**Swayam Course - Analytical Techniques**

**Week 12, Tutorial 30 - Sequencing Techniques of Nucleic Acids**

**Content Writer - Dr. Ashikh Seethy, Senior Resident, Department of Biochemistry, AIIMS, New Delhi**

## 1. Objectives

At the end of this module, you should be able to:

a. List the applications of nucleic acid sequencing.
b. Differentiate between first, second and third generation sequencing techniques.
c. Explain the principle of commonly used sequencing techniques.
d. List the applications of RNA sequencing and its advantages over microarray.

## 2. Introduction

Nucleic acid sequencing refers to the process by which the precise order of the bases in a strand of nucleic acid can be deciphered. Knowing the sequence of the particular molecule of nucleic acid under study is important in the following contexts:

1. Diagnosis of genetic diseases, especially the monogenic diseases (diseases that occur due to mutations in a single gene).
2. Certain variations in gene sequences are associated with aberrant responses to certain drugs (*Pharmacogenomics*). E.g. Anti-malarial drug primaquine can lead to hemolysis in individuals harbouring mutations in Glucose-6-phosphate dehydrogenase.
3. In the field of evolutionary biology, sequencing can aid in the study of homologous proteins, conserved sequences, divergent evolution and construction of an evolutionary tree.
4. A prior knowledge of the gene sequence is important in various molecular biology techniques like genetic engineering, gene editing, polymerase chain reaction, and cloning.
5. RNAseq (Sequencing of mRNA) is a powerful tool for studying gene expression and discovering new genes.
6. Immunoprecipitaion of proteins that can potentially bind to DNA/RNA followed by sequencing of the DNA/RNA eluted from these proteins is important in the study of nucleic acid binding proteins and their consensus sequences.
7. Nucleic acid sequencing is an important tool in microbiology for identification and characterisation of various strains of organism.
8. Sequencing may have applications in the field of Forensic Medicine where identity can be established using DNA sequencing.

The history of nucleic acid sequencing dates back to late 1960s and as the time goes, new methods of sequencing DNA are taking precedence. The techniques for sequencing DNA can be crudely classified into 1st generation, 2nd generation and 3rd generation sequencing techniques:

| Generation | Examples | Characteristics |
|---|---|---|
| First | Sanger Sequencing<br>Maxam Gilbert Sequencing | ▪ Only a single type of DNA molecule is analysed at a time.<br>▪ Average read length is 300- 700 bases |
| Second | Illumina<br>Pyrosequencing<br>Ion Torrent<br>SOLiD | ▪ Parallel sequencing of DNA fragments.<br>▪ Sequencing by synthesis is used in most these platforms (except SOLiD)<br>▪ Read lengths are 100 – 1000 bases. |
| Third | Nanopore sequencing<br>Single Molecule Real Time Sequencing (SMRT) | Long read lengths (in the range of kilobases). |

Since most of the sequencing techniques (including the RNA sequencing methodologies) sequence DNA, we will restrict the initial part of this discussion to DNA sequencing. A description of RNA sequencing techniques will then follow.

### 3. First Generation Sequencing Techniques:

Among the early DNA sequencing methods discovered were Maxam Gilbert sequencing and Sanger's sequencing. Both these techniques analyse only one fragment of DNA at a time.

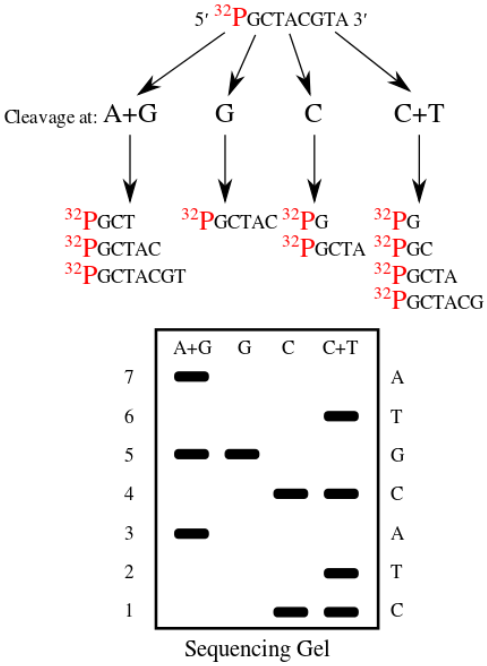
#### *a. Maxam Gilbert sequencing:*

This technique is based on cleavage of single stranded DNA at specific bases by different chemicals. For this, the dsDNA is denatured and the 5' end is radiolabelled with $^{32}P$. The radiolabelled ssDNA is then subjected to treatment with the following chemicals in different reactions:

| Chemicals | Cleavage sites |
|---|---|
| Piperidine + Dimethyl sulphate | G, A |
| Piperidine + Dimethyl sulphate + HCl | G |
| Piperidine + Hydrazine | C, T |
| Piperidine + Hydrazine + NaCl | C |

*Dimethyl sulphate modifies purines while hydrazine modifies pyrimidines. The modified bases are cleaved by piperidine.*

The chemically treated DNA is then subjected to electrophoresis in a polyacrylamide gel, each lane containing a different combination of chemicals as mentioned above. The electrophoretically separated fragments are then visualized using autoradiography. Note that a cleavage after guanine is mediated by (Piperidine + Dimethyl sulphate) as well as (Piperidine + Dimethyl sulphate + HCl) while cleavage after adenine is mediated only by (Piperidine + Dimethyl sulphate). Similar is the case with thymine and cytosine. This helps in distinguishing the bases. As an example, in the gel shown below, cleavage by (Piperidine + Dimethyl sulphate) has occurred at positions 3, 5 and 7 while cleavage by (Piperidine + Dimethyl sulphate + HCl) has occurred at positions 3 and 7. Thus positions 3 and 7 are adenines and position 5 is guanine. Similarly, other bases can also be identified.




**Maxam Gilbert sequencing**

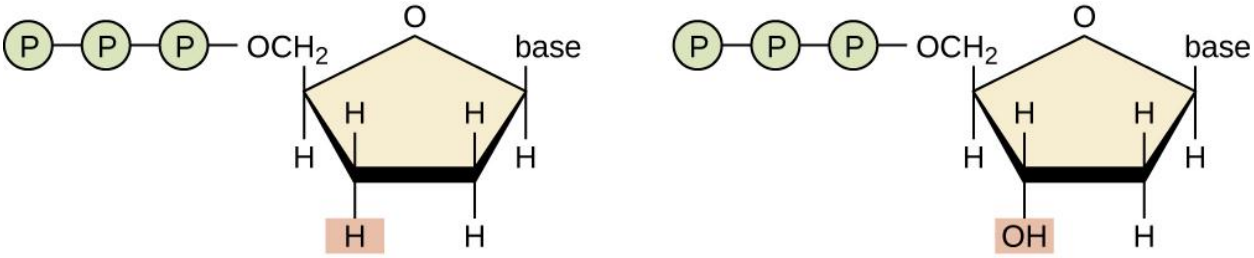
Source: *By Shakiestone [CC BY-SA 3.0 (https://creativecommons.org/licenses/by-sa/3.0) or GFDL (http://www.gnu.org/copyleft/fdl.html)], from Wikimedia Commons*

Maxam Gilbert Sequencing was soon replaced by other methods due to its disadvantages like usage of radioactive labels, use of hydrazine which is neurotoxic, cumbersome procedure and the limitation of sequence read lengths.

***b. Sanger sequencing:***

Invented by Frederick Sanger in circa 1977, Sanger sequencing is based on controlled termination of replication. Sanger, along with Walter Gilbert (co-developer of Maxam Gilbert sequencing), were awarded Nobel Prize in Chemistry in 1980 "for their contributions concerning the determination of base sequences in nucleic acids."

In Sanger sequencing, controlled termination of replication is achieved by the use of dideoxy nucleotides. Dideoxynucleotides lack oxygen at both 2' and 3' positions of ribose sugar.




dideoxynucleotide (ddNTP) deoxynucleotide (dNTP)

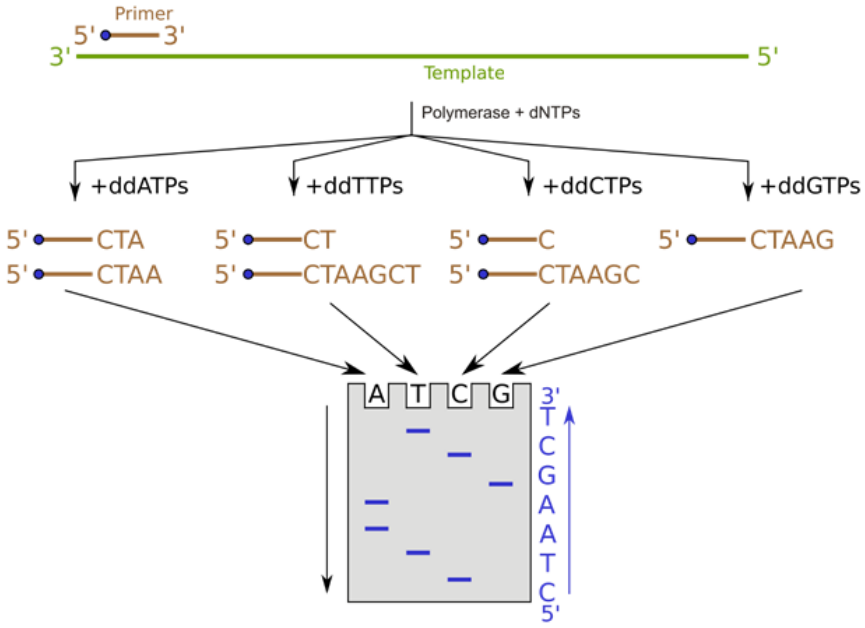
Source: *By CNX OpenStax (https://cnx.org/contents/5CvTdmJL@4.4) [CC BY 4.0 (https://creativecommons.org/licenses/by/4.0)], via Wikimedia Commons*

During synthesis of nucleic acids, for example during replication, transcription or PCR, the incoming base is added to the 3' OH group of the pre-existing nucleotide sequence via 3'-5' phosphodiester linkages. If a dideoxy nucleotide gets incorporated into the DNA polymer which is getting synthesized, the incoming nucleotide cannot form a 3'-5' phosphodiester linkage, and the synthesis of DNA is terminated.

Unlike Maxam Gilbert sequencing where DNA can be used directly, Sanger sequencing requires multiple copies of the input DNA for sequencing, and this, nowadays is achieved by polymerase chain reaction. Also for Sanger sequencing, the flanking sequence of the region to be sequenced should be known.

A typical Sanger sequencing reaction will contain copies of the DNA fragment to be sequenced (template), a primer flanking the region to be sequenced, all the 4 deoxynucleotides (dNTPs), DNA polymerase, buffer and $Mg^{2+}$. Each of the 4 dideoxynucleotides (ddNTPs) are added to separate reactions, hence there will be 4 different reactions happening. The ddNTPs will be used at a concentration much lower than that of dNTPs, i.e., around 1%. This ensures that the chain termination happens at a relatively low frequency.

The primer binds to the template and the nucleotides are added by the polymerase in accordance with the Watson-Crick base pairing. Whenever a dideoxynucleotide is added, the synthesis of the strand is terminated. For instance, in the reaction comprising of ddCTPs, the binding of ddCTP to the strand being synthesized will occur in positions where a G is present in the template, and this will lead to termination. Since the concentration of ddCTP is much lower than that of dCTP, termination will not happen at those Gs in the template where dCTP rather than ddCTP is incorporated. Thus a mixture of terminated strands of different lengths will be produced, based on the positions of G in the template. These products are then run on a polyacrylamide gel and the length of the products are indicative of the bases that are present in the template. An example is shown below:

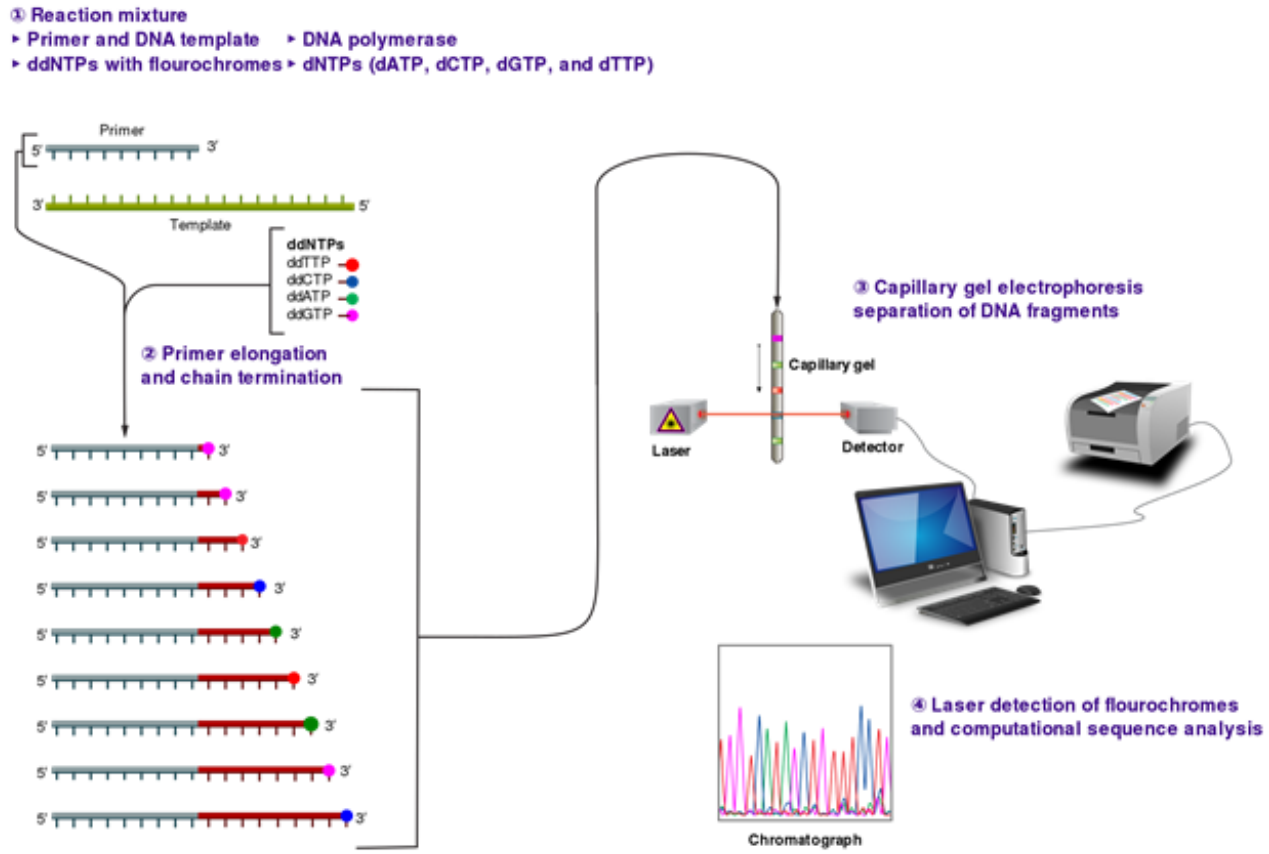



Source: *By Christoph Goemans (modifiziert) CC BY-SA 3.0 (https://creativecommons.org/licenses/by-sa/3.0)], via Wikimedia Commons*

Here the reaction containing ddCTP gives two products of length 1 nucleotide long and 6 nucleotide long beyond the primer. This means that the newly synthesized strand is comprised of C at 1$^{st}$ and 6$^{th}$ position beyond the primer. Hence the template strand will contain G at the corresponding positions. Similarly the positions of other bases can be determined by analysing the reactions containing other ddNTPs. Thus the sequence of the template in the above case will be 5'AGCTTAG3'.

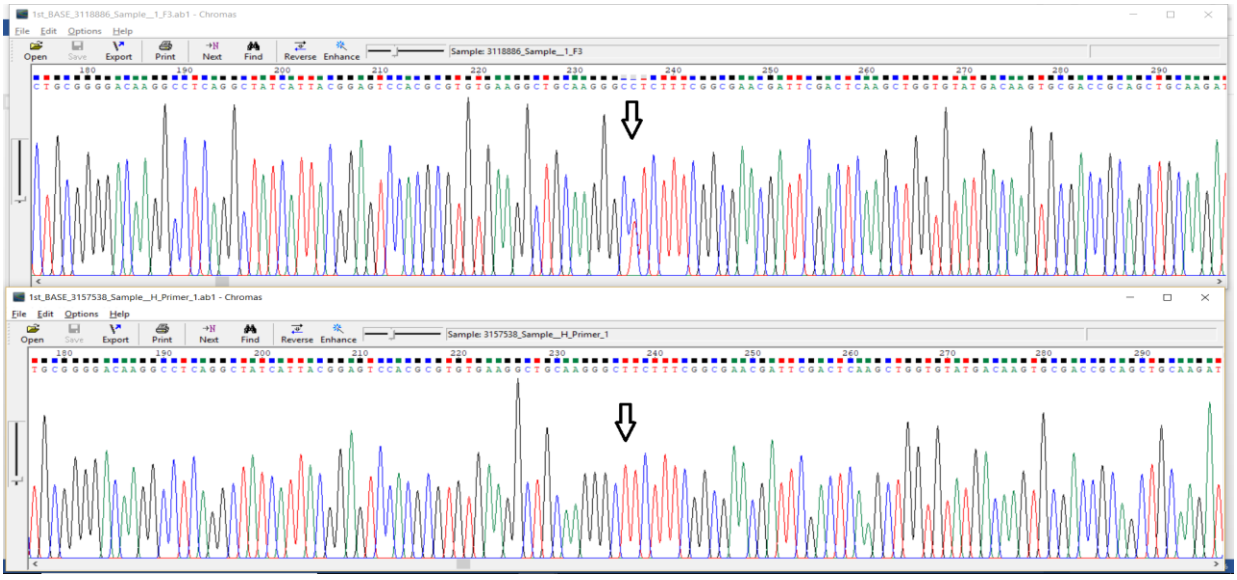
Running an acrylamide gel for sequencing analysis had some inherent problems. Large gels were to be casted for analysis of longer fragments, visualisations required radiolabelling which is hazardous, and the resolution between fragments decreased as the length of the fragment to be sequenced increased – i.e., the resolution between a 10 bp fragment and an 11 bp fragment is much higher than that between a 400 bp fragment and 401 bp fragment.

A way to overcome these limitations and to accelerate the process of sequencing was to use fluorescently labelled ddNTPs, each of the ddNTP being labelled with a different fluorescent tag. All the four types of labelled ddNTPs can be added to the same reaction mixture, thus abating the need for multiple reactions. The reaction, once completed, is subjected to capillary electrophoresis where based on the fluorescence pattern the sequence of the newly synthesised strand can be deciphered and the sequence of the template strand will be complementary to this, in the 5' to 3' direction.




Source: *By Estevezj [CC BY-SA 3.0 (https://creativecommons.org/licenses/by-sa/3.0)], from Wikimedia Commons*

The diagram below shows the chromatogram (read out from capillary electrophoresis) of a sequence in PPARα gene. Note that the base in the position 236 is C/T in the chromatogram on top while it is T in the chromatogram below (arrows). On comparison with the reference sequences, it was concluded that the sample on top is harbouring a C>T mutation in 50% of samples.



The Sanger sequencing methodologies that employ fluorescent nucleotides and capillary electrophoresis, even-though a technique that originated in the 90s, is still being commonly used and can give sequencing reads up to 700 – 800 nucleotides long. Sanger sequencing was also used in one of the most significant studies in genomics – the Human Genome Project.

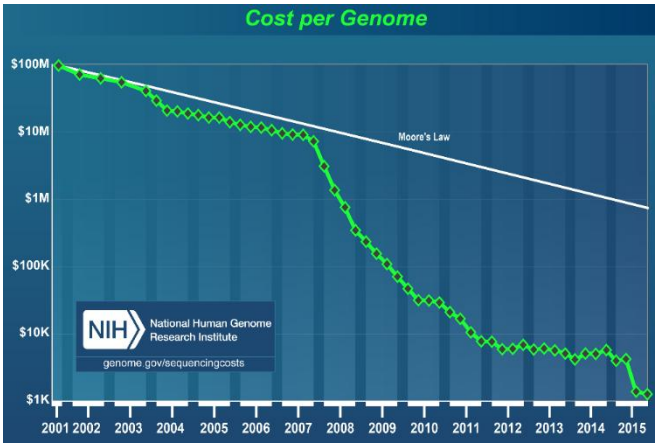
**4. Second Generation Sequencing Techniques:**

One disadvantage of the first generation sequencing techniques is the time taken for sequencing the DNA. The human haploid genome is ~3 billion base pairs long and sequencing the entire genome using a technique like Sanger sequence is an expensive albeit arduous task. The Human Genome Project which employed Sanger's sequencing took 10 years for completion of the first draft and a whopping $3 Billion budget. The technique employed was called shotgun sequencing where the large DNA molecule is fragmented into smaller pieces and are sequenced. The sequencing was performed in multiple centres and the reads were finally aligned.

It is not practical to have such a tedious and expensive process for analysing genome of individuals for various applications mentioned in the beginning of this module. However, the same results could be achieved at the expense of minimal time and money, thanks to the development of Next Generation Sequencing (NGS) techniques.

NGS techniques comprise of 2$^{nd}$ and 3$^{rd}$ generation sequencing techniques and relies on massive parallel sequencing of DNA. The DNA to be sequenced in fragmented into smaller pieces and sequencing is performed using various chemistries, all the fragments being analysed almost simultaneously in the same device. Thus what used to be a multicentre process in Human Genome Project is now performed in a single flow-cell of an instrument. The reads are then aligned to obtain the complete sequence. As expected, the aligning of the small fragments

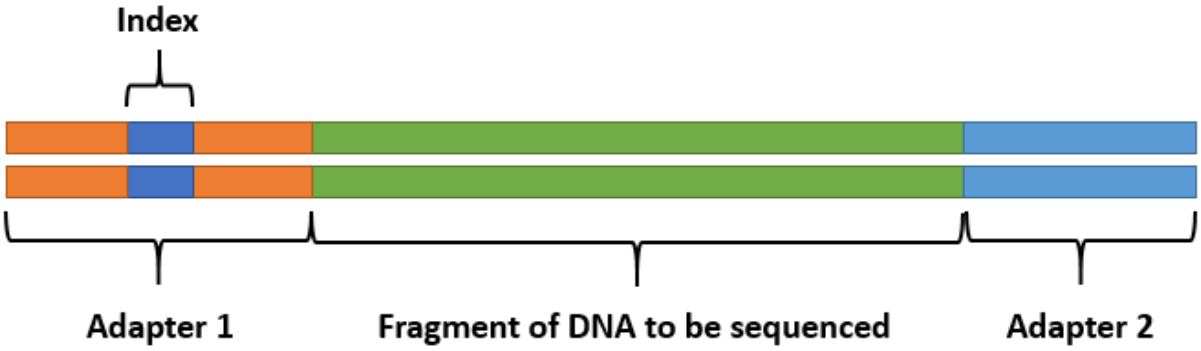
is computationally challenging, and the development of NGS techniques occurred in parallel with the development of computational capabilities, in accordance with what is known as Moore's law (number of transistors in integrated circuits per square inch doubles every year), or even outpacing it. This has led to a significant reduction in the sequencing costs as well.




Source: *By National Human Genome Research Institute (NHGRI) Web site genome.gov (http://www.genome.gov/sequencingcosts/) [CC0], via Wikimedia Commons*

Here is the outline of NGS techniques:

1. DNA to be sequenced is isolated.
2. The DNA is then fragmented to a particular size depending on the sequencing platform employed. Various strategies for fragmentation include sonication, shearing by centrifugal force, needle base shearing, employing restriction endonucleases for strand breaks etc.
3. Adapters are then ligated to the ends of the DNA fragments.




Adapters are known nucleotide sequences that serve multiple purposes like binding to the flow cell, template for attachment of sequencing primers, and some adaptors have an index sequence that helps in identifying the reads. Adapters also facilitate multiplexing, as similar adapters can be added to all the fragments from a single sample, thus helping in distinguishing different samples.

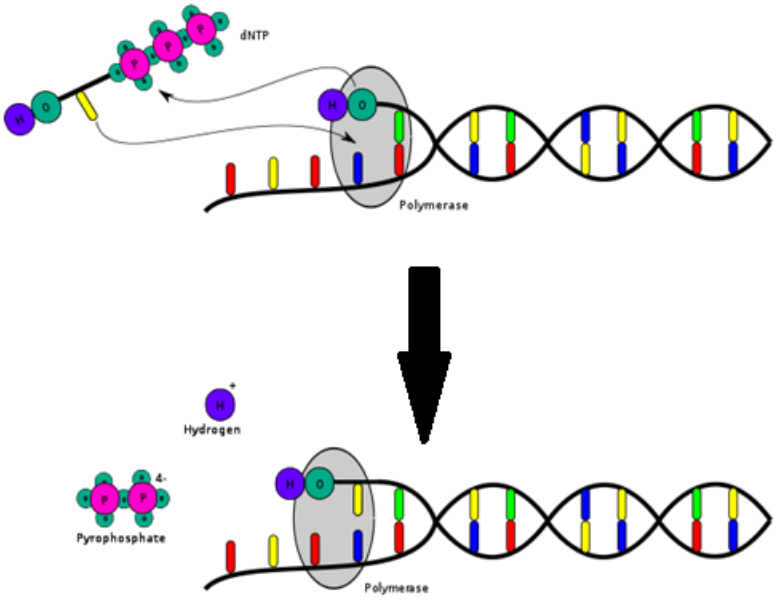
4. The adapter containing DNA is then attached to a flow cell where the sequencing reaction happens. Other components required for sequencing like polymerase, dNTPs, buffer etc. are also present in the flow cell. The nature of detection of bases differs in different sequencing techniques, but most of the second generation sequencing techniques utilizes what is known as sequencing by synthesis, where the dNTPs are added to the growing strand and the bases are identified (base calling) as the process of DNA synthesis is happening.
5. The reads are now obtained from the sequencer, their quality can be checked by various tools, and then they are subjected to processing and alignment.

   Let us now see some of the second generation sequencing techniques in detail:

***a. Roche 454/ Pyrosequencing:***

In Pyrosequencing, also known as 454 sequencing or Roche 454 sequencing (based on the firms that developed or acquired this technology), the DNA is fragmented to sizes up to 1 kB. Adaptors are ligated and the fragments are immobilized on the surface of beads such that there is one DNA fragment per bead. Sequencing happens by synthesis.

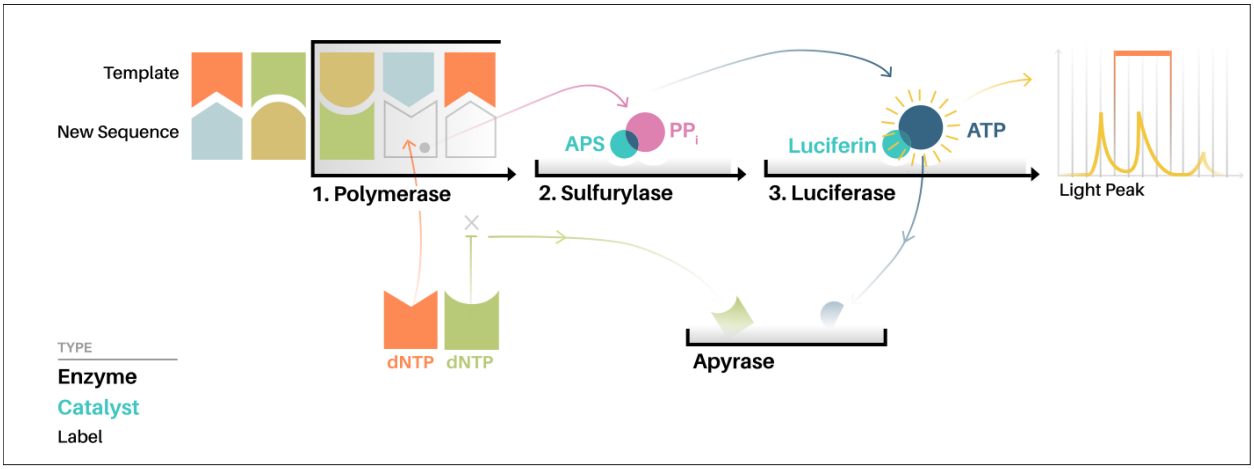
For each round of read, one of the 4 dNTPs is added. If a complementary base is present in the sequence, the dNTP is incorporated and a pyrophosphate ($PP_i$) and a proton is released. If the complementary base is not present, there will not be any release of the pyrophosphate or proton.




Source: *By David Tack [CC BY-SA 3.0 (https://creativecommons.org/licenses/by-sa/3.0) or GFDL (http://www.gnu.org/copyleft/fdl.html)], from Wikimedia Commons*

The $PP_i$ reacts with adenosine phosphosulfate (APS) in presence of an enzyme ATP sulfurylase to form ATP. ATP generation can then be monitored using luciferase mediated oxidation of luciferin

which generates light in proportion to the ATP generated. The light generation is detected as base calls.




Source: *By "Jacopo Pompilii, DensityDesign Research Lab". [CC BY-SA 4.0 (https://creativecommons.org/licenses/by-sa/4.0)], from Wikimedia Commons*

Note that instead of dATP to be added initially, pyrosequencing uses dATPαS (2'-Deoxyadenosine-5'-α-thio-triphosphate) which is not a substrate for luciferase, thus reducing noise and giving accurate calls for adenine.

The linearity of light generation is lost after the sequencer encounters a repeat of 5-6 identical nucleotides.

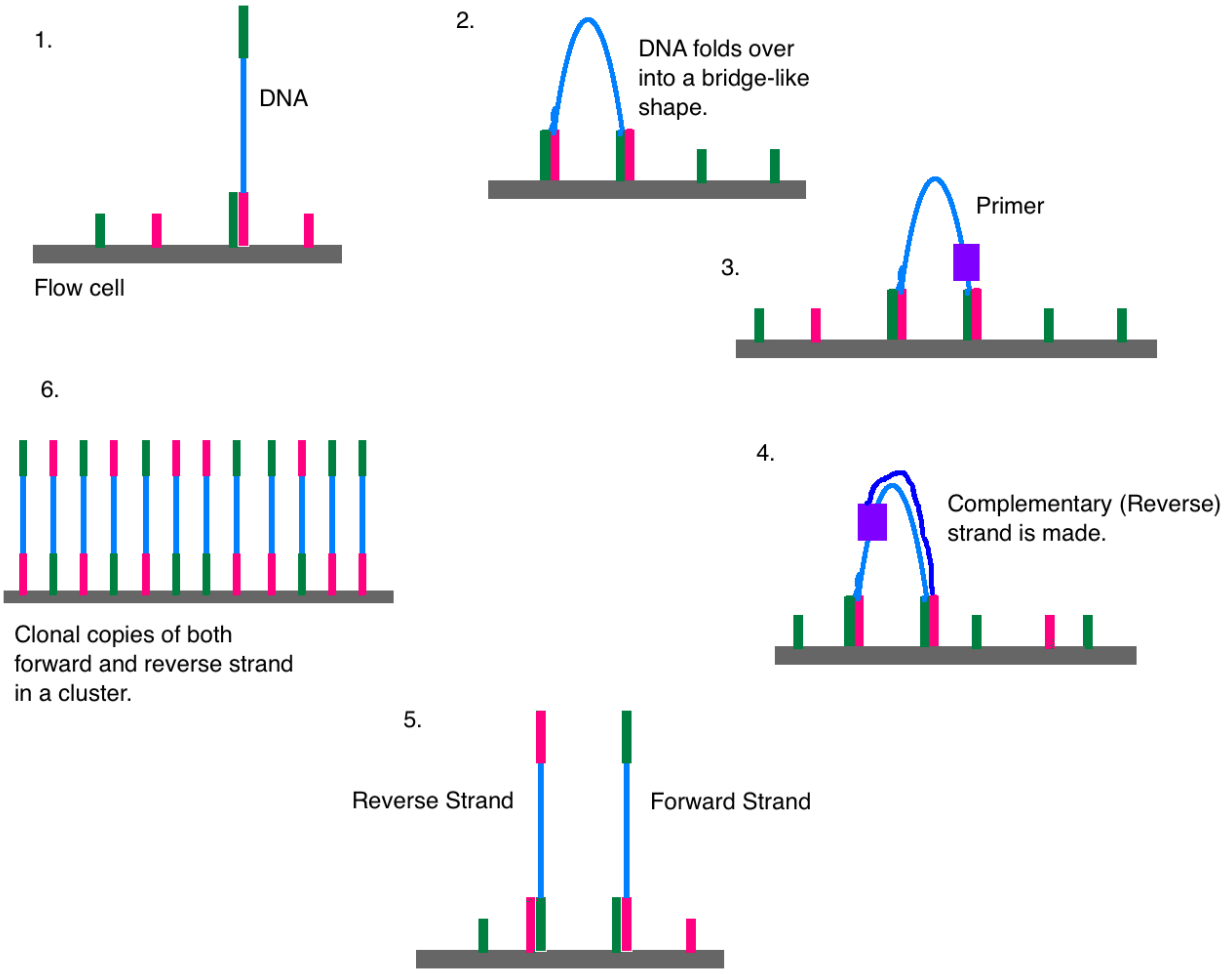
Another disadvantage of pyrosequencing was that all the reagents were to be washed after each base incorporation, to remove the ATP and thus light production, in the so called solid phase pyrosequencing. This has been modified with the use of an enzyme known as apyrase, which degrades unincorporated dNTPs and ATP (liquid phase pyrosequencing).

Pyrosequencing can generate reads of 400-500 in length.

***b. Illumina sequencing:***

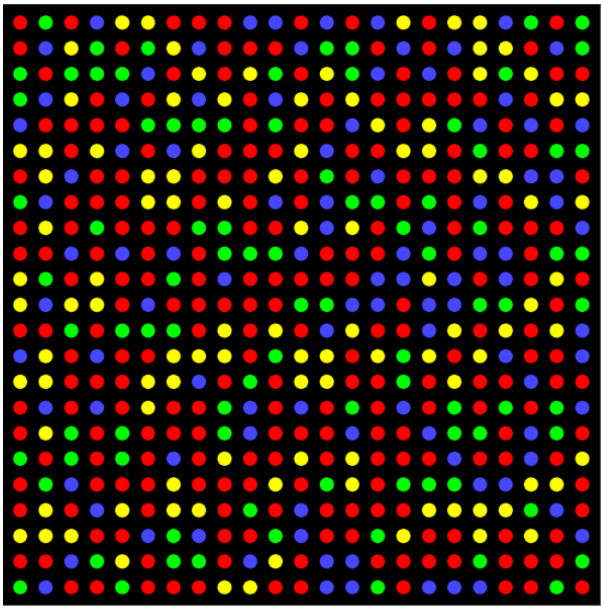
Perhaps the most widely used sequencing platform used at present is based on the Illumina sequencing. In Illumina sequencing, the DNA is fragmented to 150-200 bp and after adapter ligation, the DNA is immobilized on the surface of a flow cell with the help of sequences that are complementary to the adapter sequence.

Primers are then added, which also bind to the adapter to amplify the DNA, which creates spots in the flow cell with multiple copies of the same DNA fragment. This is achieved by a process known as bridge PCR (See the figure below). The double stranded DNA fragments thus generated are denatured and are now subjected to sequencing by synthesis.



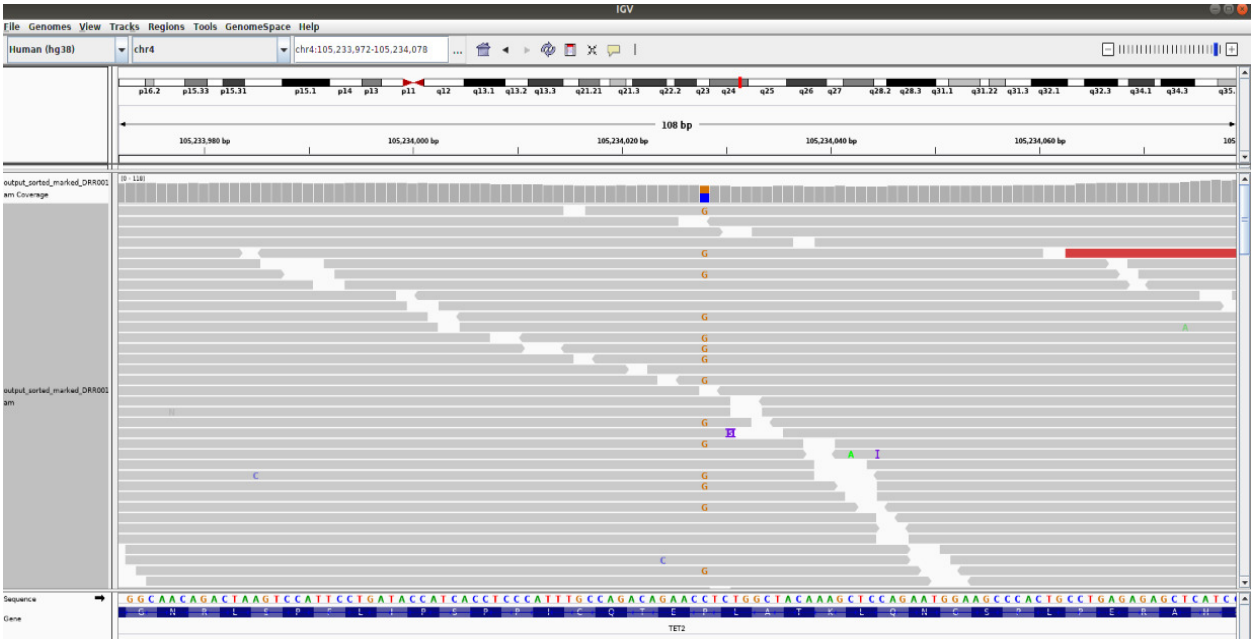

Source: *By DMLapato [CC BY-SA 4.0 (https://creativecommons.org/licenses/by-sa/4.0)], from Wikimedia Commons*

Illumina uses fluorescently labelled dNTPs for base calling. The dNTPs are also chemically modified in such a way that only one nucleotide is added at a time (by addition of a chemical moiety known as terminator; Illumina is thus also known as reversible terminator sequencing). The picture below is the snapshot of an Illumina flow cell. After addition of dNTPs and removal of the excess dNTPs, the fluorescent label at each spot represents the incorporation of the corresponding base to each of the spot and as discussed earlier, each dot represents a clone of DNA fragment being synthesised. Thus the base calling is done.



Source: *Автор Thomas shafee [CC BY 4.0 (https://creativecommons.org/licenses/by/4.0)], с Викисклада*

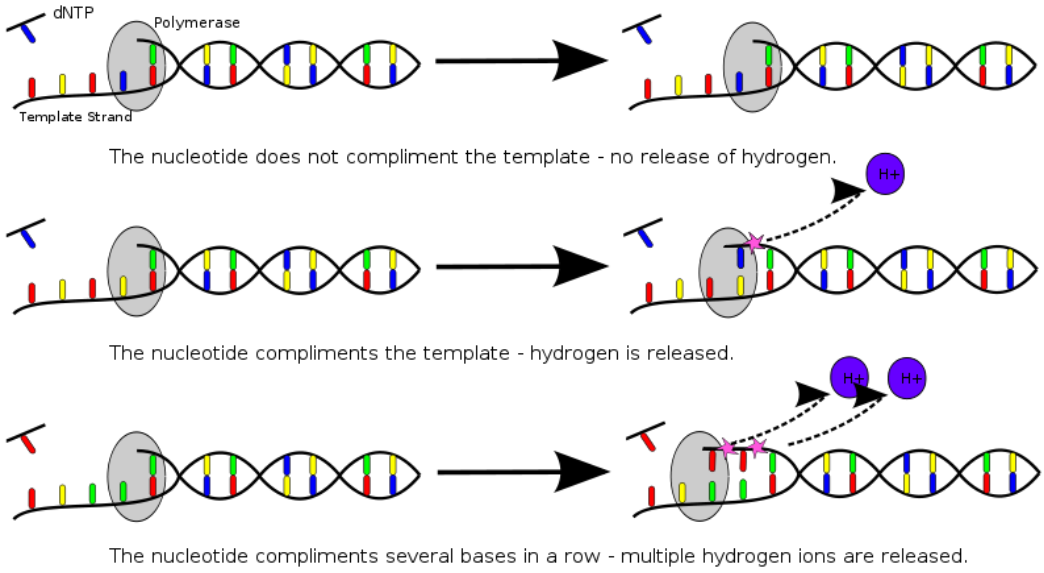
In the next cycle, the terminator moieties are removed from the incorporated nucleotides and this cycle is repeated. The reads are then obtained by the analysis of all the snapshots.




The image shows processed Illumina sequencing data. A heterozygous mutation (C/G) can also be observed.

### *c. Ion Torrent sequencing:*

The chemistry here is similar to pyrosequencing, but instead of pyrophosphate, proton release is measured. The DNA is fragmented to 200 bp, adapters are attached and the fragment is immobilised on a bead. The fragment is amplified by a technique known as emulsion PCR (*or droplet PCR – in droplet PCR, the fragmented DNA which is adapter ligated, along with primers, DNA polymerase and other PCR reaction components are prepared in a solution that is later subjected to emulsification. The resulting droplets are prepared in such a way that only a single molecule of DNA will be present per droplet. Thus we have thousands of unique micro-reactions happening simultaneously, while the mixture is subjected to PCR, and these reactions can be monitored individually*). Each bead is then immobilised on the surface of a slide and the reads are carried out. After addition of each dNTP, the pH change is noted, which is indicative of the incorporation of that particular dNTP.




Source: *By David Tack [CC BY-SA 3.0 (https://creativecommons.org/licenses/by-sa/3.0) or GFDL (http://www.gnu.org/copyleft/fdl.html)], from Wikimedia Commons*

After the base calling, the dNTPs are washed off from the slide, and a fresh cycle begins. In case of sequences where a particular base is repeating, there will be a linear increase in pH, compared to incorporation of a single base, for up to 10 repeats.

***d. SOLiD sequencing:***

SOLiD refers to Sequencing by Oligonucleotide Ligation and Detection. SOLiD also uses emulsion PCR as mentioned above. Unlike other second generation sequencing techniques, SOLiD is not based on sequencing by synthesis, as it is based on sequencing by ligation. SOLiD employs short oligonucleotide probes that are fluorescently labelled, and these probes carry a specific dinucleotide sequence at the end. Competitive binding of the specific probe to the template DNA results in ligation and the signal generation. This process is repeated several times. In the next step, a new primer which is similar to the initial primer except for the absence of the last

nucleotide is used and the process is repeated. This helps in interrogating the base that is one nucleotide prior to the nucleotide interrogated in the first step. The process continues till the primers are reset by 5 nucleotides, and this helps in covering the entire template sequence. Usual read lengths in SOLiD sequencing is around 50 bases, and SOLiD sequencing has accuracy up to 99.99%.

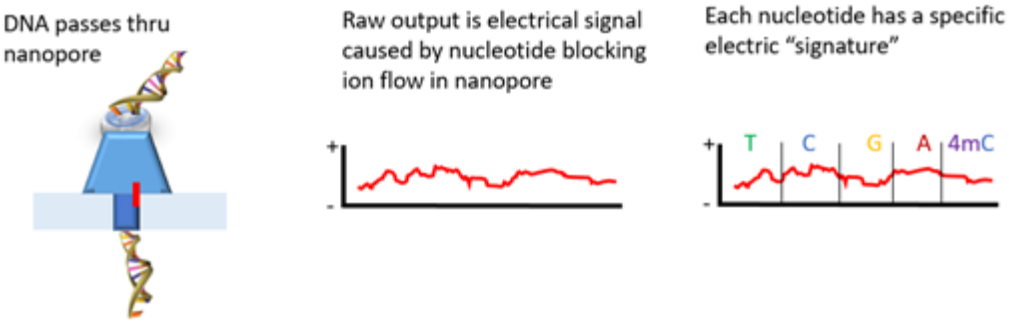
**5. Third Generation Sequencing Techniques:**

The major distinguishing feature between second and third generation sequencing techniques is the read length and these often are in the range of kilobases in the latter. Thus alignment of the reads obtained is relatively easy in third generation platforms. Two widely used 3rd generation sequencing platforms are Oxford nanopore sequencing and Pacific Biosciences (PacBio) Single Molecule Real Time (SMRT) sequencing. Both the platforms can also be used to detect base modifications in DNA (like methyl cytosine).

***a. Nanopore sequencing:***

In nanopore sequencing, as the name implies, there is a pore of nanometer dimensions. Earlier, molecules like porins or alpha-hemolysin were used to form the pore, but the latest devices use synthetic nanopores. Nanopore sequencing, unlike the second generation sequencing techniques, can be used for analysis of RNA also, directly without conversion to cDNA.

Nanopore sequencing is based on the movement of the nucleic acid molecules through the nanopore. Since the properties of individual bases differ, they can be distinguished as they move across the pore. The nucleic acid molecules are in the environment of ions, and the ions continuously move across the pore. Presence of different bases in the pore, as the molecule of nucleic acid moves across the pore, alters the movement of ions, and hence the current generated.




Source: *By Nivretta Thatra [CC BY-SA 4.0 (https://creativecommons.org/licenses/by-sa/4.0)], from Wikimedia Commons*

In nanopore sequencing, the nucleic acid molecule is fragmented into 10 kB or more fragments. Adapters are attached to these fragments. The adapters, apart from the nucleic acid sequences also contain tethering proteins that help in the binding of the fragmented nucleic acid molecule

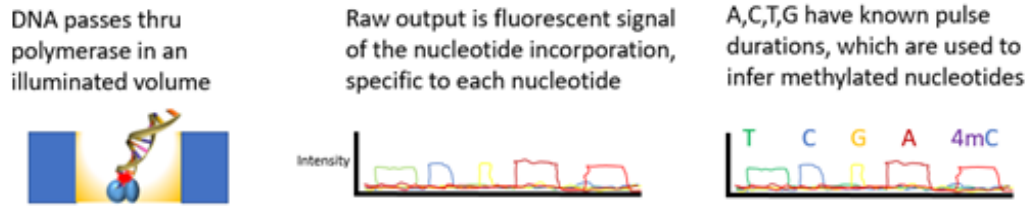
to the nanopore. Motor proteins also aid in the movement of the nucleic acid across the nanopore.

Nanopore can be even performed directly on non-amplified DNA; hence performing a PCR before nanopore sequencing is not obligatory. Latest nanopore devices are also convenient in the aspect of miniaturisation of the equipment, where the device is only a few inches in dimension and can be plugged in to a USB.

***b. PacBio SMRT sequencing:***

PacBio SMRT involves visualisation of fluorescent labelled nucleotides getting incorporated into the growing DNA strand. The heart of an SMRT based sequencer is the SMRT flow cell, also known as ZMWs (zero mode waveguide). A ZMW is cavity of nanometer dimensions that has been made on a metallic surface coated on glass. The dimensions of ZMW are lower than the wavelength of visible light. Hence a LASER beam pointed at the ZMW will illuminate only a part of it and this allows a sensitive detection of light signal. Millions of ZMW are present in a single flow cell, allowing massive parallel sequencing.

Attached to the surface of the ZMW is a DNA polymerase. Once the reaction consisting of the strand to be sequenced, along with the fluorescent labelled nucleotides and other PCR components are added, the nucleotides will start getting incorporated into the newly synthesized strand. This allows the nucleotide which is complementary to the one of template strand that is occupying the catalytic site of DNA polymerase to bind to the DNA polymerase. The incorporation of the base is detected as the sustaining light signal produced at the vicinity of DNA polymerase. The nucleotides that are present in solution in the ZMW acts as a background noise.



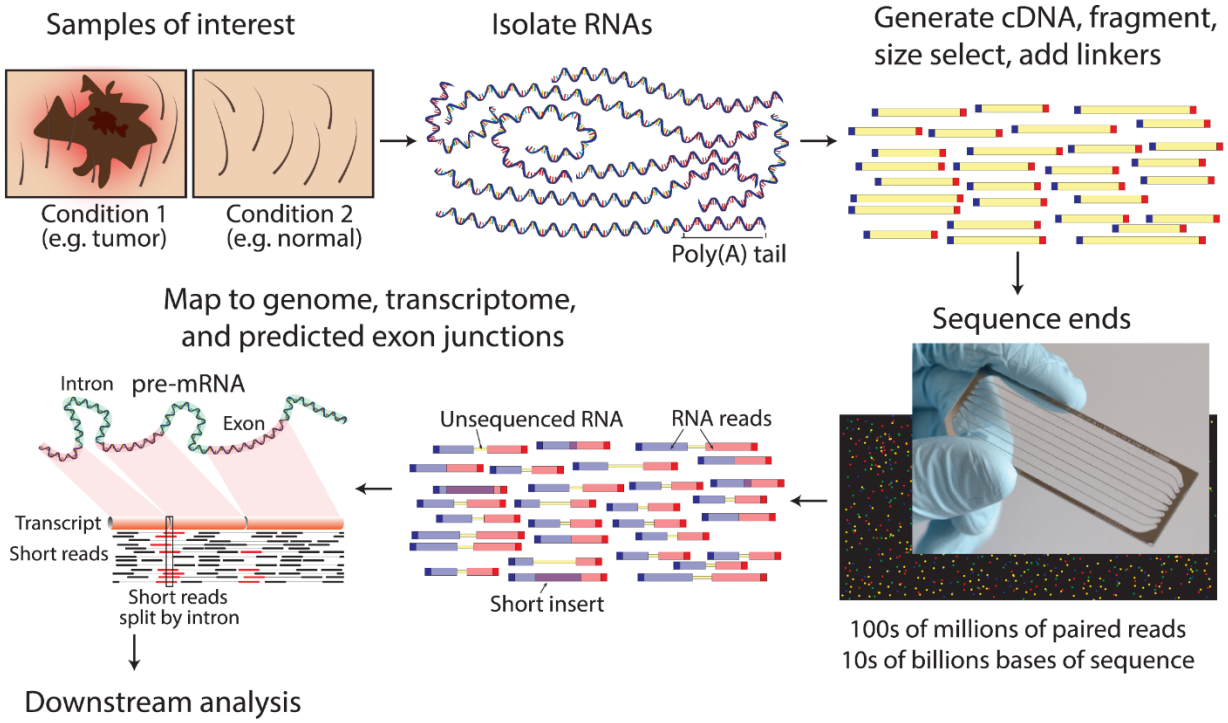

Source: *By Nivretta Thatra [CC BY-SA 4.0 (https://creativecommons.org/licenses/by-sa/4.0)], from Wikimedia Commons*

The time taken by the nucleotides to get incorporated also varies depending on any pre-existing DNA modifications (as a part of epigenetic modulations) in the template strand. Thus, SMRT sequencing can also be used for studying base modifications in DNA.

**6. RNA-Seq and its applications:**

With the advent of second and third generation sequencing, the sequencing of RNA has also become popular. In RNA sequencing, RNA isolated from the cells/tissue/biological sample are

subjected to further purification to remove ribosomal RNA which constitutes the majority of cellular RNAs (using affinity column purification) and genomic DNA (by DNase treatment). The mRNAs thus isolated are fragmented and complementary DNA (cDNA) is synthesized from the mRNA using random hexamer primers and reverse transcriptase. The ds-cDNA is then purified and adapters are ligated and are subjected to sequencing by any of the next generation sequencing techniques mentioned above.




Source: *By Malachi Griffith, Jason R. Walker, Nicholas C. Spies, Benjamin J. Ainscough, Obi L. Griffith [CC BY 2.5 (https://creativecommons.org/licenses/by/2.5)], via Wikimedia Commons*

Some techniques like nanopore sequencing give an option of directly sequencing RNA also, which avoids the need to conversion to cDNA.

RNA-Seq has unique applications.

1. The RNA sequencing can be used to decipher the structure of protein coding genes as it can show the presence of novel transcripts that are not otherwise documented in the data-bases. It can also provide information like transcription start site, untranslated regions, polyadenylation sites and intron-exon boundaries.
2. RNA-seq is commonly used for quantification and comparison of gene expression. This can be done in various scenarios like diseased vs. healthy tissue, treated vs. untreated cells, or cells at various time points of treatment.
3. Variation in genes in the form of single nucleotide variants can be studied using RNA-seq. A similar study can be done in case of DNA in the form of whole exome sequencing, where all the protein coding genes are sequenced.
4. Micro-RNAs and other non-coding RNAs can be studied by RNA-seq.

5. Fusion genes, which have a role in the pathogenesis of cancer and other diseases can be studied using RNA-seq.
6. Single cell sequencing and single cell transcriptomics: This also has become possible with the advancement of sequencing techniques like nanopore sequencing. The transcriptomics and gene expression patterns of single cells can be studied by isolating the RNA from single cell and subjecting it to sequencing. This provides information regarding a single type of cell from a heterogeneous population, thus opening avenues to understand the biology of various cellular processes.

A widely used technique for studying the gene expression is microarray. RNA-seq has distinct advantages over microarray that in the former, novel isoforms of mRNAs and novel mRNAs itself can be identified, whereas microarray relies on the hybridization of mRNA (or cDNA from it) to the probes that are synthesised based on the prior knowledge about existing transcripts.

**Summary:**

- First generation sequencing techniques involves single short reads.
- Next generation sequencing involved massive parallel sequencing.
- Third generation sequencing techniques generates longer reads.
- Various chemistries provide distinct sequencing techniques.
- RNA can be sequenced directly or after conversion to cDNA.
- DNA modifications can be detected directly using third generation sequencing.
- RNA-seq has advantages over microarray.