Polymerase Chain Reaction (PCR)
## Principle, Procedure, Types, Applications and Animation

Polymerase Chain Reaction (PCR) is a powerful method for amplifying particular segments of DNA, distinct from cloning and propagation within the host cell. This procedure is carried out entirely biochemically, that is, in vitro. PCR was invented by Kary Mullis in 1983. He shared the Nobel Prize in chemistry with Michael Smith in 1993.

Principle of PCR

PCR uses the enzyme DNA polymerase that directs the synthesis of DNA from deoxynucleotide substrates on a single-stranded DNA template. DNA polymerase adds nucleotides to the 3` end of a custom-designed oligonucleotide when it is annealed to a longer template DNA. Thus, 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 its 3` end to generate an extended region of double stranded DNA.

Procedure/Steps of PCR

Polymerase chain reaction - PCR

1. Denaturation at 94-96°C
2. Annealing at ~68°C
3. Elongation at ca. 72 °C
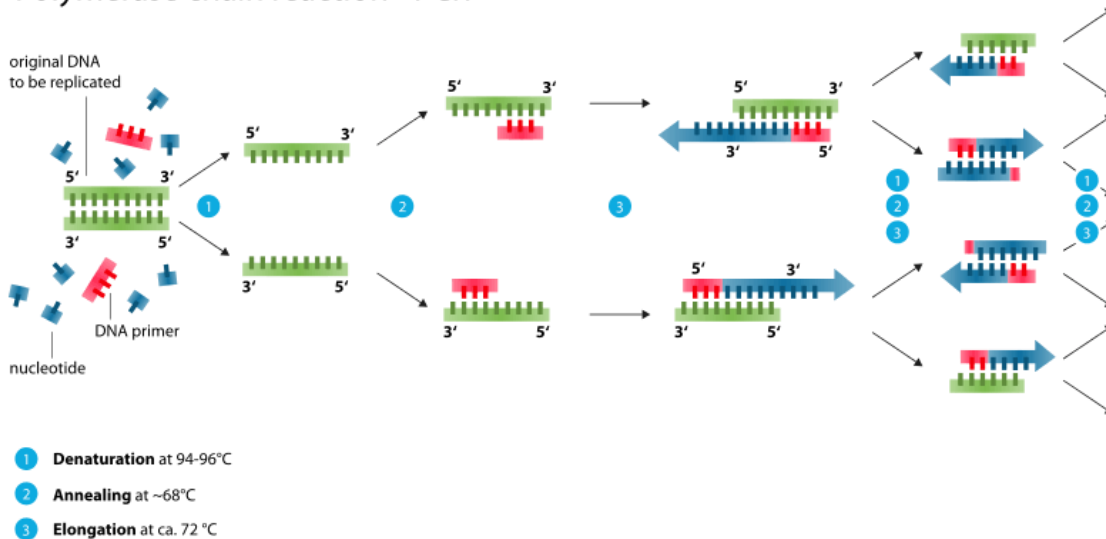
1. Denaturation
The DNA template is heated to 94° C. This breaks the weak hydrogen bonds that hold DNA strands together in a helix, allowing the strands to separate creating single stranded DNA.

2. Annealing

The mixture is cooled to anywhere from 50-70° C. This allows the primers to bind (anneal) to their complementary sequence in the template DNA.

3. Extension
The reaction is then heated to 72° C, the optimal temperature for DNA polymerase to act. DNA polymerase extends the primers, adding nucleotides onto the primer in a sequential manner, using the target DNA as a template.
With one cycle, a single segment of double-stranded DNA template is amplified into two separate pieces of double-stranded DNA. These two pieces are then available for amplification in the next cycle. As the cycles are repeated, more and more copies are generated and the number of copies of the template is increased exponentially.

Types of PCR

1. Real-time PCR
2. Quantitative real time PCR (Q-RT PCR)
3. Reverse Transcriptase PCR (RT-PCR)
4. Multiplex PCR
5. Nested PCR
6. Long-range PCR
7. Single-cell PCR
8. Fast-cycling PCR
9. Methylation-specific PCR (MSP)
10. Hot start PCR
11. High-fidelity PCR
12. In situ PCR
13. Variable Number of Tandem Repeats (VNTR) PCR
14. Asymmetric PCR
15. Repetitive sequence-based PCR
16. Overlap extension PCR
17. Assemble PCR
18. Intersequence-specific PCR(ISSR)
19. Ligation-mediated PCR
20. Methylation –specifin PCR
21. Miniprimer PCR
22. Solid phase PCR
23. Touch down PCR, etc

Applications of PCR

1. PCR is used in analyzing clinical specimens for the presence of infectious agents, including HIV, hepatitis, malaria, anthrax, etc.
2. PCR can provide information on a patient's prognosis, and predict response or resistance to therapy. Many cancers are characterized by small mutations in certain genes, and this is what PCR is employed to identify.
3. PCR is used in the analysis of mutations that occur in many genetic diseases (e.g. cystic fibrosis, sickle cell anaemia, phenylketonuria, muscular dystrophy).

4. PCR is also used in forensics laboratories and is especially useful because only a tiny amount of original DNA is required, for example, sufficient DNA can be obtained from a droplet of blood or a single hair.
5. PCR is an essential technique in cloning procedure which allows generation of large amounts of pure DNA from tiny amount of template strand and further study of a particular gene.
6. The Human Genome Project (HGP) for determining the sequence of the 3 billion base pairs in the human genome, relied heavily on PCR.
7. PCR has been used to identify and to explore relationships among species in the field of evolutionary biology. In anthropology, it is also used to understand the ancient human migration patterns. In archaeology, it has been used to spot the ancient human race. PCR commonly used by Paleontologists to amplify DNA from extinct species or cryopreserved fossils of millions years and thus can be further studied to elucidate on.

# *Restriction Fragment Length Polymorphism (RFLP)*

- Restriction Fragment Length Polymorphism (RFLP) is a technique in which organisms may be differentiated by analysis of patterns derived from cleavage of their **DNA**.
- It is a technique that exploits variations in homologous DNA sequences.
- A restriction fragment length polymorphism is defined by the existence of alternative alleles associated with restriction fragments that differ in size from each other. Simply, the variations in the restriction DNA fragments length between individuals of a species is called RFLP.
- The basic technique of identifying such restriction fragment length polymorphisms involve fragmenting a sample of DNA by a restriction enzyme, which can recognize and cut DNA wherever a specific short sequence occurs, in a process known as a restriction digest.
- The resulting DNA fragments are then separated by length through a process known as agarose gel electrophoresis, and transferred to a membrane via the Southern blot procedure. Hybridization of the membrane to a labeled DNA probe then determines the length of the fragments which are complementary to the probe.
- An RFLP occurs when the length of a detected fragment varies between individuals.
- Although now largely obsolete due to the rise of inexpensive DNA sequencing technologies, RFLP analysis was the first DNA profiling technique inexpensive enough to seek widespread application.
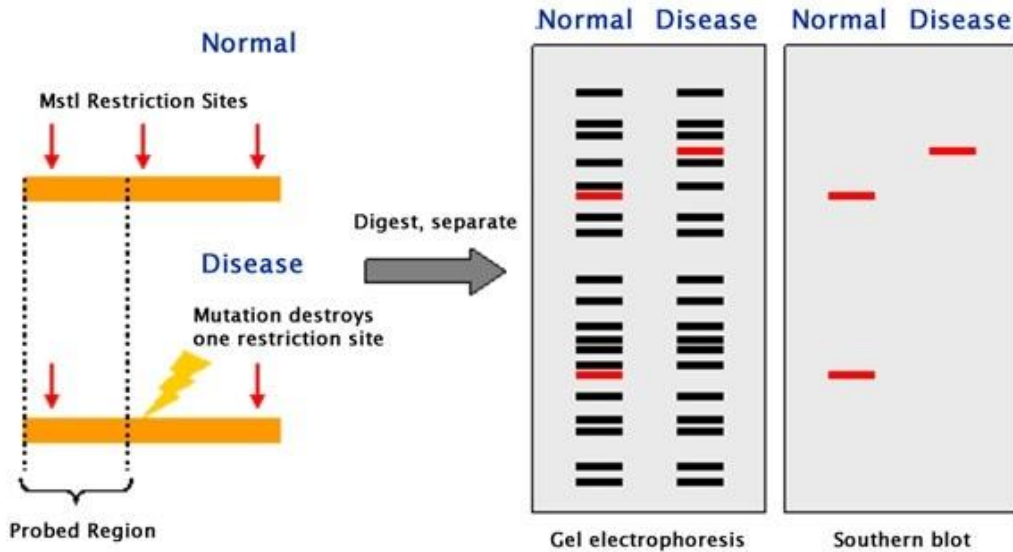
Principle of Restriction Fragment Length Polymorphism (RFLP)

If two organisms differ in the distance between sites of cleavage of a particular restriction endonuclease, the length of the fragments produced will differ when the DNA is digested with a restriction enzyme. The similarity and differences of the patterns thus generated can be used to differentiate species (and even strains) from one another.

Steps Involved in Restriction Fragment Length Polymorphism (RFLP)
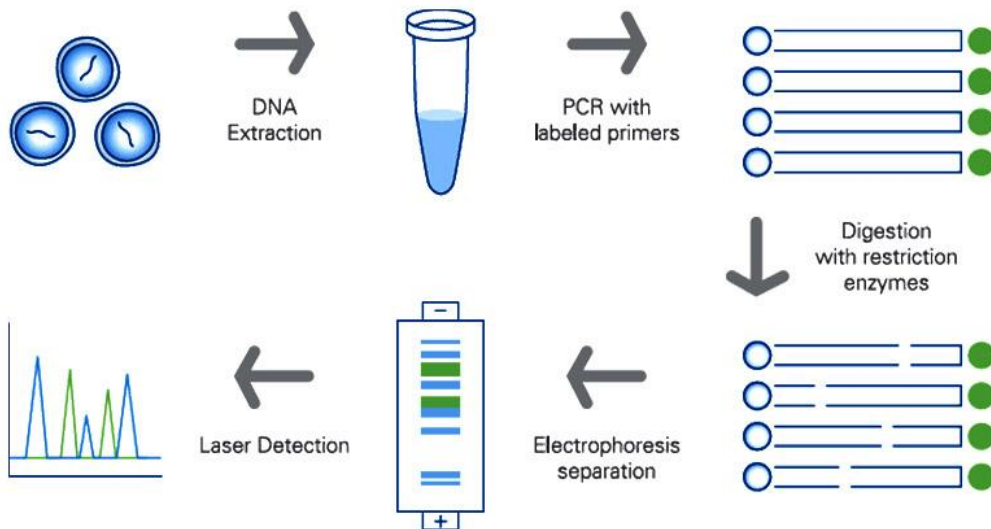
1. The first step in this process is to **isolate the DNA** from the target.
2. Once the the DNA is isolated from the sample it is **subjected to restriction digestion** using restriction enzymes.
3. The digested DNA sample is then subjected to **gel electrophoresis**, in which the DNA is separated based on its size. Many DNA fragments with slight differences in length are produced.

4. The gel is then exposed to a chemical to denature double-stranded DNA to become single-stranded.
5. This is followed by southern blotting where DNA is transferred from gel to nylon membrane.
6. The nylon membrane is then exposed to solution with radioactive complementary nucleotide probes that hybridize to specifically chosen DNA sequences on nylon membrane.
7. The membrane is then placed against X- ray film, where hybridized radioactive probes cause exposure of X-ray film, producing an autoradiogram.
8. RFLP analysis is carried out to detect differences in pattern to confirm polymorphisms.

Applications of Restriction Fragment Length Polymorphism (RFLP)

RFLP analysis was formerly an important tool in genome mapping, localization of genes for genetic disorders, determination of risk for disease, and paternity testing.

RFLP can be used in many different settings to accomplish different objectives:

1. In paternity cases or criminal cases to determine the source of a DNA sample. (i.e. it has forensic applications).
2. Determining the disease status of an individual. (e.g. it can be used in the detection of mutations)
3. To measure recombination rates which can lead to a genetic map with the distance between RFLP loci.
4. In the characterization of genetic diversity or breeding patterns in animal populations.
5. RFLP has been developed for chromosomes mapping of humans, mice, maize, tomato, rice, etc.

Advantages of Restriction Fragment Length Polymorphism (RFLP)

- The main advantage of RFLP analysis over PCR-based protocols is that no prior sequence information, nor oligonucleotide synthesis, is required.
- Results are based on reliable genotypic characteristics' rather than on phenotypes.
- RFLP based Genetic Marker
- RFLP is the co dominant marker thus can estimate heterozygosity.
- RFLP & is very useful study in Genomic DNA Sequence.
- Highly robust methodology with good transferability between laboratories.

Limitations of Restriction Fragment Length Polymorphism (RFLP)

- Slow
- Cumbersome
- Requires a large amount of sample DNA.
- Automation not possible
- Low levels of polymorphism in some species
- Few loci detected per assay
- Need a suitable probe library
- Needing the combined process of probe labeling, DNA fragmentation, electrophoresis, blotting, hybridization, washing, and autoradiography.

RANDOM AMPLIFIED POLYMORPHIC DNA

Introduction

The quality of DNA, obtained by various protocols and subjected to various storage conditions can be judged by comparing the PCR amplification profiles. The most commonly used technique that can be accessed by almost any laboratory is RAPD . Random amplified polymorphic DNA (RAPD) is a PCR based technique for identifying genetic variation. It involves the use of a single arbitrary primer in a PCR reaction, resulting in the amplification of many discrete DNA products. The technique was developed independently by two different laboratories (Williams et. al., 1990; Welsh and McClelland, 1990) and called as RAPD and AP-PCR (Arbitary primed PCR) respectively. This procedure detects nucleotide sequence polymorphisms in a DNA amplification-based assay using only a single primer of arbitrary nucleotide sequence. In this reaction, a single species of primer binds to the genomic DNA at two different sites on opposite strands of the DNA template. If these priming sites are within an amplifiable distance of each other, a discrete DNA product is produced through thermocylic amplification. The polymorphisms between individuals result from sequence differences in one or both of the primer binding sites, and are visible as the presence or absence of a particular RAPD band. Such polymorphisms thus behave as dominant genetic markers.

Principle

The RAPD technique is based on the polymerase chain reaction (PCR). A target DNA sequence is exponentially amplified with the help of arbitrary primers, a thermostable DNA polymerase, dideoxy nucleotide tri - phosphates, magnesium and reaction buffer. The reaction involves repeated cycles, each consisting of a denaturation, a primer annealing and an elongation step (fig 1). In the first step C (denaturation). In the second°the DNA is made single stranded by raising the temperature to 94°C results in annealing of the primer to their°step, lowering of the temperature to about 40 to 65 target sequences on the template DNA (annealing step). In the third cycle, temperature is chosen where the activity of the thermostable Taq DNA polymerase is optimal, i.e., usually 72°C.

The polymerase now extends the 3' ends of the DNA-primer hybrids towards the other primer binding site. Since this happens at both primer-annealing sites on both the DNA strands, the target fragment is completely replicated. Repeating these three step cycles 40 to 50 times results in the exponential amplification of the target between the 5' ends of the two primer binding sites. Amplification products are separated by gel electrophoresis and visualized by ethidium bromide staining. The most crucial factors that need to be optimized in a RAPD reaction are the magnesium concentration, enzyme concentration, DNA concentration and annealing temperature of the primer.

The G+C content of the primers should be 40-60% and care should be taken to avoid sequences that produce internal secondary structures. The annealing temperature for a PCR cycle is generally 3-5 0 below the melting temperature (Tm) of the primer. The Tm is calculated as Na+[(log10•Tm = 81.5 +16.6 (%G+C) – 675/n•) + 0.41] Na+[Where K[is the molar salt concentration ;] + Na+[=]] and n = number of base in the oligonucleotide As a thumb rule the following formula can be used Tm = 2 (A+T) + 4(G+C) Other useful formulae are Nanogram of primer = picomole of primer x 0.325 x # bases L) in which the primer isµMicroMolar concentration of primer = picomoles of primer/ volume (dissolved. Material and reagents Instruments: PCR machine (Perkin Elmer 9600), microcentrifuge, 100V power supply, Gel electrophoresis tank, gel mould and slot former, UV transilluminator, Camera, autopipettes, vortex.
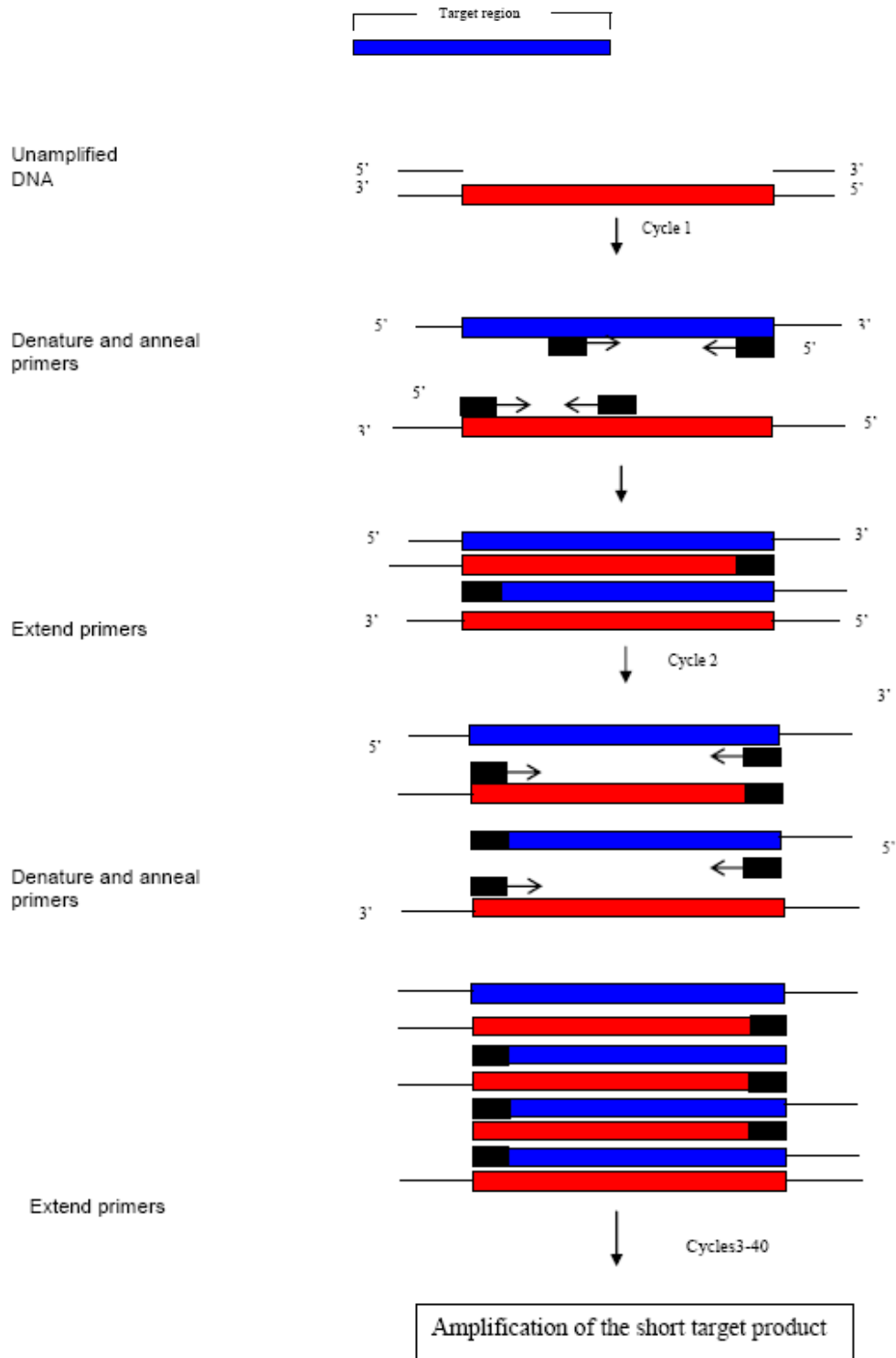
Figure 1. Schematic diagram of the PCR process

Reagents:
 1. Taq DNA polymerase
2. Genomic DNA (5ng/ μL)
3. dNTP mix (2mM each of dATP, dCTP, dGTP and dTTP)
4. MgCl2 (25mM)
5. Buffer for DNA polymerase
6. 10-mer oligonuleotide primers(5μM)
7. Sterile distilled water
8. Electrophoresis grade agarose
9. 0.5X TBE buffer
10. Ethidium bromide solution (10mg/ml)
11. DNA length marker
12. Loading buffer
Miscellaneous: Thin walled PCR tubes, tips, tissue paper.

Protocol

1. Each 10μl of reaction mix contains

Component	Volume	Final Concentration
Genomic DNA	3.0μl	15ng
Buffer	1.0μl	1X
dNTPs	1.0μl	0.2mM
Primer	0.6μl	0.6μM
Taq DNA pol	0.2μl	1 unit
MgCl$_2$	1.0μl	2.5mM
Water		to 10μl

Prepare a mastermix (for all samples + control) that contains all the above components except the DNA.

➢ Thaw all components completely
➢ Vortex the MgCl$_2$ vigorously
➢ Vortex the master mix to mix all components before aliquoting

2. Aliquot into PCR tubes and add the template DNA. Mix well.
3. Place the PCR plate carrying the reaction tubes in the sample block of the thermocycler.
4. Carry out an initial denaturation step at 94°C for 4min followed by 40 cycles with the following cycle parameters:
Step 1: 94°C for 1min
Step 2: 35°C for 1min
Step 3: 72°C for 2min
Extend the 72°C step of the final cycle by 5min
5. When the amplification has finished, add 3μl of the loading dye to each sample.
6. Prepare a 1.2% agarose gel in 0.5X TBE buffer containing ethidium bromide (5μg/ml of gel).Load the DNA length marker and the samples. Run the gel in 0.5 x TBE buffer at 55V for 4 h.
7. Visualize the gel on a UV transilluminator. If required the gel can be photographed using Polaroid 665 or 667 film and analysed further.

Fluorescence In Situ Hybridization (FISH) protocol

FISH Principle

Fluorescence in situ hybridization (FISH) is a technique that uses fluorescent probes which bind to special sites of the chromosome with a high degree of sequence complementarity to the probes. The fluorescent probes are nucleic acid labeled with fluorescent groups and can bind to specific DNA/RNA sequences. Thus, we can understand where and when a specific DNA sequences exist in cells by detecting the fluorescent group. It was developed in the early 1980s. Fluorescence microscopy can be used to find out where the fluorescent probe is bound to the chromosomes and flow cytometry can be used to detect the binding quantitatively. This FISH protocol is for a Cy5 and FAM labeled probe used in flow cytometry detection and fluorescence microscopy detection.
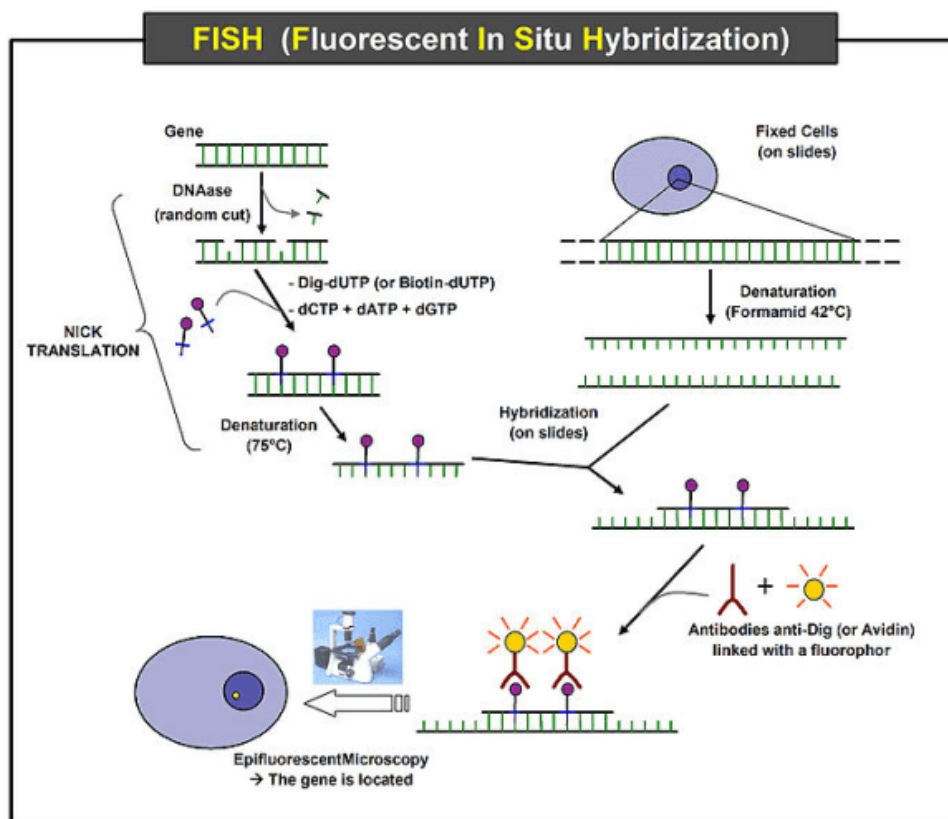

Fig. The schematic diagram of fluorescence in situ hybridization (FISH). The resource come from Wikipedia.

FISH protocol

For flow cytometry detection

1. Probes are synthesized and end-labeled at the 5'-end with either Cy5 or 6-carboxyfluorescein (FAM) and purified via HPLC. Cy5-labeled probes are used in experiments involving conventional flow cytometry and FAM-labeled probes are used for experiments involving microscopy or imaging cytometry. The probes is used as a dual probe cocktail at a total probe concentration of 5 ng/μl (2.5 ng/μl each probe).
2. Cell samples are pelleted for 5 minutes at 2,000 x and fixed for 30 min at 25°C with 500 μl 10% neutral buffered formalin before FISH procedure.
3. After fixation, the formalin is discarded and samples are washed once in 1x PBS.
4. Fixed samples are harvested via centrifugation (5 min, 2,000 x g), the supernatant is discarded and cell pellets are resuspended in 0.5 ml of cell storage solution (a 50:50 mix of PBS:absolute ethanol) and then samples can be stored at -20?C until analyzed.
5. For each hybridization reaction, 100 μl of fixed samples are pelleted (5 min, 2,000 x g) and resuspended in 100 μl of hybridization buffer (0.7 M NaCl, 0.1 M Tris [pH 8.0], 0.1% SDS, 10 mM EDTA) containing the dual probe cocktail.
6. Samples are hybridized for 30 min on a heat block set to 55ºC, followed by a 30 min wash step at the same temperature in 500 μl hybridization buffer without probe.
7. Hybridized samples are pelleted and resuspended in a 50:50 mixture of PBS and absolute ethanol, cooled to -20ºC and ship on ice via overnight.
8. Once received, samples are placed and held at -20ºC until used, for up to a week. Samples remained liquid under these storage conditions due to the ethanol content.
9. Prior to analysis, tubes are vortexed to resuspend cells and break up loosely associated flocs or aggregates.
10. 100 uL samples are pelleted via centrifugation (2,000 x g, 5 min).
11. Wash samples with 100 μl PBS + 0.5% (w/v) bovine serum albumin (BSA).
12. Centrifuge samples again at 2,000 x g, 5 min.
13. Resuspend samples in 50 μl PBS + 0.5% BSA prior to running on the flow cytometer.

For fluorescence microscopy detection

1. Prior to FISH, tape-bound cells are fixed for 30 min at 25°C by covering the sample contact area with 500 μl 10% neutral buffered formalin.
2. After fixation, the formalin is discarded and the tape is washed once in 1x phosphate buffered saline (PBS).
3. Then samples are dehydrated in ethanol (a 50, 80 and 95% (v/v) series, exposure for 3 min to 300 μl ethanol at each concentration) prior to hybridization.
4. Samples on tapes are hybridized for 15 minutes at 55?C using a moisture-sealed slide incubation chamber. Briefly, 500 μl volumes of hybridization buffer (0.7 M NaCl, 0.1 M Tris [pH 8.0], 0.1% sodium dodecyl sulfate, 10 mM EDTA, containing probe, preheated to 55 ºC) are applied to the surface of the tape and the chamber's lid is sealed, creating a moist, temperature controlled environment within the chamber.
5. After 15 min, the lid is removed and samples are briefly rinsed with probe-free hybridization buffer, preheated to 55ºC.
6. Hybridized cells on tapes are counterstained for 10 minutes in the dark with ~30 μl mounting medium containing 1.5 μg ml-1 4',6-diamidino-2-phenylindole (DAPI).
7. Then tapes are mounted with a coverslip and examined using a fluorescence microscope.

DNA Sequencing Methods

Maxam–Gilbert and Sanger Dideoxy Method

- DNA sequencing refers to methods for determining the order of the nucleotides bases adenine,guanine,cytosine and thymine in a molecule of **DNA**.
- The first DNA sequence was obtained by academic researchers, using laboratories methods based on 2- dimensional chromatography in the early 1970s.
- By the development of dye based sequencing method with automated analysis, DNA sequencing has become easier and faster.
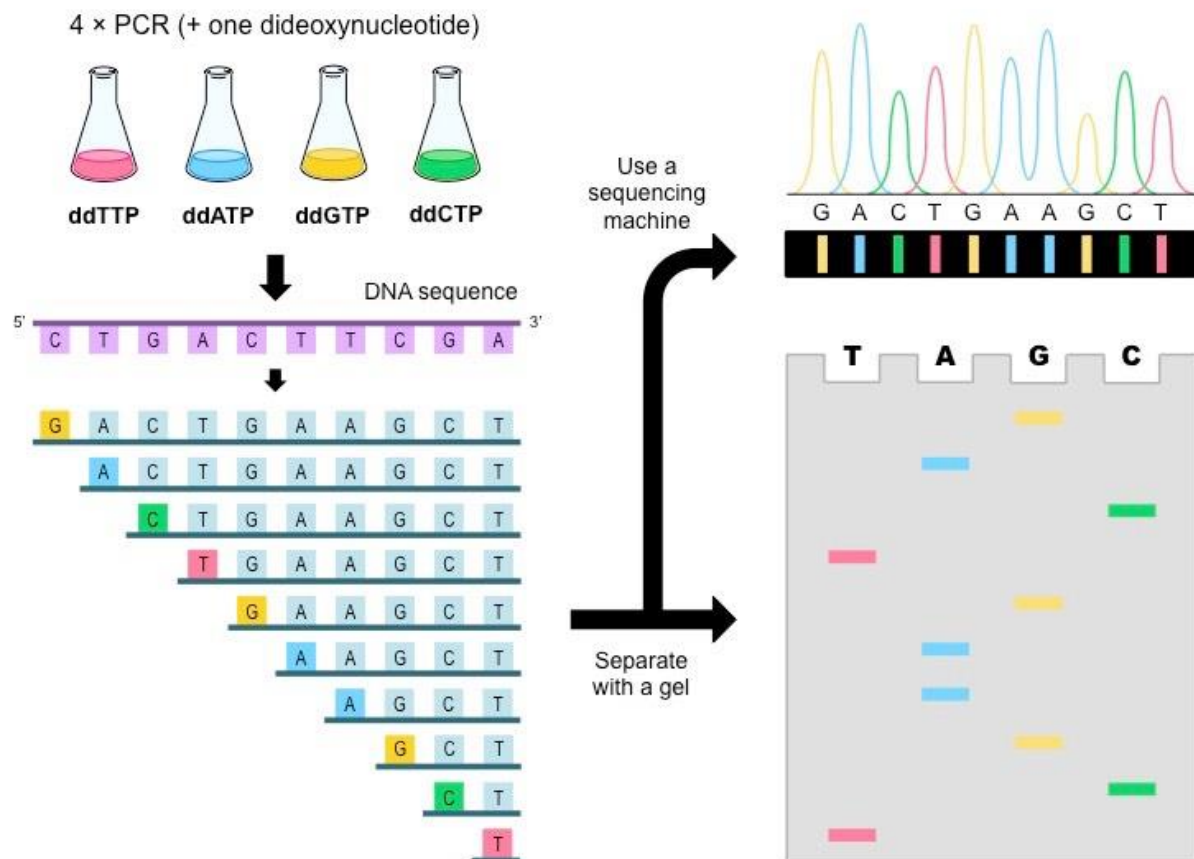
Two main methods are widely known to be used to sequence DNA:
1. **The Chemical Method** (also called the Maxam–Gilbert method after its inventors).
2. **The Chain Termination Method** (also known as the Sanger dideoxy method after its inventor).

- Maxam–Gilbert technique depends on the relative chemical liability of different nucleotide bonds, whereas the Sanger method interrupts elongation of DNA sequences by incorporating dideoxynucleotides into the sequences.
- The chain termination method is the method more usually used because of its speed and simplicity.

Chemical Cleavage Method (Maxam–Gilbert Method)

- In 1976-1977, Allan Maxam and Walter Gilbert developed a DNA sequencing method based on chemical modification of DNA and subsequent cleavage at specific bases.
- The method requires radioactive labelling at one end and purification of the DNA fragment to be sequenced.
- Chemical treatment generates breaks at a small proportions of one or two of the four nucleotide based in each of four reactions (G,A+G, C, C+T).
- Thus a series of labelled fragments is generated, from the radiolabelled end to the first 'cut' site in each molecule.
- The fragments in the four reactions are arranged side by side in gel electrophoresis for size separation.
- To visualize the fragments, the gel is exposed to X-ray film for autoradiography, yielding a series of dark bands each corresponding to a radiolabelled DNA fragment, from which the sequence may be inferred.

Key Features

- Base-specific cleavage of DNA by certain chemicals
- Four different chemicals, one for each base
- A set of DNA fragments of different sizes
- DNA fragments contain up to 500 nucleotides

Advantages

- Purified DNA can be read directly
- Homopolymeric DNA runs are sequenced as efficiently as heterogeneous DNA sequences
- Can be used to analyze DNA protein interactions (i.e. footprinting)
- Can be used to analyze nucleic acid structure and epigenetic modifications to DNA

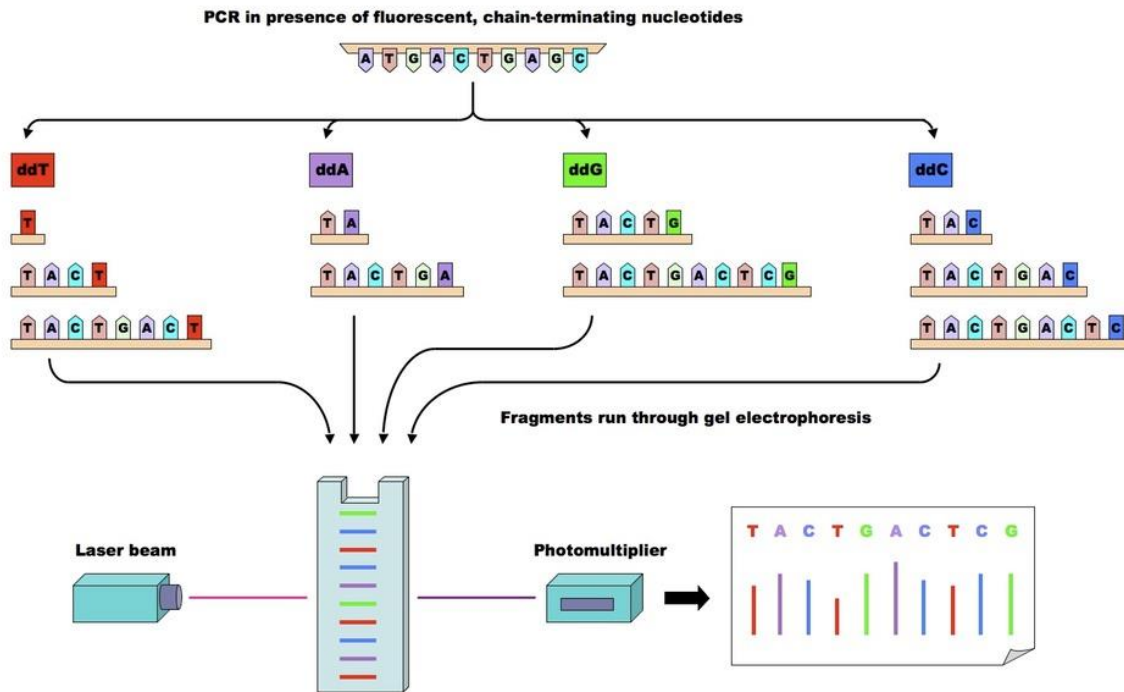## Disadvantages

- It requires extensive use of hazardous chemicals.
- It has a relatively complex set up / technical complexity.
- It is difficult to "scale up" and cannot be used to analyze more than 500 base pairs.
- The read length decreases from incomplete cleavage reactions.
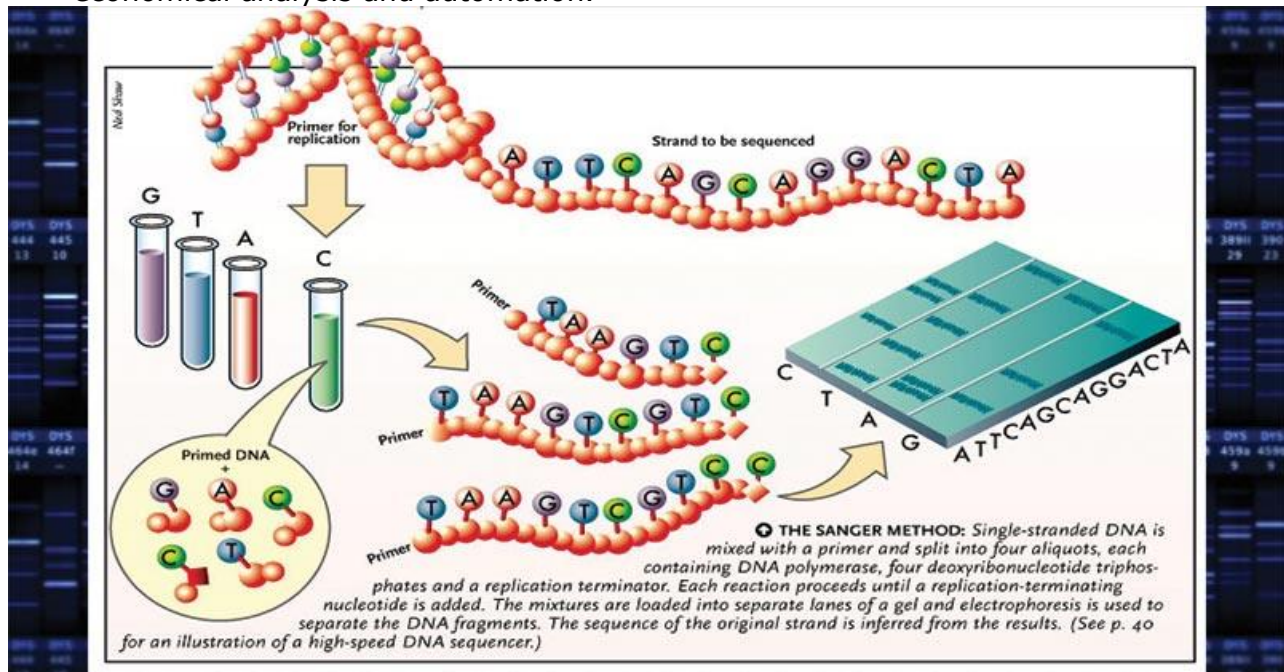- It is difficult to make Maxam-Gilbert sequencing based DNA kits.

Chain Termination Method (Sanger Dideoxy Method)

PCR in presence of fluorescent, chain-terminating nucleotides

A T G A C T G A G C

ddT | ddA | ddG | ddC

ddT:
T
T A C T
T A C T G A C T

ddA:
T A
T A C T G A

ddG:
T A C T G
T A C T G A C T C G

ddC:
T A C
T A C T G A C
T A C T G A C T C

Fragments run through gel electrophoresis

Laser beam

Photomultiplier

T A C T G A C T C G

Fluorescent fragments detected by laser and represented on a chromatogram

- The chain terminator method is more efficient and uses fewer toxic chemicals and lower amount of radioactivity than the method of Maxam and Gilbert.
- The key principle of the Sanger method was the use of dideoxynucleotide triphosphates (ddNTPs) as DNA chain terminators.
- The chain termination method requires a single-stranded DNA template, a DNA primer, a DNA polymerase, radioactively or fluorescently labelled nucleotides, and modified nucleotides that terminate DNA strand elongation.
- The DNA sample is divided into four separate sequencing reactions, containing all four of the standard deoxynucleotides (dATP, dGTP, dCTP, dTTP) and the DNA polymerase.
- To each reaction is added only one of the four dideoxynucleotide (ddATP, ddGTP, ddCTP, ddTTP) which are the chain terminating nucleotides, lacking a 3'-OH group required for the formation of a phosphodiester bond between two nucleotides, thus terminating DNA strand extension and resulting in DNA fragments of varying length.
- The newly synthesized and labelled DNA fragments are heat denatured, and separated by size by gel electrophoresis on a denaturing polyacrylamide-urea gel with each of the four reactions run in one of the four individual lanes (lanes A, T, G, C).
- The DNA bands are then visualized by autoradiography or UV light, and the DNA sequence can be directly read off the X-ray film or gel image.
- A dark band in a lane indicates a DNA fragment that is result of chain termination after incorporation of a dideoxynucleotide (ddATP, ddGTP, ddCTP, or ddTTP).
- The relative position of the different bands among the four lanes are then used to read (from bottom to top) the DNA sequence.

- The technical variations of chain termination sequencing include tagging with nucleotides containing radioactive phosphorus for labelling, or using a primer labelled at the 5' end with a fluorescent dye.
- Dye- primer sequencing facilitates reading in an optical system for faster and more economical analysis and automation.

THE SANGER METHOD: Single-stranded DNA is mixed with a primer and split into four aliquots, each containing DNA polymerase, four deoxyribonucleotide triphosphates and a replication terminator. Each reaction proceeds until a replication-terminating nucleotide is added. The mixtures are loaded into separate lanes of a gel and electrophoresis is used to separate the DNA fragments. The sequence of the original strand is inferred from the results. (See p. 40 for an illustration of a high-speed DNA sequencer.)

Key Features
- Uses dideoxy nucleotides to terminate DNA synthesis.
- DNA synthesis reactions in four separate tubes
- Radioactive dATP is also included in all the tubes so the DNA products will be radioactive.
- Yielding a series of DNA fragments whose sizes can be measured by electrophoresis.
- The last base in each of these fragments is known.

Advantage
Chain termination methods have greatly simplified DNA sequencing.

Limitations
- Non-specific binding of the primer to the DNA, affecting accurate read-out of the DNA sequence.
- DNA secondary structures affecting the fidelity of the sequence.

Significance of DNA Sequencing
- Information obtained by DNA sequencing makes it possible to understand or alter the function of genes.
- DNA sequence analysis demonstrates regulatory regions that control gene expression and genetic "hot spots" particularly susceptible to mutation.
- Comparison of DNA sequences shows evolutionary relationships that provide a framework for definite classification of microorganisms including viruses.

- Comparison of DNA sequences facilitates identification of conserved regions, which are useful for development of specific hybridization probes to detect microorganisms including viruses in clinical samples.
- DNA sequencing has become sufficiently fast and inexpensive to allow laboratory determination of microbial sequences for identification of microbes. Sequencing of the 16S ribosomal subunit can be used to identify specific bacteria. Sequencing of viruses can be used to identify the virus and distinguish different strains.
- DNA sequencing shows gene structure that helps research workers to find out the structure of gene products.

Southern and Northern blot hybridization

Southern blot hybridization refers to the detection of specific DNA fragments that have been separated by gel electrophoresis (Figure 1). After the electrophoresis the separated DNA fragments are denatured and transferred to a nitrocellulose (or nylon) membrane sheet by blotting. In the blotting the gel is supported on a sponge in a bath of alkali solution, and buffer is sucked through the gel and the sheet by paper towels stacked on top of the nitrocellulose sheet. The buffer denaturates the DNA and transfers the single stranded fragments from the gel to the surface of the sheet, where they adhere firmly. The nitrocellulose sheet containing the bound single-stranded DNA fragments is pealed off the gel and placed in a sealed plastic bag or a box together with buffer containing labelled DNA probe specific for the target DNA sequence. The sheet is exposed to the probe under conditions favouring hybridization. After the hybridization, the sheet is removed from the bag, washed thoroughly to remove unhybridized probes and viewed using autoradiography or ultraviolet light depending on the labels used (radioactive of fluorescent). An adaptation of Southern blotting is Northern blotting, in which RNA molecules are electrophoresed through the gel instead of DNA.
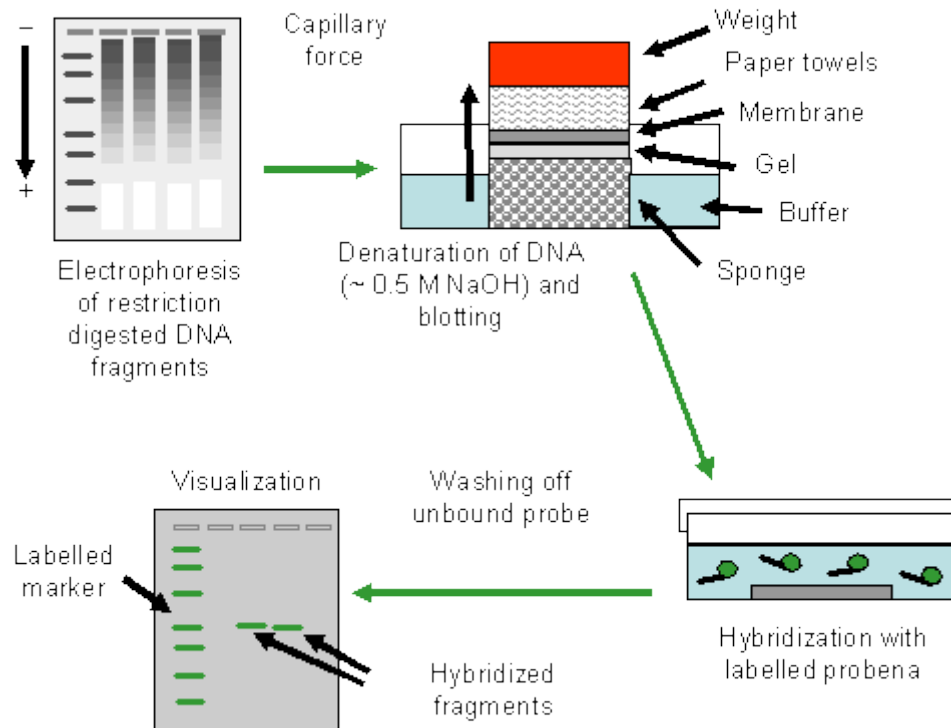
Figure 1. Principle of Southern blotting.

DNA Microarray

Introduction:

A DNA microarray (also commonly known as **gene** or **genome** chip, DNA chip, or gene array) is a collection of microscopic DNA spots, commonly representing single genes, arrayed on a solid surface by covalent attachment to chemically suitable matrices.

DNA arrays are different from other types of microarray, only in that they either measure DNA or use DNA as part of its detection system.

Qualitative or quantitative measurements with DNA microarrays utilize the selective nature of DNA-DNA or DNA-RNA hybridization under high-stringency conditions and fluorophore-based detection. DNA arrays are commonly used for expression profiling, i.e., monitoring expression levels of thousands of genes simultaneously, or for comparative genomic hybridization.

Arrays of DNA can either be spatially arranged, as in commonly known gene or genome chip, DNA chip, or gene array, or can be specific DNA sequences tagged or labelled such that they can be independently identified in solution. The traditional solid-phase array is a collection of microscopic DNA spots attached to a solid surface, such as glass, plastic or silicon chip.

The affixed DNA segments are known as probes (although some sources will use different nomenclature such as reporters), thousands of which can be placed in known locations on a single DNA microarray. Microarray technology evolved from Southern blotting, whereby fragmented DNA is attached to a substrate and then probed with a known gene or fragment.

Applications of these Arrays include:
1. mRNA or gene expression profiling:

Monitoring expression levels for thousands of genes simultaneously is relevant to many areas of biology and medicine, such as studying treatments, disease, and developmental stages. For example, microarrays can be used to identify disease genes by comparing gene expression in diseased and normal cells.

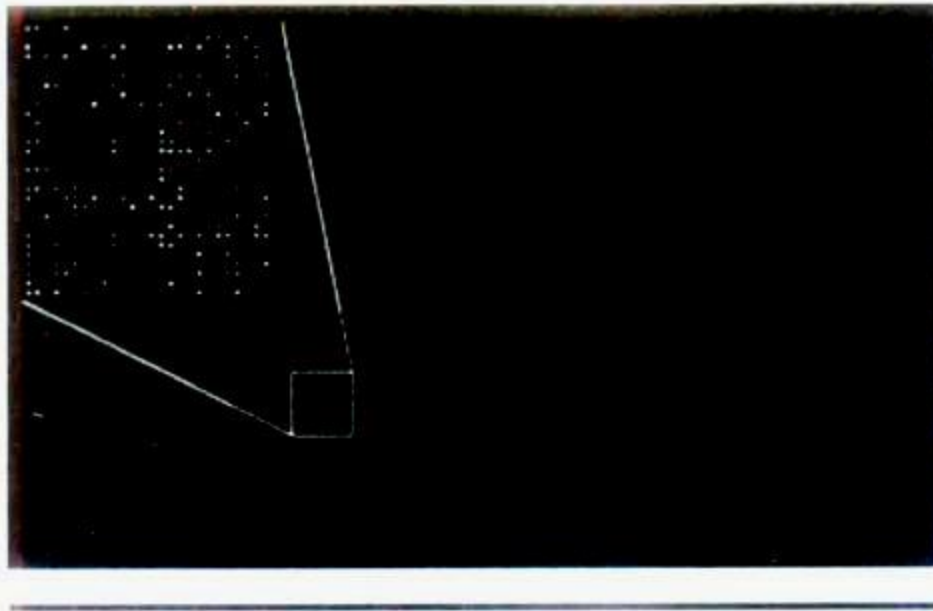### 2. Comparative genomic hybridization (Array CGH):
Assessing large genomic rearrangements.

3. SNP detection arrays:
Looking for single nucleotide polymorphism in the genome of populations.

4. Chromatin immunoprecipitation (ChIP) studies:
Determining protein binding site occupancy throughout the genome, employing ChlP-on-chip technology.

Fig 16.1: Example of an approximately 40,000 probe spotted oligo microarray with enlarged inset to show in details.

Fabrication:

Microarrays can be fabricated using a variety of technologies, including printing with fine-pointed pins onto glass slides, photolithography using pre-made masks, photolithography using dynamic micro-mirror devices, ink-jet printing, or electrochemistry on microelectrode arrays.

DNA microarrays can be used to detect RNAs that may or may not be translated into active proteins. Scientists refer to this kind of analysis as "expression analysis" or expression profiling. Since there can be tens of thousands of distinct probes on an array, each microarray experiment can accomplish the equivalent number of genetic tests in parallel.

Arrays have, therefore, dramatically accelerated many types of investigations. The use of microarrays for gene expression profiling was first published in 1995 (Science) and the first complete eukaryotic genome {Saccharomyces cerevisiae) on a microarray was published in 1997 (Science).

1. Spotted Microarrays:

In spotted microarrays (or two-channel or two-colour microarrays), the probes are oligonucleotides, cDNA or small fragments of PCR products that correspond to mRNAs and are spotted onto the microarray surface. This type of array is typically hybridized with cDNA from two samples to be compared (e.g., diseased tissue versus healthy tissue) that are labelled with two different fluorophores (e.g., Rhodamine (Cyanine 5, red) and Fluorescein (Cyanine 3, green)).

The two samples are mixed and hybridized to a single microarray that is then scanned in a microarray scanner to visualize fluorescence of the two fluorophores. Relative intensities of each fluorophore are then used to identify up-regulated and down-regulated genes in ratio-based analysis. Absolute levels of gene expression cannot be determined in the two-colour array, but relative differences in expression among different spots (= genes) can be estimated with some oligonucleotide arrays.
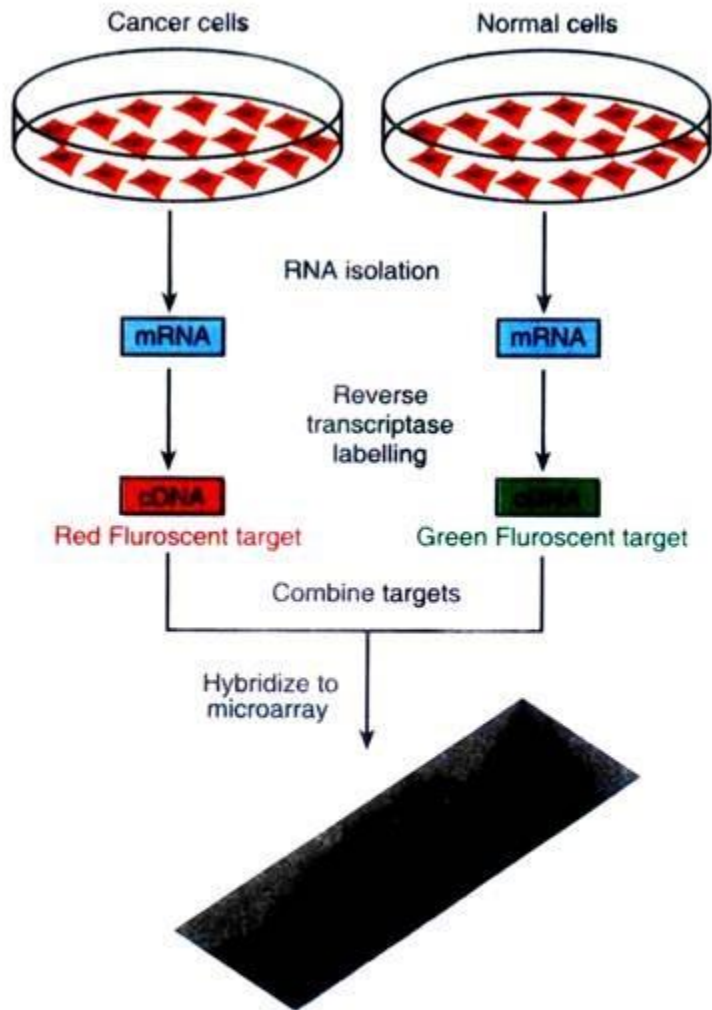
Fig. 16.2: Diagram of typical dual-colour microarray experiment

2. Oligonucleotide Microarrays:

In oligonucleotide microarrays (or single-channel microarrays), the probes are designed to match parts of the sequence of known or predicted mRNAs. There are commercially available designs that cover complete genomes from companies such as GE Healthcare, Affymetrix, Ocimum Bio-solutions, or Agilent. These microarrays give estimations of the absolute value of gene expression and, therefore, the comparison of two conditions requires the use of two separate micro- arrays.

Fig. 16.3: Two Affymetrix chips

Oligonucleotide Arrays can be either produced by piezoelectric deposition with full length oligonucleotides or in situ synthesis. Long Oligonucleotide Arrays are composed of 60-mers, or 50-mers and are produced by ink-jet printing on a silica substrate. Short Oligonucleotide Arrays are composed of 25-mer or 30-mer and are produced by photolithographic synthesis (Affymetrix) on a silica substrate or piezoelectric deposition (GE Healthcare) on an acrylamide matrix.

More recently, Maskless Array Synthesis from NimbleGen Systems has combined flexibility with large numbers of probes. Arrays can contain up to 390,000 spots, from a custom array design. New array formats are being developed to study specific pathways or disease states for a systems biology approach.

Oligonucleotide microarrays often contain control probes designed to hybridize with RNA spike-ins. The degree of hybridization between the spike-ins and the control probes is used to normalize the hybridization measurements for the target probes.

Genotyping Microarrays:

DNA microarrays can also be used to read the sequence of a genome in particular positions. SNP microarrays are a particular type of DNA microarrays that are used to identify genetic variation in individuals and across populations.

Short oligonucleotide arrays can be used to identify the single nucleotide polymorphisms (SNPs) that are thought to be responsible for genetic variation and the source of susceptibility to genetically caused diseases. Generally termed genotyping applications, DNA microarrays may be used in this fashion for forensic applications, rapidly discovering or measuring genetic predisposition to disease, or identifying DNA-based drug candidates.

These SNP microarrays are also being used to profile somatic mutations in cancer, specifically loss of heterozygosity events and amplifications and deletions of regions of DNA. Amplifications and deletions can also be detected using comparative genomic hybridization, or aCGH, in conjunction with microarrays, but may be limited in detecting novel Copy Number Polymorphisms, or CNPs, by probe coverage.

Re-sequencing arrays have also been developed to sequence portions of the genome in individuals. These arrays may be used to evaluate germ line mutations in individuals, or somatic mutations in cancers. Genome tiling arrays include overlapping oligonucleotides designed to blanket an entire genomic region of interest. Many companies have successfully designed tiling arrays that cover whole human chromosomes.

Microarrays and Bioinformatics:
1. Experimental Design:
Due to the biological complexity of gene expression, the considerations of experimental design that are discussed in the expression profiling article are of critical importance if statistically and biologically valid conclusions are to be drawn from the data.
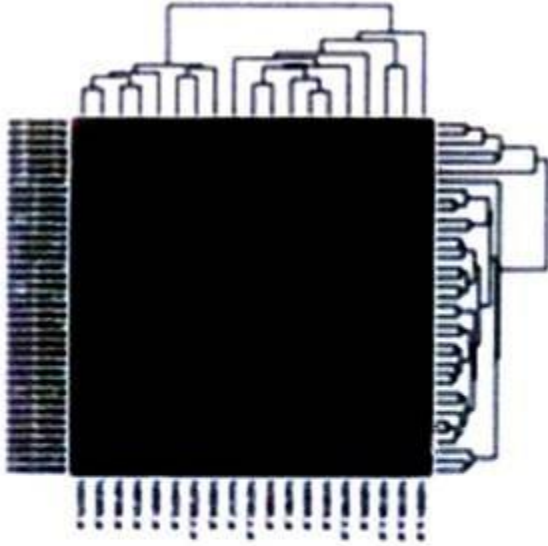
Fig. 16.4: Gene expression values from microarray experiments can be represented as heat maps to visualize the result of data analysis

There are three main elements to consider when designing a microarray experiment.

First, replication of the biological samples is essential for drawing conclusions from the experiment.

Second, technical replicates (two RNA samples obtained from each experimental unit) help to ensure precision and allow for testing differences within treatment groups. The technical replicates may be two independent RNA extractions or two aliquots of the same extraction.

Third, spots of each cDNA clone or oligonucleotide are present at least as duplicates on the microarray slide, to provide a measure of technical precision in each hybridization. It is critical that information about the sample preparation and handling is discussed in order to help identify the independent units in the experiment as well as to avoid inflated estimates of significance.

2. Standardization:

The lack of standardization in arrays presents an interoperability problem in bioinformatics, which hinders the exchange of array data. Various grass-roots open-source projects are attempting to facilitate the exchange and analysis of data produced with non-proprietary chips.

1. The "Minimum Information about a Microarray Experiment" (MIAME) checklist helps define the level of detail that should exist and is being adopted by many journals as a requirement for the submission of papers incorporating microarray results. MIAME describes the minimum required information for complying experiments, but not its format. Thus, as of 2007, whilst many formats can support the MIAME requirements there is no format which permits verification of complete semantic compliance.

2. The "MicroArray Quality Control (MAQC) Project" is being conducted by the FDA to develop standards and quality control metrics which will eventually allow the use of MicroArray data in drug discovery, clinical practice and regulatory decision-making.

3. The MicroArray and Gene Expression (MAGE) group is working on the standardization of the representation of gene expression data and relevant annotations.

3. Statistical Analysis:

The analysis of DNA microarrays poses a large number of statistical problems, including the normalization of the data. There are dozens of proposed normalization methods in the published literature; as in many other cases where authorities disagree, a sound conservative approach is to try a number of popular normalization methods and compare the conclusions reached; how sensitive are the main conclusions to the method chosen?

From a hypothesis-testing perspective, the large number of genes present on a single array means that the experimenter must take into account a multiple testing problem; even if the statistical P-value assigned to a given gene indicates that it is extremely unlikely that differential expression of this gene was due to random rather than treatment effects, the very high number of

genes on an array makes it likely that differential expression of some genes represents false positives or false negatives.

Statistical methods tailored to microarray analyses have recently become available that assess statistical power based on the variation present in the data and the number of experimental replicates, and can help minimize type I and type II errors in the analyses.

A basic difference between microarray data analysis and much traditional biomedical research is the dimensionality of the data. A large clinical study might collect 100 data items per patient for thousands of patients. A medium-size microarray study will obtain many thousands of numbers per sample for perhaps a hundred samples. Many analysis techniques treat each sample as a single point in a space with thousands of dimensions, then attempt by various techniques to reduce the dimensionality of the data to something humans can visualize.

4. Relation between Probe and Gene:
The relation between a probe and the mRNA that it is expected to detect is problematic. On the one hand, some mRNAs may cross-hybridize probes in the array that are supposed to detect another mRNA. On the other hand, probes that are designed to detect the mRNA of a particular gene may be relying on genomic EST information that is incorrectly associated with that gene.

Public Databases of Microarray Data:

Database	Microarray Experiment Sets	Sample Profiles	As of Date
Gene Expression Omnibus - NCBI	5366	134669	April 1, 2007
Stanford Microarray database	12742	?	April 1, 2007
UNC Microarray database	~31	2093	April 1, 2007
MUSC database	~45	555	April 1, 2007
ArrayExpress at EBI	1643	136	April 1, 2007
caArray at NCI	41	1741	November 15, 2006
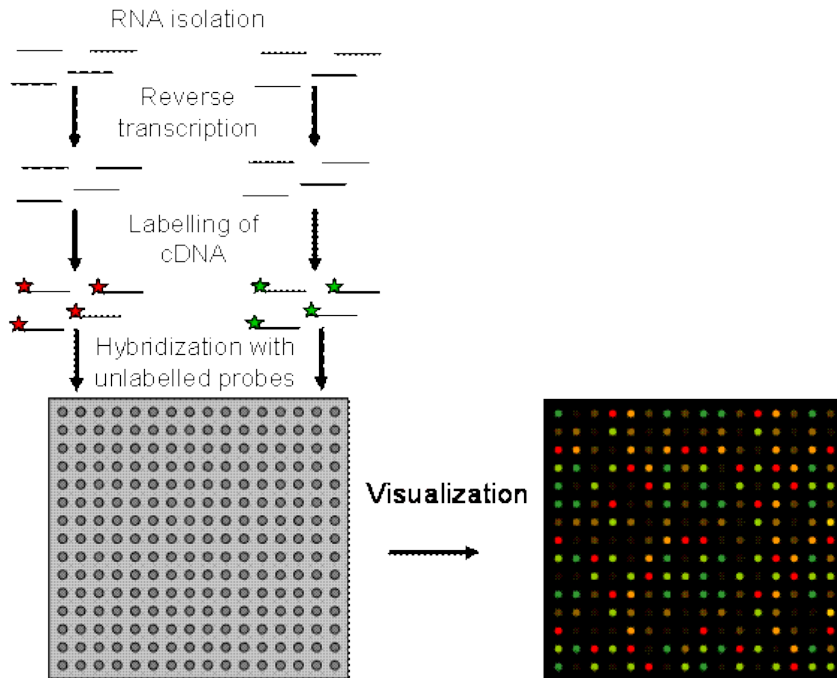
Microarray analysis of RNA from two samples using two different labels.

Microarray is fast and cost effective tool to detect specific microorganisms and study gene expression, since thousands of hybridizations can be performed on one chip. The results are well comparable, but the optimization of hybridization conditions for a large number of probes may be challenging. Also the designing of probes and arrays is very time consuming.

Online Microarray Data Analysis Programs and Tools:

Several Open Directory Project categories list online microarray data analysis programs and tools:

i. Bioinformatics: Online Services:

Gene Expression and Regulation at the Open Directory Project

ii. Gene Expression:

Databases at the Open Directory Project

iii. Gene Expression:

Software at the Open Directory Project

iv. Data Mining:

Tool Vendors at the Open Directory Project

v. Bio-conductor:

Open source and open development software project for the analysis and comprehension of genomic data

vi. Genevestigator:

Web-based database and analysis tool to study gene expression across large sets of tissues, developmental stages, drugs, stimuli, and genetic modifications.

SAGE - Serial Analysis of Gene Expression

Briefly, SAGE experiments proceed as follows:

1. The mRNA of an input sample (e.g. a tumour) is isolated and a reverse transcriptase and biotinylated primers are used to synthesize cDNA from mRNA.

2. The cDNA is bound to Streptavidin beads via interaction with the biotin attached to the primers, and is then cleaved using a restriction endonuclease called an anchoring enzyme (AE). The location of the cleavage site and thus the length of the remaining cDNA bound to the bead will vary for each individual cDNA (mRNA).

3. The cleaved cDNA downstream from the cleavage site is then discarded, and the remaining immobile cDNA fragments upstream from cleavage sites are divided in half and exposed to one of two adaptor oligonucleotides (A or B) containing several components in the following order upstream from the attachment site: 1) Sticky ends with the AE cut site to allow for attachment to cleaved cDNA; 2) A recognition site for a restriction endonuclease known as the tagging enzyme (TE), which cuts about 15 nucleotides downstream of its recognition site (within the original cDNA/mRNA sequence); 3) A short primer sequence unique to either adaptor A or B, which will later be used for further amplification via PCR.

4. After adaptor ligation, cDNA are cleaved using TE to remove them from the beads, leaving only a short "tag" of about 11 nucleotides of original cDNA (15 nucleotides minus the 4 corresponding to the AE recognition site).

5. The cleaved cDNA tags are then repaired with DNA polymerase to produce blunt end cDNA fragments.

6. These cDNA tag fragments (with adaptor primers and AE and TE recognition sites attached) are ligated, sandwiching the two tag sequences together, and flanking adaptors A and B at either end. These new constructs, called ditags, are then PCR amplified using anchor A and B specific primers.

7. The ditags are then cleaved using the original AE, and allowed to link together with other ditags, which will be ligated to create a cDNA concatemer with each ditag being separated by the AE recognition site.

8. These concatemers are then transformed into bacteria for amplification through bacterial replication.

9. The cDNA concatemers can then be isolated and sequenced using modern high-throughput DNA sequencers, and these sequences can be analysed with computer programs which quantify the recurrence of individual tags.

10. The output of SAGE is a list of short sequence tags and the number of times it is observed. Using sequence databases a researcher can usually determine, with some confidence, from which original mRNA (and therefore which gene) the tag was extracted.

11. Statistical methods can be applied to tag and count lists from different samples in order to determine which genes are more highly expressed. For example, a normal tissue sample can be compared against a corresponding tumor to determine which genes tend to be more (or less) active.