GEL ELECTROPHORESIS

By the technique of gel electrophoresis. Linear DNA molecules separate according to size when subject to an electric field through a gel matrix, an inert, jello -like porous material. Because DNA is negatively charged, when subject to an electrical field it migrates through the gel toward the positive pole. Because DNA molecules are flexible and occupy an effective volume. Pores in the gel matrix sieve the DNA molecules according to this volume.

Once the gels have been "run" for a given time, molecules of different sizes are separated because they have moved different distances through the gel. After electrophoresis is complete, the DNA molecules can be visualized by staining the gel with fluorescent dyes, such as ethidium.

Two alternative kinds of gel matrices are used: polyacrylamide and agarose. **Polyacrylamide** has high resolving capability but can separate DNAs only over a narrow size range. **Agarose** gel is the opposite.

Very long DNA molecules when performed electrophoresis run following a sort of zig-zag pattern, to avoid this, in the case of very long molecules is used the **PULSED-FIELD ELECTROPHORESIS** where the electric field is applied in pulses that are orthogonally oriented between each other.

Electrophoresis also separate molecules according to their shape and topological domain i.e. supercoiled DNA are faster than normal and circular or nick DNA molecules are slower in respect to the normal linear molecules.

Treatment of molecules such as RNA with **glyoxal** prevents the formation of secondary structures

RESTRICTION ENZYMES

Large DNA molecules must be broken down to be analyzed, this is done by **RESTRICTION ENDONUCLEASES** which can cleave at particular sites:

EcoRI: was found in certain strains of Escherichia coli, This enzyme recognizes and cleaves the sequence 5 '-GAATTC-3'

HindIII: recognizes a 6 bp target, but of a different sequence (5'-AAGCTT-3')

Sau3A1: recognize tetrameric sequences (5 '-GATC-3') and so cut DNA more frequently, approximately once every 250 bp.

NotI: recognizes an octameric sequence (5 '-GCGGCCGC-3') and cuts, on average, only once every 65 kilobases

It's clear that the frequencies of cut depends by the length of the recognized sequence

The molecule would have been cut at different positions and generated fragments of different sizes. Thus, the use of multiple enzymes allows different regions of a DNA molecule to be isolated

PROBE AND LABELLING

The capacity of denatured DNA to reanneal (that is, to re-form base pairs between complementary strands) allows for the formation of hybrid molecules, This process of base-pairing between complementary single - stranded polynucleotides from two different sources is known as **hybridization**. Many techniques rely on the specificity of hybridization between two DNA molecules of complementary sequence.

One of the molecules is a **probe** of defined sequence, probe is used to search mixtures of nucleic acids for molecules containing a complementary sequence. This probe must be labeled so it can be readily located once it has found.

The labeled precursors are most commonly nucleotides modified with either a fluorescent moiety or radioactive atoms.

DNA labeled with **fluorescent** precursors can be detected by irradiating the DNA sample with appropriate wavelength **UV light** and monitoring the longer wavelength light that is emitted in response

Radioactive DNA can be detected by exposing the sample of interest to **X-ray film**.

SOUTHERN BLOT

It allows you to identify within the smear the size of the particular fragment containing your gene of interest. In this procedure. It is often desirable to monitor the abundance or size of a particular DNA or RNA molecule in a population of many other similar molecules. For example, this can be useful when determining the amount of a specific mRNA that is expressed in two different cell types;

1) Cutting with enzymes

2) The cut DNA that has been separated by gel electrophoresis is soaked in alkali to denature the double-stranded DNA fragments.

3) Those fragments are then transferred to a positivelycharged membrane to which they adhere, creating an imprint, or blot. The DNA fragments are bound to the membrane in positions comparable to where they migrated in the gel during electrophoresis.

4) The DNA bound to the membrane is then incubated with probe DNA containing a sequence complementary to a sequence within the gene of interest.

This probing is done under conditions of salt concentration and temperature close to those at which nucleic acids denature and renature. Under these conditions, the probe DNA will only hybridize tightly to its exact complement

Where on the blot the probe hybridizes can be detected by a variety of films or other media that are sensitive to the light or electrons emitted by the labeled DNA.

NORTHERN BLOT

It can be used to identify a particular **mRNA** in a population of RNAs. Because mRNAs are relatively short (typically less than 5 kb) there is no need for them to be digested with any enzymes.

1) gel electrophoresis

2) The separated mRNAs are transferred to a positively-charged membrane and probed with a radioactive DNA of choice.

The rest are the same of the southern.

The objectives of this blot can be:

Identify a particular mRNA in a population of RNAs, ascertain the amount of a particular mRNA present in a sample rather than its size (This measure is a reflection of the level of expression of the gene that encodes that mRNA), ask how much more mRNA of a specific type is present in a cell treated with an inducer of the gene in question compared to an uninduced cell ecc.

WESTERN BLOT

The first step in a western blot is to prepare the protein sample by mixing it with a detergent called **sodium dodecyl sulfate**, which makes the proteins unfold into linear chains and charges it negatively.

Next, the protein molecules are separated according to their sizes using gel electrophoresis

Following separation, the proteins are transferred from the gel onto a blotting membrane.

Once the transfer is complete, the membrane carries all of the protein bands originally on the gel. Next, the membrane goes through a treatment called blocking, which prevents any nonspecific reactions from occurring. The membrane is then incubated with an antibody called the primary antibody, which specifically binds to the protein of interest.

Following incubation, any unbound primary antibody is washed away, and the membrane is incubated yet again, but this time with a secondary antibody that specifically recognizes and binds to the primary antibody. The secondary antibody is linked to a reporter enzyme that produces color or light, which allows it to be easily detected and imaged. These steps permit a specific protein to be detected from among a mixture of proteins.

USES)

A western blot is a laboratory method used to detect specific protein molecules from among a mixture of proteins. This mixture can include all of the proteins associated with a particular tissue or cell type. Western blots can also be used to evaluate the size of a protein of interest, and to measure the amount of protein expression. This procedure was named for its similarity to the previously invented method known as the Southern blot.

CHEMICALLY SYNTHESIZED OLIGONUCLEOTHIDES

Short, custom-designed segments of DNA known as oligonucleotides are critical for several techniques. The most common methods of chemical synthesis are performed on solid supports using machines that automate the process.

The precursors used for nucleotide addition are chemically protected molecules called **phosphoamidines**. Growth of the DNA chain is by addition to the 5' end of the molecule, in contrast to the direction of chain growth used by DNA polymerases. Chemical synthesis of DNA molecules up to 30 bases long is efficient and accurate, and takes only a few hours.

But as the synthetic molecules get longer, the final product is less uniform due to the inherent failures that occur during any cycle of the process.

The rather short DNA molecules that can readily be made, however, are well suited for many purposes. For example, a custom-designed oligonucleotide harboring a mismatch to a segment of cloned DNA can be used to create a directed mutation in that cloned DNA. This method, called **site-directed mutagenesis**

LIBRARIES

It is trivial to generate a specific clone if the starting donor DNA is simple, This is harder to do if the starting DNA is more complex.

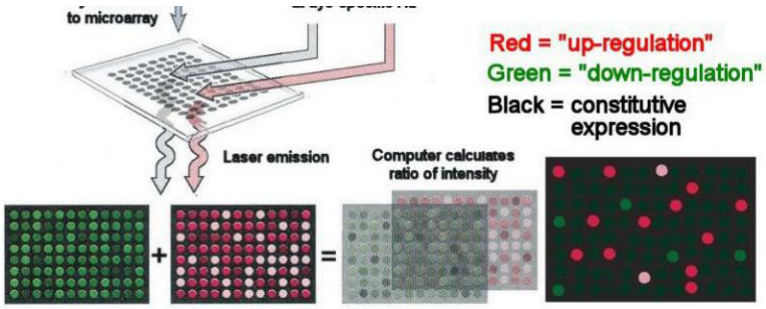
A **DNA library** is a population of identical vectors that each contains a different DNA insert. To construct a DNA library, the target DNA s digested with a restriction enzyme that gives a desired average insert size. The insert size can be of any size. The cleaved DNA is then mixed with the appropriate vector cut with the

same restriction enzyme in the presence of ligase. This creates a large collection of vectors with different DNA inserts. Different kinds of libraries are made using insert DNA from different sources.

When attempting to clone a gene, a common step is to identify fragments of that gene among clones in a library. This can be achieved using a DNA probe whose sequence matches part of the gene of interest.

The process by which a labeled DNA probe is used to screen a library is called **colony hybridization**

Rna – Dna hybridization can be utilized for detecting the transcription profile of a cell compared to another inside Microarrays. Labelling mRNA of a specific gene A differently for the two cell and putting that labelled RNA in a given site of the microarray you can see if one is more transcribed than the other or not simply by checking what kind of labelling is more present in that specific box (Red cell compared to green cell



CLONING

Much of the molecular analysis of genes and their function requires the separation of specific segments of DNA from much larger DNA molecules, and their selective amplification.

The ability to construct recombinant DNA molecules and maintain them in cells is called DNA cloning. This process typically involves a **vector** that provides the information necessary to propagate the cloned DNA in the cell and an **insert DNA** that is inserted within the vector and includes the DNA of interest, a particular insert DNA can be both purified from other DNAs and amplified to produce large quantities.

Once the DNA is cleaved into fragments by a restriction enzyme, it typically needs to be inserted into a vector for propagation. That is, the DNA fragment must be inserted within that second DNA molecule (the vector) to be replicated in a host organism

Vector DNAs typically have three characteristics.

1. They contain an origin of replication that allows them to replicate independently of the chromosome of the host.

2. They contain a selectable marker that allows cells that contain the vector (and any attached DNA) to be readily identified.

3. They have single sites for one or more restriction enzymes. This allows DNA fragments to be inserted at a defined point within an otherwise intact vector.

Vector DNA Can Be Introduced into Host Organisms by **Transformation**

PCR

This procedure is carried out entirely biochemically, that is, in vitro. PCR uses the enzyme DNA polymerase that directs the synthesis of DNA from deoxynucleotide substrates on a ssDNA template.

if a synthetic oligonucleotide is annealed to a single-stranded template that contains a region complementary to the oligonucleotide, DNA polymerase can use the oligonucleotide as a primer and

elongate it in a 5' to 3' direction to generate an extended region of double-stranded DNA.

PREPARATION)

Two synthetic, single-stranded oligonucleotides are synthesized. One is complementary in sequence to the 5' end of one strand of the DNA to be amplified, the other complementary to the 5' end of the other strand.

FIRST CYCLE)

The DNA to be amplified is then denatured and the oligonucleotides annealed to their target sequences. At this point, DNA polymerase and deoxynucleotide substrates are added to the reaction and the enzyme extends the two primers. This reaction generates double-stranded DNA over the region of interest on both

of the strands of DNA. Thus two double -stranded copies of the starting fragment of DNA are produced in this, the first, cycle of the PCR reaction.

SECOND CYCLE)

the DNA is subject to another round of denaturation and DNA synthesis using the same primers. This generates four copies of the fragment of interest. In this way, additional repeated cycles of denaturation and primer-directed DNA synthesis amplify the region between the two primers in a geometric manner (2, 4, 8, 16, 32, 64, and so forth).

CONSIDERATIONS)

So a fragment of DNA that was originally present in vanishingly small amounts is amplified into a relatively large quantity of a double-stranded DNA, In a sense, DNA cloning and the polymerase chain reaction (PCR) rely on the same concept: repeated rounds of DNA duplication whether carried out by cycles of cell division or cycles of DNA synthesis in vitro-amplify tiny samples of DNA into large quantities.

In cloning, however, we often rely on a selective reagent or other device to locate the amplified sequence in an already existing library of clones, whereas in PCR, the selective reagent, the pair of oligonucleotides, limits the amplification process to the particular DNA sequence of interest from the beginning

CHAIN TERMINATION METHOD (SANGER SEQUENCING)

We determine the entire sequence of nucleotides for a genome, and this permits us to find any specific sequence with great rapidity and accuracy through the use of a computer and appropriate algorithms. In other words, our "selective reagent" when dealing with nucleotide sequences is a string of bases that we feed into a computer. The increasing availability of large numbers of genome sequences makes it possible to search with high precision. Obviously, nucleotide sequencing generates extraordinarily powerful databases.

The underlying principle of DNA sequencing is based on the separation, by size, of **nested sets** of DNA molecules. Each of the DNA molecules starts at a common 5’ end, and terminates at one of several alternative 3’ endpoints.

for one set, the molecules all and with a G, for another a C, for a third an A, and for the final set a T. Molecules within a given set vary in length depending on where the particular G at their 3’ end lies in the sequence. The different lengths of these fragments can be determined by electrophoresis through a polyacrylamide gel.

The four nested sets can be run out on the gel side-by-side, generating four ladders and revealing where there are Gs, Cs, As, and Ts within the sequence. Comparing the positions of the rungs in these four ladders reveals the entire sequence of the starting DNA molecule. Alternatively, the four nested sets can be differentially labeled with distinct fluorophores, allowing them to be subjected to electrophoresis as a single mixture and distinguished later using fluorometry.

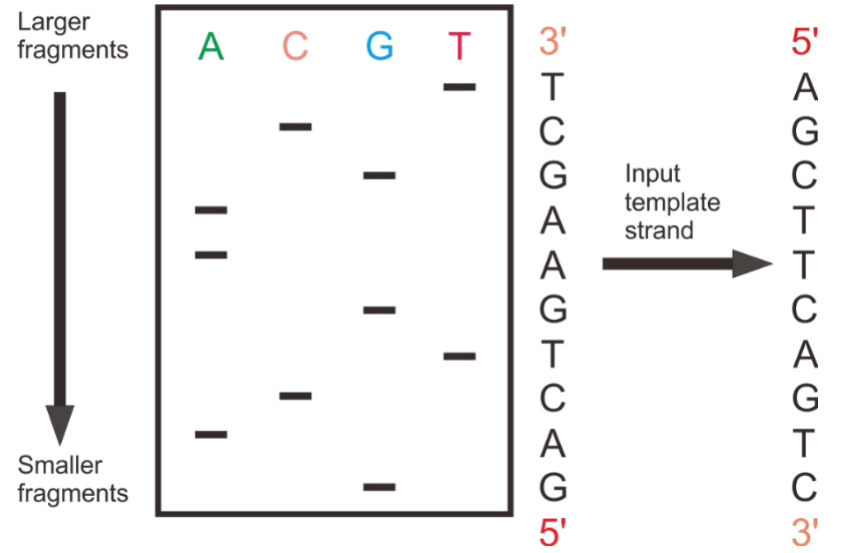
How are nested sets of DNA molecules created? Two methods were invented for doing this.

In one, DNA molecules are radioactively labeled at their 5' termini and are then subjected to four different regimens of chemical treatment that cause them to break preferentially at Gs, Cs, Ts, or As. (NO LONGER IN USE)

In the **chain termination method**, DNA is copied by DNA polymerase from a DNA template starting from a fixed point specified by the use of an oligonucleotide primer. The chain termination method employs special, modified substrates called 2 '-,3'-dideoxynucleotides (ddNTPs), which lack the 3 '-hydroxyl group on

their sugar moiety. DNA Polymerase will incorporate a 2'-,3'-dideoxynucleotide at the 3' end of a growing polynucleotide chain but once incorporated, the presence of the modified nucleotide causes elongation to terminate. The reason for this is the absence at the 3' end of the growing chain of a 3'-hydroxyl, which is needed for nucleophilic attack on the next incoming substrate molecule.

suppose that we "spike" a cocktail of the nucleotide substrates with the modified substrate 2'-,3'-dideoxyguanosine triphosphate (ddGTP) at a ratio of one ddGTP molecule to 100 2'- deoxy -GTP molecules (dGTP). This will cause DNA synthesis to abort at a frequency of one in one hundred every time the DNA polymerase encounters a C on the template strand (Figure 20-14a). Because all of the DNA chains commence growth from the same point, the chain-terminating nucleotides will generate a nested set of polynucleotide fragments, all sharing the same 5’ end but differing in their lengths and hence their 3’ ends. The length of the fragments, therefore, specifies the position of Cs in the template strand. If the fragments are labelled at their 5’ end through the use of a radioactively labeled primer, a primer that had been tagged with a fluorescent adduct, or at their 3' end with fluorescently labeled derivatives of ddGTP, then upon electrophoresis through a polyacrylamide gel the nested set of fragments would yield a ladder of fragments, each rung of the ladder representing a C on the template strand. If we similarly spike DNA synthesis reactions with ddCTP, ddATP, and ddTTP, then in toto we will generate four nested sets of fragments, which together provide the full nucleotide sequence of the DNA.



REPORTER GENE

Reporter genes are able to easily quantify the gene expression of a area of DNA. In this methodology, we have to find a reporter gene that is activated by the same promoter as the gene we are studying. They give us interesting information about how the promoter responds to changing environments in vivo. In the lab, the coding region for the reporter gene must be inserted into the coding region after the promoter in question(A) to study the promoter. To study the stability of the mRNA, the reporter gene must be inserted into the gene coding region without disturbing the mRNA(C), and ultimately we can study the localization of the protein(B) by inserting the reporter gene after the localization signal.

When expressed in a mammalian cell, a typical yeast activator can stimulate transcription. This is tested using a reporter gene. The reporter gene consists of binding sites for the yeast activator inserted upstream of the promoter of a gene whose expression level is readily measured

We have a gene and we want to insert this gene is another organism. To be sure that the promoter of the gene works and how strongly it works (how much of the gene is encoded) we need something called **reporters** which expression is readly measured**.**

There are different kind of reporters: FLUORESCENTS, CHROMOPROTEINS, CHROMOGENS AND LUMINESCENT PROTEINS

CHROMOPROTEINS: are just protein with a colour, once inserted inside a gene of interest, the chromoprotein encoding sequence in the gene will eventually produce this coloured protein. The specific amount of protein produced will be determined by the strenght of the promoter. This will give us an idea of how much strong the promoter is and how much quantity of gene of interest is going to be encoded

CHROMOGENS: the must common chromogen is a molecule called **X-gal** which is "eaten" by beta-galactosidase and produce glucose and a blue pigment

X-Gal + beta-galactosidase → glucose + blue pigment

The quantity of blue pigment produced is a measure of how strong the promoter is. We need to insert the gene coding for beta-galactosidase in another region of the genome.

This chromogen is also a good method to see if the gene has been correctly inserted inside the new organism

LUMINESCENT PROTEINS: these proteins respond to the presence of ATP undergoing to a conformational change breaking the ATP to ADP. When ADP + P are released the protein come back to its original conformation and in doing so it releases a wavelength of light that can be measured

FLUORESCENT PROTEINS: (the most used method)

Green**F**luorescent**P**rotein is derived from jellyfish. GFP responds to blue light, blue light is able to trigger a conformational change in the protein and in doing so is put in an unfavourable state so when it switches back to normal conformation it releases light. Since some of the energy released is dissipated as heat the wavelength increase and the colour emitted is green (not blue anymore)

EMSA

An electrophoretic mobility shift assay (EMSA) or mobility shift electrophoresis, also referred as a gel shift assay, gel mobility shift assay, band shift assay, or gel retardation assay, is a common affinity electrophoresis technique used to study protein–DNA or protein–RNA interactions. This procedure can determine if a protein or mixture of proteins is capable of binding to a given DNA or RNA sequence, and can sometimes indicate if more than one protein molecule is involved in the binding complex. Gel shift assays are often performed in vitro concurrently with DNase footprinting, primer extension, and promoter-probe experiments when studying transcription initiation, DNA replication, DNA repair or RNA processing and maturation, as well as pre-mRNA splicing.

PRINCIPLE)

A mobility shift assay is electrophoretic separation of a protein–DNA or protein–RNA mixture on a polyacrylamide or agarose gel for a short period (about 1.5-2 hr for a 15 - to 20-cm gel). The speed at which different molecules (and combinations thereof) move through the gel is determined by their size and charge, and to a lesser extent, their shape (see gel electrophoresis). The control lane (DNA probe without protein present) will contain a single band corresponding to the unbound DNA or RNA fragment. However, assuming that the protein is capable of binding to the fragment, the lane with a protein that binds present will contain another band that represents the larger, less mobile complex of nucleic acid probe bound to protein which is 'shifted' up on the gel (since it has moved more slowly).

Under the correct experimental conditions, the interaction between the DNA (or RNA) and protein is stabilized and the ratio of bound to unbound nucleic acid on the gel reflects the fraction of free and bound

probe molecules as the binding reaction enters the gel. This stability is in part due to a "caging effect", in that the protein, surrounded by the gel matrix, is unable to diffuse away from the probe before they recombine. If the starting concentrations of protein and probe are known, and if the stoichiometry of the complex is known, the apparent affinity of the protein for the nucleic acid sequence may be determined. Unless the complex is very long lived under gel conditions, or dissociation during electrophoresis is taken into account, the number derived is an apparent Kd. If the protein concentration is not known but the complex stoichiometry is, the protein concentration can be determined by increasing the concentration of DNA probe until further increments do not increase the fraction of protein bound. By comparison with a set of standard dilutions of free probe run on the same gel, the number of moles of protein can be calculated.

VARIANTS)

An antibody that recognizes the protein can be added to this mixture to create an even larger complex with a greater shift. This method is referred to as a supershift assay, and is used to unambiguously identify a protein present in the protein – nucleic acid complex. Often, an extra lane is run with a competitor oligonucleotide to determine the most favorable binding sequence for the binding protein. The use of different oligonucleotides of defined sequence allows the identification of the precise binding site by competition (not shown in diagram). Variants of the competition assay are useful for measuring the specificity of binding and for measurement of association and dissociation kinetics. Thus, EMSA might also be used as part of a SELEX experiment to select for oligonucleotides that do actually bind a given protein. Once DNA-protein binding is determined in vitro, a number of algorithms can narrow the search for identification of the transcription factor. Consensus sequence oligonucleotides for the transcription factor of interest will be able to compete for the binding, eliminating the shifted band, and must be confirmed by supershift. If the predicted consensus sequence fails to compete for binding, identification of the transcription factor may be aided by Multiplexed Competitor EMSA (MC-EMSA), whereby large sets of consensus sequences are multiplexed in each reaction, and where one set competes for binding, the individual consensus sequences from this set are run in a further reaction. For visualization purposes, the nucleic acid fragment is usually labelled with a radioactive, fluorescent or biotin label. Standard ethidium bromide staining is less sensitive than these methods and can lack the sensitivity to detect the nucleic acid if small amounts of nucleic acid or single-stranded nucleic acid(s) are used in these experiments. When using a biotin label, streptavidin conjugated to an enzyme such as horseradish peroxidase is used to detect the DNA fragment. While isotopic DNA labeling has little or no effect on protein binding affinity, use of non-isotopic labels including flurophores or biotin can alter the affinity and/or stoichiometry of the protein interaction of interest. Competition between fluorophore- or biotin-labeled probe and unlabeled DNA of the same sequence can be used to determine whether the label alters binding affinity orstoichiometry.

MAXAM GILBERT SEQUENCING

Similarly to Singer, Maxam Gilbert ‘s aim is to detect DNA sequence

Denaturation of DNA is the first step, DNA is labelled and subjected to 4 different chemical treatments in four different tubes:

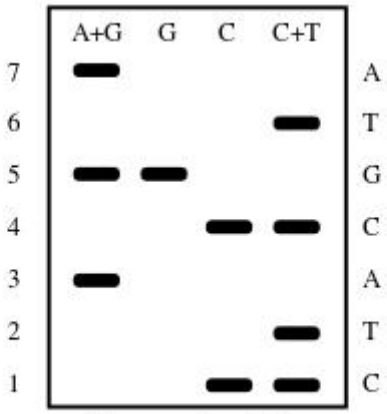
-**Formic acid** produce an apurinic site by cleavage of purines (A and

G) -**dimethyl sulfide** lead to the methylation and removal of guanines

(G) -**Hydrazine** yield the removal of pyrimidines

-**hydrazine + sodium chloride** lead to the removal of only Cytosines (C)

Using piperidine in all the for tubes we obtain cleavage where the base is missing, performing gel electrophoresis you obtain:

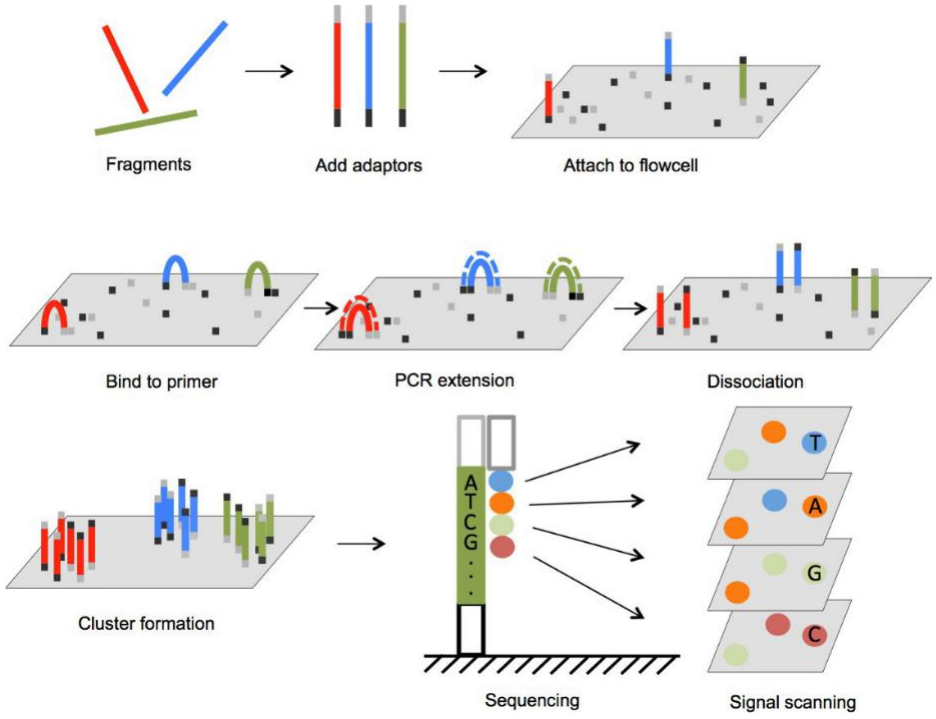


Sequencing Gel

SOLEXA ILLUMINA DYE SEQUENCING

Solexa illumine dye sequencing is one of the next generation sequencing and gives us information using instead of gel electrophoresis DNA amplification (PCR and Bridge amplification) the nature of the bases are detected thorough fluorescent signal. The advantage of this technique is that it gives you a huge amount of data.

The addition of adaptors (i.e. small portion of oligonucleotide) allow you to have the DNA clustered on a flowcell, once your sample is ready you have the synthesis of labelled oligonucleotide that release a fluorescent signal. At the end of the process you obtain a great number of reads to perform your analysis.



NUCLEAR RUN ON

The aim of nuclear run on is to see the rate of transcription of a specific gene in different condition

Imagine to have two flasks, one containing the activator for your gene of interest one lacking it

What you do is performing DNA transcription adding radioactive NTP in order to spot the transcription results.

Using **heparin** you make sure you have just one round of transcription since heparin is an ionic polysaccharide that bound to RNA polymerase doesn't allow further transcription (DNA is negatively charged as well as heparin).

So now you have to distinguish between your gene of interest and the others gene that had performed transcription as well, this is done thorough Dot blot also called slot blot.

- **Dot blot** consist in in having a filter containing the complementary strand of your gene of interest in order to isolate and compare the two transcriptions rate

CRISPS

CRISPR technology is a simple yet powerful tool for editing genomes. It allows researchers to easily alter DNA sequences and modify gene function. Its many potential applications include correcting genetic defects, treating and preventing the spread of diseases and improving crops. However, its promise also raises ethical concerns.

In popular usage, "CRISPR" (pronounced "crisper") is shorthand for "CRISPR-Cas9." CRISPRs are specialized stretches of DNA. The protein Cas9 (or "CRISPR-associated") is an enzyme that acts like a pair of molecular scissors, capable of cutting strands of DNA.

CRISPR technology was adapted from the natural defense mechanisms of bacteria and archaea (the domain of single-celled microorganisms) . These organisms use CRISPR-derived RNA and various Cas proteins, including Cas9, to foil attacks by viruses and other foreign bodies. They do so primarily by chopping up and destroying the DNA of a foreign invader. When these components are transferred into other, more complex, organisms, it allows for the manipulation of genes, or "editing."

CRISPR - **Cas9**: The key players

CRISPRs: "CRISPR" stands for "**clusters of regularly interspaced short palindromic repeats**". It is a specialized region of DNA with two distinct characteristics: the presence of nucleotide repeats and spacers. Repeated sequences of nucleotides — the building blocks of DNA — are distributed throughout a CRISPR region. Spacers are bits of DNA that are interspersed among these repeated sequences.

In the case of bacteria, the spacers are taken from viruses that previously attacked the organism. They serve as a bank of memories, which enables bacteria to recognize the viruses and fight off future attacks.

This was first demonstrated experimentally by a team of researchers. The researchers used Streptococcus thermophilus bacteria as their model. They observed that after a virus attack, new spacers were incorporated into the CRISPR region. Moreover, the DNA sequence of these spacers was identical to parts of the virus genome. They also manipulated the spacers by taking them out or putting in new viral DNA

sequences. In this way, they were able to alter the bacteria's resistance to an attack by a specific virus. Thus, the researchers confirmed that CRISPRs play a role in regulating bacterial immunity.

CRISPR RNA (crRNA): Once a spacer is incorporated and the virus attacks again, a portion of the CRISPR is transcribed and processed into CRISPR RNA, or "crRNA." The nucleotide sequence of the CRISPR acts as a template to produce a complementary sequence of single-stranded RNA. Each crRNA consists of a nucleotide repeat and a spacer portion.

Cas9: The Cas9 protein is an enzyme that cuts foreign DNA. The protein typically binds to two RNA molecules: crRNA and another called tracrRNA (or "trans -activating crRNA"). The two then guide Cas9 to the target site where it will make its cut. This expanse of DNA is complementary to a 20-nucleotide stretch of the crRNA.

Using two separate regions, or "domains" on its structure, Cas9 cuts both strands of the DNA double helix, making what is known as a "double-stranded break," according to the 2014 Science article.

There is a built-in safety mechanism, which ensures that Cas9 doesn't just cut anywhere in a genome. Short DNA sequences known as PAMs ("protospacer adjacent motifs") serve as tags and sit adjacent to the target DNA sequence. If the Cas9 complex doesn't see a PAM next to its target DNA sequence, it won't cut.

CRISPR-Cas9 as a genome-editing tool

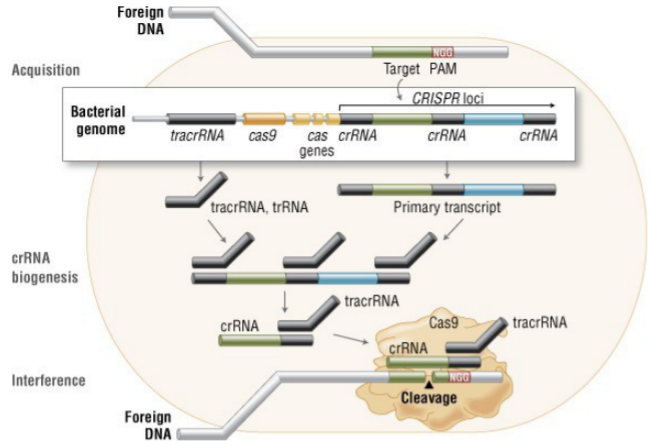
The genomes of various organisms encode a series of messages and instructions within their DNA sequences. Genome editing involves changing those sequences, thereby changing the messages. This can be done by inserting a cut or break in the DNA and tricking a cell's natural DNA repair mechanisms into introducing the changes one wants. CRISPR-Cas9 provides a means to do so.

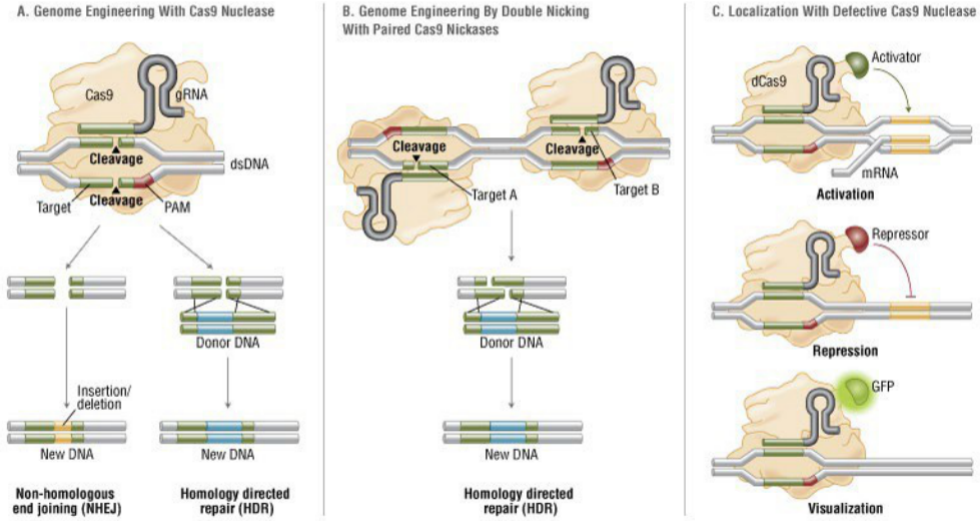
The studies, conducted by separate groups, concluded that Cas9 could be directed to cut any region of DNA. This could be done by simply changing the nucleotide sequence of crRNA, which binds to a complementary DNA target

Once the DNA is cut, the cell's natural repair mechanisms kick in and work to introduce mutations or other changes to the genome. There are two ways this can happen.

One repair method involves gluing the two cuts back together. This method, known as "non-homologous end joining," tends to introduce errors.

In the second method, the break is fixed by filling in the gap with a sequence of nucleotides. In order to do so, the cell uses a short strand of DNA as a template. Scientists can supply the DNA template of their choosing, thereby writing-in any gene they want, or correcting a mutation





CHIP (CHROMATIN IMMUNO PRECIPITATION)

Chromatin = DNA + protein
Immuno = antibody
Chip utilizes antibody that selectively recognise and binds protein, including histones, histone modification, transcription factor and so on in order to detect DNA-protein interaction.

1)**Crosslink** , when protein react with nucleotides of DNA, forming a covalent linkage between them, in this step you stabilize and lock DNA-protein reaction
Formaldehyde is used for crosslinking, glycine is instead used to stop the reaction , is important not to have over fixation because it reduce the fragmentation efficiency and inhibit the binding with the antibody
The cell membrane is permeabilized by lysis solution, cellular components are liberated and cytosolic protein are removed

2)**Chromatin fragmentation**, cleavage of chromatin in smaller fragment
Fragmentation obtained through sonication nuclease activity and enzymatic digestion
Fragment size = 200 / 800 bp long
3)**Immunoprecipitation**, recognition of the protein of interest through antibodies
Antibodies are specific for the protein of interest

4) **DNA purification and analysis**
Once DNA is purified we have to control that all the DNA we have correspond to the fragment we wanted and so we have to control we can do it with PCR because of course different results in a PCR experiment depends on

the abundance (and technically the'unwanted DNA' is less) or we can use Microarrays (same idea)
The process that consider the use of microarray acter the chip is called ChIP-Chip and has two limitations :
-the resolution is limited
-only protein for witch we have an antibody are "usable"
Another approach is the Chip-Seq that simply skip the array control and goes directly to the DNA-sequencing

Gene regulation and proteomics

The least wasteful choice for regulation is transcriptional regulation, where no cell materials are 'wasted'. There is however other important mechanism of regulation, deemed post-transcriptional, and post translational regulations which generally act faster than transcriptional regulation. Post transcriptional regulation specifically focuses on the regulation of stability, transport, maturation, and translation of mRNA, while post translational regulations focus on turning on or turning off the activity of the protein, even permanently through proteolysis.

The most powerful way of studying all of these influences is still through the study of RNA. For bioinformatics, this means studying metabolome, the fluctuation in phenotype. While the metabolome (Metabolic molecules in use) is often hard to study in a lab, the proteome (all possible proteins in the genome) can be easier to study but is still less global than transcriptome (all mRNA and tRNA in the cell). For this reason, we often choose to study transcriptomics for predicting phenotypes due to its global nature. Because of the possible regulation steps such as post-transcriptions and post-translational, the results from the transcriptome may not necessarily correlate to what we would expect, that is the mRNA is not guaranteed to translate to a protein, nor can we assume a protein that is created will be functional causing a change in metabolome.

Proteomics has evolved from gel electrophoresis, to so called high-throughput proteomics. Proteins HTP are digested into a peptide mixture, which is introduced into a mass spectrometry/gas analysis, from which the peptides can be identified as peaks in the output and match them to elements in a database. In reality, because of the massive number of proteins present in higher organisms, the peaks are too small to properly identify a protein.

For this reason, transcriptomics is more often studied, in which DNA microarrays are synthesized in a way that we can visualize each RNA that will be synthesized in each pane of the microarray. These microarrays let us study gene regulation, gene expression changes, or even diagnostic microarrays used as biomarkers for sick vs healthy tissue. They can also be used to visualize genome organization, chromatin organization and mutations. The best sequences to assemble microarrays is by using very short gene sequences, not always possible.

Reverse transcription can be used, along with DNA denaturation in such a way that the singe stranded DNA binds to the reversely transcribed elements. These double stranded regions are much more stable than the single stranded regions, and have a much better resistance to denaturation, allowing the target regions of DNA to be separated.

FGS (First Generation Sequencing)

DNA sequences can be read using special gel electrophoreses, but the DNA must be cut, labeled and separated, as well as having to read the output manually, making it an extremely time-

consuming process. Another method was later proposed in which a DNA Pol starting from a dimer would copy a length of DNA using labeled nucleotides molecularly modified called Chain Terminators to block addition of another base after attached. These chain terminators need to be mixed in correct proportions in order to get enough reading while still allowing the chain to elongate. One of the four nucleotides must be labeled so it will appear in the gel. In modern labs, four different fluorochromes are used to distinguish the bases.

NGS

New generation techniques instead fragment and sequence without closing, increasing the efficiency from 700 to billions sequences read per experiment. Present sequencing called third gene sequencing allows us to work with single DNA molecules, without the need to fragment. This has drastically decreased the cost of DNA sequencing, allowing much more data to be available.

DNA sequencing with commercially available NGS platforms is generally conducted with the following steps. First, DNA sequencing libraries are generated by clonal amplification by PCR in vitro. Second, the DNA is sequenced by synthesis, such that the DNA sequence is determined by the addition of nucleotides to the complementary strand rather than through chain-termination chemistry. Third, the spatially segregated, amplified DNA templates are sequenced simultaneously in a massively parallel fashion without the requirement for a physical separation step.