Next Generation Sequencing (NGS) techniques

Next generation sequencing (NGS) represents a set of modern technologies that revolutionized the field of genomics, allowing for DNA and RNA sequencing in a much faster, more efficient and accessible way.

Evolution

Before the NGS, Sanger sequencing (developed in the ‘70s) was considered to be the standard in the field. This method had significant limitations: it could process only a fragment of DNA at a given moment and it was expensive and slow for large-scale analyses. The development of NGS started in the 00’s, when companies like 454 Life Sciences, Illumina and Applied Biosystems introduced the first commercial platforms that allowed massively parallel sequencing

Fundamental principles

What makes NGS different is the capacity to sequence millions or even billions of DNA fragments at the same time. Although there are multiple NGS platforms, they all follow a similar workflow:

1. Library preparation – DNA is fragmented into small pieces and adapters (short, known sequences) are attached to the ends of the fragments.
2. Amplification – DNA fragments are multiplied to created detectable signals. This step may involve emulsion PCR (Polymerase Chain Reaction), bridge PCR, or other methods, depending on the platform used.
3. Sequencing – this step varies between platforms, but typically involves detecting signals (optical, chemical, or electrical) as nucleotides are incorporated into the amplified fragments.
4. Data Analysis – raw data is processed by bioinformatic algorithms to generate sequences and align the to a reference genome or assemble them ‘de novo’

Main NGS Platforms

Illumina (Sequencing by Synthesis)

The most popular NGS technology is offered by Illumina, which uses “sequencing by synthesis”. In this approach:

➔ DNA fragments are immobilized on a solid surface
➔ Fluorescently labeled nucleotides are incorporated sequentially
➔ After each incorporation, an optical scan occurs to identify the added nucleotide

Ion Torrent (Semiconductor Sequencing)

➔ Uses the detection of pH changes caused by the release of hydrogen ions during nucleotide incorporation
➔ Doesn't need fluorescence or optical imaging

PacBio (Single Molecule Real-Time Sequencing -SMRT)

➔ It sequences individual DNA molecules in real time
➔ Can generate long sequences

Oxford Nanopore

➔ Uses protein pores through which DNA molecules pass, generating changes in the electrical current.
➔ The changes are interpreted to determine the nucleotide sequence

Applications in bioinformatics

NGS opened multiple opportunities in bioinformatics and genomic research

Sequencing the entire genome (WGS): allows complete sequencing of the genome.

Sequencing the exome (WES): focuses on protein-coding regions (exons), reducing the cost and complexity of analysis.

Sequencing the transcriptome (RNA-Seq): allows the allows the analysis of all RNAs present in a cell at a given moment.

Epigenic sequencing: includes methods for studying protein-DNA interactions, DNA methylation, providing information about the regulation of gene expression.

Metagenomics: studies all the genetic material present in an environmental sample, without isolating and culturing individual microorganisms.

Challenges in analyzing NGS data

Next-generation sequencing generates huge volumes of data, creating significant challenges for bioinformatics:

1. Storing data – A single experiment can generate terabytes of data, making it need a proper structure of storing.
2. Computational power – Processing and analyzing data needs substantial computational resources
3. Quality Assurance – Evaluation and filtering of low-quality data are essential for reliable results.
4. Genome Assembly – Reconstructing the genome from short fragments is an algorithmic challenge, especially for repetitive regions.
5. Variant Analysis – Identifying and interpreting genetic variations requires sophisticated algorithms and reference databases

Trends and future directions

NGS field continues to rapidly evolve:

- Long-read sequencing – Technologies as PacBio and Oxford Nanopore are gaining popularity due to their ability to generate longer sequences.
- Single-cell sequencing – Allows the study of cellular heterogeneity in tissues and tumors.
- Direct RNA sequencing – Eliminates the need for conversion to cDNA, reducing potential errors.
- Real-time sequencing – Allows monitoring of biological processes as they occur.
- Portable sequencing – Devices such as the Oxford Nanopore MinION facilitate sequencing in the field or at the point of care.

Conclusion

New-generation sequencing techniques have fundamentally transformed biological and medical research, allowing for genomic studies at an unprecedented scale. Using their capacity to generate massive volume of genomic data rapidly and with accessible costs, NGS opened new horizons in understanding the life's genetic code. These technologies continue to evolve, offering more and more advanced possibilities for single molecule sequencing, real-time analysis and custom applications. As computational and data analysis challenges are addressed, the impact of NGS in biodiversity conservation, agriculture, medicine, and many other fields will continue to expand, propelling science toward new fundamental discoveries about the nature of life.