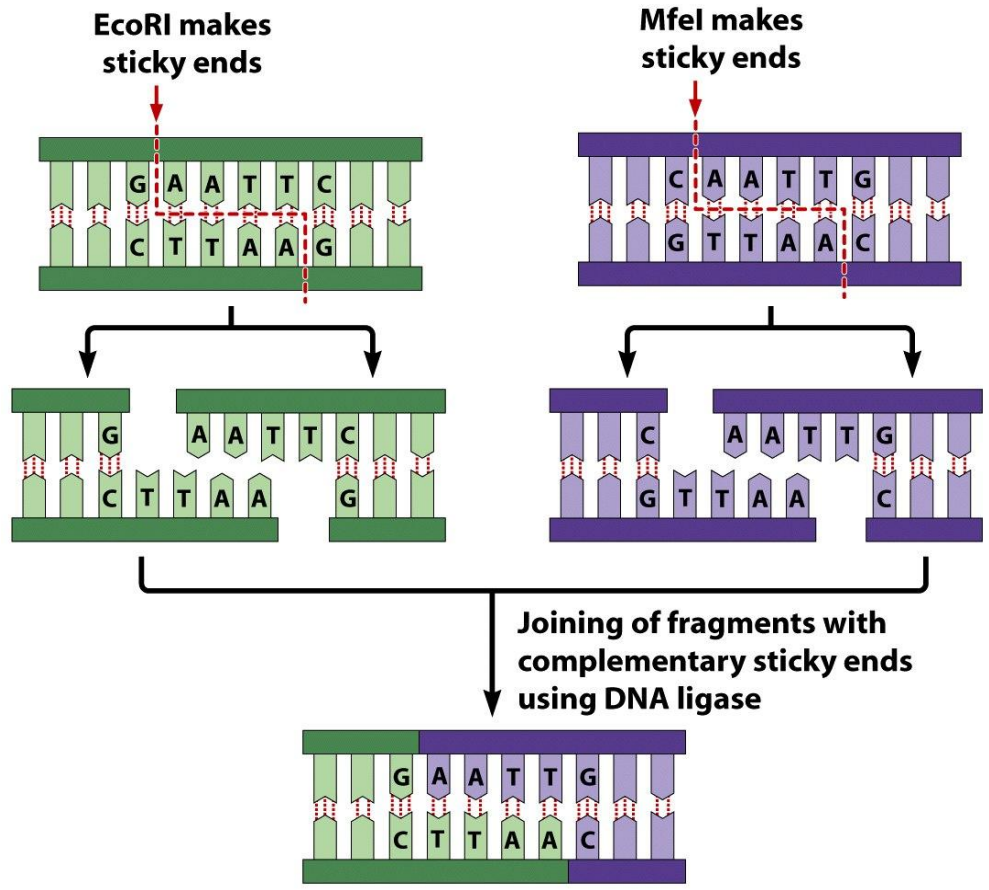
1. Enzymes in Biotechnology

DNA Structure



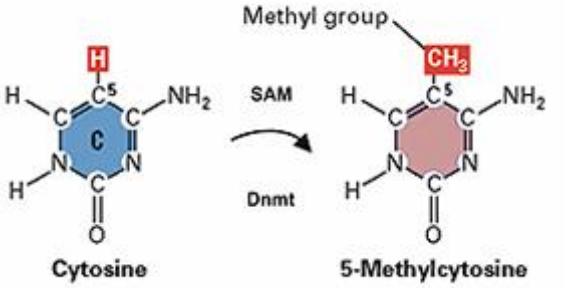
Restriction Endonucleases

- Isoschizomers: recognise the same sequence, but cut at different places



Methylation and Enzymes

- Bacteria possess methylase with the same sequence specificity as the restriction enzyme activity
- Foreign DNA is non-methylated and hence digested by the nucleases, while the cell's own DNA is methylated and so not digested
- Restriction enzymes recognise the absence of methyl groups at appropriate sites



Types of DNA Polymerases

- Standard is 5'->3' template-dependent synthesis
- Many polymerases also have 3'-> 5' exonuclease activity, which removes one base at a time
- Some enzymes also have a 5'-> 3' exonuclease activity
- Template independent DNA Polymerases add 5'->3' without needing a template
- Reverse transcriptase uses an RNA template, and is used in cDNA synthesis

- DNA ligase catalyses the formation of phosphodiester bond between 5’-P and 3’-OH groups in duplex DNA - will join blunt double stranded DNA

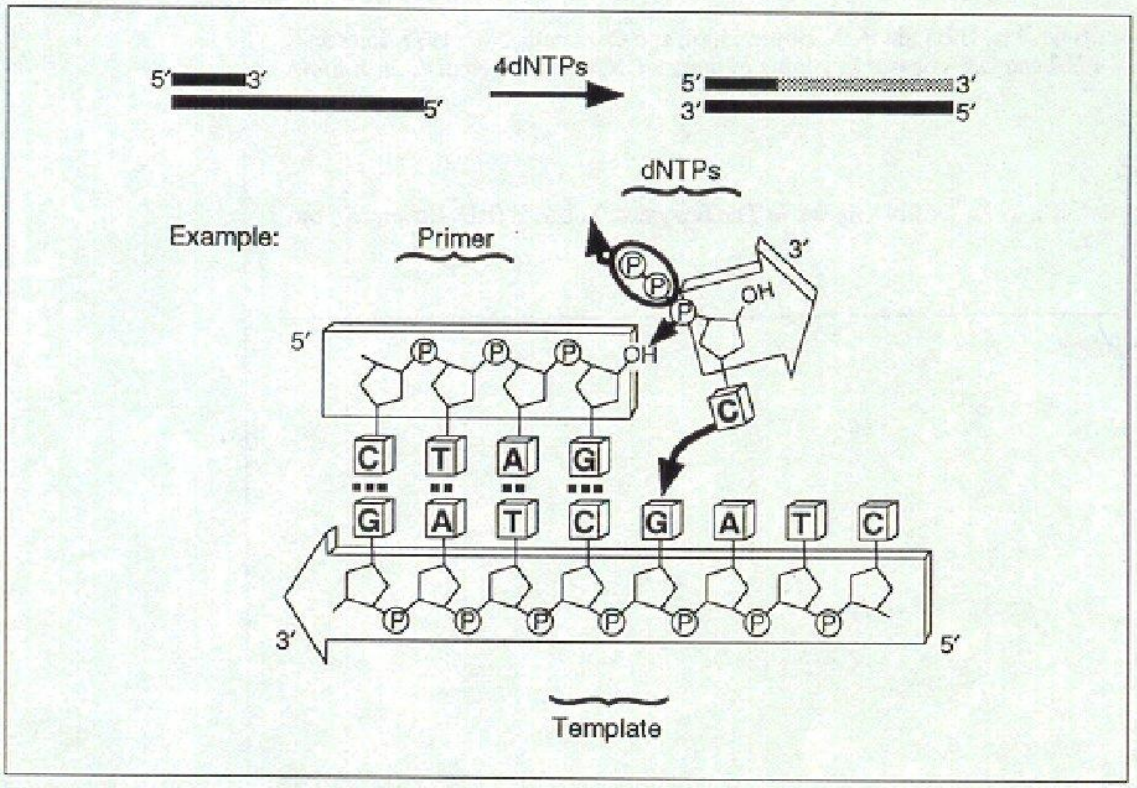


Figure 3.5.1 DNA polymerase 5′→3′ polymerase activity.

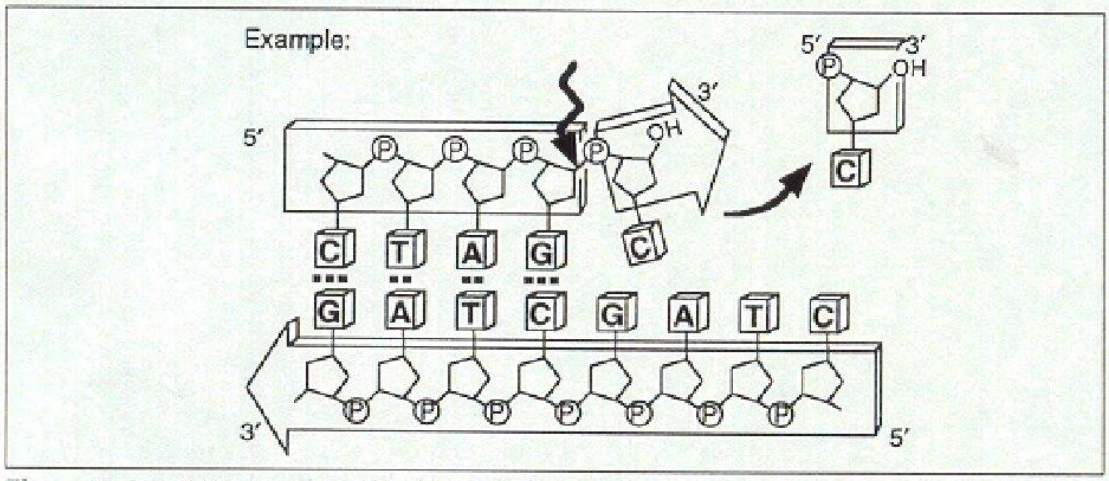


Figure 3.5.2 DNA polymerase 3′→5′ exonuclease activity.

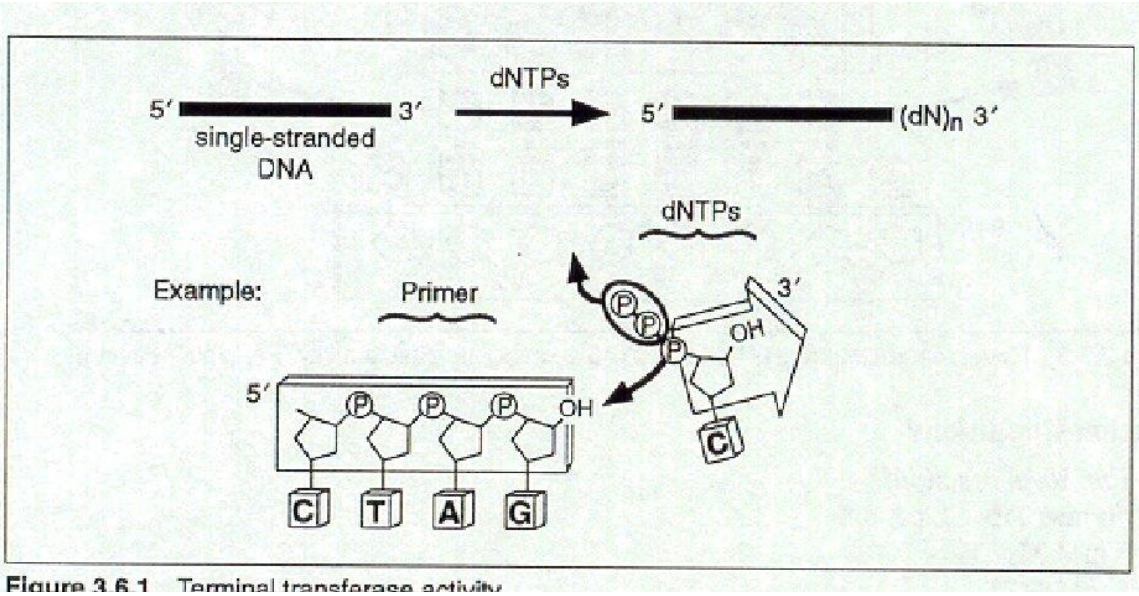


Figure 3.6.1 Terminal transferase activity

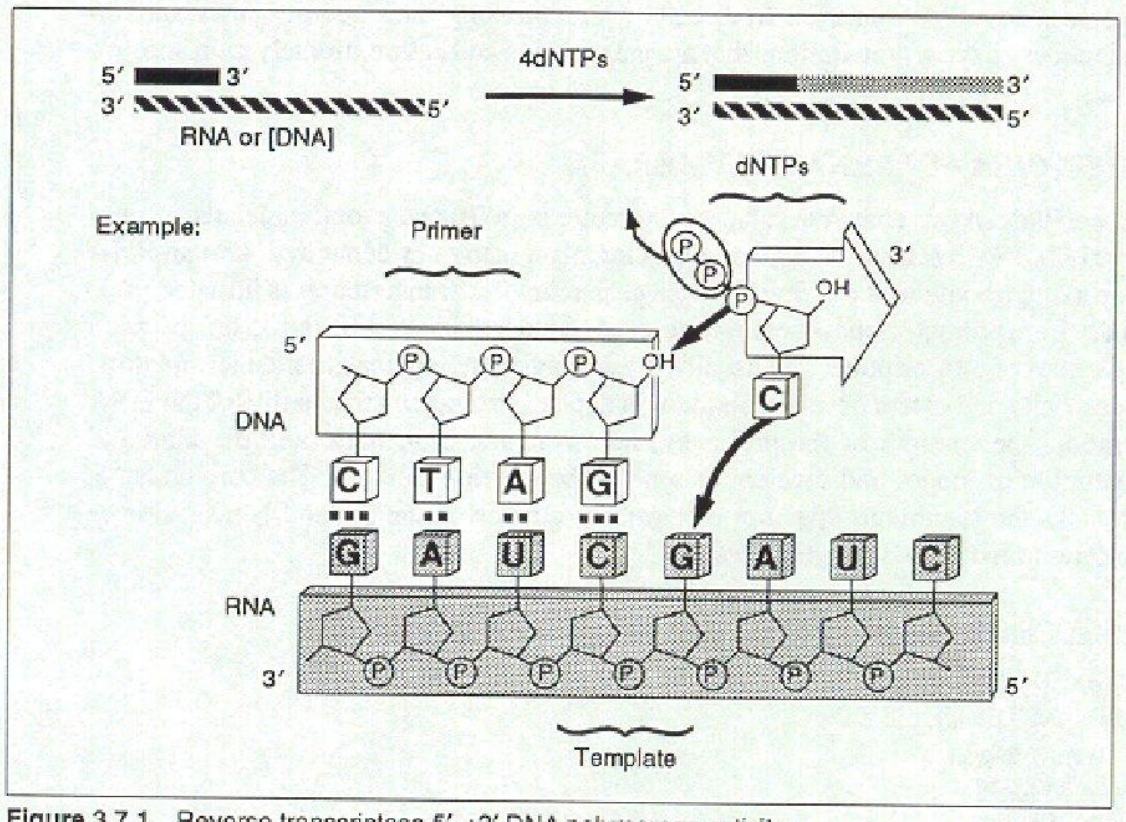


Figure 3.7.1 Reverse transcriptase 5′→3′ DNA polymerase activity.

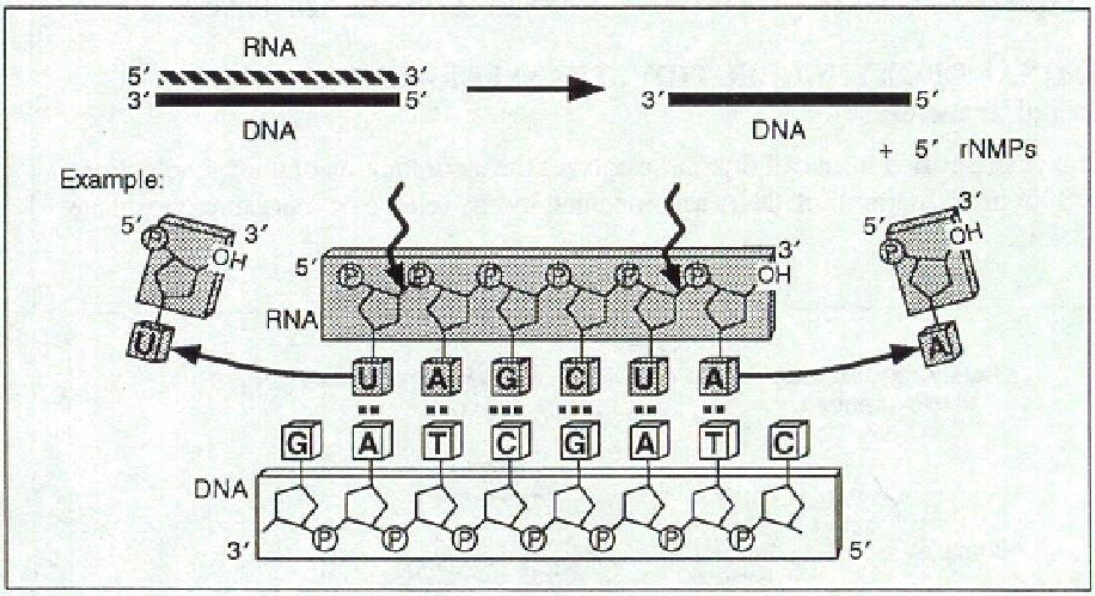


Figure 3.7.2 Reverse transcriptase 5′→3′ and 3′→5′ exoribonuclease activity (RNase H activity).

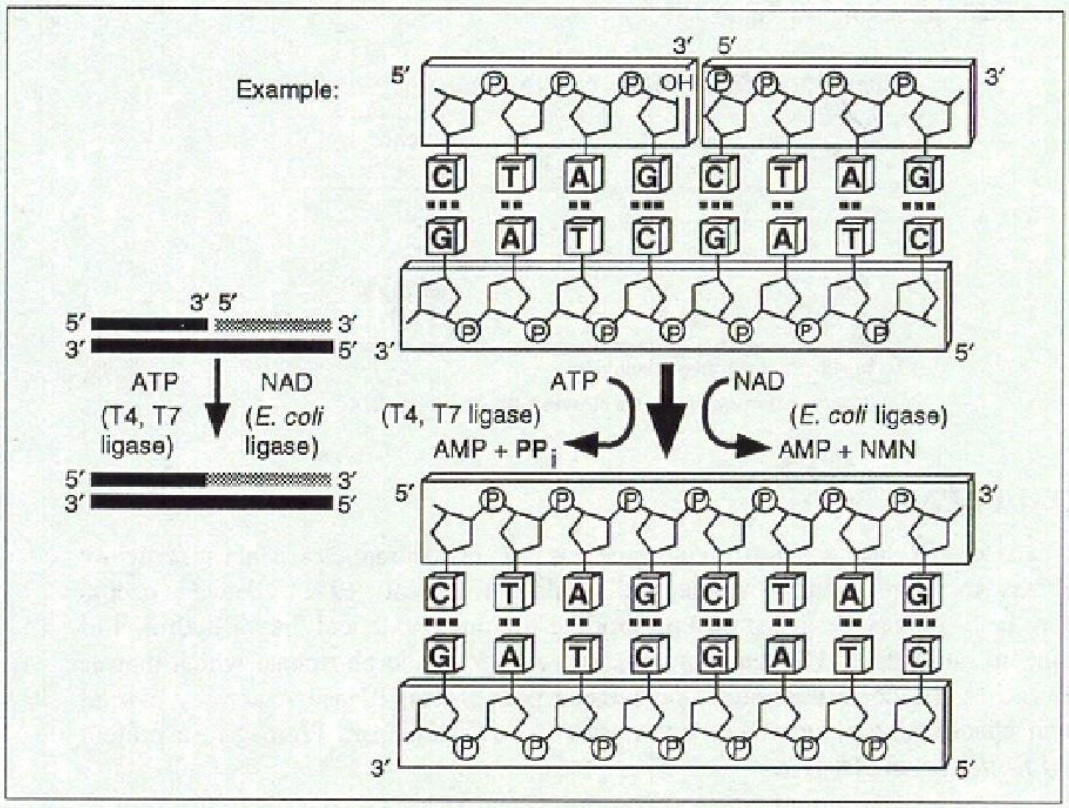
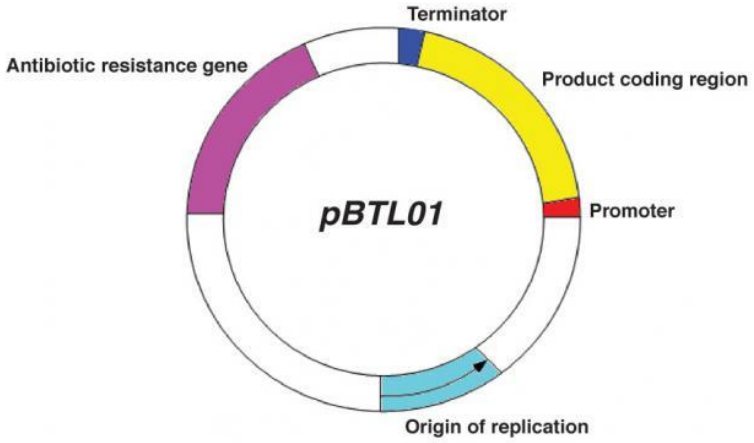


Figure 3.14.1 DNA ligase activity at a nick.

2. Cloning Vectors and Libraries

Requirements for Cloning

- A cloning vector is a small piece of DNA, taken from a virus, a plasmid, or the cell of a higher organism, that can be stably maintained in an organism, and into which a foreign DNA fragment can be inserted for cloning purposes
- All commonly used cloning vectors in molecular biology have key features necessary for their function
- Cloning site: a place where genes can be inserted, such as a multiple cloning site
- Selectable market: something carried by the vector to enable identification of transformed cells, such as antibiotic resistance or auxotrophic selection markets
- Reporter genes: allow easy identification of clones by producing visible products like breaking down particular sugars or emitting light (GFP)
- Original of replication: must be able to independently replicate in the host cell
- Ideally vectors should also have a high copy number per cell



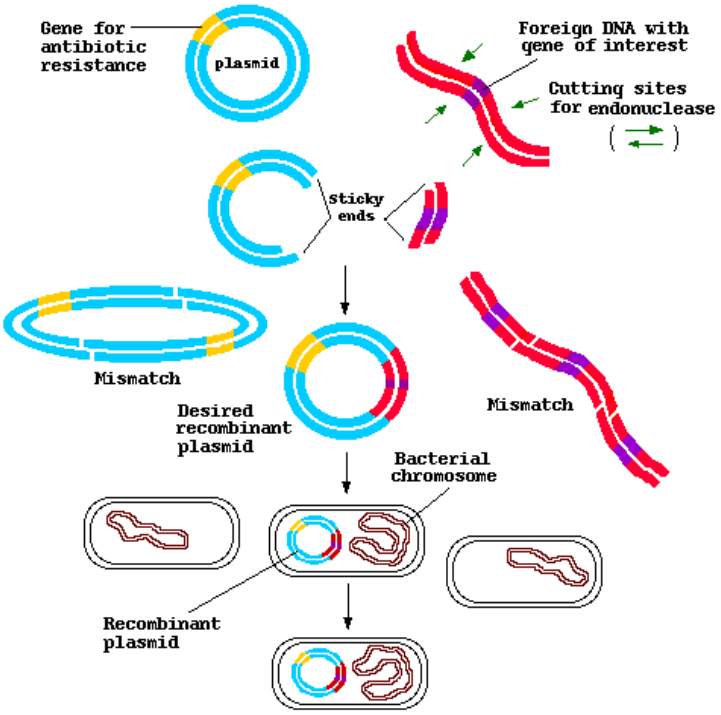
Plasmid Vectors

- Plasmid is an autonomously replicating circular extra-chromosomal DNA. They are the standard cloning vectors and the most commonly used
- Most general plasmids may be used to clone DNA insert of up to 15 kb in size

TABLE 4-2. Cloning capacity of commonly used vectors

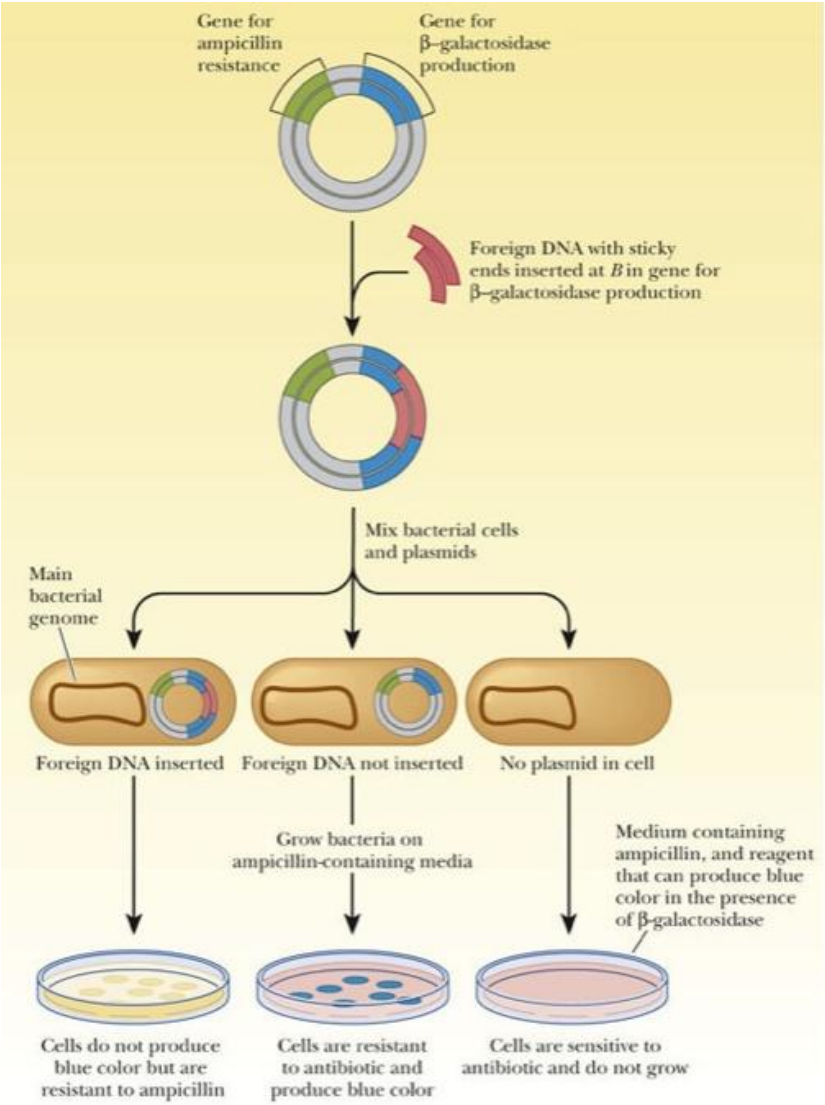
Vector	Insert size range (kb)
Plasmid	<10
Phage	<23
Cosmid	30–46
P1 artificial chromosome (PAC)	130–150
Bacterial artificial chromosome (BAC)	<300
Yeast artificial chromosome (YAC)	200–2000

Plasmid Insertion



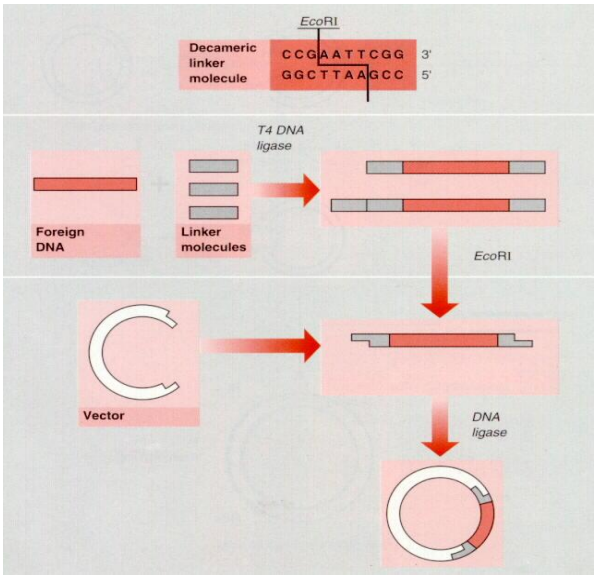
Blue/White Selection

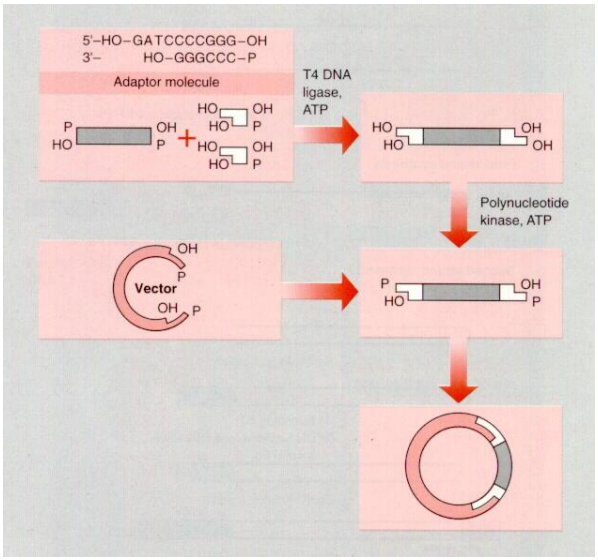
- MCS is at the 5'end of the α fragment of the β-galactosidase gene of the lac operon
- α fragment will combine with mutant β-galactosidase protein to generate a functional protein, which can metabolise X-Gal into a blue precipitate
- If inactive does not metabolise white colonies
- This process is known as insertional inactivation - cells transformed with vectors containing recombinant DNA will produce white colonies; cells transformed with non-recombinant plasmids (i.e. only the vector) grow into blue colonies



Linkers and Adapters

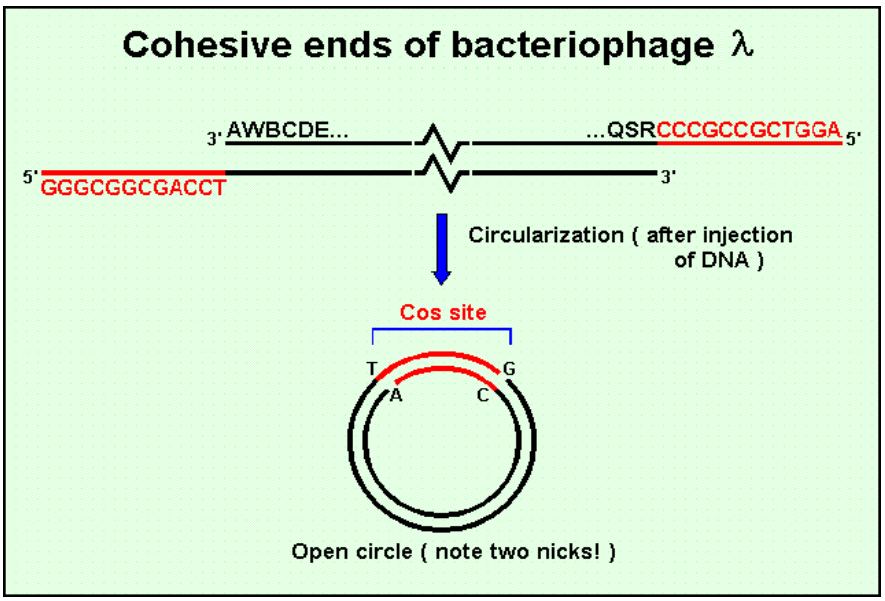
- Linkers are short stretches of double stranded DNA of length 8-14 bp and have recognition site for 3-8 RE. These linkers are ligated to blunt end DNA by ligase
- Because of the high concentration of these small molecules present in the reaction, the ligation is every efficient when compared with blunt-end ligation of large molecules
- Adaptors are linkers that have been cut





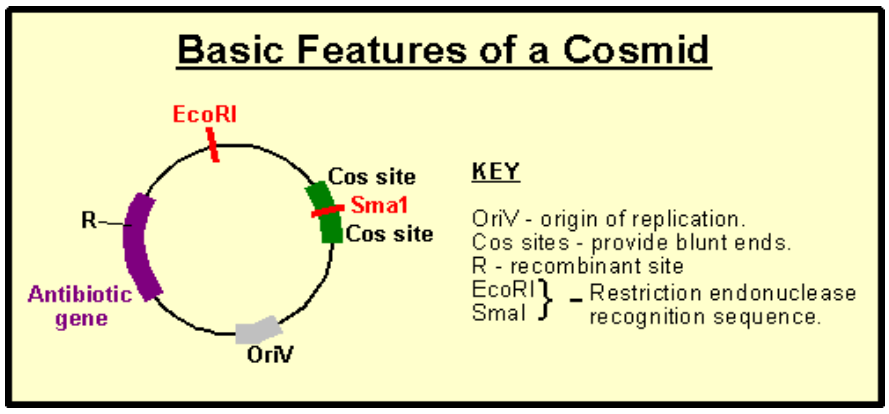
Phage Lambda

- Phage lambda is a double stranded DNA virus with a genome 50kbp
- linear molecule with single stranded complementary cohesive ends
- genome circularises and transcribes as a circle
- enters either lytic or lysogenic pathway
- major advantage of lambda is that DNA can be packaged into phages *in vitro* - more efficient than plasmid transformation



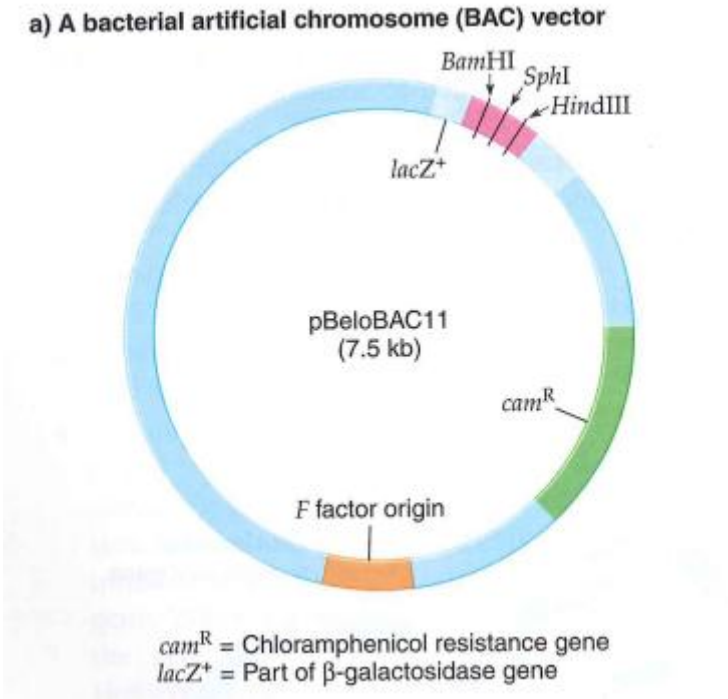
Cosmids

- A **cosmid** is a type of hybrid plasmid that contains a Lambda phage *cos sequence*
- They can replicate as plasmids if they have a suitable origin of replication
- Unlike plasmids, they can also be packaged in phage capsids, which allows the foreign genes to be transferred into or between cells by transduction
- These two single-stranded segments are the "sticky ends" of what is called the *cos* site. The *cos* site circularizes the DNA in the host cytoplasm



Bacterial Artificial Chromosome

- A DNA construct, based on a functional fertility plasmid (or F-plasmid), used for transforming and cloning in bacteria, usually *E. coli*
- Being based on the F-factor, the copy number of such chromosomes is always one
- Include the following elements:
    - *repE:* for plasmid replication and regulation of copy number
    - *parA and parB:* for partitioning F plasmid DNA to daughter cells during division and ensures stable maintenance of the BAC
    - A selectable marker: for antibiotic resistance; some BACs also have lacZ at the cloning site for blue/white selection
    - *T7 & Sp6:* phage promoters for transcription of inserted genes

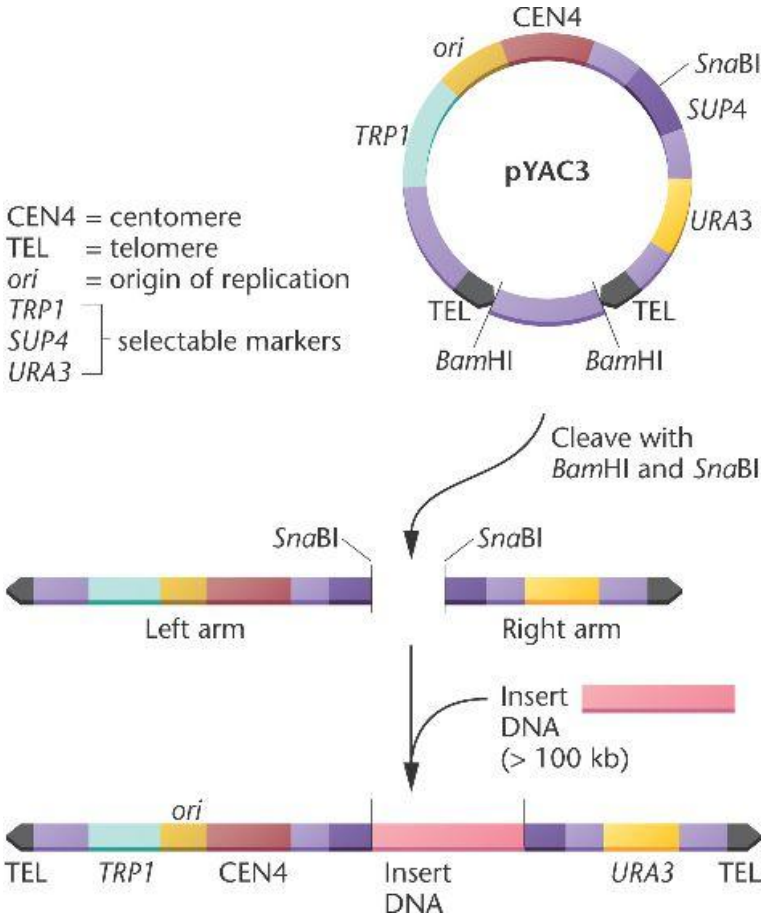


Yeast Artificial Chromosome

- Genetically engineered chromosomes derived from the DNA of the yeast, *Saccharomyces cerevisiae*, which is then ligated into a bacterial plasmid
- minimal requirements of a linear eukaryote chromosome are:
    - origin of replication (*ars*)

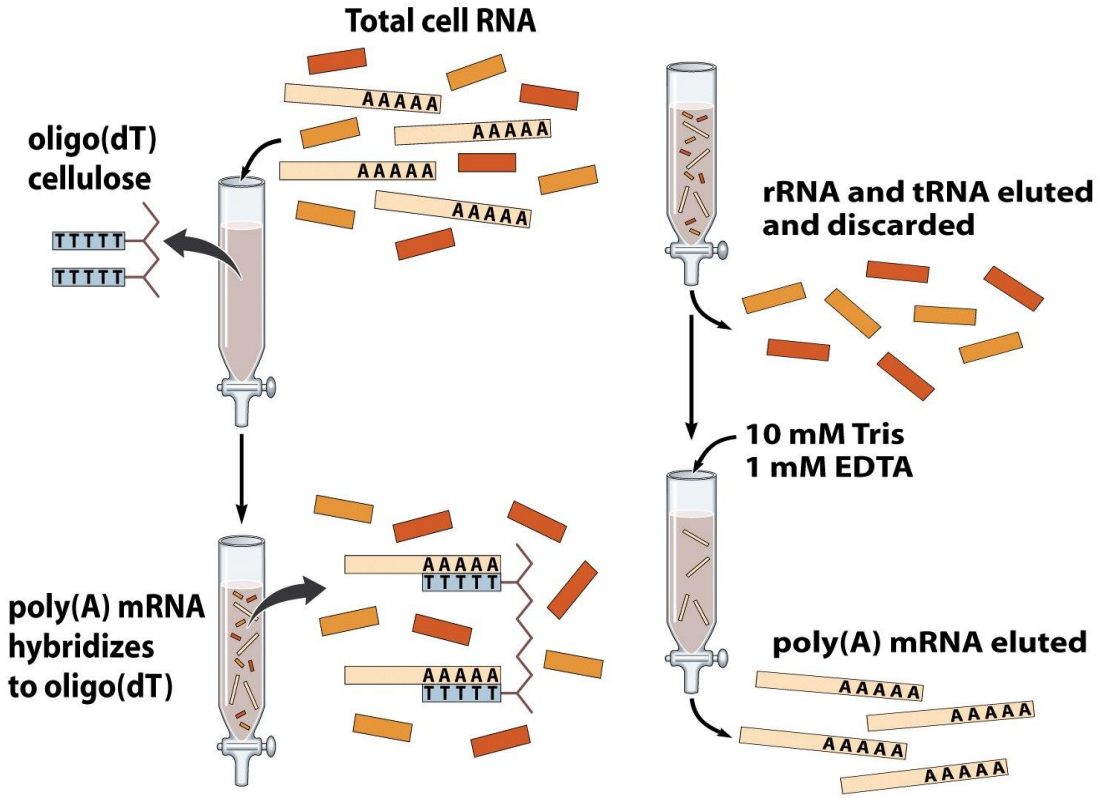
- telomeres
- a centromere

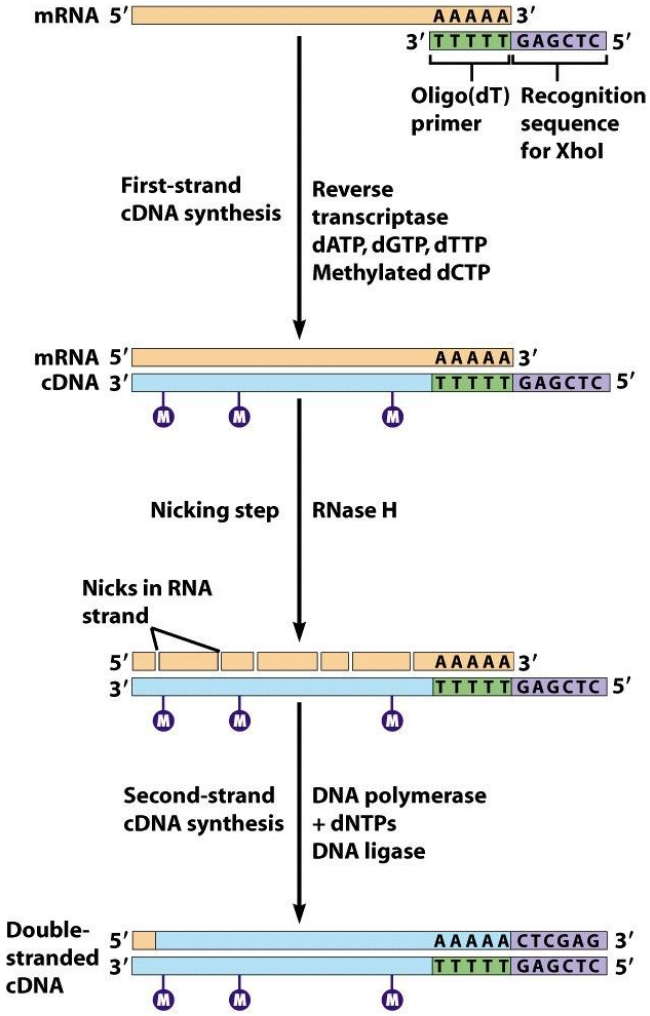
- Can hold much larger inserts, but are less stable than BACs

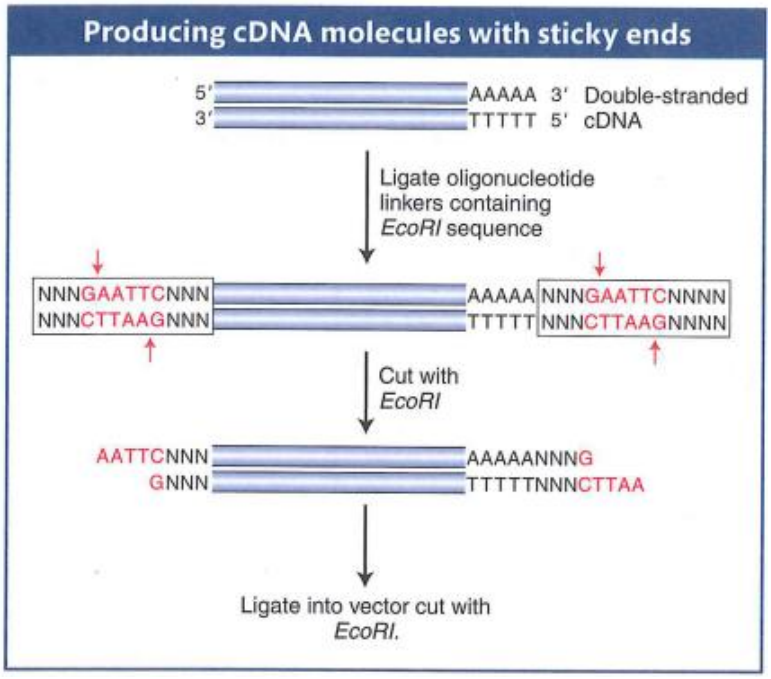


cDNA Libraries

- A **cDNA library** is a combination of cloned cDNA (complementary DNA) fragments inserted into a collection of host cells, which together constitute some portion of the transcriptome of the organism
- cDNA is produced from fully transcribed mRNA found in the nucleus and therefore contains only the expressed genes of an organism
- Firstly, the mRNA is obtained and purified from the rest of the RNAs. Several methods exist for purifying RNA such as trizol extraction and column purification. Column purification is done by using oligomeric dT nucleotide coated resins where only the mRNA having the poly-A tail will bind. The rest of the RNAs are eluted out
- Once mRNA is purified, *oligo-dT* is tagged as a complementary primer which binds to the poly-A tail providing a free 3'-OH end that can be extended by reverse transcriptase to create the complementary DNA strand
- Now, the mRNA is removed by using a RNAse enzyme leaving a single stranded cDNA. This sscDNA is converted into a double stranded DNA with the help of DNA polymerase
- Restriction endonucleases and DNA ligase are then used to clone the sequences into bacterial plasmids

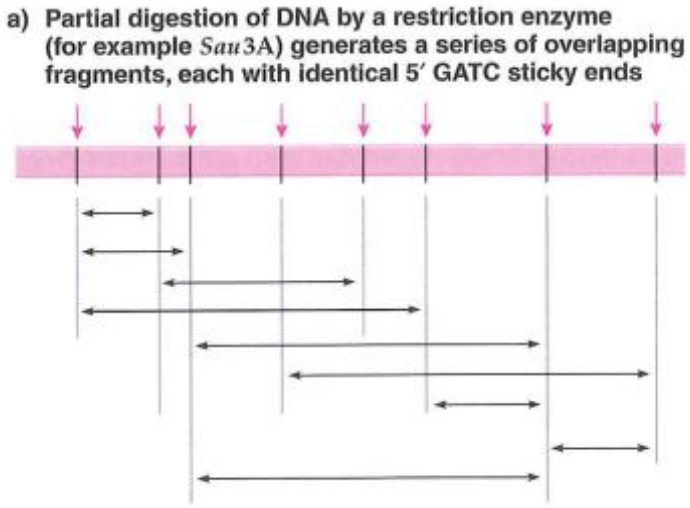

Total cell RNA
oligo(dT)
cellulose
TTTTT
TTTTT
AAAAA
AAAAA
AAAAA
AAAAA
rRNA and tRNA eluted
and discarded
10 mM Tris
1 mM EDTA
poly(A) mRNA
hybridizes
to oligo(dT)
AAAAA
TTTTT
AAAAA
TTTTT
poly(A) mRNA eluted
AAAAA
AAAAA
AAAAA
AAAAA


mRNA 5′ AAAAA 3′
3′ TTTTT GAGCTC 5′
Oligo(dT) primer
Recognition sequence for XhoI
First-strand cDNA synthesis
Reverse transcriptase
dATP, dGTP, dTTP
Methylated dCTP
mRNA 5′ AAAAA 3′
cDNA 3′ TTTTT GAGCTC 5′
M
Nicking step
RNase H
Nicks in RNA strand
5′ AAAAA 3′
3′ TTTTT GAGCTC 5′
Second-strand cDNA synthesis
DNA polymerase
+ dNTPs
DNA ligase
Double-stranded cDNA
5′ AAAAA CTCGAG 3′
3′ TTTTT GAGCTC 5′



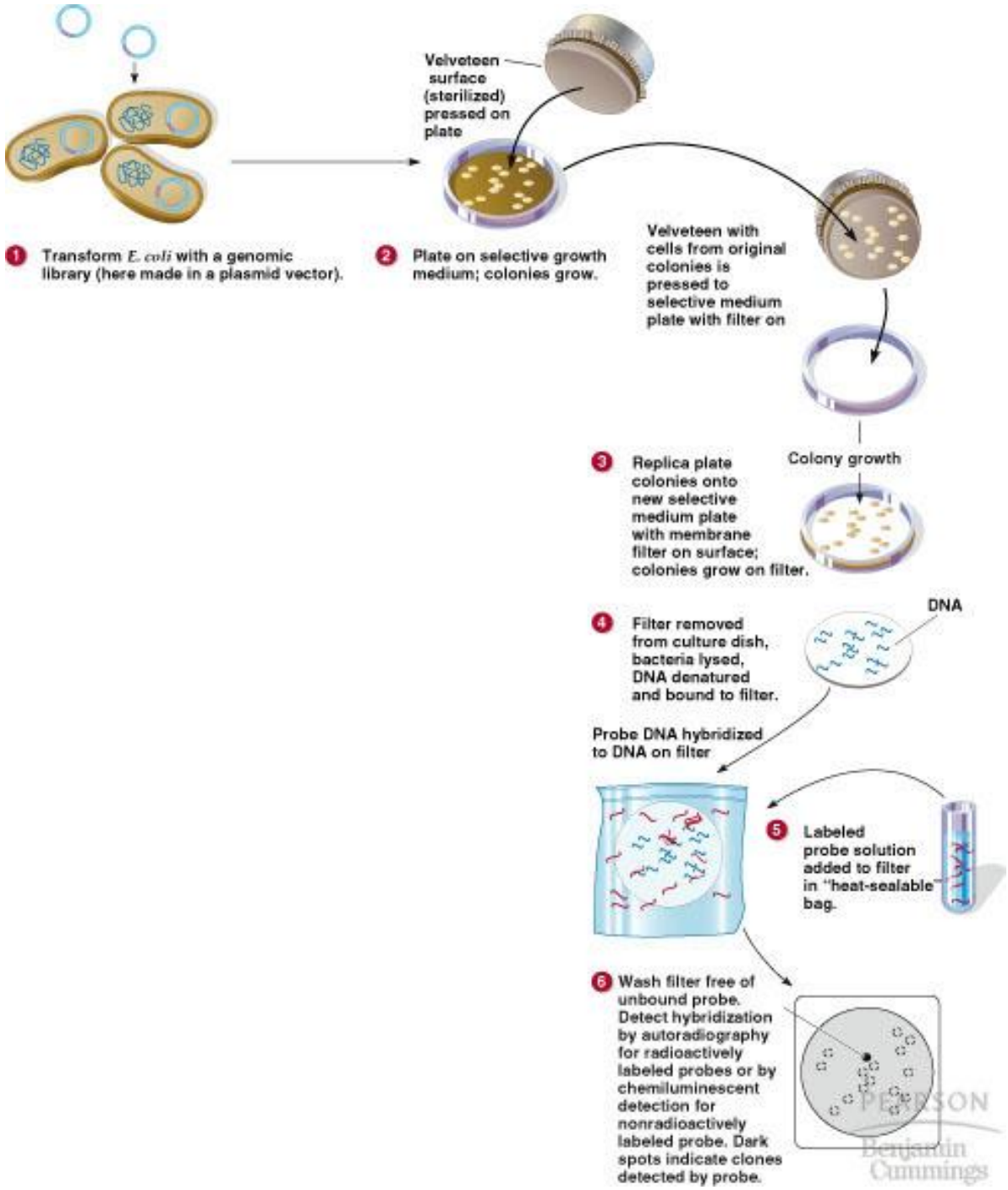
Genomic Libraries

- A genomic library is a collection of the total genomic DNA from a single organism. The DNA is stored in a population of identical vectors, each containing a different insert of DNA
- The DNA is cut by partial digestion so as to create a sequence of overlapping fragments
- Agorose gel electrophoresis is typically used to select only the fragments of the correct size for cloning
- Not all genome regions are clonable in this way. In particular, some sequences may inhibit their host cells from replicating, so these will not be preserved in the library
- Rather than whole genome libraries, chromosome libraries (one for each chromosome) are typically used for more complex organisms



$$N = \frac{ln(1-P)}{ln(1-f)}$$

Sequence-Based Screening



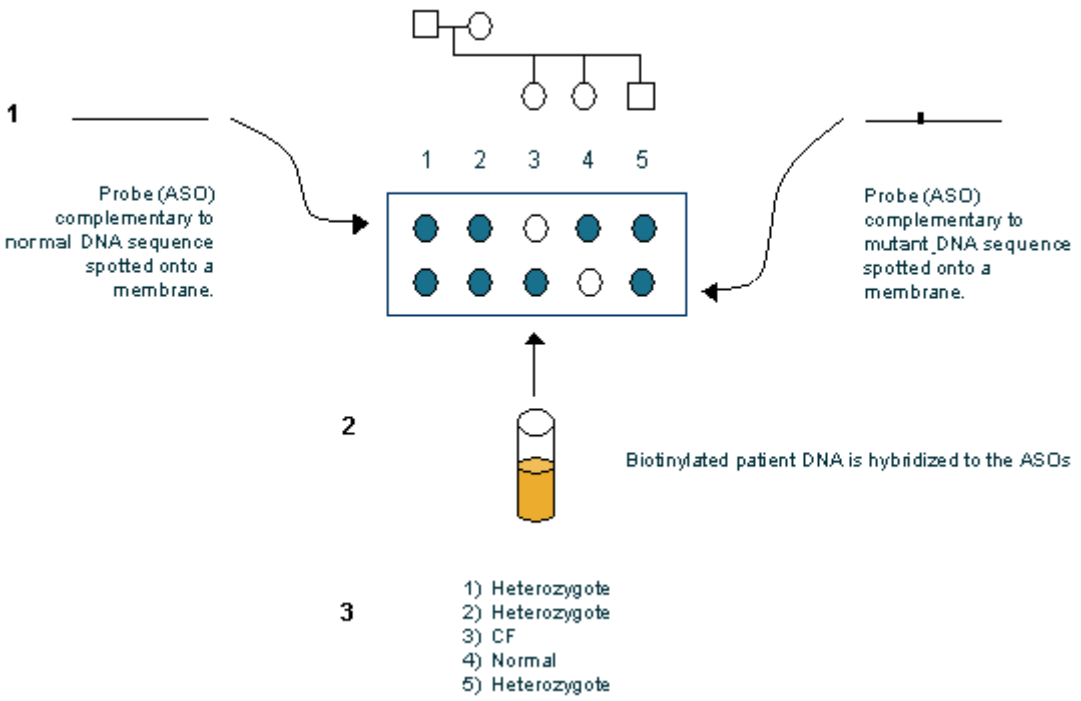
3. Gene Expression Analysis

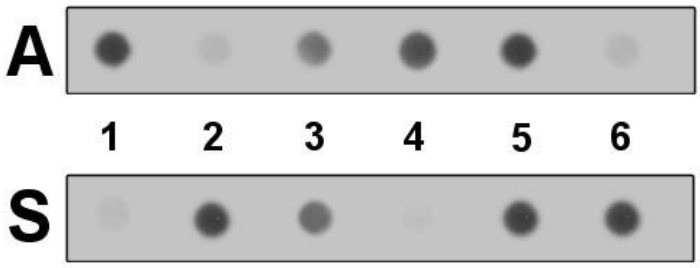
Purpose of Expression Analysis

- Gene expression analysis experiments can focus on a subset of relevant target genes, or analyze thousands of genes at once to create a global picture of cell function
- These profiles can, for example, distinguish between cells that are actively dividing, or show how the cells react to a particular treatment. Many experiments of this sort measure an entire genome simultaneously, that is, every gene present in a particular cell

Dot Blots

- A technique in molecular biology used to detect biomolecules, and for detecting, analyzing, and identifying proteins
- In a dot blot the biomolecules to be detected are not first separated by electrophoresis. Instead, a mixture containing the molecule to be detected is applied directly on a membrane as a dot, and then is spotted through circular templates directly onto the membrane or paper substrate
- This is then followed by detection by either nucleotide probes (for a northern blot and southern blot) or antibodies (for a western blot)
- Dot blots therefore can only confirm the presence or absence of a biomolecule or biomolecules which can be detected by the DNA probes or the antibody



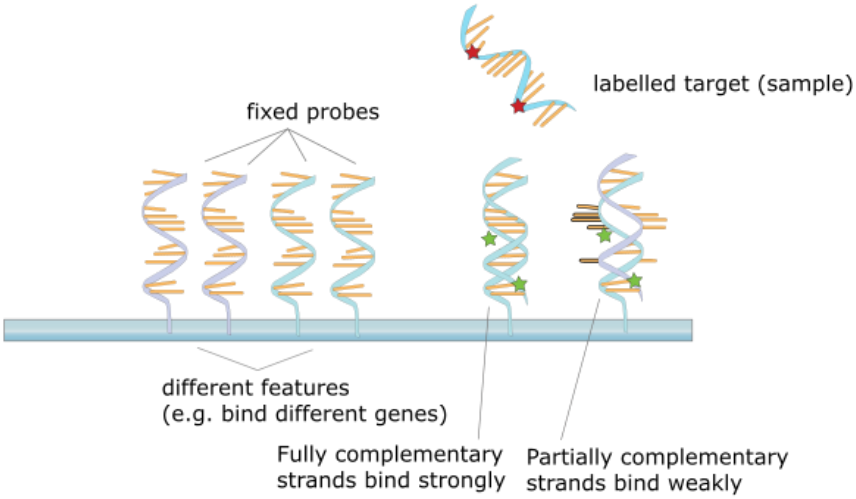


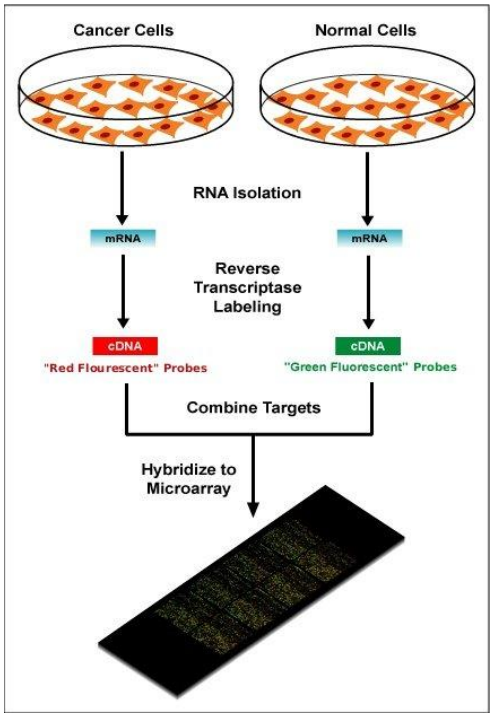
Microarrays

- Principally similar to dot blots, but on a much larger scale. Includes two subtypes: spotted arrays and photolithographic arrays
- A high-throughput method used to track the interactions and activities of proteins, and to determine their function, and determining function on a large scale. Its main advantage lies in the fact that large numbers of proteins can be tracked in parallel. The chip consists of a

support surface such as a glass slide, nitrocellulose membrane, bead, or microtitre plate, to which an array of capture proteins is bound

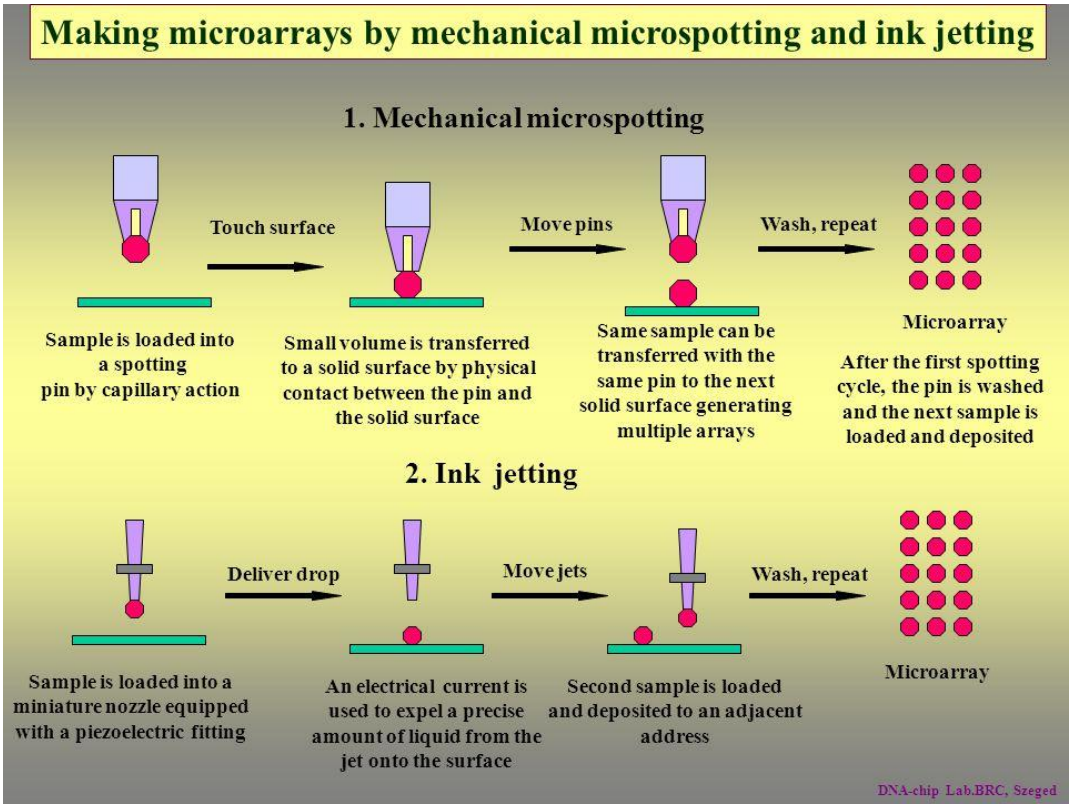
- Chips allow true parallelism, miniaturization, multiplexing and automation, and these key features provide a set of performance specifications 2–6 that cannot be achieved with the earlier technologies
- Expressed sequences account for only ~3% of the genome, hybridization-based transcript analysis effectively reduces the complexity of the human genome by ~30-fold
- DNA microarrays and quantitative PCR exploit the preferential binding or "base pairing" of complementary nucleic acid sequences, and both are used in gene expression profiling
- While high throughput DNA microarrays lack the quantitative accuracy of qPCR, it takes about the same time to measure the gene expression of a few dozen genes via qPCR as it would to measure an entire genome using DNA microarrays
- Thus, often semi-quantitative microarray experiments identify candidate genes, then qPCR is performed on some of the most interesting candidate genes

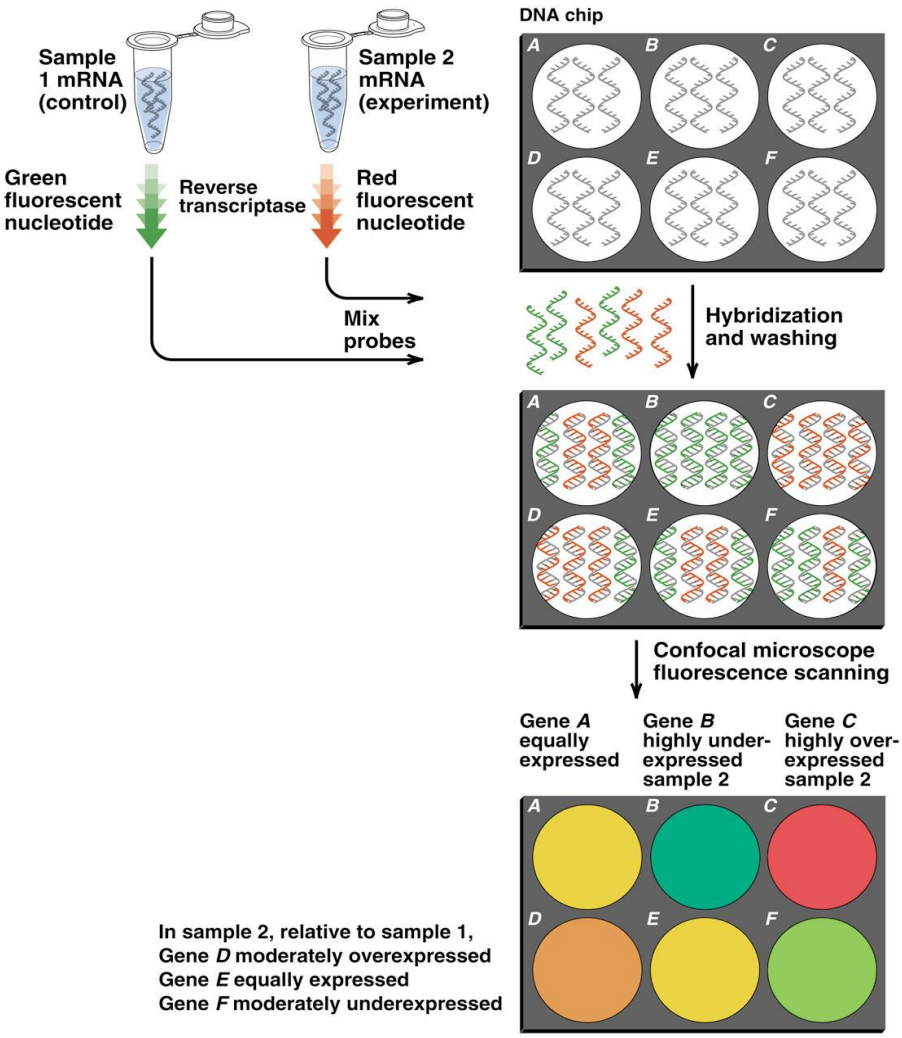




Spotted Arrays

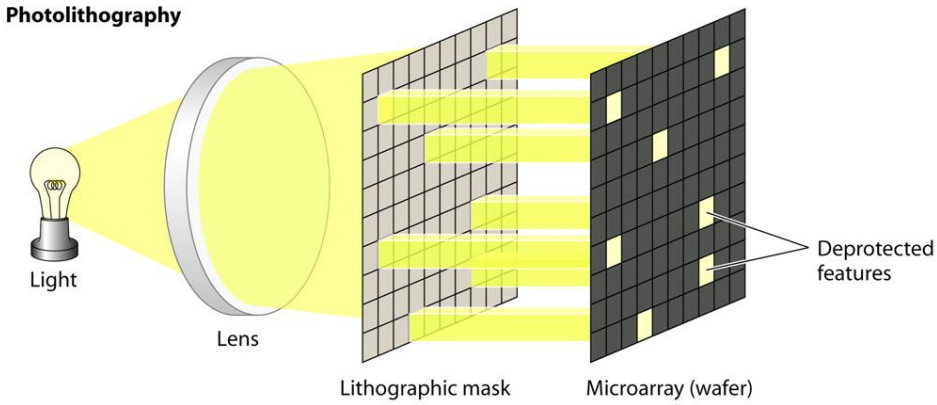
- Genes to be assayed are obtained as set of plasmid cDNA clones, which are PCR amplified, and dsDNA fragments are denatured using DMSO
- ssDNA fragments spotted onto glass slide using robotic device
- nanolitres of DNA deposited using split pins, so the liquid dries very quickly and the DNA becomes attached to the glass surface
- cDNA arrays can be prepared directly from existing cDNA libraries, a large number of which are in the public domain
- not as high density as photolithography, but DNA segments are longer so binding is more specific
- Spotted array experiments usually use a control and an experimental sample, each hydridised with a different colour phourescent label. The relative quantities of each in the microarray in each cell can then be measured in a way that automatically controls for differences in baseline concentration of each gene

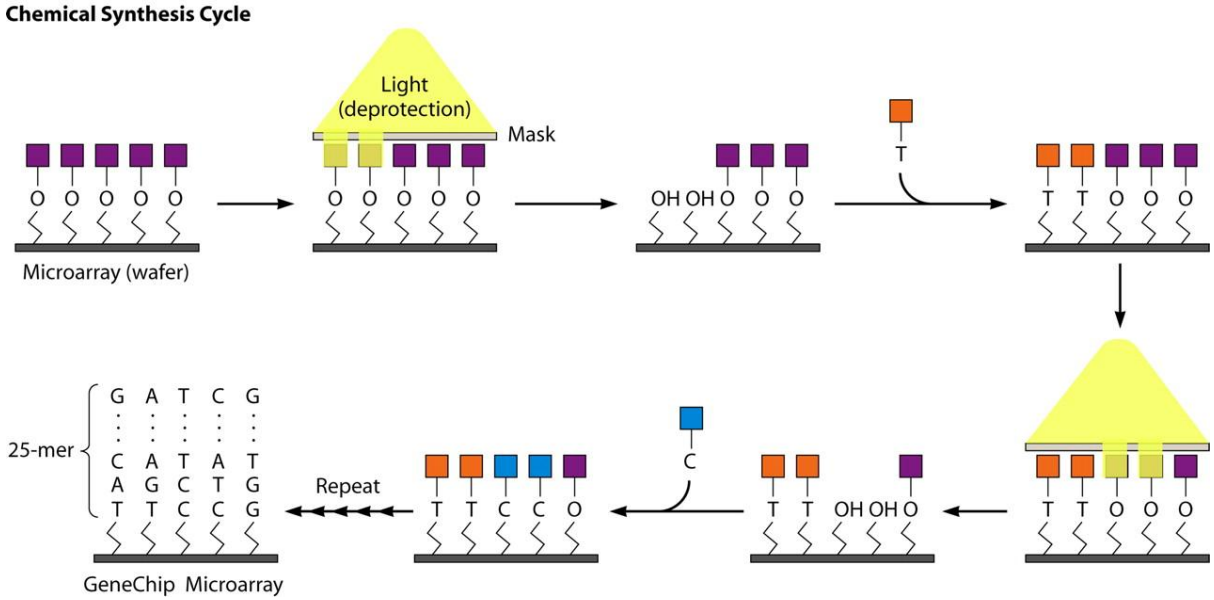


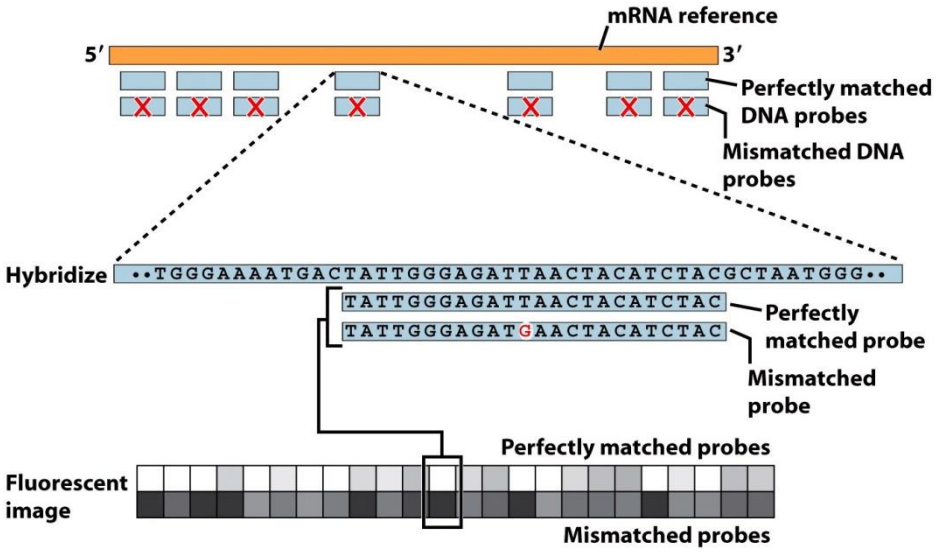


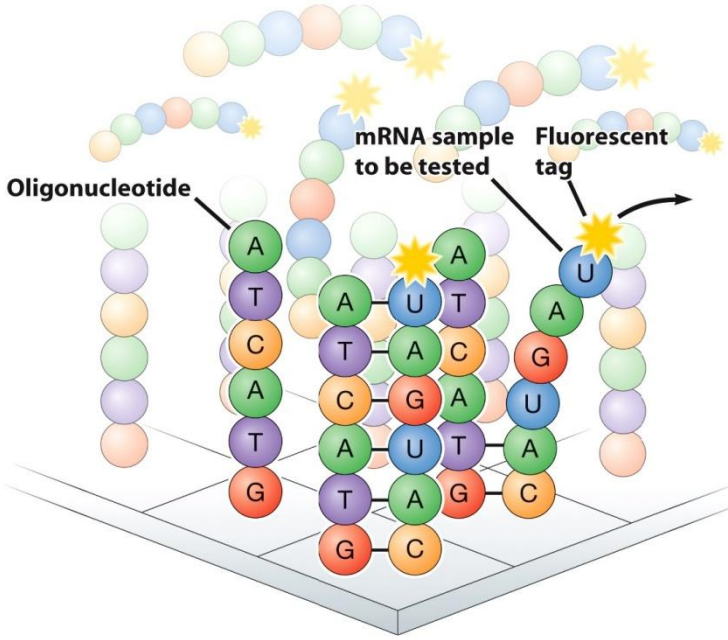
Photolithographic Arrays

- Light is used in association with photomasks, opaque plates with holes or transparencies that allow light to shine through in a defined pattern. A series of chemical treatments then enables deposition of the protein in the desired pattern upon the material underneath the photomask
- Single Chip has more than a million regularly spaced oligonucleotide sequences called features
- Such arrays typically contain 22 different oligonucleotides that average 25 nucleotides in length for each gene to be tested - 11 match and 11 mismatch probes per gene (centre nucleotide replaced in mismatch probes)
- The mismatch probes help establish background signal for each correctly matching probe
- Signal of corresponding mismatch probe is subtracted from matching probe and then average signal for each gene calculated from 11 probe sets, with the 11 providing some redundancy and control for non-specific binding


Photolithography
Light
Lens
Lithographic mask
Microarray (wafer)
Deprotected features


Chemical Synthesis Cycle
Light (deprotection)
Mask
Microarray (wafer)
25-mer
Repeat
GeneChip Microarray


mRNA reference
5′
3′
Perfectly matched DNA probes
Mismatched DNA probes
Hybridize
Perfectly matched probe
Mismatched probe
Perfectly matched probes
Fluorescent image
Mismatched probes

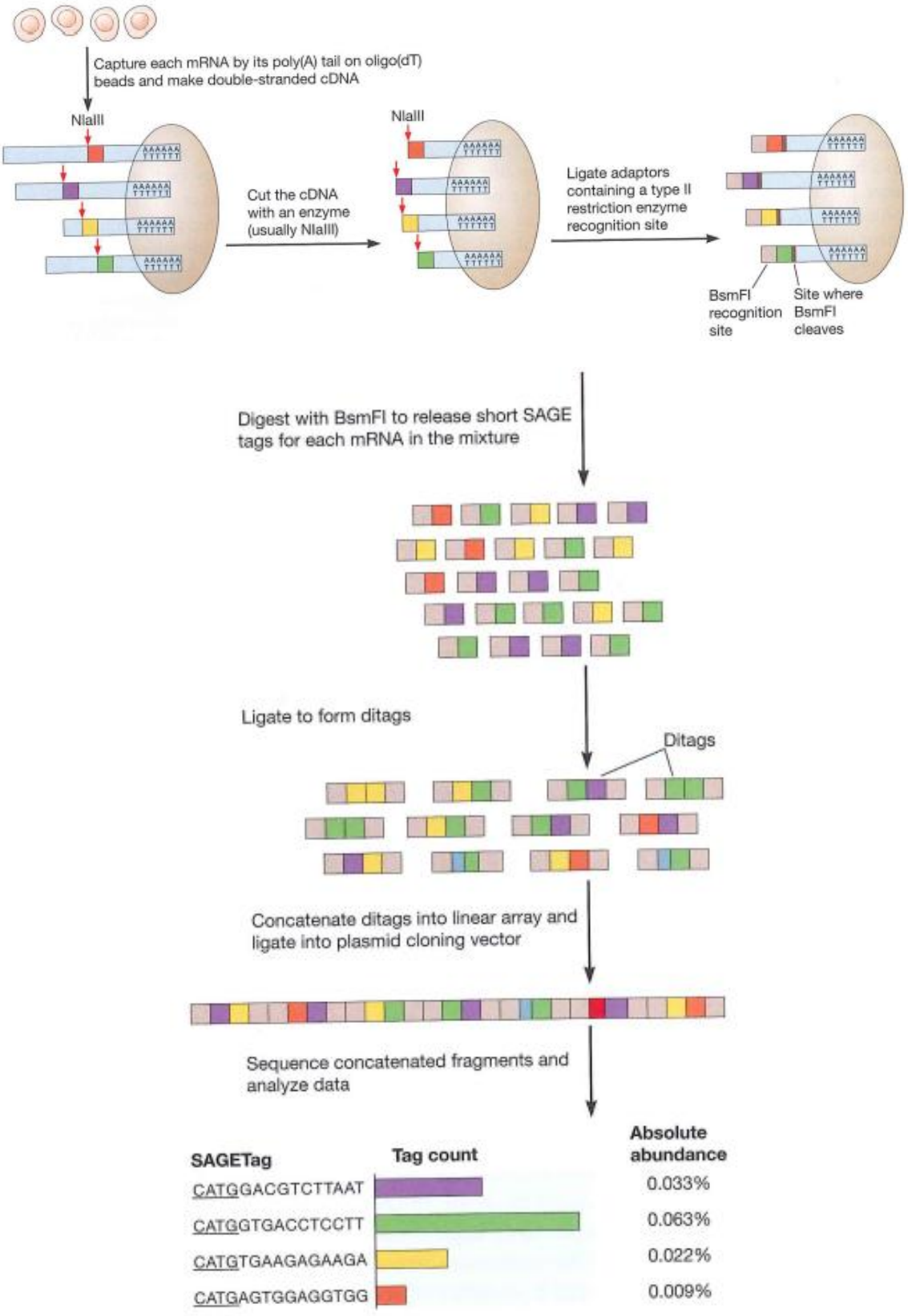


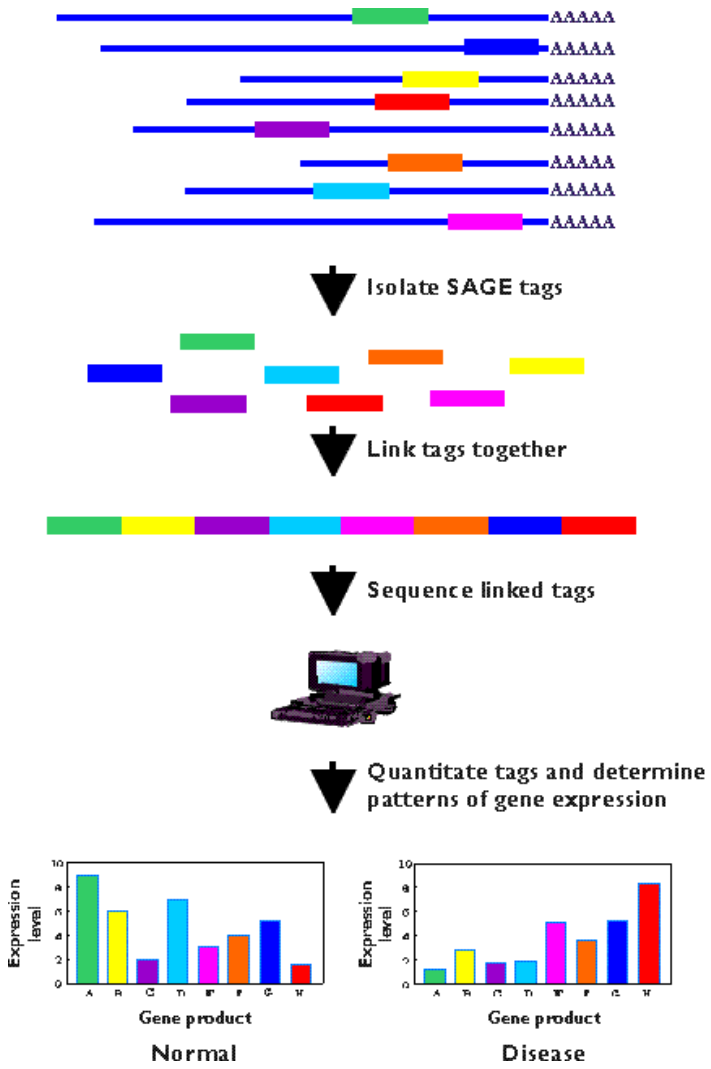
Analysing the Data

- Single experiment with 20,000 genes and 2 time points produces hundreds of data points one of the most useful and commonly used way to look at the data is hierarchical clustering
- Genes ranked in table according to similarity of expression levels in single or multiple experiments under different experimental conditions
- It should also be noted that replicate measurements from a pool of tissue only provides information about variability stemming from measurement error, and provides no information about variability that stems from population heterogeneity

SAGE

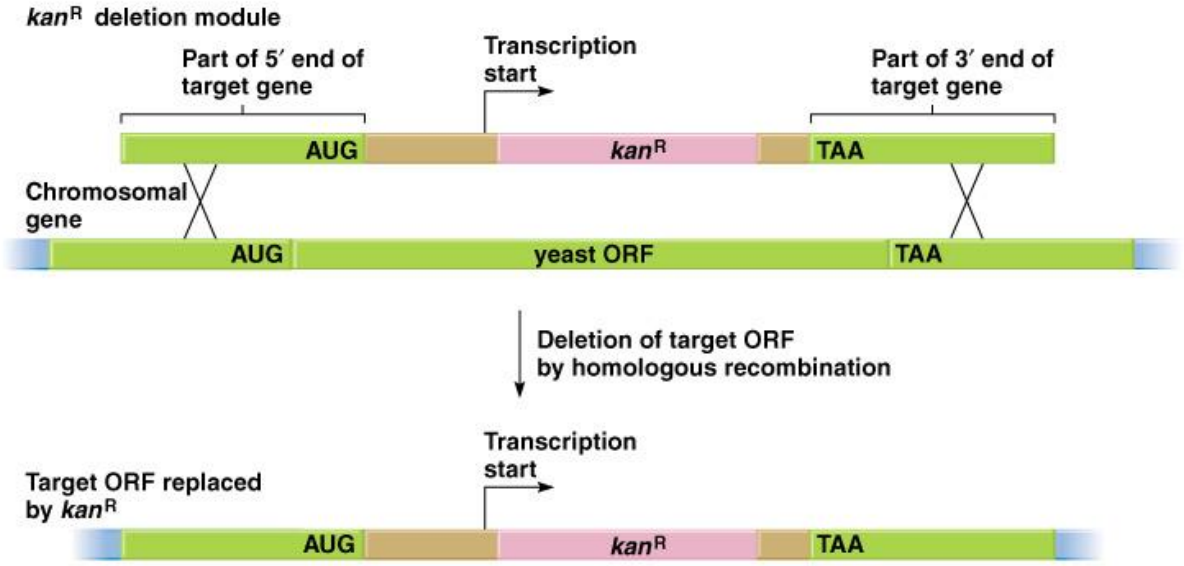
- Serial Analysis of Gene Expression which uses plasmid libraries are made from mRNA pool of a cell and large number of clones are sequenced the number of times a sequence is present for a particular gene is an indication of its abundance in original mRNA pool
- Feasibility depends on how much DNA sequencing can be done (research budget) - large number of reads is needed for transcripts expressed at low levels
- Polyadenylated mrnas are captured by binding them to oligo(dt) containing beads
- Mrnas are converted to cdnas with RT, and then cdnas cleaved with frequent cutter
- Short dsdna linkers containing site for bsmfi are attached to these cdnas
- Bsmfi recognizes particular site but cleaves several bases away from it thus releasing tags from the beads
- Two tags are ligated tail-to-tail to form ditags
- Ditags are amplified by PCR and ligated to form concatemers with 20-25 ditags
- Pool of such long concatemers is cloned into a plasmid vector and sequenced
- Individual tag sequences are mapped back to the genome, associated with a transcript, and added to a tally
- Abundance of each tag is calculated


Capture each mRNA by its poly(A) tail on oligo(dT) beads and make double-stranded cDNA
NlaIII
Cut the cDNA with an enzyme (usually NlaIII)
NlaIII
Ligate adaptors containing a type II restriction enzyme recognition site
BsmFI recognition site
Site where BsmFI cleaves
Digest with BsmFI to release short SAGE tags for each mRNA in the mixture
Ligate to form ditags
Ditags
Concatenate ditags into linear array and ligate into plasmid cloning vector
Sequence concatenated fragments and analyze data
SAGETag
Tag count
Absolute abundance
CATGGACGTCTTAAT
0.033%
CATGGTGACCTCCTT
0.063%
CATGTGAAGAGAAGA
0.022%
CATGAGTGGAGGTGG
0.009%



Yeast Gene Knockouts

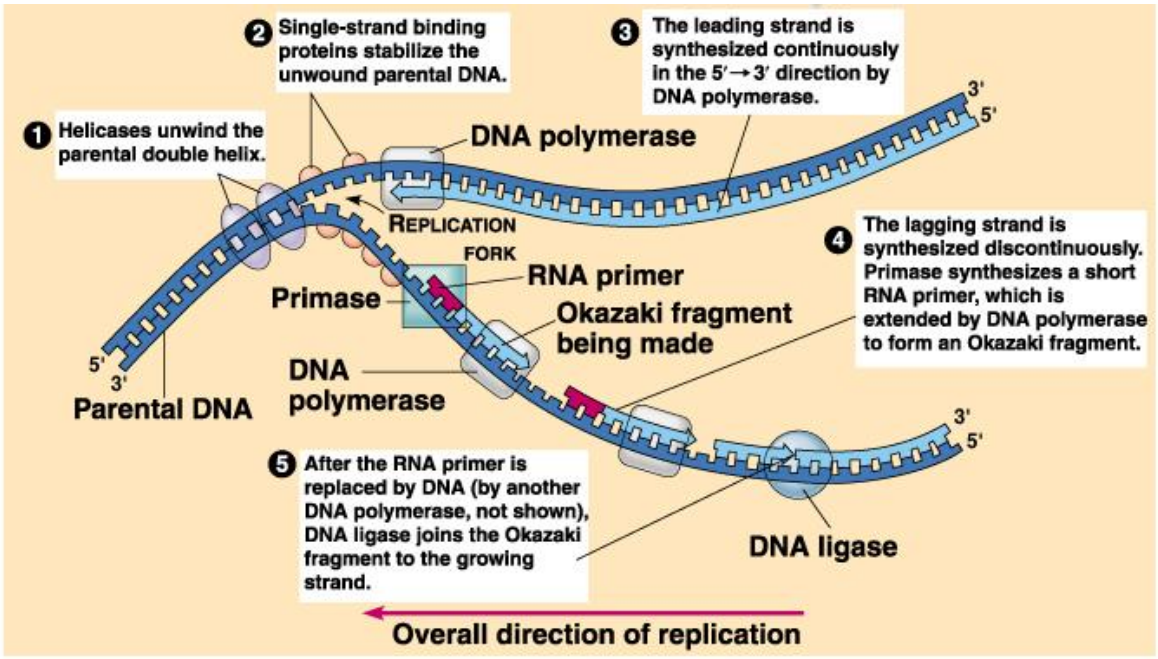
- The complete sequence of gene to be knocked is required, and based on the sequence, an artificial linear deletion module is constructed & amplified
- This module will have part of upstream gene sequence including the start codon, DNA fragment coding for selectable marker (*e.g.* kanamycin resistance), and part of downstream gene sequence including the stop codon
- This artificial module is transformed into yeast cells
- During normal DNA replication, the artificial module replaces target gene ORF by homologous recombination
- Yeast cells carrying the artificial module (deleted ORF) can be selected by screening for inserted marker (*e.g.* growing on kanamycin)
- Since major part of target gene ORF in these transformants has been deleted, they will lose the ability to perform target gene function; thus creating a 'knockout'

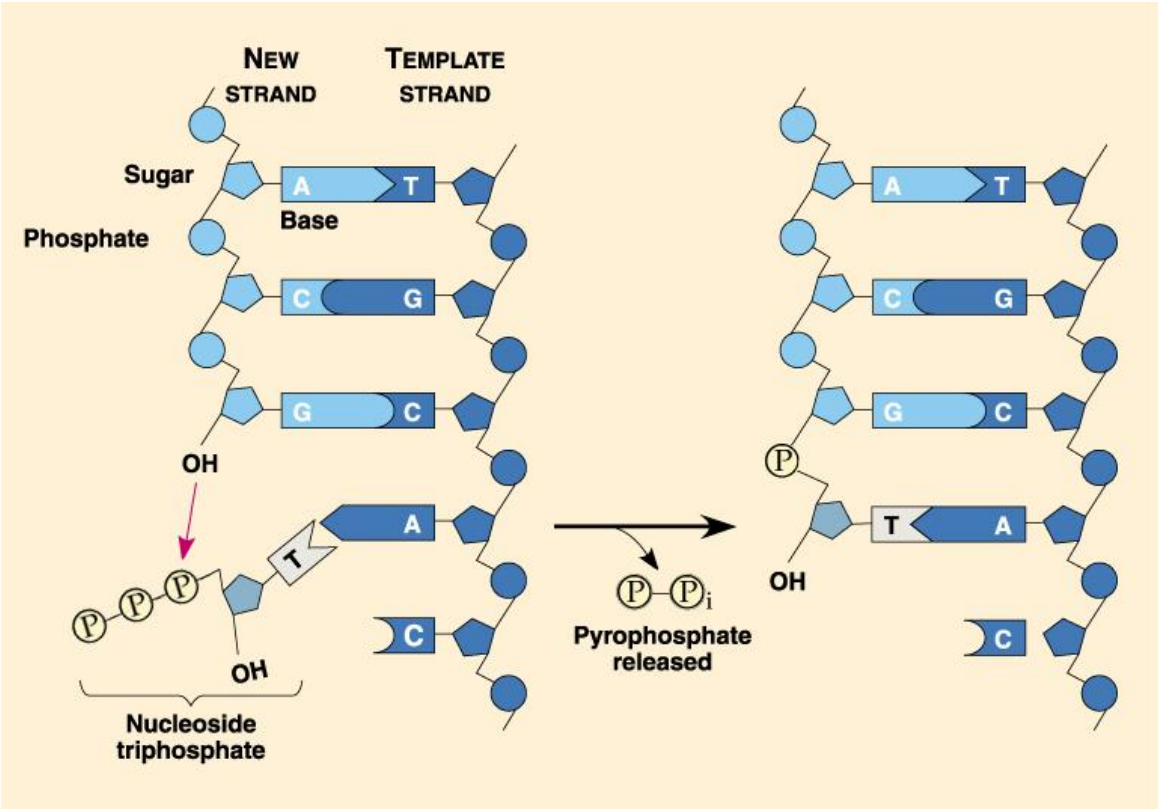


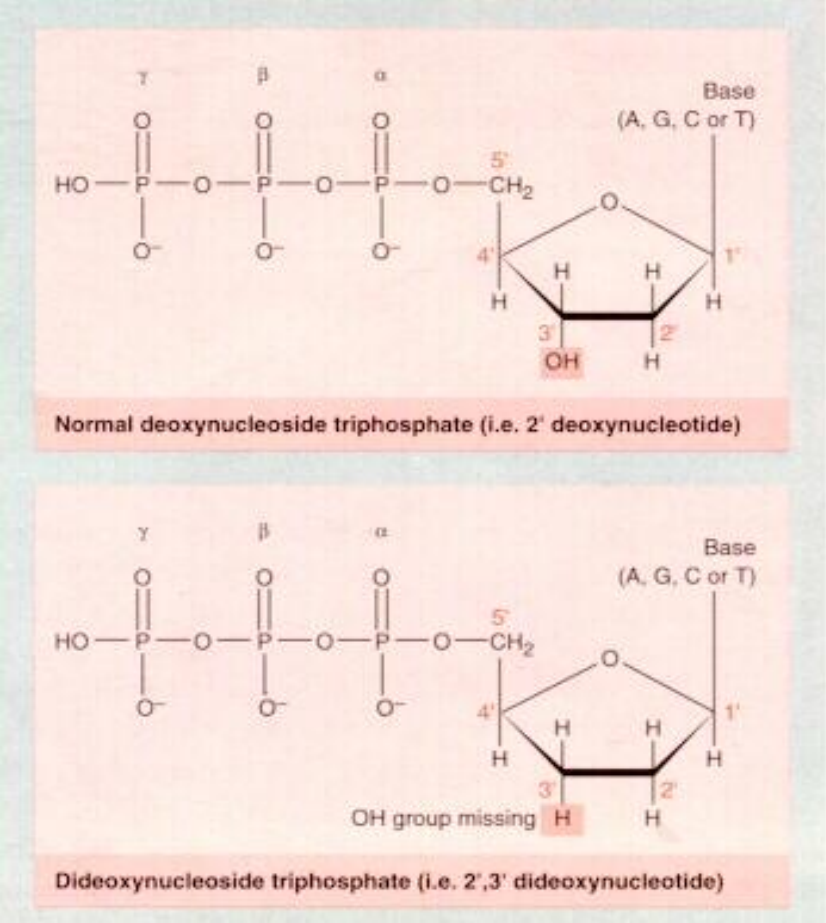
4. Next Generation Sequecing

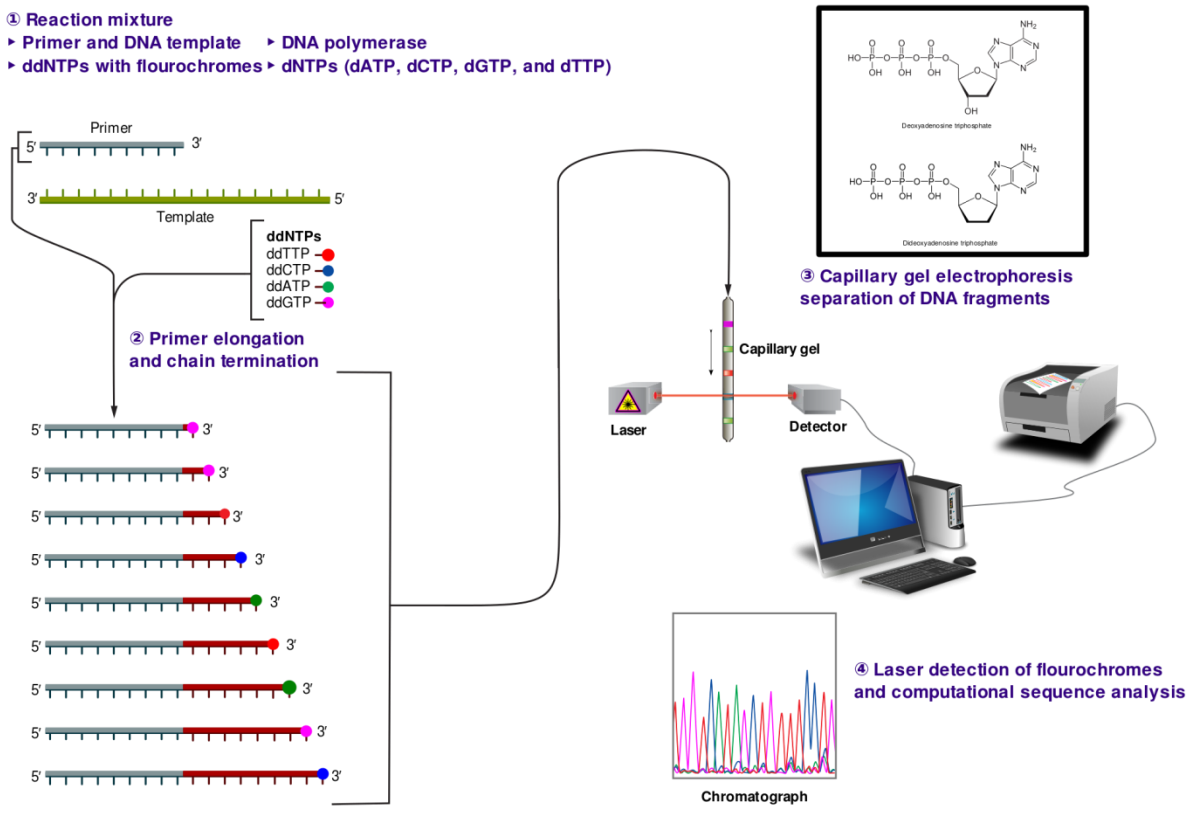
Sanger Sequencing

- The classical chain-termination method requires a single-stranded DNA template, a DNA primer, a DNA polymerase, normal deoxynucleosidetriphosphates (dNTPs), and modified di-deoxynucleosidetriphosphates (ddNTPs), the latter of which terminate DNA strand elongation
- These chain-terminating nucleotides lack a 3'-OH group required for the formation of a phosphodiester bond between two nucleotides, causing DNA polymerase to cease extension of DNA when a modified ddNTP is incorporated
- The ddNTPs may be radioactively or fluorescently labeled for detection in automated sequencing machines
- The DNA sample is divided into four separate sequencing reactions, containing all four of the standard deoxynucleotides and the DNA polymerase. To each reaction is added only one of the four dideoxynucleotides, while the three other nucleotides are ordinary ones
- Following rounds of template DNA extension from the bound primer, the resulting DNA fragments are heat denatured and separated by size using gel electrophoresis


2 Single-strand binding proteins stabilize the unwound parental DNA.
3 The leading strand is synthesized continuously in the 5′→3′ direction by DNA polymerase.
1 Helicases unwind the parental double helix.
DNA polymerase
3′
5′
REPLICATION FORK
RNA primer
Primase
Okazaki fragment being made
4 The lagging strand is synthesized discontinuously. Primase synthesizes a short RNA primer, which is extended by DNA polymerase to form an Okazaki fragment.
5′
3′
Parental DNA
DNA polymerase
3′
5′
5 After the RNA primer is replaced by DNA (by another DNA polymerase, not shown), DNA ligase joins the Okazaki fragment to the growing strand.
DNA ligase
Overall direction of replication


NEW STRAND
TEMPLATE STRAND
Sugar
Phosphate
Base
A
T
C
G
G
C
OH
A
T
P
P
P
OH
Nucleoside triphosphate
C
P
Pi
Pyrophosphate released
A
T
C
G
G
C
P
T
A
OH
C


γ
β
α
Base
(A, G, C or T)
HO
5'
CH2
4'
1'
3'
2'
OH
Normal deoxynucleoside triphosphate (i.e. 2' deoxynucleotide)
OH group missing
H
Dideoxynucleoside triphosphate (i.e. 2',3' dideoxynucleotide)

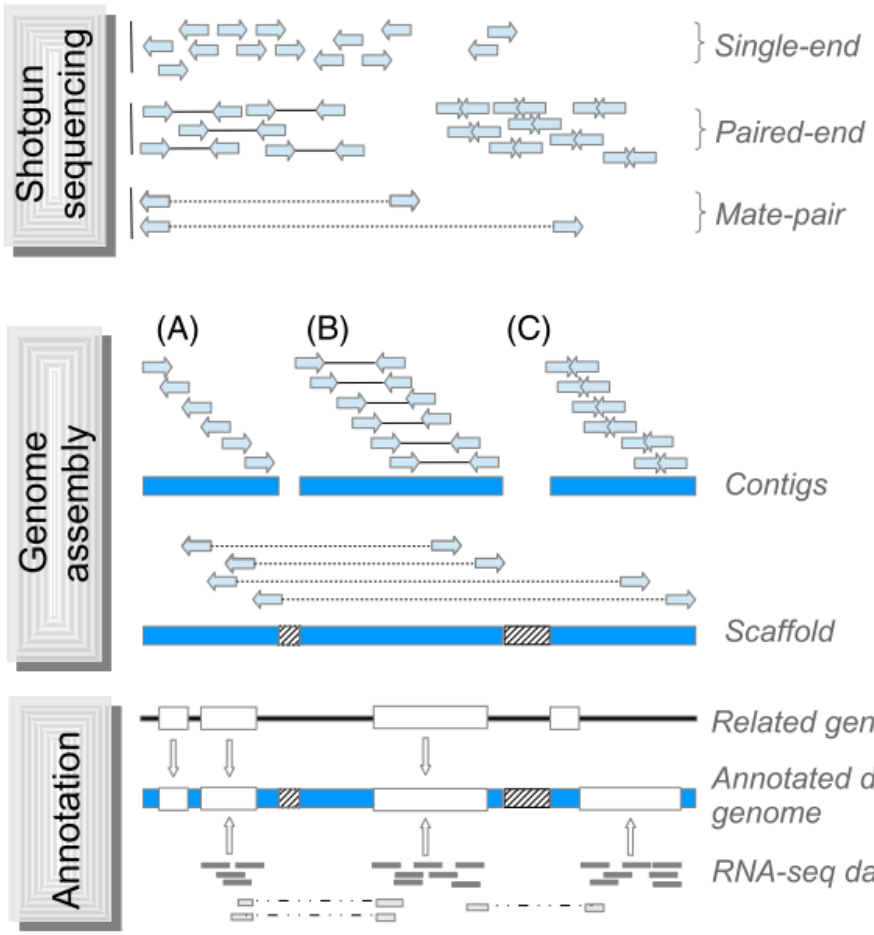

① Reaction mixture
‣ Primer and DNA template
‣ DNA polymerase
‣ ddNTPs with flourochromes
‣ dNTPs (dATP, dCTP, dGTP, and dTTP)
Primer
5'
3'
Template
ddNTPs
ddTTP
ddCTP
ddATP
ddGTP
② Primer elongation
and chain termination
Deoxyadenosine triphosphate
Dideoxyadenosine triphosphate
③ Capillary gel electrophoresis
separation of DNA fragments
Capillary gel
Laser
Detector
Chromatograph
④ Laser detection of flourochromes
and computational sequence analysis

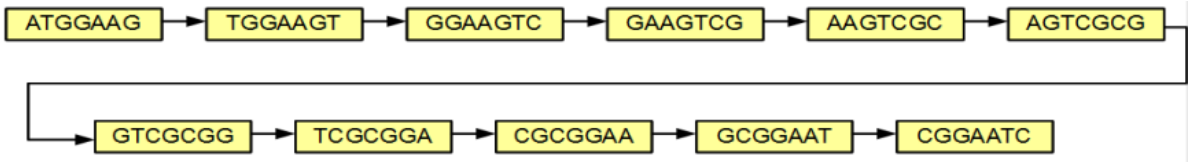
Characteristics of New Gen Processes

- Preparation of libraries occurs in a cell free system, not requiring bacterial cloning
- Many millions of sequencing reactions are performed in parallel
- Sequencing output is directly detected without needing to use electrophoresis
- Generally have shorter reads than Sanger, so require more sophisticated alignment algorithms
- Much lower sample quantities needed - often nanograms are sufficient
- Dramatic reduction in cost: Sanger sequencing $500/Mb (USD), NGS $0.50/Mb

Technology Overview and Assembly

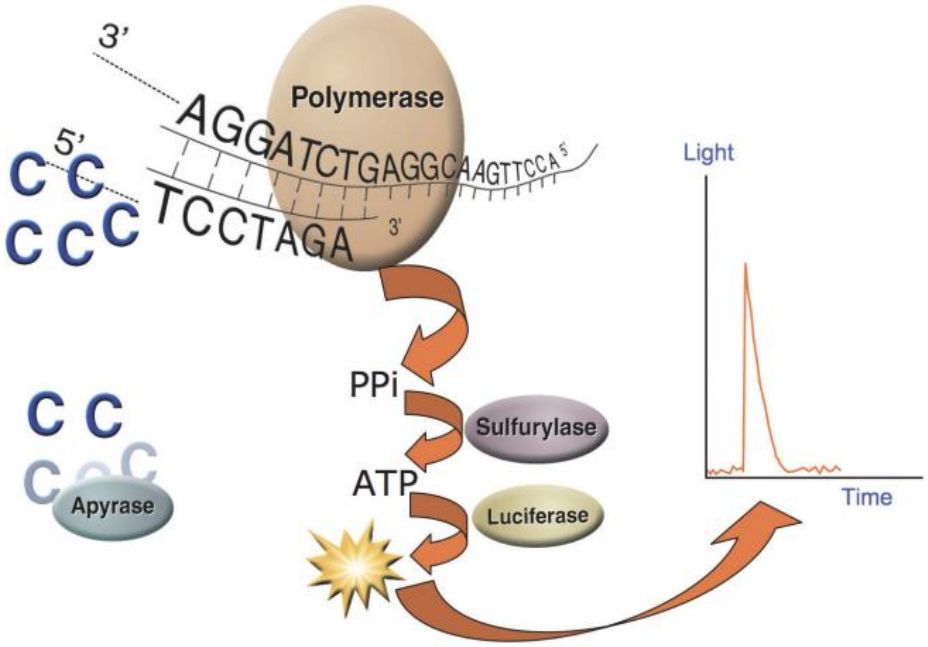
- Currently, most genome projects use a shotgun sequencing strategy for genome sequencing
- Firstly, genomic DNA is sheared into small random fragments. Then, depending on the technology, these are sequenced independently to a given length
- Usually, longer fragments (several hundred base pairs) are sequenced from both ends (paired-end sequencing) to provide additional information
- Computer algorithms are then utilized to piece the resulting sequence reads back together into longer continuous stretches of sequence (contigs), a process called de novo assembly
- Naturally, for longer sequence reads, more overlap can be expected, reducing the required raw read depth
- Next, libraries from long DNA fragments spanningseveral kilobases (kb) of sequence in the genome are prepared and their endpoints sequenced. Depending on the technology, these libraries are (somewhat confusingly) called, paired-end, mate-pair or jump libraries
- If the endpoint sequences of several independent fragments come to lie on two different contigs, they are joined into a scaffold. Any gaps are filled with the generic 'N' character
- In a last step, the resulting scaffolds are often joined into linkage groups or placed on chromosomes
- As a rough guideline for mammalian genomes, it has been proposed to use at least 45 9 coverage of short-insert paired-end libraries, 45 times coverage of medium-sized insert libraries (3 – 10 kb) and 1 – 5 times coverage of long-insert libraries (10 – 40 kb)
- Use de bruijn graph method: reads are partitioned into k-mers (substrings of the read sequence of length k)-form the nodes of the graph (network) and are linked when sharing a k-1 mer





Sequencing Methods

- Pyrosequencing: based on detecting the activity of DNA polymerase (a DNA synthesizing enzyme) with another chemoluminescent enzyme. Essentially, the method allows sequencing of a single strand of DNA by synthesizing the complementary strand along it, one base pair at a time, and detecting which base was actually added at each step. The template DNA is immobile, and solutions of A, C, G, and T nucleotides are sequentially added and removed from the reaction. Light is produced only when the nucleotide solution complements the first unpaired base of the template



The enzyme cascade of the Pyrosequencing technology.

- Illumina dye sequencing: begins with the attachment of DNA molecules to primers on a slide, followed by amplification of that DNA to produce local colonies. The four types (adenine, cytosine, guanine, and thymine) of reversible terminate bases are added, each fluorescently labeled with a different color and attached with a blocking group. The four bases then compete for binding sites on the template DNA to be sequenced and non-incorporated molecules are washed away. After each synthesis, a laser is used to excite the dyes and a photograph of the incorporated base is taken. A chemical deblocking step is then used in the removal of the 3' terminal blocking group and the dye in a single step. The process is repeated until the full DNA molecule is sequenced

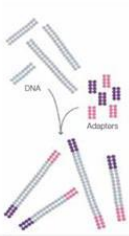


Figure 1
Randomly fragment genomic DNA and ligate adapters to both ends of the fragments.

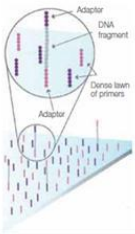


Figure 2
Bind single-stranded fragments randomly to the inside surface of the flow cell channels.

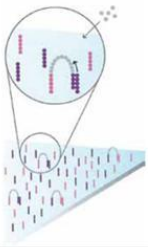


Figure 3
Add unlabeled nucleotides and enzyme to initiate solid-phase bridge amplification.

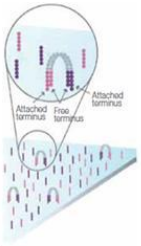


Figure 4
The enzyme incorporates nucleotides to build double-stranded bridges on the solid-phase substrate.

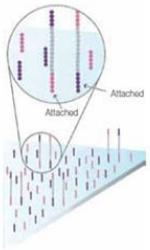


Figure 5
Denaturation leaves single-stranded templates anchored to the substrate.

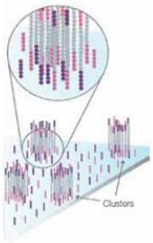


Figure 6
Several million dense clusters of double-stranded DNA are generated in each channel of the flow cell.

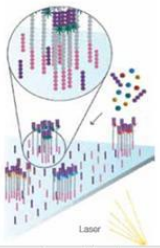


Figure 7
The first sequencing cycle begins by adding four labeled reversible terminators, primers, and DNA polymerase.

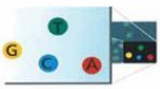


Figure 8
After laser excitation, the emitted fluorescence from each cluster is captured and the first base is identified.

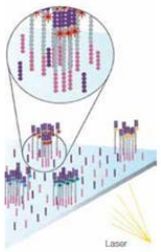


Figure 9
The next cycle repeats the incorporation of four labeled reversible terminators, primers, and DNA polymerase.

Figure 10
After laser excitation, the image is captured as before, and the identity of the second base is recorded.

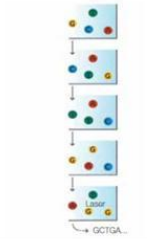


Figure 11
The sequencing cycles are repeated to determine the sequence of bases in a fragment, one base at a time.

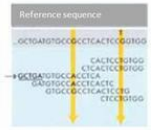
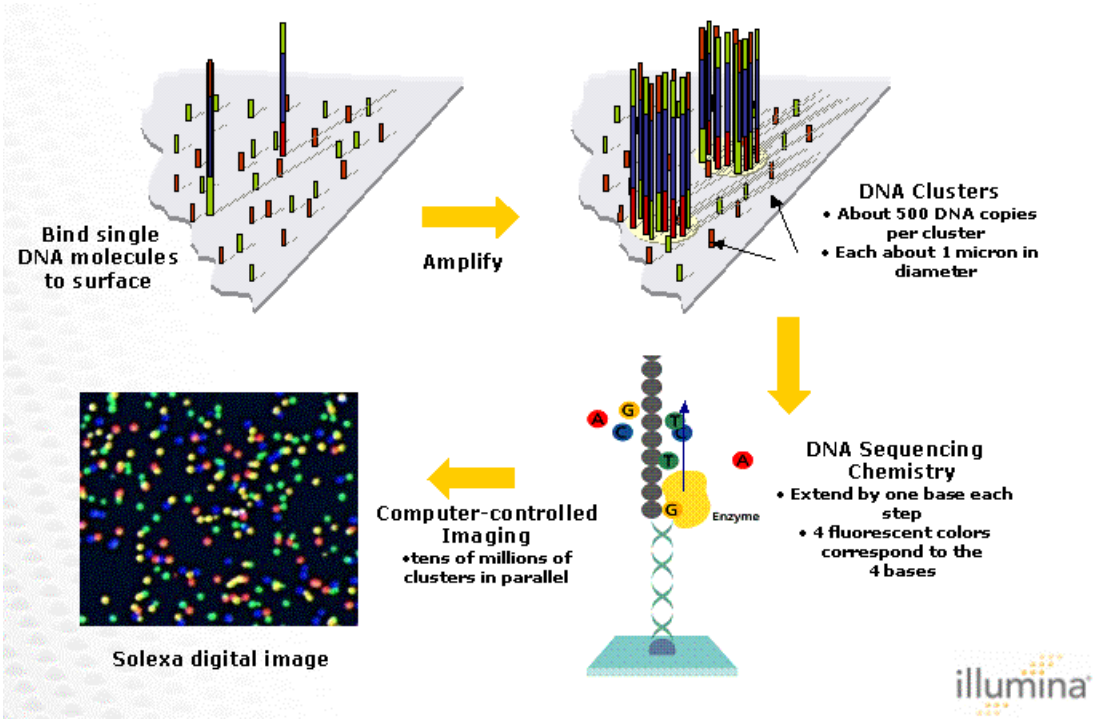
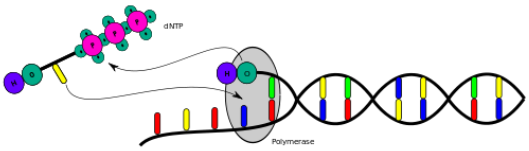


Figure 12
The data are aligned and compared to reference, and sequencing differences are identified.

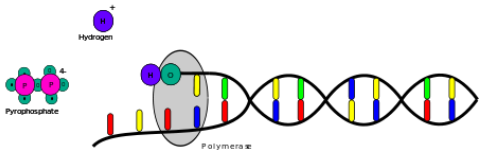


- Ion semiconductor sequencing: a microwell containing a template DNA strand to be sequenced is flooded with a single species of deoxyribonucleotide triphosphate (dNTP). If the introduced dNTP is complementary to the leading template nucleotide, it is incorporated

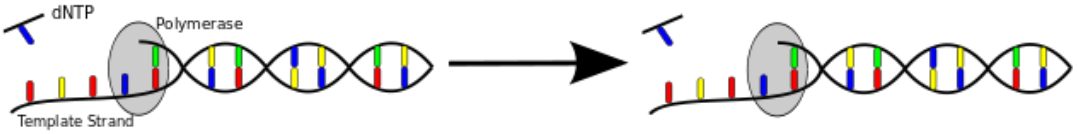
into the growing complementary strand. This causes the release of a hydrogen ion that triggers an ISFET ion sensor, which indicates that a reaction has occurred. If homopolymer repeats are present in the template sequence, multiple dNTP molecules will be incorporated in a single cycle. This leads to a corresponding number of released hydrogens and a proportionally higher electronic signal.



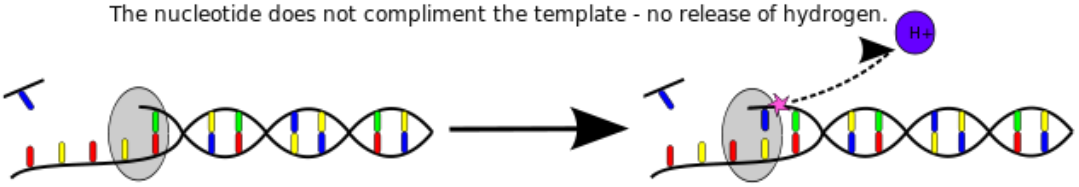
Polymerase integrates a nucleotide.



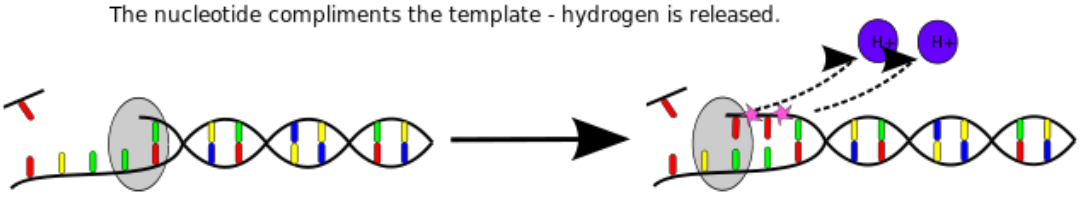
Hydrogen and pyrophosphate are released.



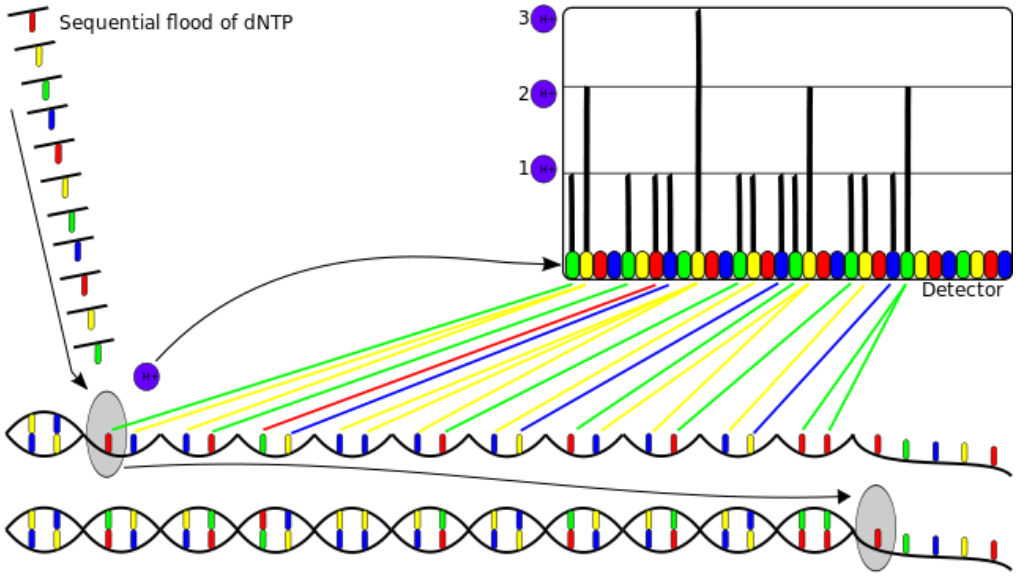
The nucleotide does not compliment the template - no release of hydrogen.



The nucleotide compliments the template - hydrogen is released.



The nucleotide compliments several bases in a row - multiple hydrogen ions are released.



Applications of NGS

- More sequences: de novo genome sequencing from the model organisms such as yeast, E. coli, Drosophila, Arabidopsis, mouse and human genome to the genomes from endangered or even extinct species like the panda, mammoth and early humans
- Metagenomics: the study of organisms in a microbial community based on analyzing the DNA within an environmental sample. Human microbiome studies is to understand the role of microbes in health and disease
- Non-invasive prenatal testing: to detect trisomies, in particular trisomy-21 or Down's syndrome
- Expression analysis: detection of low expressed genes using microarrays is limited by background noise, sensitivity of sequence based studies is predominantly limited by the depth of sequencing
- Single cell genomics: study the cell lineage tree of an organism by analysing the somatic mutations progressing through cell lines
- Human disease and health: cancers, drug resistance, drug discovery

Quality Scores

- A quality score is a prediction of the probability of an error in a base pair
- Measured by the Q score: $Q = -10 \log P$

Quality scores and base calling accuracy

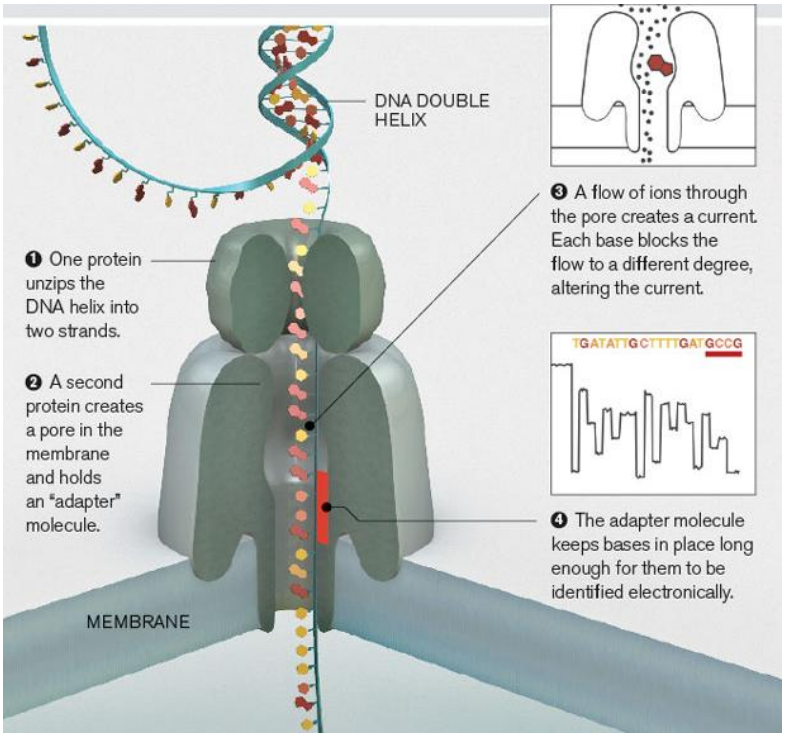
Phred Quality Score	Probability of incorrect base call	Base call accuracy
10	1 in 10	90%
20	1 in 100	99%
30	1 in 1000	99.9%
40	1 in 10,000	99.99%

Genome Annotation

- Computationally process attaching biologically relevant information to genome sequence data, generally by referencing a comparison sequence
- Accurate annotation of a larger genome requires a more contiguous genome assembly in order to avoid splitting genes across scaffolds
- A standard metric to evaluate assembly contiguity is the N50 statistic: by definition, 50% of the assembled nucleotides are found in contigs (contig N50) or scaffolds (scaffold N50) of at least this length. The N50 statistic thus describes a kind of median of assembled sequence lengths
- ab initio algorithms trained on gene models from related species can be used for baseline prediction of coding sequence by looking for open reading frames
- Although these tools generally provide good results, qualitative validation is important (e.g. by assessing the length of open-reading frames). Visual inspection of the annotation is another vital component to detect systematic issues such as intron leakage (introns being annotated as exons due to the presence of premRNA) or gene fusion

Future Improvements

- Ability to sequence single DNA or RNA molecules without any amplification
- Long read lengths
- No GC bias
- Higher read accuracy
- Cheap to acquire and run, easy to operate
- Short run times
- Simple or no library preparation steps
- Nanopore sequencing: thread DNA through a membrane pore channel, and determine order of bases by the affect they have on the flow of ions through the pore



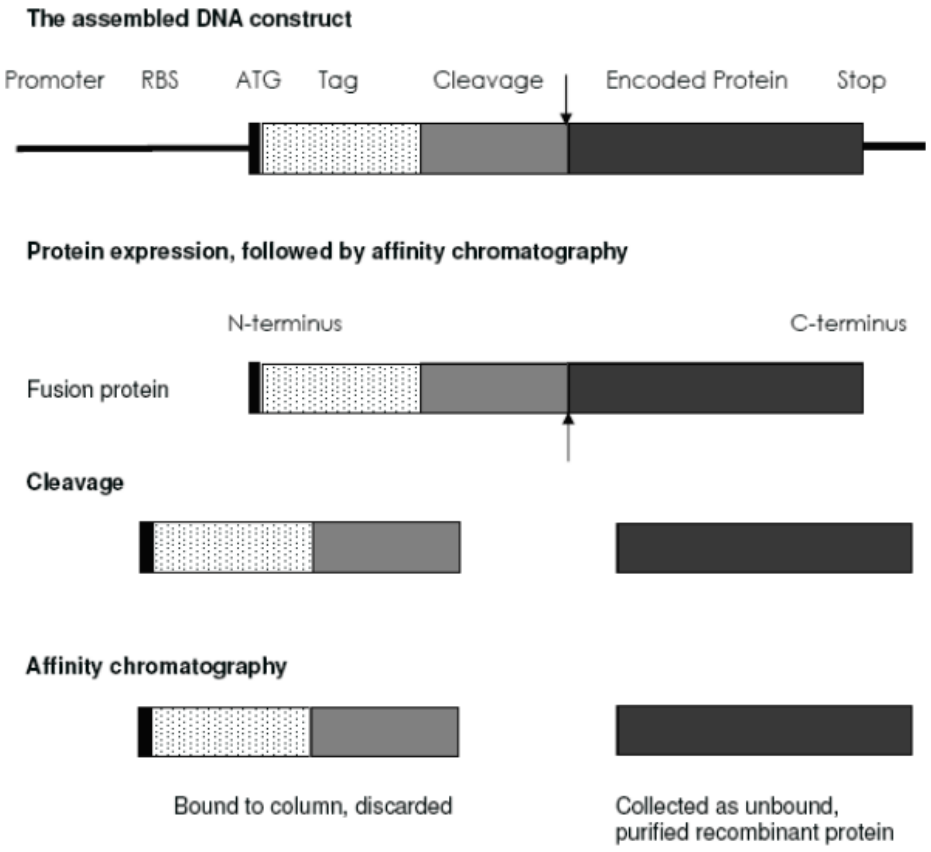
5. Recombinant Protein Vaccinations

Recombinant Proteins

- Formed by laboratory methods of genetic recombination (such as molecular cloning) to bring together genetic material from multiple sources, creating sequences that would not otherwise be found in biological organisms
- Usually proteins are first extracted from organisms of interest and amplified by PCR
- If the organism and the host have different codon usage it is necessary to construct a synthetic, codon-optimised gene

Host Organisms

- All rely on the same principles of foreign gene expression: clone gene downstream of an active promoter, and usually have an induction step
- E coli: easy to grow and simple to scale up, however won't form disulfide bonds or glycosolate proteins. Insoluble proteins will accumulate in acclusion bodies and not fold correctly
- Yeast: ease of microbial growth and cultivation on inexpensive growth media, while adding the ability to perform many post-translational modifications such as O- or N- linked glycosylation, phosphorylation, disulphide bridge formation, proteolytic processing and folding in a eukaryotic system. However, there are several complex post-translational modifications such as prolyl hydroxylation and amidation that yeasts cannot perform. One undesirable attribute of S. cerevisiae is the potential to hyperglycosylate proteins
- Mammalian cell: will correctly fold pretty much anything, overcoming problem of hyperglycosylation. However, much more difficult to work with
- The ultimate use of the protein will also determine the expression system used. For example, if a protein is being used as a diagnostic target for humoral immunity, it may not matter if the protein is not in native conformation, as although some conformational epitopes may be lost, there will exist linear epitopes that can be recognised



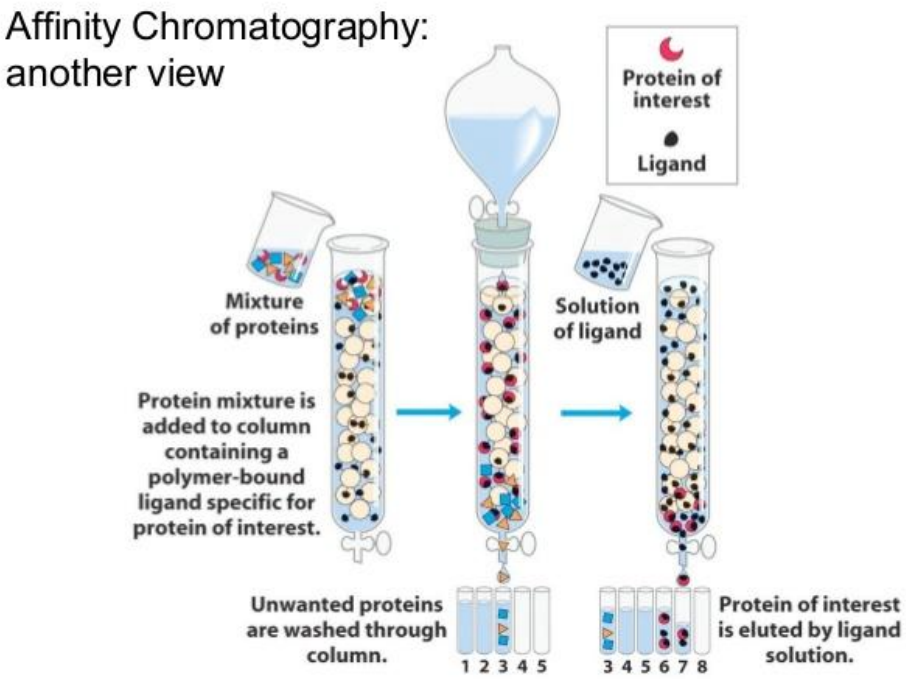
Promoters

- Insert foreign genes downstream of a strong promoter- this ensures maximal transcription
- Commonly use the lac promoter from the E. coli lactose operon
- Constituitive expression places a strain on the organism leading to negative selective pressures, and so is to be avoided if possible
- The lac promoter is repressed in the absence of lactose
- Induced by the addition of lactose (or more commonly IPTG)
- Can therefore induce production of protein when we need to. This allows high levels of expression in a short time

Purification

- Series of processes intended to isolate one or a few proteins from a complex mixture, usually cells, tissues or whole organisms
- Polyhistidine-tags are often used for affinity purification of polyhistidine-tagged recombinant proteins expressed in *Escherichia coli*
- Bacterial cells are harvested via centrifugation and the resulting cell pellet lysed either by physical means or by means of detergents and enzymes such as lysozyme or any combination of these
- At this stage raw lysate contains the recombinant protein among many other proteins originating from the bacterial host. This mixture is incubated with an affinity resin containing bound nickel or cobalt ions, which the polyhistidine-tag binds with micromolar affinity
- The resin is then washed with phosphate buffer to remove proteins that do not specifically interact with the cobalt or nickel ion

- Affinity purification using a polyhistidine-tag usually results in relatively pure protein when the recombinant protein is expressed in prokaryotic organisms



Protein Engineering

- Often, a recombinant protein derived from natural gene is sufficient
- In other cases, however, we may need to modify the protein, for example to inactivate a toxin, change the enzyme specificity, or to create a soluable protein
- For example, Cysteine proteases are difficult to express in yeast and bacteria as they are toxic to the cells.

Generations of Vaccinations

- Vaccination: the inoculation of an individual with antigenic components in order to induce or increase immunity
- First generation: attenuated and killed vaccines, generally gave good immune responses, but important to ensure complete attenuation or killing
- Second generation: subunit vaccines, which rely on recombinant DNA technology to design. Not as good as killed or attenuated vaccines, and often difficult to produce, however completely safe and a well defined molecule
- Third generation: DNA vaccines

Methods of Attenuation

- Historically the approach was to use related non-virulent organisms (e.g. cowpox against smallpox)
- An alternate approach was to grow the organism under unnatural conditions, like altered temperature of pH
- More recent molecular approaches seek to directly knockout specific genes to eliminate virulence and eliminate possibiity of reversion

- Good virulence genes to target include flagella, adhesins, mucinases and chemotaxis

DNA Vaccination

- A technique for protecting an animal against disease by injecting it with genetically engineered DNA so cells directly produce an antigen, resulting in a protective response

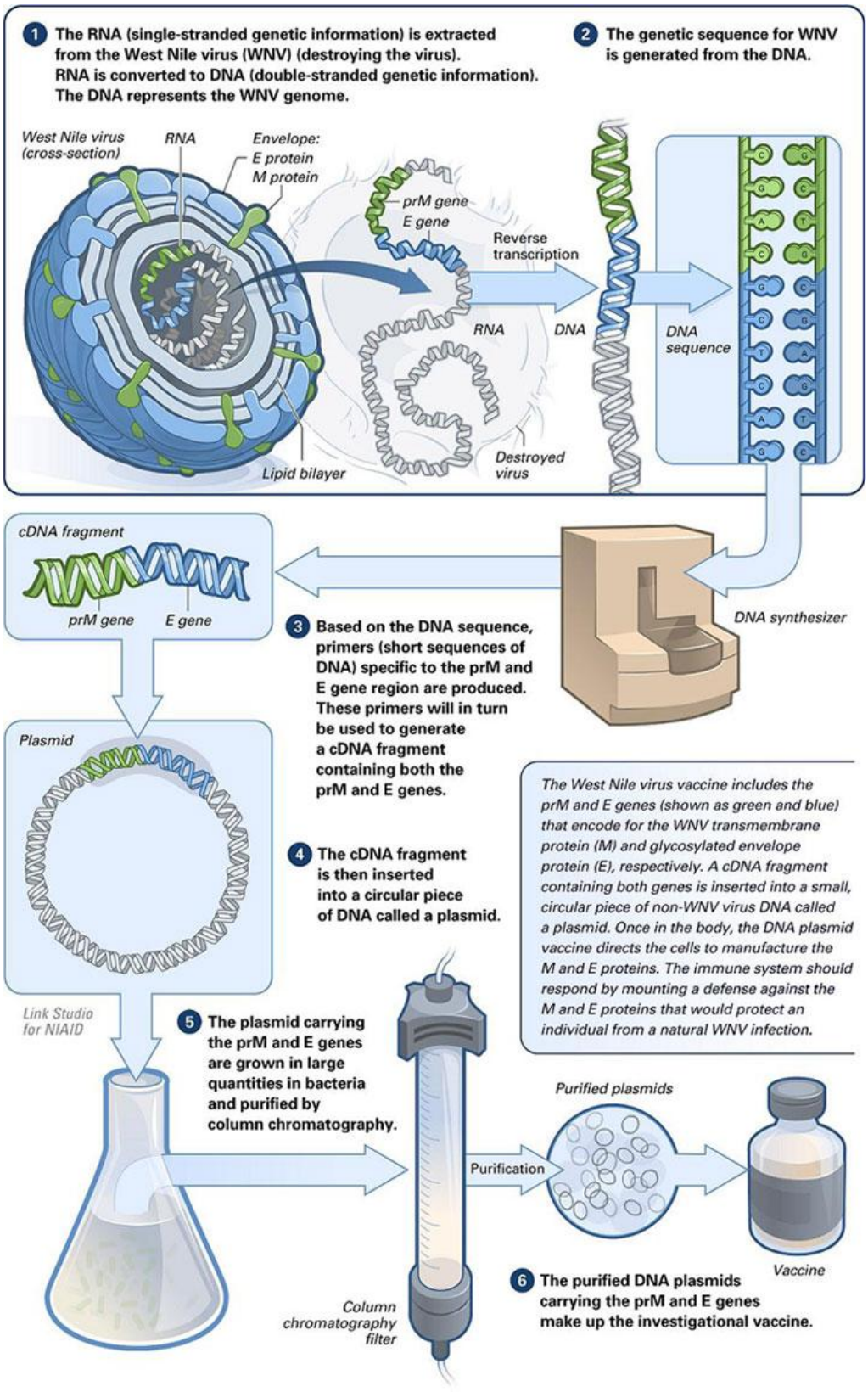


Table 2 | **Current licensed DNA therapies**

Vaccine target	Product name	Company involved	Date licensed and country	Target organisms	Benefits
West Nile virus	West Nile Innovator	Centers for Disease Control and Prevention and Fort Dodge Laboratories	2005 USA	Horses	Protects against West Nile virus infection
Infectious haematopoietic necrosis virus	Apex-IHN	Novartis	2005 Canada	Salmon	Improves animal welfare, increase food quality and quantity
Growth hormone releasing hormone	LifeTide-SW5	VGX Animal Health	2007 Australia	Swine and food animals*	Increases the number of piglets weaned in breeding sows; significantly decreases perinatal mortality and morbidity
Melanoma	Canine Melanoma Vaccine	Merial, Memorial Sloan–Kettering Cancer Center and The Animal Medical Center of New York	2007 USA, conditional license	Dogs	Treats aggressive forms of cancer of the mouth, nail bed, foot pad or other areas as an alternative to radiation and surgery

Advantages

- No risk of reversion of attenuated vaccines to virulent form
- Avoids risk of manufacturing killed vaccine which could be contaminated with live organism
- No adverse reactions or dangers reported in large number of trials done to date
- 'As DNA vaccine plasmids are non-live, non-replicating and non-spreading, there is little risk of either reversion to a disease-causing form or secondary infection'
- Stability of vaccine for storage and shipping, no need for cold chain
- Easy and cheap to produce

Disadvantages

- Still working on techniques to develop necessary high levels of immunity
- Concern that DNA vaccines could integrate into cellular DNA owing to optimized expression plasmids, resulting in insertional mutagenesis, chromosomal instability, or activation or inactivation of tumour suppressor genes
- Production process involves selection of bacterial cells using antibiotic resistance, which is conferred by a plasmid gene
- Concern about possible development of autoantibodies against immune adjuvants

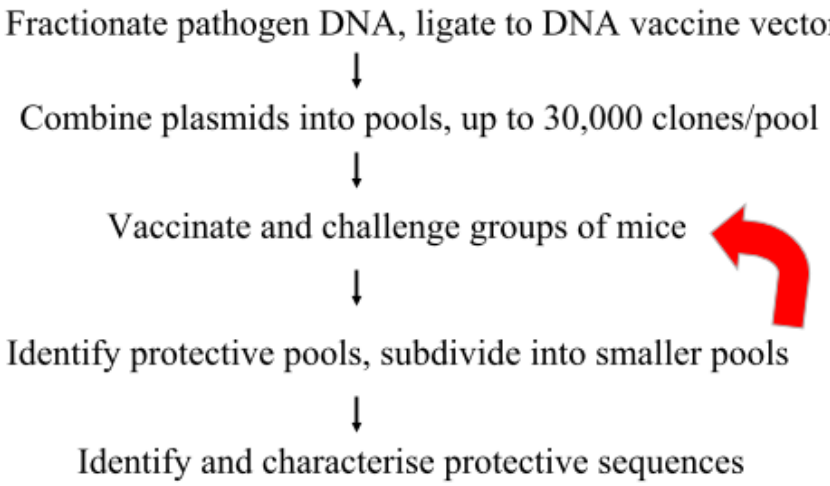
Mechanism of Immunity

- Using the host cellular machinery, the plasmid enters the nucleus of transfected local cells, including resident antigen presenting cells (APCs)
- Here, expression of plasmid-encoded genes is followed by generation of foreign antigens as proteins that have been converted to peptide strings
- These host-synthesized antigens can become the subject of immune surveillance in the context of both MHC class I and class II molecules of APCs in the vaccinated host
- Myocytes are the origin of protein expression after IM vaccination, however they do not contain all the necessary co-stimulatory molecules for antigen presentation
- It was shown that presentation is due to bone-marrow derived cells, therefore the muscle cells are acting as a protein “factory”

Improving Immune Responses

- Find improved mammalian promoters
- Use of next-generation delivery methods (gene gun with DNA bound to gold particles)
- Application of codon optimization methods
- Responses to secreted protein are much higher than to non-secreted

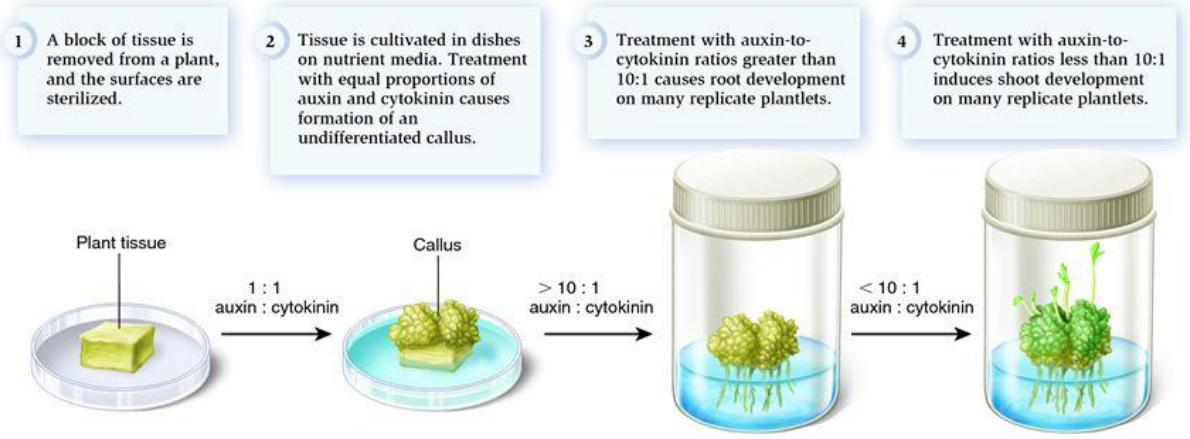
- RNA optimization can also lead to more efficient translation through several important modifications, including removal of instability elements that lower expression
- In addition to optimizing the DNA antigenic plasmid construct, immune modulatory genes can be included s as part of the plasmid vaccines cocktail. Co-injection of plasmids encoding cytokines, chemokines or co-stimulatory molecules can have a substantial effect on the immune response to plasmid-encoded antigen
- Prime boost: priming with DNA and boosting with viral vector vaccines encoding multiple common antigens, used to induce broad and T-cell immunity



6. Plant Biotechnology

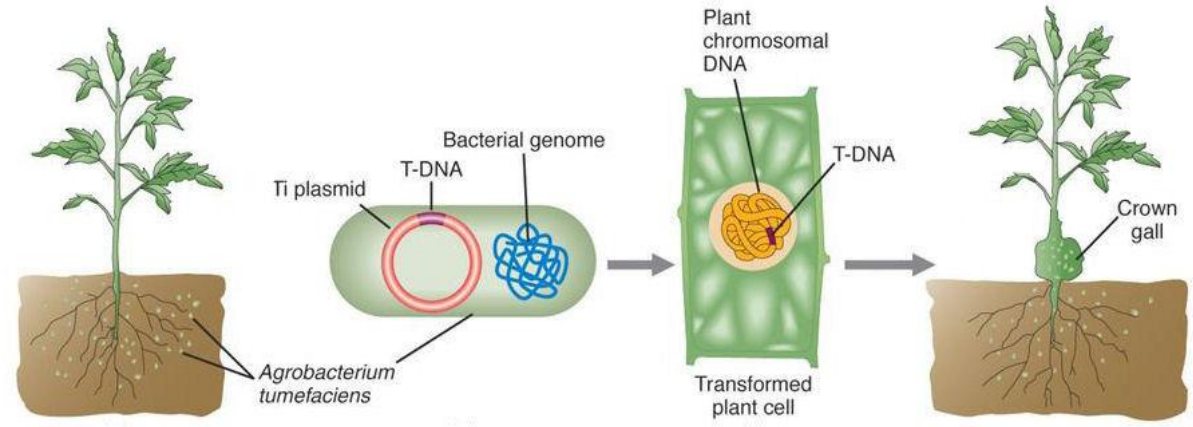
Plant Regeneration

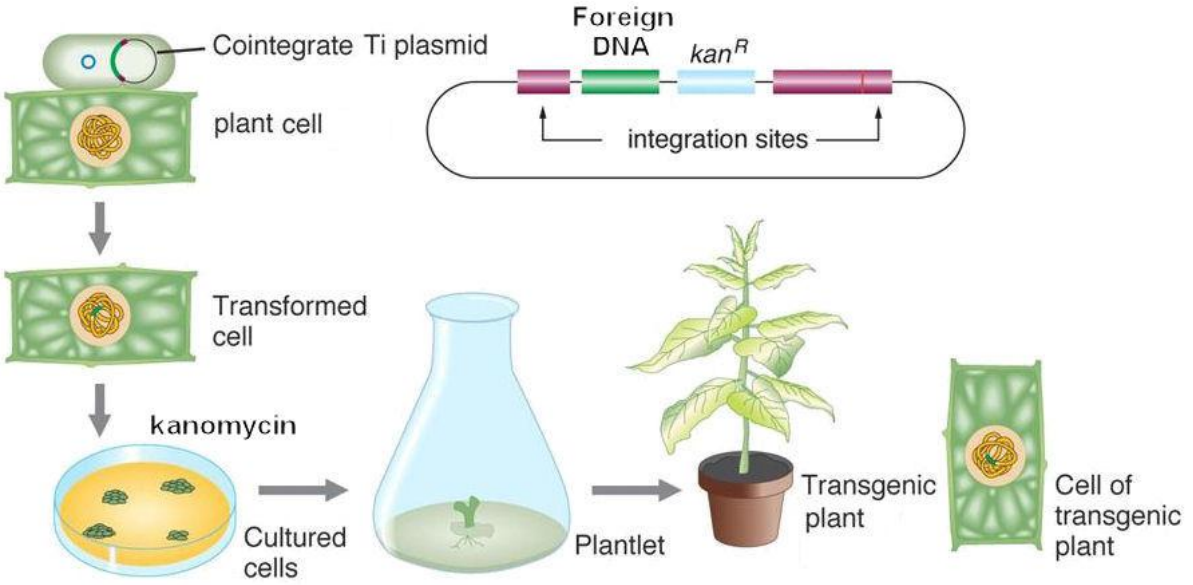
- Most plant cells are totipotent, meaning that an entire plant can be regenerated from a single plant cell
- The tissue obtained from a plant to be cultured is called an explant
- A totipotent explant can often (though not always) be taken from any part of a plant including portions of shoots, leaves, stems, flowers, roots and single, undifferentiated cells
- Tissue cultured plants are clones. If the original mother plant used to produce the first explants is susceptible to a pathogen or environmental condition, the entire crop would be susceptible to the same problem. Any positive traits would remain within the line also



Agrobacterium tumefaciens

- A rod-shaped, Gram-negative soil bacterium which infects a large number of plant species causing the formation of crown galls
- Symptoms are caused by the insertion of a small segment of DNA (known as the T-DNA, for 'transfer DNA'), from a plasmid, into the plant cell, which is incorporated at a semi-random location into the plant genome
- To be virulent, the bacterium must contain a tumour-inducing plasmid (Ti plasmid or pTi), of 200 kb, which contains the T-DNA and all the genes necessary to transfer it to the plant cell
- Bacteria usually enter plant through wounds, usually at or near soil level

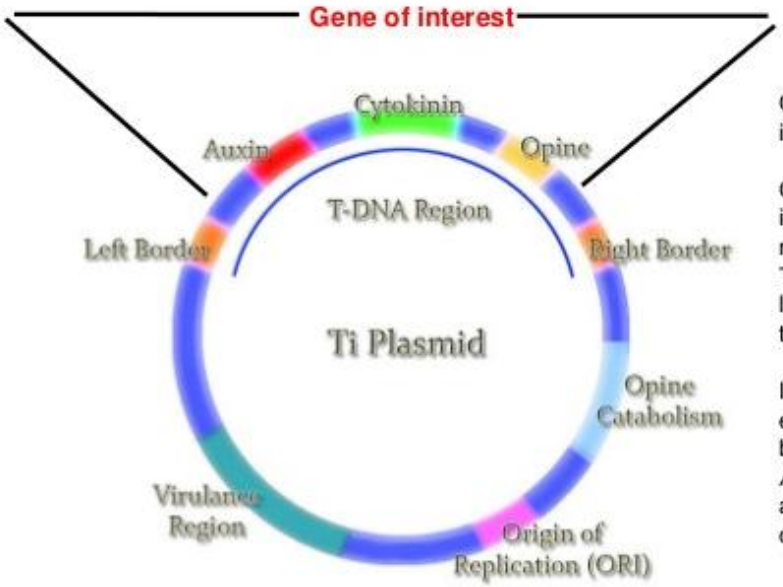




T-DNA

- The **transfer DNA** (abbreviated **T-DNA**) is the transferred DNA of the tumor-inducing (Ti) plasmid of some species of bacteria such as *Agrobacterium tumefaciens*
- It derives its name from the fact that the bacterium transfers this DNA fragment into the host plant's nuclear DNA genome
- The T-DNA is bordered by 25-base-pair repeats on each end. Transfer is initiated at the right border and terminated at the left border and requires the *vir* genes of the Ti plasmid

- By transferring the T-DNA into the plant genome, the bacterium essentially reprograms the plant cells to grow into a tumor and produce a unique food source for the bacteria
- The synthesis of the plant hormones auxin and cytokinin enables the plant cell to grow uncontrollably, thus forming the crown gall tumors
- The opines are amino acid derivatives used by the bacterium as a source of carbon and energy
- To insert into the plant genome, first the bacterium nicks at 25 bp T-DNA borders repeats, then the linear single stranded DNA transfers and integrates into the plant genome

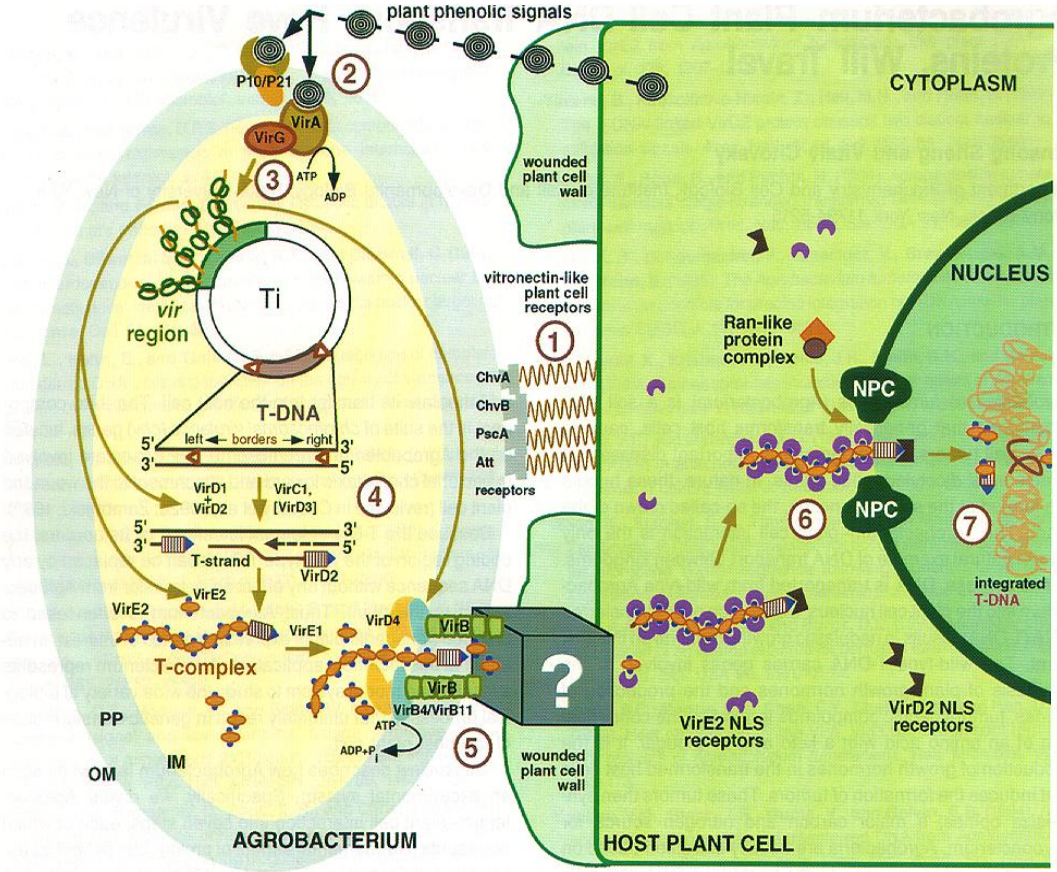


Strategy

Clone an ori for *E. coli* into the Ti plasmid

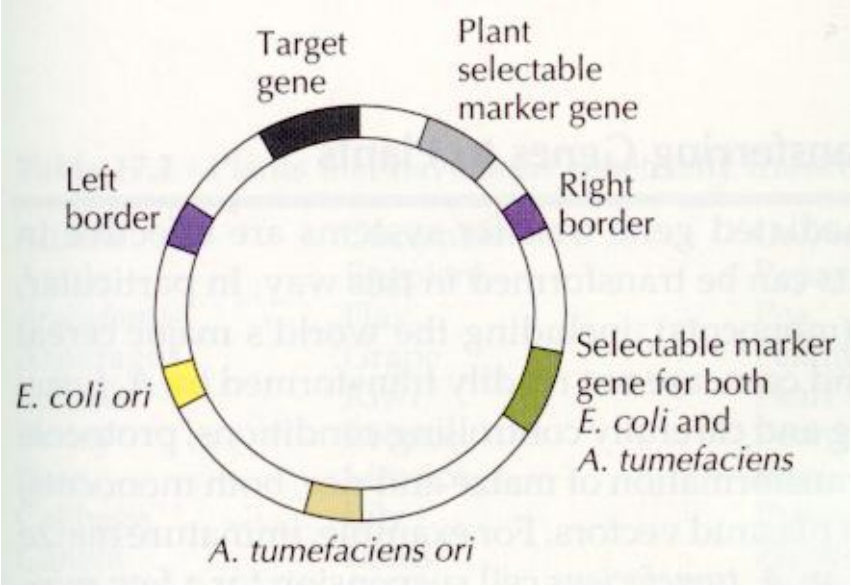
Clone gene of interest into Ti plasmid by replacing genes in the T-DNA (but keeping the left and right borders of the T-DNA intact).

Introduce the engineered Ti plasmid back into *Agrobacterium* and allow it to infect plant cells in culture.



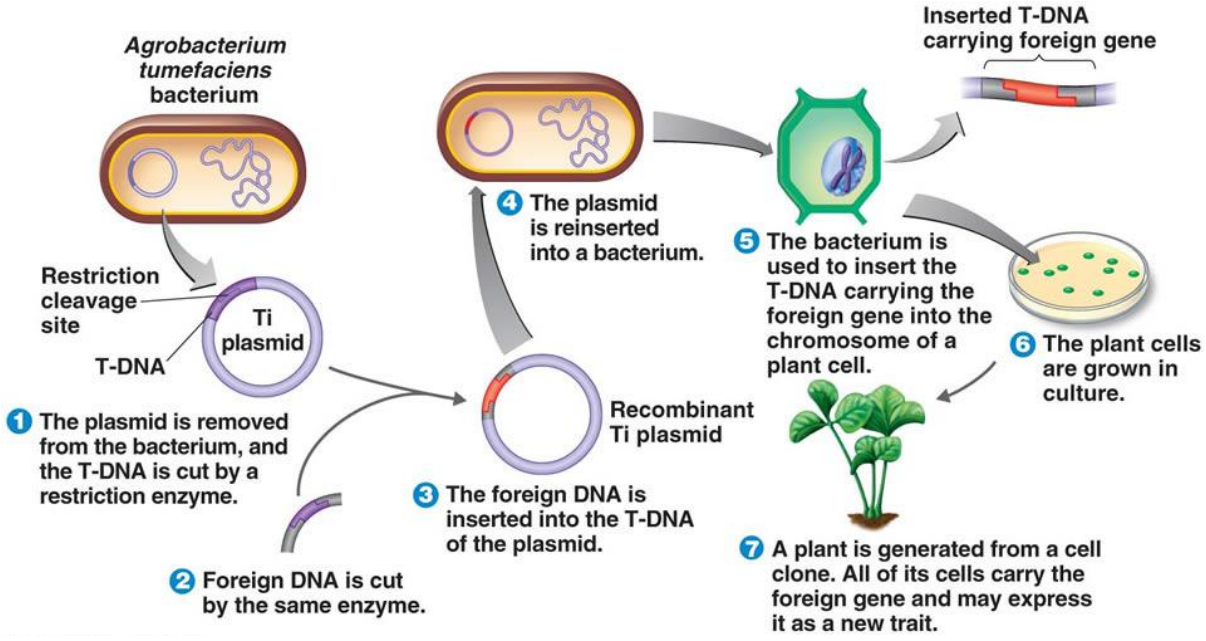
Binary Vectors

- A transfer DNA (T-DNA) binary system is a pair of plasmids consisting of a binary plasmid and a helper plasmid. The two plasmids are used together to produce genetically modified plants
- Left and right border T-DNA sequences retained, but in between the auxin and cytokinin hormone genes are removed, as is that for opine synthesis (useless to the plant)
- In their place is inserted genes for a plant selectable marker gene (e.g. Antibiotic resistance Herbicide resistance), as well as the gene of interest. GFP can also be used as reporter gene
- An E coli origin of replication is also typically added (outside the T region) for ease of propogating the plasmids in E coli cultures
- Typically are smaller than Ti plasmids as non-essential components removed
- In this system, the binary cloning vector itself lacks the vir genes so it cannot be inserted into plants. For this to occur, it must first be introduced into A. tumefaciens cells which have been engineered to have a defective version of the plasmid (vir but no T-DNA)
- Thus, the defective TI plasmid synthesises the transfer vir proteins that the binary vector needs to be inserted into the plant genome



Generating Transgenic Plants

- Cut plant tissue and incubate with *Agrobacterium*
- Plate onto plant selectable marker- e.g. kanamycin
- Media has another antibiotic to kill *Agrobacterium*
- Regenerant plants will be transgenics, while non-transformed/cocultivated cells die



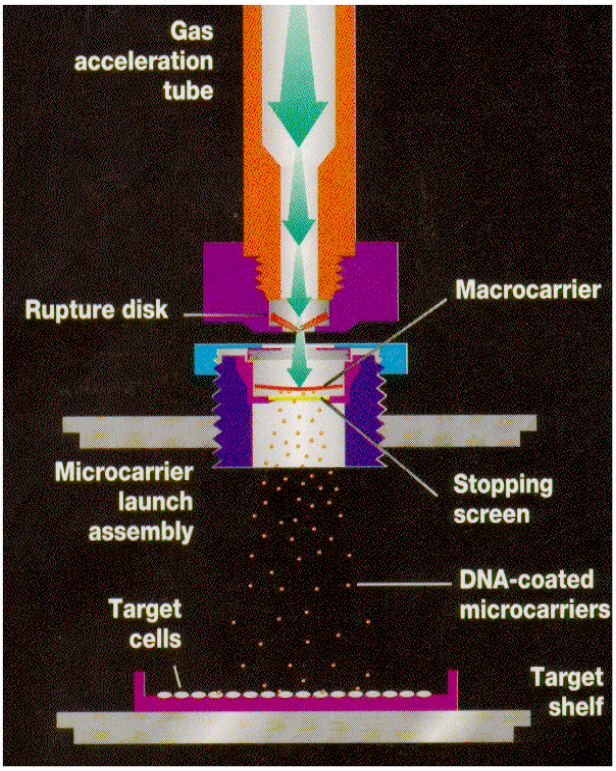
Direct DNA Transfer

- A tumefaciens-mediated transfer does not work for all plant species, so other techniques for gene transfer have been developed
- Microprojectile bombardment is a method by which foreign substances such as DNA are introduced into living cells and tissues via high-velocity microprojectiles
- In particle bombardment, particles, mostly heavy metals, are used to break the cell wall. If the particles are coated in DNA, the foreign DNA will be brought into the cell through the breaks in the cell wall

- If the cell walls are floating in a solution of DNA, a few will be picked up by the particles as they move through the solution, forcing the DNA into the cell.

Table 17.1 Plant cell DNA-delivery methods

Method	Comment
Ti plasmid-mediated gene transfer	An excellent and highly effective system that is limited to a few kinds of plants
Microprojectile bombardment	Used with a wide range of plants and tissues; easy and inexpensive
Viral vectors	Not an effective way to deliver DNA to plant cells
Direct gene transfer into plant protoplasts	Can be used only with plant cell protoplasts that can be regenerated into viable plants
Microinjection	Has limited usefulness because only one cell can be injected at a time; requires the services of a highly skilled individual
Electroporation	Generally limited to plant cell protoplasts that can be regenerated into viable plants
Liposome fusion	Can be used only with plant cell protoplasts that can be regenerated into viable plants



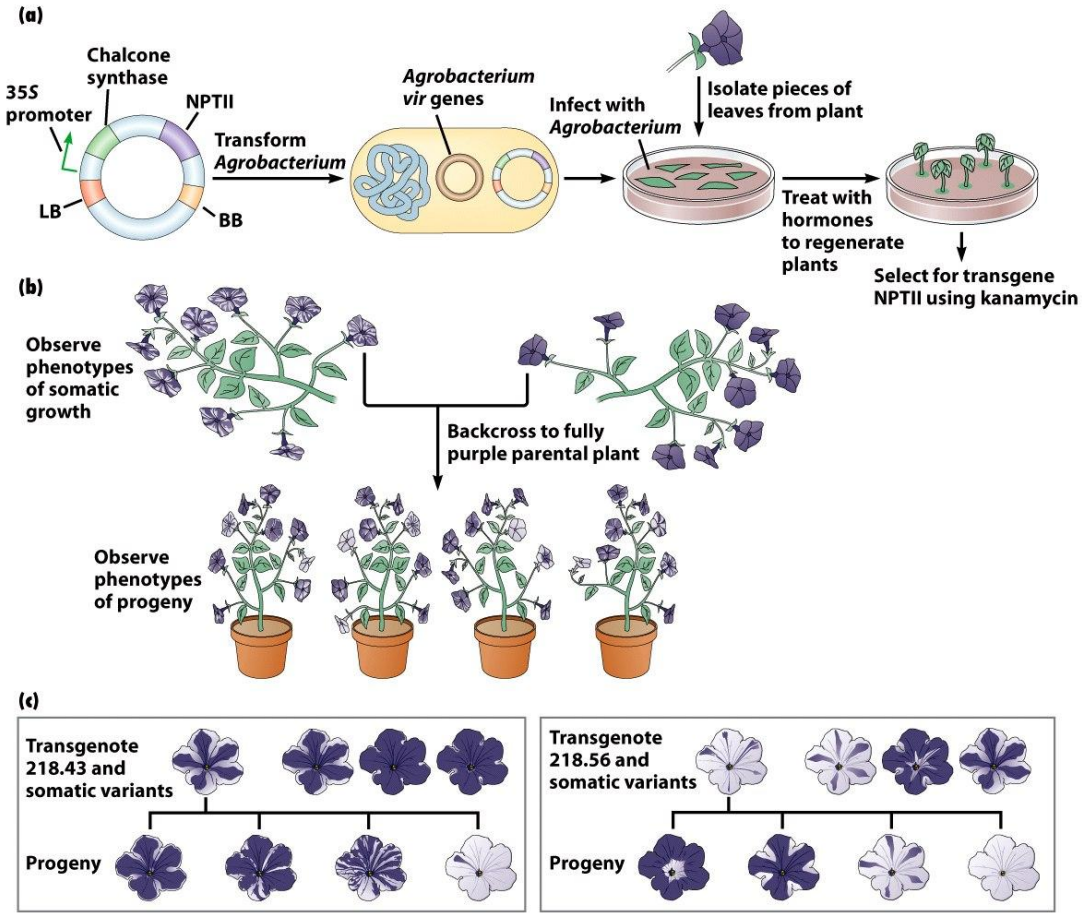
Biotechnology Applications

- Herbicide tolerance (roundup ready)
- Insect/virus resistance
- Nutritional improvements (e.g. golden rice)
- Use of plants as biofactories for pharmaceuticals, hormones, etc

7. RNA Interference

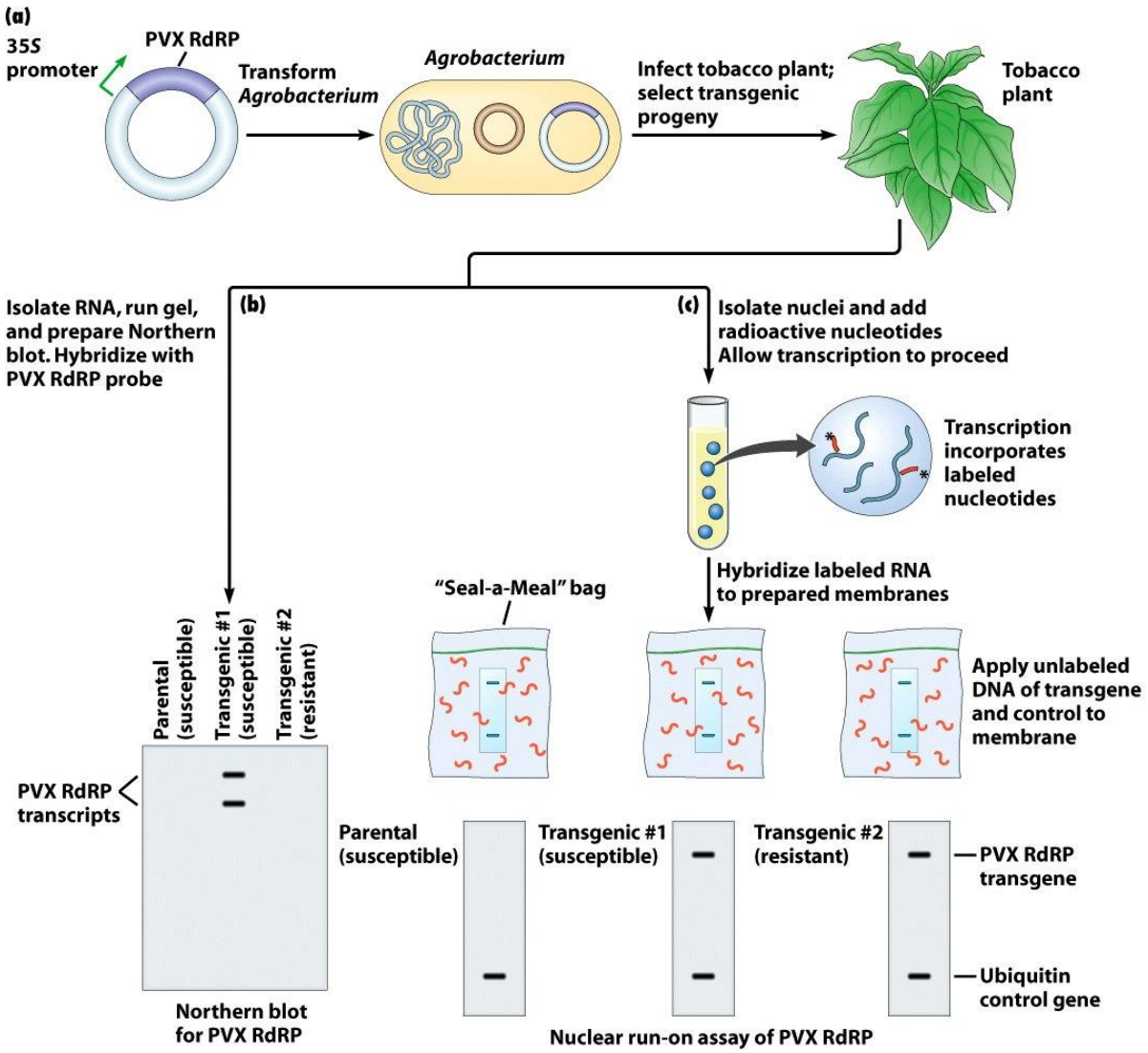
Chalcone Synthase Experiments

- In an attempt to alter flower colors in petunias, researchers introduced additional copies of a gene encoding chalcone synthase, a key enzyme for flower pigmentation into petunia plants of normally pink or violet flower color
- The overexpressed gene was expected to result in darker flowers, but instead produced less pigmented, fully or partially white flowers, indicating that the activity of chalcone synthase had been substantially decreased
- In fact, both the endogenous genes and the transgenes were downregulated in the white flowers



Potato virus X Replicase Experiment

- In this experiment, an altered phage replicase protein competes with normal infecting replicase, preventing phage replication and granting plants viral immunity
- A northern blot to measure steady-state RNA levels indicates none of the relevant transgene, however a radioactive analysis of RNA during transcription shows that the transgene RNA is being expressed - just being degraded quickly thereafter
- Replicase constructs with frameshifts also showed resistance - clearly the protein itself wasn't doing anything
- This phenomenon was called co-suppression of gene expression, but the molecular mechanism remained unknown



C. elegans Mutation Experiment

- After these initial observations in plants, laboratories searched for this phenomenon in other organisms
- *par-1* mutants of C. elegans fail to make an asymmetric cleavage in first cell division
- In investigating the regulation of muscle protein production, they observed that neither mRNA nor antisense RNA injections had an effect on protein production, but double-stranded RNA successfully silenced the targeted par-1 gene
- Thus, injecting worms with the par-1 gene that they needed produced the same phenotypic effects as a mutation of that gene, but only when injected double-stranded
- Indeed, it was found that all that was needed were fragments of the gene - RNAi works in *C.elegans* even if worms are fed bacteria expressing dsRNA

TABLE 9-1. Various sense and antisense RNAs were injected into *Caenorhabditis elegans* and scored for induction of embryonic lethality

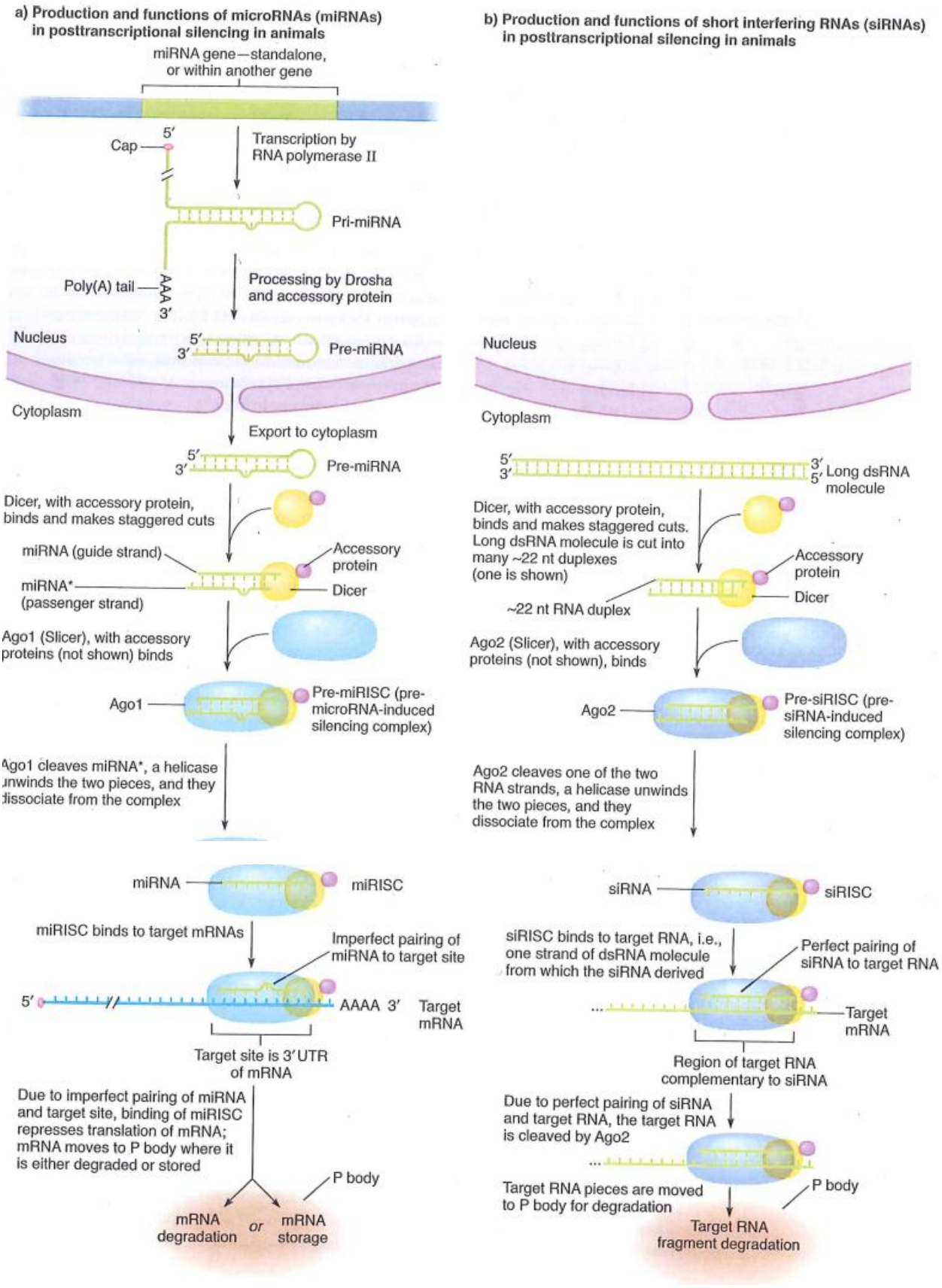
Molecule injected	No. of worms injected		Embryonic lethality (%)
par-1 antisense	16	Blocking function of par gene	52
par-1 sense	12		54
Drosophila cofilin gene, antisense	8		0
Two unrelated genes			
Caenorhabditis elegans zygotic gene, antisense	8		0
Water	4		0

microRNA and siRNA

- Two types of small ribonucleic acid (RNA) molecules – microRNA (miRNA) and small interfering RNA (siRNA) – are central to RNA interference
- MicroRNAs (miRNAs) are genomically encoded non-coding RNAs that help regulate gene expression, particularly during development
- The phenomenon of RNA interference, broadly defined, includes the endogenously induced gene silencing effects of miRNAs as well as silencing triggered by foreign dsRNA. Mature miRNAs are structurally similar to siRNAs produced from exogenous dsRNA, but before reaching maturity, miRNAs must first undergo extensive post-transcriptional modification
- siRNAs have a well-defined structure: a short (usually 20 to 24-bp) double-stranded RNA (dsRNA) with phosphorylated 5' ends and hydroxylated 3' ends with two overhanging nucleotides. The Dicer enzyme catalyzes production of siRNAs from long dsRNAs and small hairpin RNAs.[4] siRNAs can also be introduced into cells by transfection
- siRNAs derived from long dsRNA precursors differ from miRNAs in that miRNAs, especially those in animals, typically have incomplete base pairing to a target and inhibit the translation of many different mRNAs with similar sequences. In contrast, siRNAs typically base-pair perfectly and induce mRNA cleavage only in a single, specific target
- The evolutionary function of miRNAs is predoninantly gene regulation - they silence other genes. In contrast, siRNAs function to inhibit insertional elements in the genome

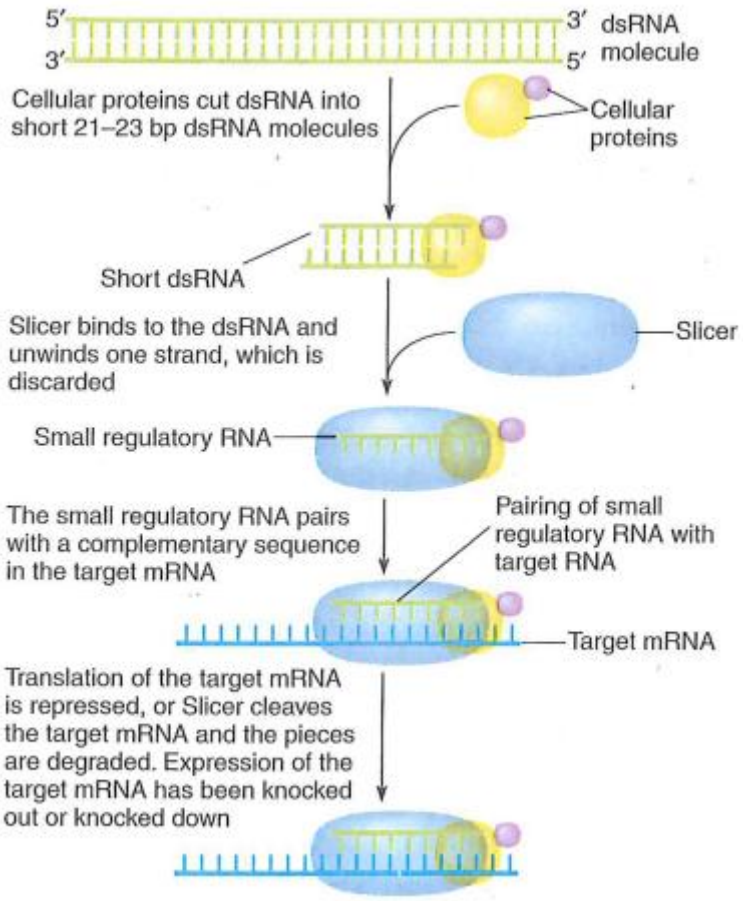
Figure 18.15

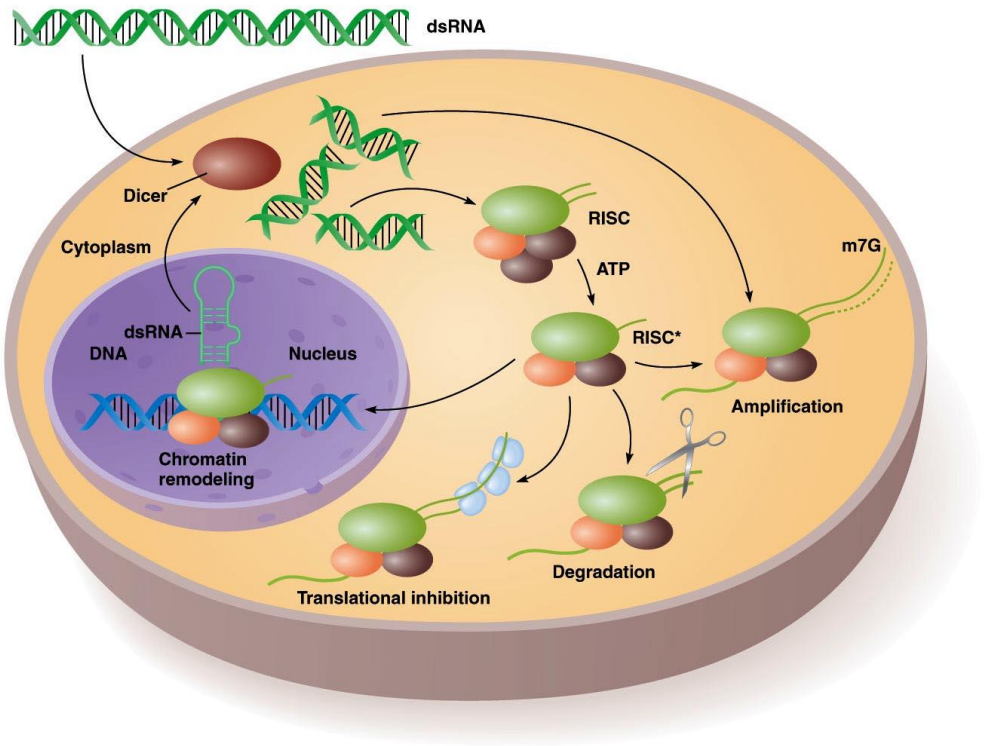
RNA interference (RNAi) by small regulatory RNAs.



Mechanism of Action

- The RNAi pathway is found in many eukaryotes, including animals, and is initiated by the enzyme Dicer, which cleaves long double-stranded RNA (dsRNA) molecules into short double stranded fragments of ~20 nucleotide siRNAs
- Each siRNA is unwound into two single-stranded RNAs (ssRNAs), the passenger strand and the guide strand
- The passenger strand is degraded and the guide strand is incorporated into the RNA-induced silencing complex (RISC)
- The most well-studied outcome is post-transcriptional gene silencing, which occurs when the guide strand pairs with a complementary sequence in a messenger RNA molecule and induces cleavage by Argonaute, the catalytic component of the RISC complex





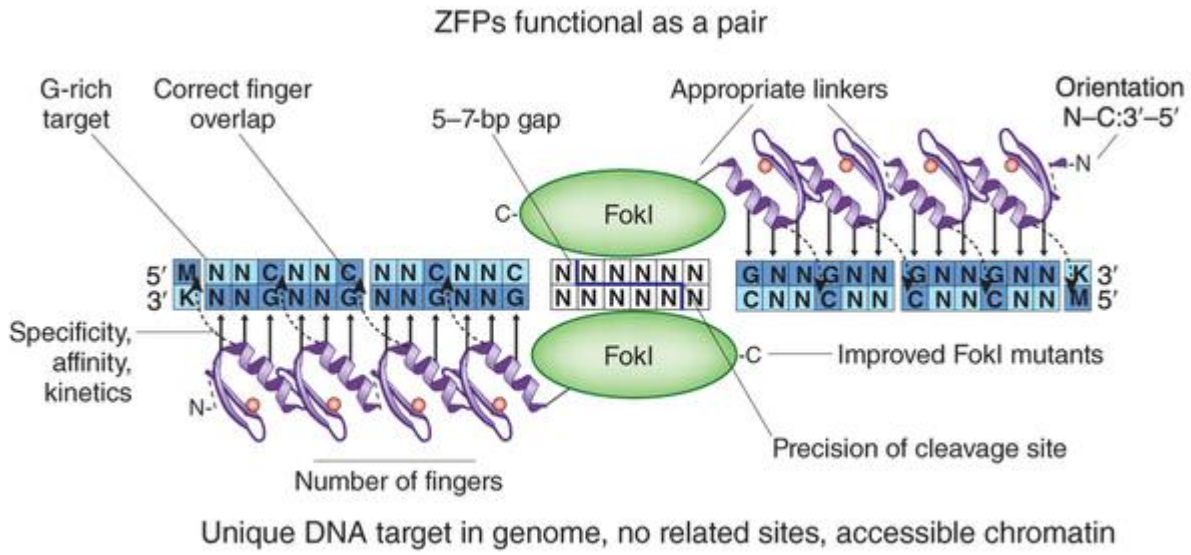
8. Targeted Gene Transfer

Genome Editing

- A type of genetic engineering in which DNA is inserted, replaced, or removed from a genome using artificially engineered nucleases
- There are currently four families of engineered nucleases being used: Zinc finger nucleases (ZFNs), Transcription Activator-Like Effector Nucleases (TALENs), the CRISPR/Cas system, and engineered meganucleases

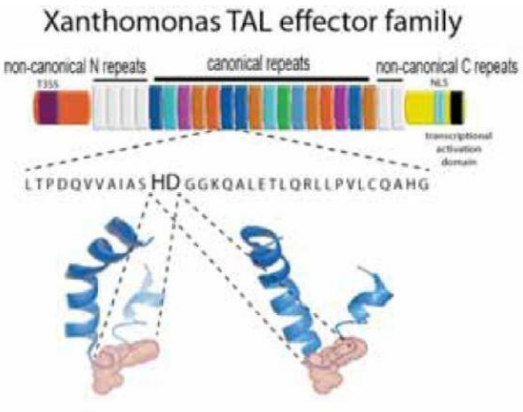
Zinc Finger Nucleases

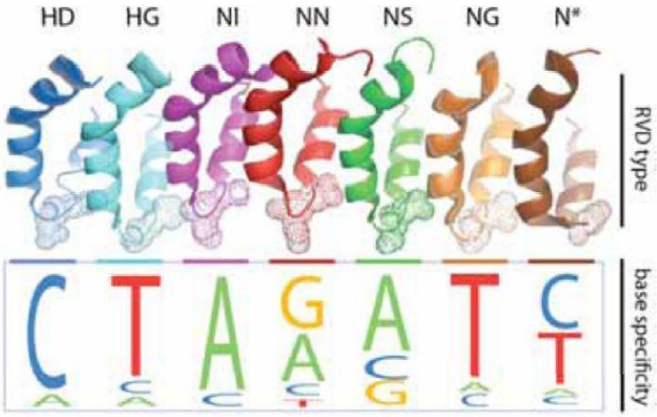
- DNA-binding domains of individual ZFNs typically contain between three and six individual zinc finger repeats and can each recognize between 9 and 18 basepairs.
- f the zinc finger domains are perfectly specific for their intended target site then even a pair of 3-finger ZFNs that recognize a total of 18 basepairs can, in theory, target a single locus in a mammalian genome
- The non-specific cleavage domain from the type IIs restriction endonuclease FokI is typically used as the cleavage domain in ZFNs



TAL Effectors

- Transcription activator-like effector nucleases are proteins secreted by Xanthomonas bacteria. The DNA binding domain contains a repeated highly conserved 33–34 amino acid sequence with the exception of the 12th and 13th amino acids
- These two locations are highly variable and show a strong correlation with specific nucleotide recognition. This simple relationship between amino acid sequence and DNA recognition has allowed for the engineering of specific DNA-binding domains by selecting a combination of repeat segments containing the appropriate RVDs
- The variable regions are found at the end of an alpha helix, so lining up alpha helicies in a row allows targeting of a long specific sequence of DNA
- By combining such an engineered TALE with a DNA cleavage domain (which cuts DNA strands), one can engineer restriction enzymes that are specific for any desired DNA sequence





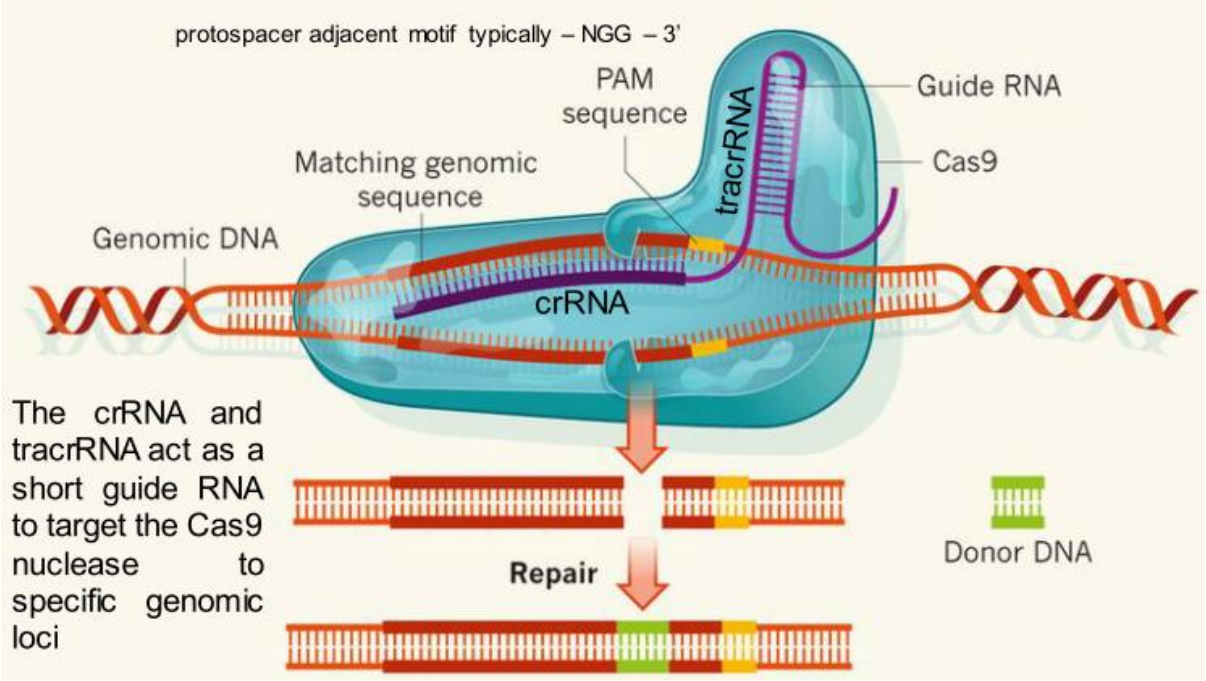
Meganuclease

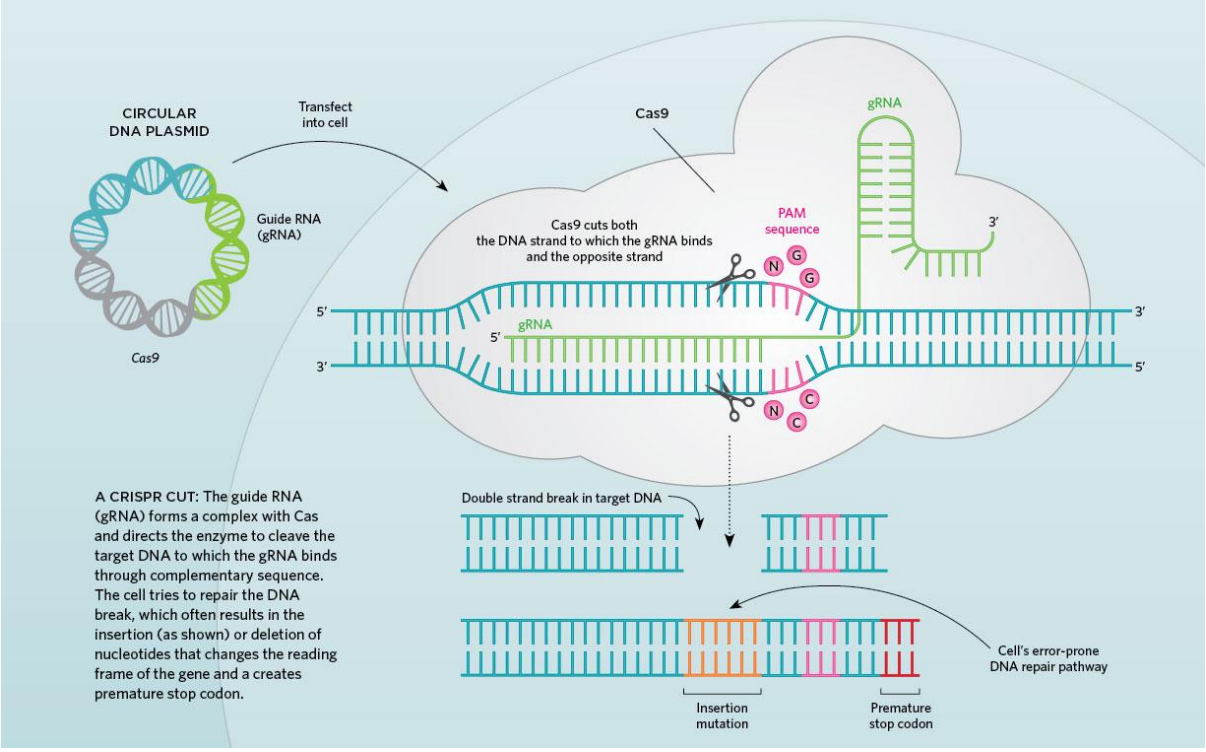
- Endodeoxyribonucleases characterized by a large recognition site (double-stranded DNA sequences of 12 to 40 base pairs); as a result this site generally occurs only once in any given genome

- To create tailor-made meganucleases, two main approaches have been adopted: modifying the specificity of existing meganucleases by introducing a small number of variations to the amino acid sequence, and associating or fusing protein domains from different enzymes

CRISPR

- Clustered regularly interspaced short palindromic repeats are segments of prokaryotic DNA containing short repetitions of base sequences
- Each repetition is followed by short segments of "spacer DNA" from previous exposures to a bacterial virus or plasmid
- Endonuclease cleavage specificity is guided by RNA sequences to cleave the DNA strands
- Editing can be directed to virtually any genomic loci by engineering the guide RNA sequence and delivering it along with the Cas endonuclease to your target cell
- Cas9 is a nuclease, an enzyme specialized for cutting DNA





Hybrid Meganuclease

ZFN

Zinc finger domains

TALEN

TALE subunits

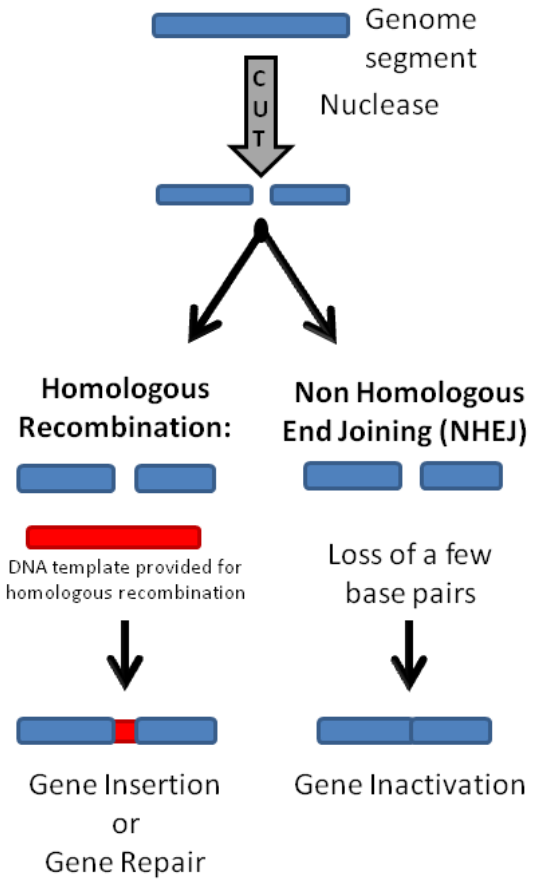
active FokI catalytic subunit heterodimer

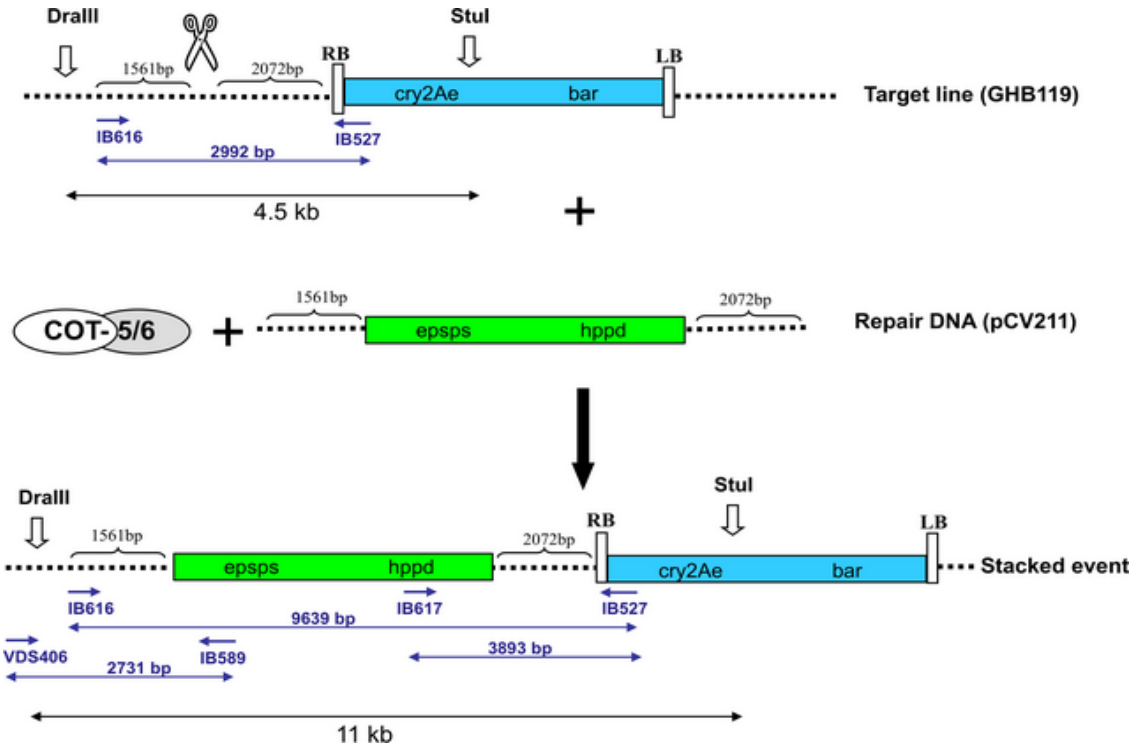
Molecular Trait Stacking

- Combining two or more traits in one variety can be performed either by conventional breeding stacks or by molecular trait stacking in a single transgene locus using a transformation vector carrying multiple-trait genes. With both technologies, the logistical challenges increase with the number of trait genes to be stacked
- Targeted molecular trait stacking could overcome many of these challenges by targeted insertion of the additional trait genes in close genetic vicinity of an already existing transgene locus

Targeted Double Strand Breaks

- Double-strand breaks in plant cells are predominantly repaired by nonhomologous recombination (NHR) or non-homologous end joining (NHEJ)
- Precise repair possible through homologous recombination (HR) by use of homologous template
- This allows DNA containing multiple traits to be inserted at specific targeted sites in the genome
- Sequence analysis after recombination showed that 60%-70% of stacked events did not contain addition insertions elsewhere in genome, and the majority of events did not show meganuclease incorporated into genome
- These stacked events can be inherited

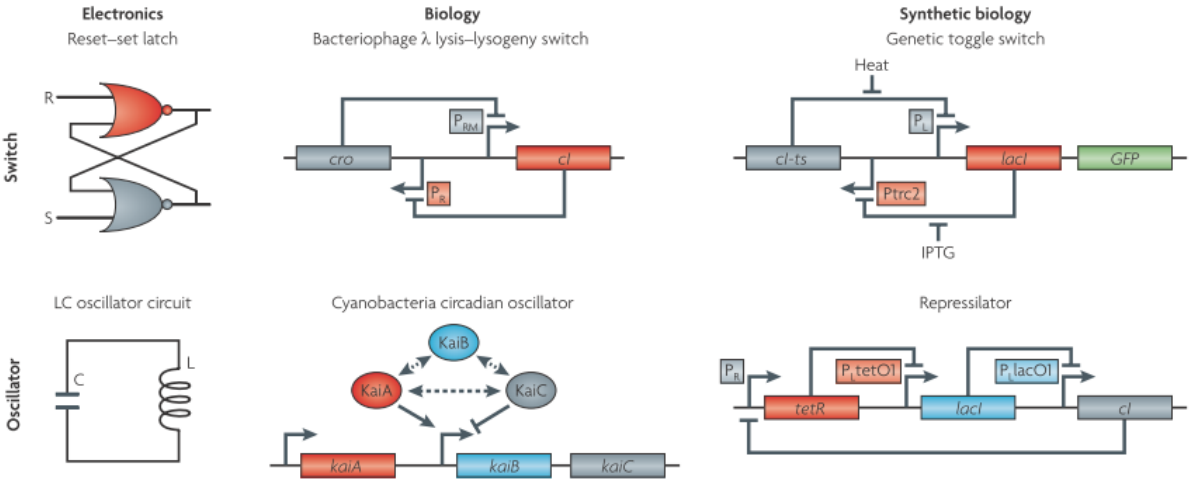


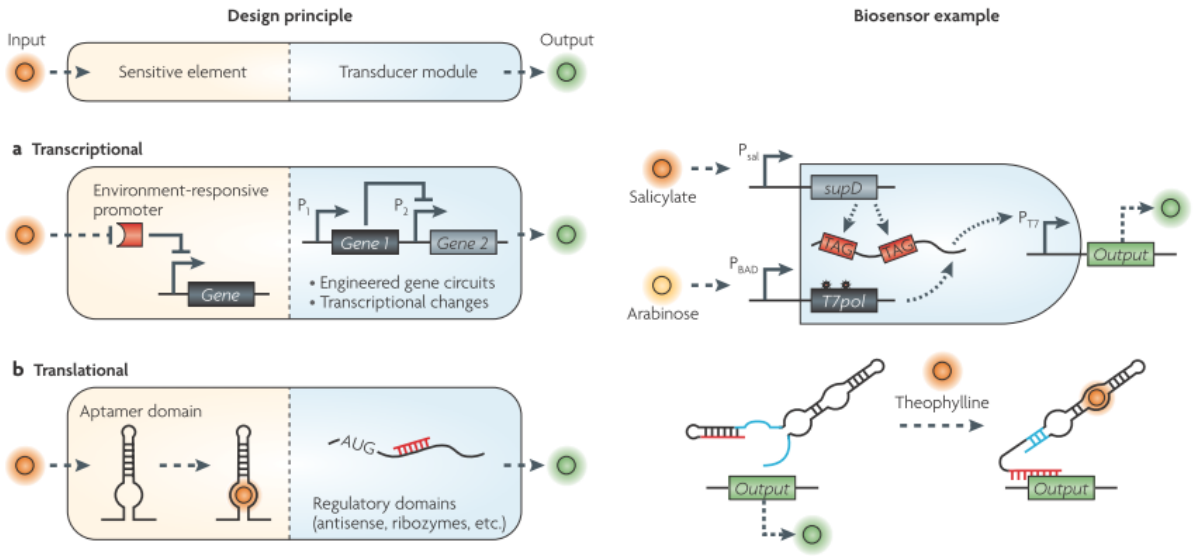


10. Synthetic Biology

What is Synthetic Biology?

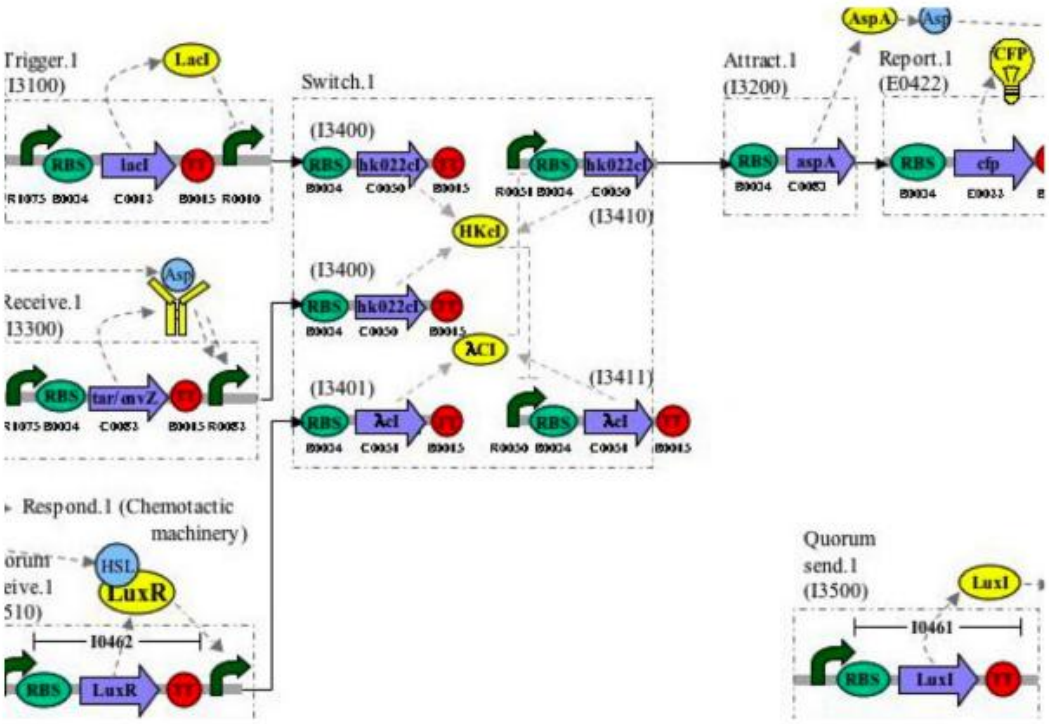
- Synthetic biology is most commonly defined as the design and construction of new biological components as well as the re-design and construction of already existing systems in nature





Bioengineering Principles

- Hierarchy: allows a simplified approach to be adopted whereby the aspects of the gene circuit are classified broadly in terms of function rather than sequence or origin. Promoters are described as parts that initiate transcription and different promoter sequences are not characterised or described separately from one another
- Modularity: a "modular" thing can be broken down into a number of components that may be mixed and matched in a variety of configurations. The components are able to connect, interact, or exchange resources in some way (such as energy), by adhering to the standardised interface
- Decoupling: separation of tasks into specialised categories in order to simplify complicated problems. Such problems can be addressed independently and later combined to deliver a working system with the necessary genes and regulatory motifs that express desired proteins and coordinate timing of their expression
- Standardisation: enables simplified thinking about interfaces, reusable parts, makes subcontracting and division of labour easier, allows for independent evolution of components and technologies, and easier comparison of results



Standard Biological Parts

- BioBrick parts are DNA sequences which conform to a restriction-enzyme assembly standar
- Genetic parts are cloned with the same restriction sites. This means the parts that are available are compatible in a standard and interchangeable manner
- This can be considered analogous to building different toys with LEGO blocks using the same assembly method; genes can also be arranged and assembled together in any order with their reading frame either maintained or disrupted
- Examples of BioBrick parts include promoters, ribosomal binding sites (RBS), coding sequences and terminators

Symbol	BioBrick parts
	Promoter
	Coding sequence
	RBS
	DNA

	Inverter
	Plasmid backbone
	Terminator

- There are three levels to the hierarchy:
  - Parts: Pieces of DNA that form a functional unit (for example promoter, RBS, etc)
  - Device: Collection set of parts with defined function. In simple terms, a set of complementary BioBrick parts put together forms a device
  - System: Combination of a set of devices that performs high-level tasks
- Systems built out of combining input sensors, computational devices, and output actuators

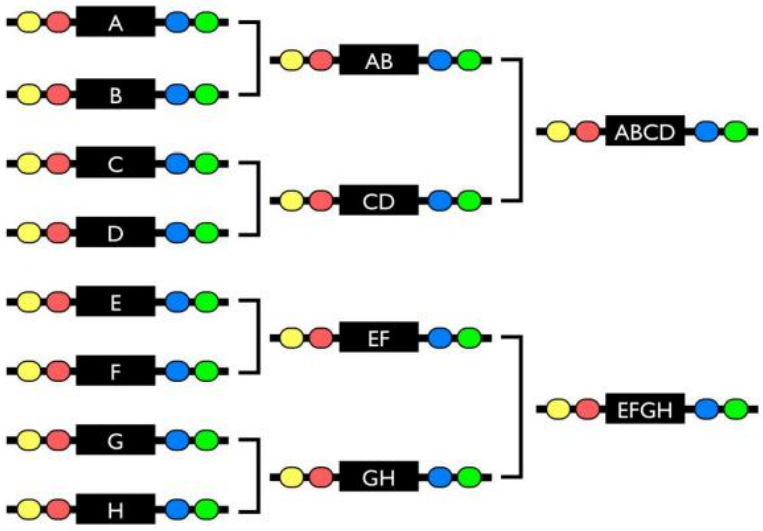
Naturally Occurring Sensor and Actuator Parts Catalog

Sensors

- Light (various wavelengths)
- Magnetic and electric fields
- pH
- Molecules
  - Autoinducers
  - H2S
  - maltose
  - serine
  - ribose
  - cAMP
  - NO
- Internal State
  - Cell Cycle
  - Heat Shock
- Chemical and ionic membrane potentials

Actuators

- Motors
  - Flagellar
  - Gliding motion
- Light (various wavelengths)
- Fluorescence
- Autoinducers (intercellular communications)
- Sporulation
- Cell Cycle control
- Membrane transport
- Exported protein product (enzymes)
- Exported small molecules
- Cell pressure / osmolarity
- Cell death



Venter Artificial Genome

- Synthetic genomics is a nascent field of synthetic biology that uses aspects of genetic modification on pre-existing life forms with the intent of producing some product or desired behavior on the part of the life form so created
- The J. Craig Venter Institute has assembled a quasi-synthetic *Mycoplasma genitalium* bacterial genome by recombination of 25 overlapping fragments in a single step
- the genome was synthesized chemically in many pieces (a synthetic method), joined together by means of molecular biological techniques (an artificial method), and

transplanted into the cytoplasm of a natural cell (after a few generations, though, the original protein content is undetectable

- 'The synthetic M. mycoides genome described in this paper was constructed from 1,078 overlapping synthetic DNA cassettes. A Not I cleavage site (GCGGCCGC) is included at the beginning and the end of each cassette sequence. Our original intention was to build the complete sequence of naturally occurring M. mycoides, with several watermark sequences added at places where insertion of additional sequence would not interfere with viability. During construction of the synthetic genome we learned that certain sequences could be deleted without affecting viability, and some of these deletions were incorporated into the synthetic genome design.'

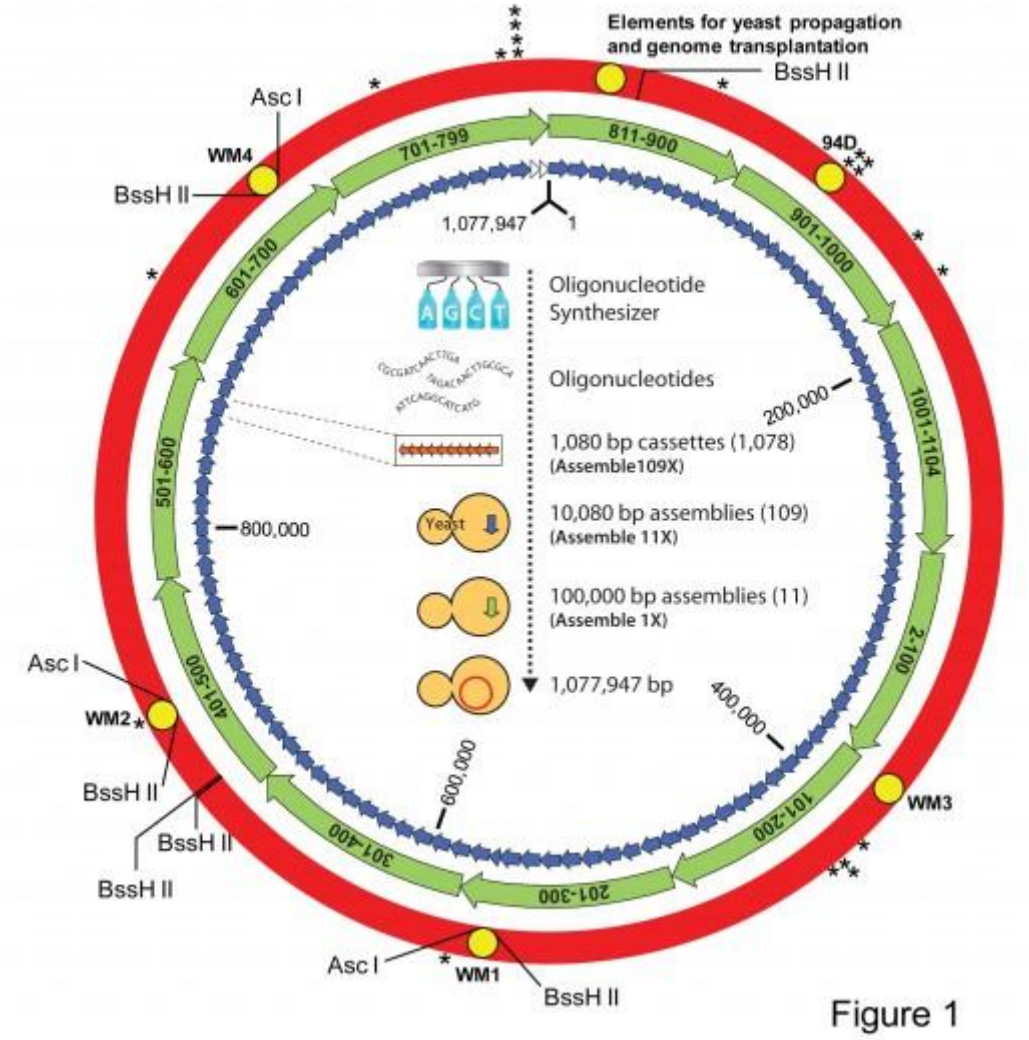
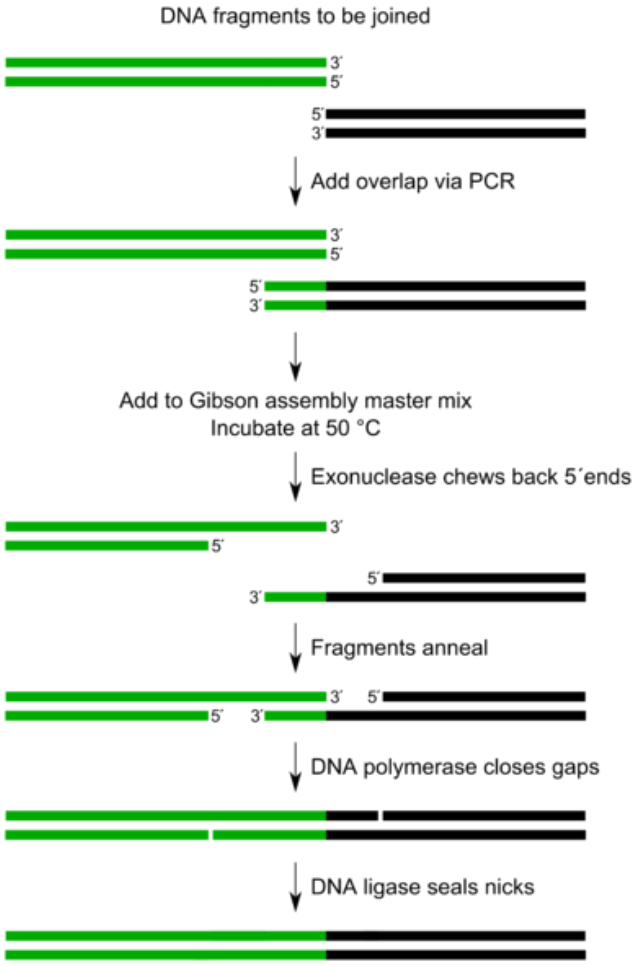


Figure 1

Assembly Methods

- These relate to how to join multiple DNA fragments
- Gibson assembly: can simultaneously combine numerous (>10) DNA fragments based on sequence identity. It requires that the DNA fragments contain ~20-40 base pair overlap with adjacent DNA fragments. These DNA fragments are mixed with a cocktail of three enzymes, along with other buffer components. The three required enzyme activities are: exonuclease, DNA polymerase, and DNA ligase



- Golden Gate cloning: type IIS restriction endonucleases (e.g. *Fok*I) cleave DNA at a defined distance from their non-palindromic asymmetric recognition site. One great advantage of this is it means that recognition sites can be removed immediately upon cutting, so the whole reaction can be done in a single step, and allows a seamless link to be formed