Journal Healthcare Treatment Development
ISSN: 2799-1148
Vol : 03 , No. 03 , April-May 2023
http://journal.hmjournals.com/index.php/JHTD
DOI: https://doi.org/10.55529/jhtd.33.35.44

Genomics: The History of Discovery, the Role, Tasks and Methods of Sequencing

Abidova Ra'no Mannapovna*

**Virologist of the highest category, head of the virological laboratory, Tashkent Scientific Research Institute of Vaccines and Serums, Tashkent Uzbekistan*

*Corresponding Email: *tolmas4th@mail.ru*

Received: 06 February 2022 **Accepted:** 23 April 2022 **Published:** 29 May 2023

Abstract: At the same time, the concept of a gene - the smallest structural and functional unit of heredity - appeared, and a new science, genetics, was formed. Until the middle of the last century, the structure of carriers of genetic information and the methods of its transmission remained unclear. The subsequent discovery of the genetic code and the development of the central dogma of molecular biology gave a powerful impetus to the development of natural sciences, primarily genetics.

Keywords: Nucleic Acid, Gene, Polymer Molecules, Sequencing, Double Helix, DNA polymerase.

1. INTRODUCTION

Almost one and a half hundred years have passed since the discovery of nucleic acids. Back in 1869, Johann Friedrich Mischer isolated a hitherto unknown substance containing nitrogen and phosphorus from the cells in the pus, which he called nuclein, and then (because of its properties) - nucleic acid. Initially, it was believed that nucleic acid molecules were a phosphorus reserve in cells, but already in the first half of the XX century, scientists proved their hereditary nature. At the same time, the concept of a gene - the smallest structural and functional unit of heredity - appeared, and a new science, genetics, was formed [1, 2].

Until the middle of the last century, the structure of carriers of genetic information and the methods of its transmission remained unclear. The model of the DNA double helix, which is included in all modern textbooks of genetics and molecular biology, was proposed in 1953 by Francis Crick and James Watson (for this, in 1962, scientists received the Nobel Prize). The subsequent discovery of the genetic code and the development of the central dogma of molecular biology gave a powerful impetus to the development of natural sciences, primarily genetics [3, 4, 5]. The main historical milestones of this process are shown in Figure 1.

Journal Healthcare Treatment Development
ISSN: 2799-1148
Vol : 03 , No. 03 , April-May 2023
http://journal.hmjournals.com/index.php/JHTD
DOI: https://doi.org/10.55529/jhtd.33.35.44

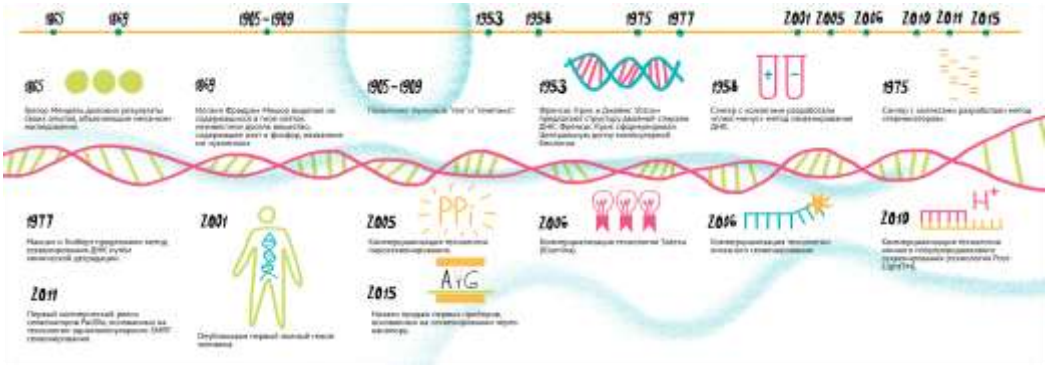

Fig. 1. History of genomic research and sequencing of nucleic acids

March 8, 1865 - Gregor Mendel reported the results of his experiments explaining the mechanism of inheritance. 1869 - Johann Friedrich Mischer isolated a hitherto unknown substance containing nitrogen and phosphorus from cells in pus, which he called nuclein. 1905-1909 - the appearance of the terms "gene" and "genetics". 1953 - Francis Crick and James Watson proposed the structure of the DNA double helix. 1958 - Francis Crick formulated the central dogma of molecular biology. 1975 - Sanger and his colleagues developed a "plus or minus" method of DNA sequencing. 1977 - Sanger's group developed a method of "terminators". 1977 - Maxam and Gilbert proposed a method of DNA sequencing by chemical degradation. 2001 - the first complete human genome is published. 2005 - commercialization of pyrosequencing technology. 2006 - commercialization of Solexa (Illumina) technology. 2006 - commercialization of ligase sequencing technology. 2010 - commercialization of ion semiconductor sequencing technology (PostLightTM technology). 2011 - the first commercial release of PacBio sequencers based on single-molecule SMRT sequencing technology. 2015 - the start of sales of the first devices based on sequencing through a nanopore [5, 6].

What does a DNA molecule look like?Polymer molecules are characterized by a primary structure, which means simply the composition of the molecule (in this case, the sequence of letters A, C, G and T, which make up the genome), a secondary structure, i.e., what chemical bonds are established between these components and what basic spatial structures are obtained as a result (in this case - double helix), and the tertiary structure, i.e. the way the secondary structure is "stacked" in space. The secondary structure of DNA is a double helix consisting of four different nucleotides. Nucleotides are designated by the nitrogenous bases contained in them: adenine (A), cytosine (C), guanine (G) and thymine (T) (there is also uracil, which replaces thymine in RNA), and in the future we will always use these letters.

Journal Healthcare Treatment Development
ISSN: 2799-1148
Vol : 03 , No. 03 , April-May 2023
http://journal.hmjournals.com/index.php/JHTD
DOI: https://doi.org/10.55529/jhtd.33.35.44

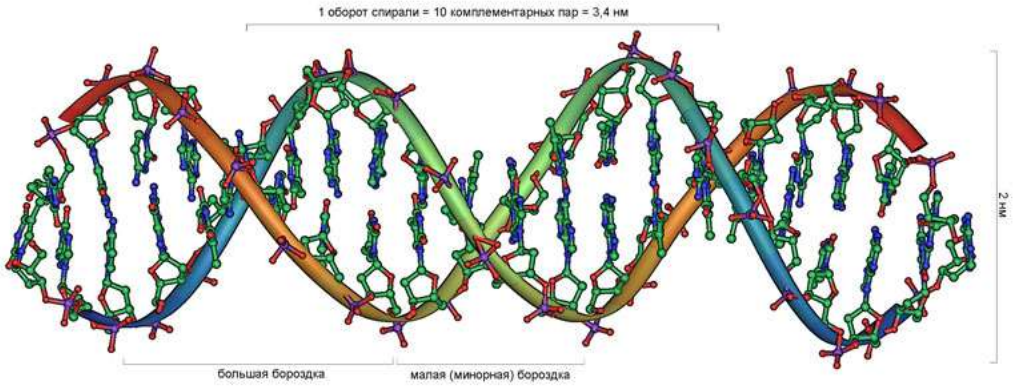

Fig. No. 2. Structure of the DNA molecule

In a double helix, these nucleotides are connected to each other by hydrogen bonds, and the connection is established according to the principle of complementarity: if there is A in one strand of DNA, then there will be T in the complementary strand, and if there is C in one strand, then there will be G in the other. This is what makes it relatively easy to replicate (copy) DNA, for example, during cell division: to do this, it is enough to simply break the hydrogen bonds by dividing the double helix into strands, after which the paired thread for each "descendant" will automatically assemble correctly. It is important to understand that DNA is two copies of the same "text" of four "letters"; the "letters" in the copies are not identical, but uniquely correspond to each other. For example,

ATGCAGAACAGACGATCAGCGACACTTTA
TACGTCTTGTCTGCTAGTCGCTGTGAAAT

Of course, it would be convenient if we could carefully "pull out" one strand of DNA and calmly, nucleotide by nucleotide, "read" this thread from beginning to end. With such an ideal method of sequencing (reading DNA), no tricky algorithms would be needed. Unfortunately, this is not possible at this stage, and we have to be content with the results of the sequencing that we have [7-10].

What is sequencing? Sequencing is a common name for methods that allow you to establish the sequence of nucleotides in a DNA molecule. Currently, there is not a single sequencing method that would work for the whole DNA molecule; they are all arranged as follows: first, a large number of small DNA sections are prepared (the DNA molecule is cloned repeatedly and "cut" in random places), and then each section is read separately.

Cloning occurs either simply by growing cells in a Petri dish, or (in cases where it would be too slow or for some reason would not work) using the so-called polymerase chain reaction. In a brief and inaccurate presentation, it works something like this: first, DNA is denatured, i.e., hydrogen bonds are destroyed, obtaining separate strands. Then the so-called primers are attached to the DNA; these are short sections of DNA to which DNA polymerase can join - a compound that, in fact, is engaged in copying (replication) of the DNA strand.

Journal Healthcare Treatment Development
ISSN: 2799-1148
Vol : 03 , No. 03 , April-May 2023
http://journal.hmjournals.com/index.php/JHTD
DOI: https://doi.org/10.55529/jhtd.33.35.44

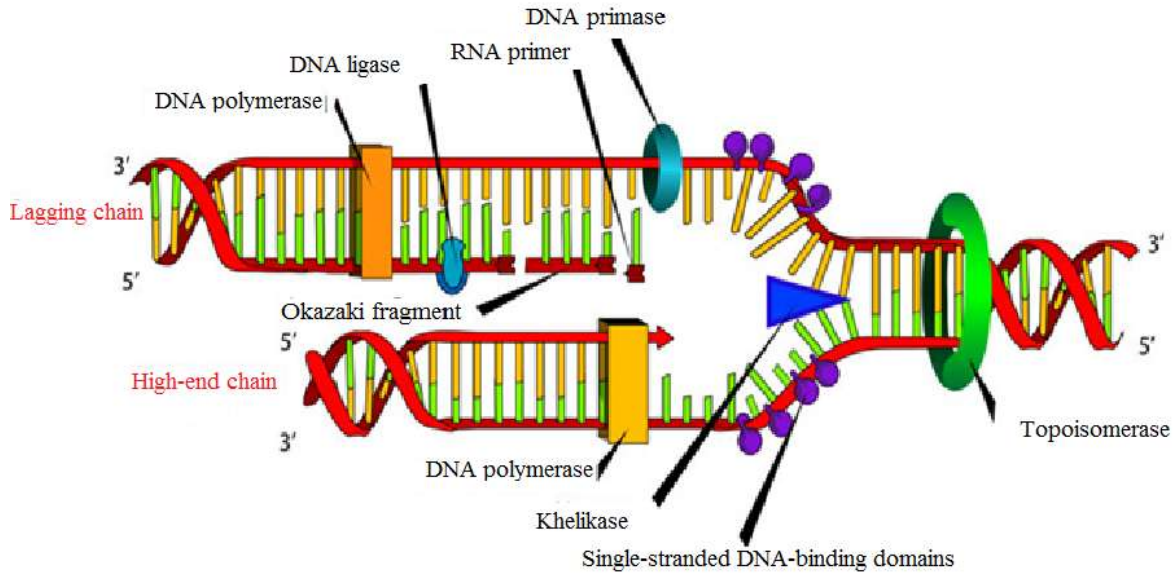

Fig. No. 3. Stages of the polymerase chain reaction process

At the next stage, the polymerase copies the DNA, after which the process can be repeated: after a new denaturation, there will be twice as many individual strands, at the third cycle - four times, and so on.

All these effects are achieved mainly by changing the temperature of a mixture of DNA, primers and polymerase; for our purposes, it is important that this is a fairly accurate process, and errors in it are rare, and the output is a large number of copies of sections of the same DNA. Different sequencing methods differ from each other not by cloning methods, but by how to then read the resulting "soup" from numerous copies of the same DNA [11-14].

Sequencing by Sanger

The first sequencing method that scientists were able to use to process entire genomes (including the human genome) was Sanger sequencing. The meaning is as follows: a section of DNA is cloned, after which the resulting mixture is divided into four parts. Each part is placed in an active environment where there are:

1. DNA polymerase, which, as we have already found out, is engaged in replication,
2. primers needed to start the replication process,
3. a mixture of all four nucleotides that will serve as "bricks" for the construction of new copies of DNA,
4. and, most importantly, special variations of one of the nucleotides (exactly one type of nucleotides for each part), which stop further copying of the DNA molecule.

Actually, the process is almost identical to DNA cloning, which we met in the previous section. The only difference is that now "false" nucleotides are mixed into one of the nucleotides; they can form exactly the same hydrogen bond, but they cannot continue their thread further.

As a result, a large number of copies of the prefixes of the DNA section under study are formed in each part, which have different lengths, but always end with the same letter - depending on when you are lucky enough to take a "false" nucleotide into the cloning process.

Journal Healthcare Treatment Development
ISSN: 2799-1148
Vol : 03 , No. 03 , April-May 2023
http://journal.hmjournals.com/index.php/JHTD
DOI: https://doi.org/10.55529/jhtd.33.35.44

For example, in a test tube where all sequences end in T, our example above would result in a mixture of the following prefixes:
ATGCAGAACAGACGATCAGCGACACTTTA (example)
AT
ATGCAGAACAGACGAT
ATGCAGAACAGACGATCAGCGACACT
ATGCAGAACAGACGATCAGCGACACTT
ATGCAGAACAGACGATCAGCGACACTTT

Results and errors of Sanger sequencing

At the output of the Sanger sequencer, short sections of DNA are obtained, the so-called reads. For bioinformatics, two things are fundamental: firstly, how long the reads are obtained, and secondly, what errors may be in them and how often.

Sanger's reeds are very good according to these criteria: reeds about a thousand nucleotides long are obtained, and the quality begins to drop noticeably only after 700-800 nucleotides. The very process of sequencing by Sanger, which we met in the previous section, determines both the effect of a drop in quality (it is more difficult to distinguish a molecule weighing 700 from a molecule weighing 701 than a mass of 5 from a mass of 6), and another unpleasant effect - if a long sequence of the same letter occurs in the genome (...AAAAAAAA ...), it is difficult to determine exactly how long it is - all intermediate masses will fall into the same tube, some of them may not meet, some merge with each other, etc. But still, Sanger sequencing gives excellent results with sufficiently long reads, which are then relatively easy to assemble.

It was with the help of Sanger sequencing that the human genome was first decoded. Sanger sequencing is still used today, but it is increasingly being replaced by other methods, and it is being used less and less [15-18].

Second generation Sequencers: Illumina

Modern sequencers are the so-called second generation sequencers (SGS, second generation sequencing). In them, sections of DNA are still repeatedly cloned, but the reading process is not arranged in the same way as Sanger's. There are many different methods that differ quite significantly, so we will consider only one of them, one of the most popular today is sequencing using the Solexa method (now Illumina; there is no need to look for a deep meaning in changing the name, just one company bought another).

How is the process of sequencing using the Illumina method.

1. DNA copies are cut in random places into a large number of small sections.

2. Special adapters are added to each site from both sides - small sequences of nucleotides known in advance.

3. Then the resulting mixture is placed on a specially prepared substrate, from which DNA regions complementary to adapters "grow" in the form of a lattice. Thus, they are able to "bind" the DNA regions equipped with adapters to these places. In addition, the adapters also contain primers, sites to which DNA polymerase can join, which carries out DNA replication.

4. In step 3, different sections of DNA randomly "stick" to different places in the lattice. Now we repeatedly clone each section around its place, thereby obtaining whole "clusters".

Journal Healthcare Treatment Development
ISSN: 2799-1148
Vol : 03 , No. 03 , April-May 2023
http://journal.hmjournals.com/index.php/JHTD
DOI: https://doi.org/10.55529/jhtd.33.35.44

This process is known as bridge amplification, because DNA binds to the substrate with two ends at once;

5. DNA sections are denatured (hydrogen bonds are destroyed) – as a result, different DNA sections consisting of one strand "grow" from the lattice nodes on the substrate.

6. The substrate is placed in a solution containing DNA polymerase and specially labeled nucleotides, which immediately finish the replication process (if you remember, these were also used in Sanger sequencing). They attach to the DNA, one to each site. Accordingly, each section is joined by the "letter" with the complementary to which it begins.

7. Then the "extra" nucleotides are washed off, and the remaining labels are read; in Illumina technology, these are fluorescent labels that can be made to glow in different colors and photographed. It is at this step that we will find out with which letter each "cluster of sections" of DNA begins.

8. After that, a radical is chemically "cut off" from the already bound nucleotides, which interfered with the further superstructure of the DNA molecule. Now you can go back to step 6 and repeat the process, reading the second letters in each sequence on the second cycle, and so on.

As a result, at each cycle we read simultaneously a very large number of nucleotides from different sequences. But we have to pay for this by the fact that the DNA sections that we can read turn out to be much shorter than in the case of Sanger sequencing - Illumina reads are usually about 100 nucleotides long.

Paired reads and problem statement

There is one more important detail. Sections of DNA "stick" to the substrate with both ends, and we can find out which sequences correspond to the same section. This means that in reality we are reading the same section, the length of which is approximately known to us, from both sides at once. As a result, the data is obtained something like this:

ATGCAGA??????????????CACTTTA,

moreover, the distance between the known lines (the number of question marks) is not exactly known. Depending on the technology, it is possible to obtain both very long unknown fragments (about 1000 nucleotides), "framed" by two 100-length reeds, and short fragments in which literally two to three dozen nucleotides between the reeds are unknown. Both can be very helpful in assembling genomes.

So, now we can formally set the task of assembling genomes. It sounds like this: for a large number of substrings of small length, restore the original long string in the alphabet from the letters A, C, G, T. In the case of Illumina sequencing, for a large number of pairs of short substrings separated in the original string by an approximately known distance. Having set this task, we can forget about biology, chemistry and medicine - we are facing a purely algorithmic task. However, before moving on to mathematics, let's make a few more remarks.

Errors and quality indicators in second-generation sequencers

As we already know, sequencing always contains errors. In Illumina sequencers and similar ones, errors usually occur at the phase when it is necessary to recognize labeled nucleotides, i.e. to understand what color and with what strength clusters of repeatedly cloned DNA sections glow.

Journal Healthcare Treatment Development
ISSN: 2799-1148
Vol : 03 , No. 03 , April-May 2023
http://journal.hmjournals.com/index.php/JHTD
DOI: https://doi.org/10.55529/jhtd.33.35.44

The problem here is that due to the imperfection of the remaining stages of the process, clusters never glow with only one color; it is always a mixture of all four colors with one intensity or another. It is necessary to identify the most intensive component and assess how likely an error is in this letter; this task is called base calling (recognition of nucleotides).

It is important for us now that, as a result, the sequencer matches each nucleotide of each reed with the probability that this nucleotide was recognized correctly. These probabilities can also be used during assembly, and sequencers issue them along with the actual reads.

As a result, a typical read in the so-called fastq format, standard for second-generation sequencers, looks something like this:

@EAS20_8_6_1_3_25/1
GCAAAAAACTTACCCCGGAACAGGCCGAGCAGATCAAAACGCTACTGCAATACA
GACCATCAAGCACCAACTCCCNNNCGTAGNNNNNNTATGTTNNNNG
+EAS20_8_6_1_3_25/1
HHHHHHHHGHHHHHHHHHHHHHHHHHHHHHEHHHHHHHHHEGHHHHGHHGHEFD?
A=A&FFBB> &: ===@&@E@E> A##########################

The first and third lines contain reed's name; the second line is the nucleotide sequence itself. Note that among the letters A, C, G, T there are also letters N - this means that the sequencer could not unambiguously determine which nucleotide was here, and gave up. And the fourth line encodes, on a logarithmic scale, the probability that a particular nucleotide is recognized correctly; for example, H here corresponds to an error probability of about one ten thousandth. As a rule, the quality deteriorates by the end of the read; in our example, as you can see, the tail of the read could not be read at all reliably [19-22].

Sanger or Illumina?

The human genome was first assembled on Sanger sequencers, and the algorithmic side of that project was worked out much less than it is now, ten years later. The algorithms used to assemble the first human genome are much simpler than those that we will talk about in the future. However, the first genome was still collected; maybe all the algorithmic progress is an unnecessary myth, and old programs would be enough?

Incredibly, but a fact: the "old" sequencers (first generation, Sanger) produce much more suitable data for assembly than the "new" ones (second generation). This is mainly expressed in the length of the reads, those sections of DNA that can be read sequentially, and which, in fact, need to be assembled into one big line. The first-generation sequencers produced reeds longer than five hundred nucleotides, usually about a thousand. Modern sequencers produce pairs of reads, each of which has a length of about one hundred nucleotides.

There are two main types of NGS platforms: second- and third-generation sequencers. Second-generation technologies can read DNA directly. After the DNA is divided into fragments, short pieces of genetic material, called adapters, are added to give each nucleotide a different color. For example, adenine is colored blue, and cytosine is red. Finally, these DNA fragments are loaded into a computer and reassembled into the entire genomic sequence.
Third-generation technologies, such as Nanopore MinIon, directly sequence DNA by passing the entire DNA molecule through an electrical pore in the sequencer. Since each pair of

Journal Healthcare Treatment Development
ISSN: 2799-1148
Vol : 03 , No. 03 , April-May 2023
http://journal.hmjournals.com/index.php/JHTD
DOI: https://doi.org/10.55529/jhtd.33.35.44

nucleotides disrupts the electric current in a certain way, the sequencer can read these changes and upload them directly to the computer. This allows doctors to sequence samples in clinical and medical institutions on the ground. However, nanopore sequences have a smaller DNA volume compared to other NGS platforms.

Although each class of sequencers processes DNA in its own way, they can all report on the millions or billions of building blocks that make up the genome in a short time - from a few hours to several days. For example, Illumina NovaSeq can sequence approximately 150 billion nucleotides, equivalent to 48 human genomes, in just three days.

The importance of carrying out genome sequencing is such an important tool in the fight against the spread of diseases and changes in the genome over time. This can be clearly seen on the example of the SARS-CoV-2 coronavirus.

Scientists have used genome sequencing to track SARS-CoV-2 almost in real time since the beginning of the pandemic. Millions of individual SARS-CoV-2 genomes have been sequenced and placed in various public repositories, such as the Global Avian Influenza Data Exchange Initiative and the National Center for Biotechnology Information.

Genomic surveillance determines public health decisions as each new option becomes available. For example, sequencing the genome of the omicron variant allowed researchers to detect more than 30 mutations in a spiked protein that allows the virus to bind to cells of the human body. This makes omicron a cause for concern, since it is known that these mutations contribute to the ability of the virus to spread. Researchers are still investigating how these mutations may affect the severity of infections caused by omicron, and how well it can avoid the use of current vaccines.

Sequencing has also helped researchers identify variants that have spread to new regions. After receiving a sample of SARS-CoV-2 taken from a traveler who returned from South Africa on November 22, 2021, researchers from the University of California, San Francisco were able to detect the presence of omicron in five hours and had almost the entire genome. sequence of eight

The rapid detection of OMICRON worldwide highlights the power of reliable genomic surveillance and the importance of sharing genomic data across the globe. Understanding the genetic makeup of the virus and its variants gives researchers and public health officials an idea of how best to update public health guidelines and maximize the allocation of resources for vaccine and drug development. By providing important information on how to contain the spread of new variants, genomic sequencing has saved and will continue to save countless lives during the pandemic [23-28].

2. CONCLUSION

Thus, At the same time, the concept of a gene - the smallest structural and functional unit of heredity - appeared, and a new science, genetics, was formed. Until the middle of the last century, the structure of carriers of genetic information and the methods of its transmission remained unclear. The subsequent discovery of the genetic code and the development of the central dogma of molecular biology gave a powerful impetus to the development of natural sciences, primarily genetics.

Journal Healthcare Treatment Development
ISSN: 2799-1148
Vol : 03 , No. 03 , April-May 2023
http://journal.hmjournals.com/index.php/JHTD
DOI: https://doi.org/10.55529/jhtd.33.35.44

3. REFERENCES

1. Phillip E. C. Compeau, Pavel A. Pevzner. Genome Reconstruction: A Puzzle with a Billion Pieces. In Bioinformatics for Biologists.
2. Heather JM, Chain B. The sequence of sequencers: The history of sequencing DNA. Genomics. 2016 Jan;107(1):1-8. doi: 10.1016/j.ygeno.2015.11.003.
3. Sanger F. 1980. Frederick Sanger — Biographical. (URL http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1980/sanger-bio.html)
4. Watson J., Crick F. Molecular structure of nucleic acids. Nature. 1953;171:709–756. (URL http://www.nature.com/physics/looking-back/crick/)
5. Pevzner P.A., Tang H., Waterman M.S. An Eulerian path approach to DNA fragment assembly. Proc. Natl. Acad. Sci. USA, 98(17):9748-9753, 2001.
6. Hutchison C. A. DNA sequencing: bench to bedside and beyond. Nucleic Acids Res. 2007;35:6227–6237.
7. M.C. Schatz, A.L. Delcher, S. Salzberg. Assembly of large genomes using second-generation sequencing. Genome Research, 20(9):1165-1173, 2010.
8. M.Chaisson, P. Pevzner, H. Tang. Fragment assembly with short reads. Bioinformatics 20 (13): 2067-2074, 2004.
9. Holley R.W., Apgar J., Merrill S.H., Zubkoff P.L. Nucleotide and oligonucleotide compositions of the alanine-, valine-, and tyrosine-acceptor soluble ribonucleic acids of yeast. J. Am. Chem. Soc. 1961;83:4861–4862. (URL http://pubs.acs.org/doi/abs/10.1021/ja01484a040)
10. Holley R.W. Structure of a ribonucleic acid. Science. 1965;147:1462–1465. (URL http://www.sciencemag.org.libproxy.ucl.ac.uk/content/147/3664/1462.full.pdf)
11. Margulies M., Egholm M., Altman W., Attiya S. Genome sequencing in microfabricated high-density picolitre reactors. Nature. 2005;437:376–380. (URL http://www.nature.com/nature/journal/v437/n7057/abs/nature03959.html)
12. Levy S. The diploid genome sequence of an individual human. PLoS Biol. 2007;5 (URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1964779&tool=pmcentrez&rendertype=abstract)
13. Quail M.A. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics. 2012;13:341. (URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3431227&tool=pmcentrez&rendertype=abstract)
14. Buermans H.P.J., den Dunnen J.T. Next generation sequencing technology: advances and applications. Biochim. Biophys. Acta. 2014;1842:1932–1941. (URL http://www.ncbi.nlm.nih.gov/pubmed/24995601)
15. Gut I.G. New sequencing technologies. Clin. Transl. Oncol. 2013;15:879–881. (URL http://www.ncbi.nlm.nih.gov/pubmed/23846243)
16. Braslavsky I., Hebert B., Kartalov E., Quake S.R. Sequence information can be obtained from single DNA molecules. Proc. Natl. Acad. Sci. U. S. A. 2003;100:39603964. (URL http://www.pnas.org/content/100/7/3960.short)

Journal Healthcare Treatment Development
ISSN: 2799-1148
Vol : 03 , No. 03 , April-May 2023
http://journal.hmjournals.com/index.php/JHTD
DOI: https://doi.org/10.55529/jhtd.33.35.44

17. Bowers J. Virtual terminator nucleotides for next-generation DNA sequencing. Nat. Methods. 2009;6:593-595. (URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2719685&tool=pmcentrez&rendertype=abstract)
18. GenomeWeb . 2012. Helicos BioSciences Files for Chapter 11 Bankruptcy Protection. (URL http://www.genomeweb.com/sequencing/helicos-biosciences-files-chapter-11-bankruptcy-protection)
19. van Dijk E.L., Auger H., Jaszczyszyn Y., Thermes C. Ten years of next-generation sequencing technology. Trends Genet. 2014;30 (URL http://linkinghub.elsevier.com/retrieve/pii/S0168952514001127)
20. Flusberg B.A. Direct detection of DNA methylation during single-molecule, real-time sequencing. Nat. Methods. 2010;7:461–465. (URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2879396&tool=pmcentrez&rendertype=abstract)
21. Haque F., Li J., Wu H.-C., Liang X.-J., Guo P. Solid-state and biological nanopore for real-time sensing of single chemical and sequencing of DNA. Nano Today. 2013;8:56–74. (URL http://www.sciencedirect.com/science/article/pii/S1748013212001454)
22. Dekker C. Solid-state nanopores. Nat. Nanotechnol. 2007;2:209–215. (URL http://www.nature.com/nnano/journal/v2/n4/abs/nnano.2007.27.html)
23. Kilianski A. Bacterial and viral identification and differentiation by amplicon sequencing on the MinION nanopore sequencer. GigaScience. 2015;4 (URL http://www.gigasciencejournal.com/content/4/1/12)
24. Madoui M.-A. Genome assembly using nanopore-guided long and error-free DNA reads. BMC Genomics. 2015;16:1–11. (URL http://www.biomedcentral.com/1471-2164/16/327)
25. Hayden E.C. Pint-sized DNA sequencer impresses first users. Nature. 2015;521:15–16.
26. Holley R.W., Apgar J., Merrill S.H., Zubkoff P.L. Nucleotide and oligonucleotide compositions of the alanine-, valine-, and tyrosine-acceptor soluble ribonucleic acids of yeast. J. Am. Chem. Soc. 1961;83:4861–4862. (URL http://pubs.acs.org/doi/abs/10.1021/ja01484a040)
27. PM, A MinION nanopore sequencing identifies the position and structure of a bacterial antibiotic resistance island. Nat. Biotechnol. 2014;33
28. Branton D. The potential and challenges of nanopore sequencing. Nat. Biotechnol. 2008;26:1146–1153. (URL http://www.nature.com/nbt/journal/v26/n10/abs/nbt.1495.html)