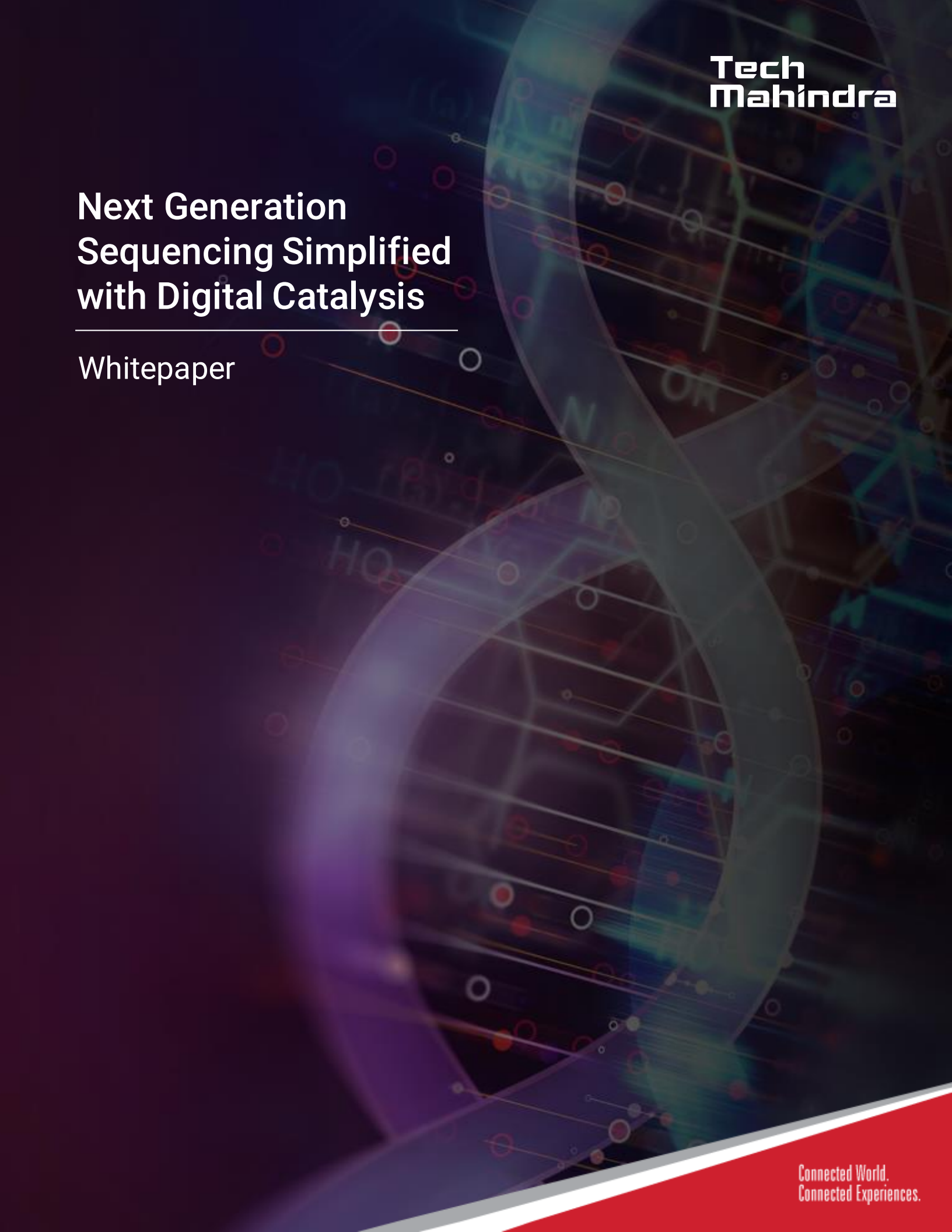

Tech Mahindra

Next Generation Sequencing Simplified with Digital Catalysis

Whitepaper

Connected World.
Connected Experiences.

Table of Content

Abstract

Gene sequencing has advanced a lot since its invention in 1977 through Sanger's chain termination method (first generation sequencing). This advancement eventually led to the discovery of Next generation sequencing or NGS (also termed as massively parallel or deep sequencing) over the last decade. NGS' commercialization in 2005 revolutionized genomic research linked to drug discovery and other biotechnological innovations. And NGS technologies enabled life sciences researchers getting higher throughput of data with lower cost and thus made population-scale genome research possible.

Today, NGS provides three key advantages, which also showcase the improvements over its predecessor first generation sequencing[1]:

- **Reduced time and lab effort:** User can avoid in-vitro cloning and make libraries for sequencing in cell-free system
- **Ability to provide wider coverage:** As the name massively parallel sequencing for NGS suggests, it can process many sequencing reactions in parallel at the same time
- **Increased accuracy and efficiency:** Detection of bases happens precisely and in parallel

Using NGS, an entire human genome can be sequenced in just a day whereas Sanger's method took almost a decade to come up with the final draft of the sequencing report for an entire human genome.[2]

This easily demonstrates the speed of NGS related research and discovery in the field of life sciences. Within a remarkably shorter period, NGS may allow us to look at thousands of people and see the critical genetic differences amongst them, to discover individual gene traits that cause cancer, autism, heart disease, antibiotic resistance, and many more, and was not technically feasible and affordable previously. Besides this, the growing demand for NGS technology applications for a future bio-based economy by global industrial biotech communities is an important agenda towards a sustainable greener climate for tomorrow.

NGS Technology and its Applications Today

Apart from whole genome sequencing (human to microbes), NGS technology is currently being used for the investigation of genome diversity (mutation detection and relation to drug discovery), metagenomics, epigenetics, discovery of non-coding RNAs, DNA-protein interaction understanding, and so on to accelerate pharma/biotech innovation.

Each of these applications is fundamental to the drug discovery process. Today, faster and easier identification of novel pathogens (bacteria, virus, fungi, or parasite) has become possible due to the integration of high throughput application of NGS technology, particularly in metagenomics. This enables biopharma companies/researchers to devise a strategy for easy target detection; and thereby, come up with suitable remedies much faster. This has opened up the doors for pharmaceutical or life sciences companies to re-strategize, narrowing down the gap between disease identification and medical remedies.

An Opportunity for Public-Private Partnership

Analysts predict the NGS market to reach approximately USD 20 billion in 2025 from the current USD 8 billion. National population DNA sequencing is gaining traction in many countries via public-private partnership models in order to link genetics and disease identification, and thereby find possible remedies.

For example, the 100,000 genome project by the National Health Service (NHS) in the UK has been completed recently by achieving 100,000 sequence completion.[3] This successful completion with sequence data will help in developing new genomic medicine services for their patients. In another instance, Thermo Fisher Scientific signed a companion diagnostic (CDx) agreement with Chugai Pharmaceutical Co. Ltd. in July-2020,[4] and applied to the Japan government to expand the use of Oncomine Dx Target Test for cancer research.

Current NGS Technology Trends

NGS technology's increasing demand across the globe in the field of healthcare and life sciences includes transcriptome sequencing (RNA-sequencing), whole-genome, and whole-exome sequencing (WGS and WXS) in genome-wide association studies (GWAS), chromatin immunoprecipitation, or methylated DNA immunoprecipitation followed by sequencing (ChIP-Seq or MeDIP-Seq) as well as various other specialised protocols.

However, the basic process flow (see Figure 1) remains the same and differs only for different types of sample processing prior to sequencing and sequencing volume or scale.[5]

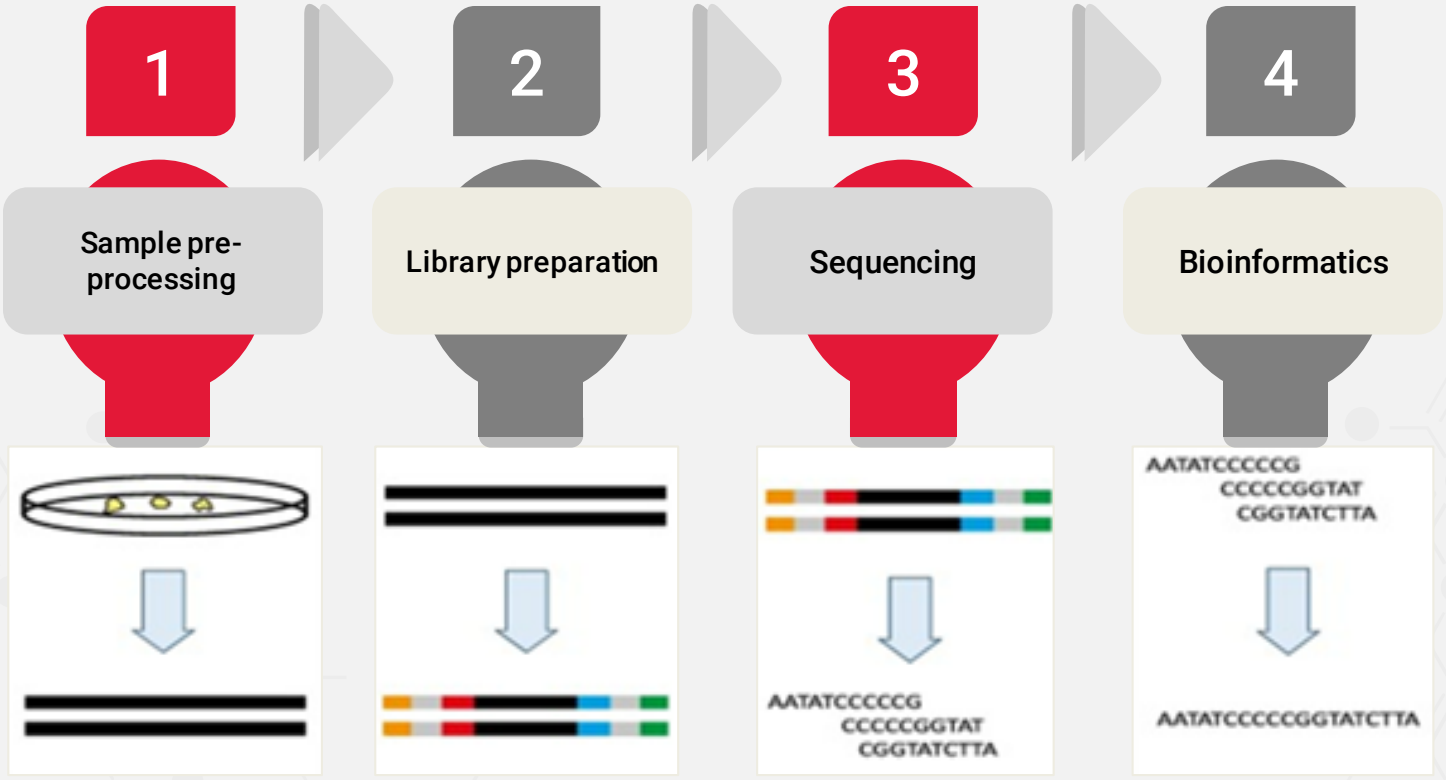

Figure 1: The Steps of NGS

Though second-generation sequencing technologies have revolutionised life science discoveries over the last decade, it, still has not been able to answer a few pertinent questions, especially when it comes to complex rearrangements or repetitive regions in the target genome. Advancements to third-generation sequencing in the late-2000s does help in finding some answers to such cases.

Digital Transformation to NGS Pathway: The Need of the Hour

Applications of NGS in precision medicine generate huge amounts of data at a base-level resolution within a very short span and at low cost. However, this massive increase in data output in NGS technology has caused life sciences companies to think beyond conventional approaches while incorporating NGS into their discovery flow in order to handle, store, share, manage and conclude on big data with no or very minimal manual intervention and to be in compliance.

Despite huge progress in the field of NGS over the past years, the set-up of NGS data analytics is still challenging (see Figure 2), particularly in a core facility environment where the target architecture must be able to crunch data from thousands of samples per year.[6,7]

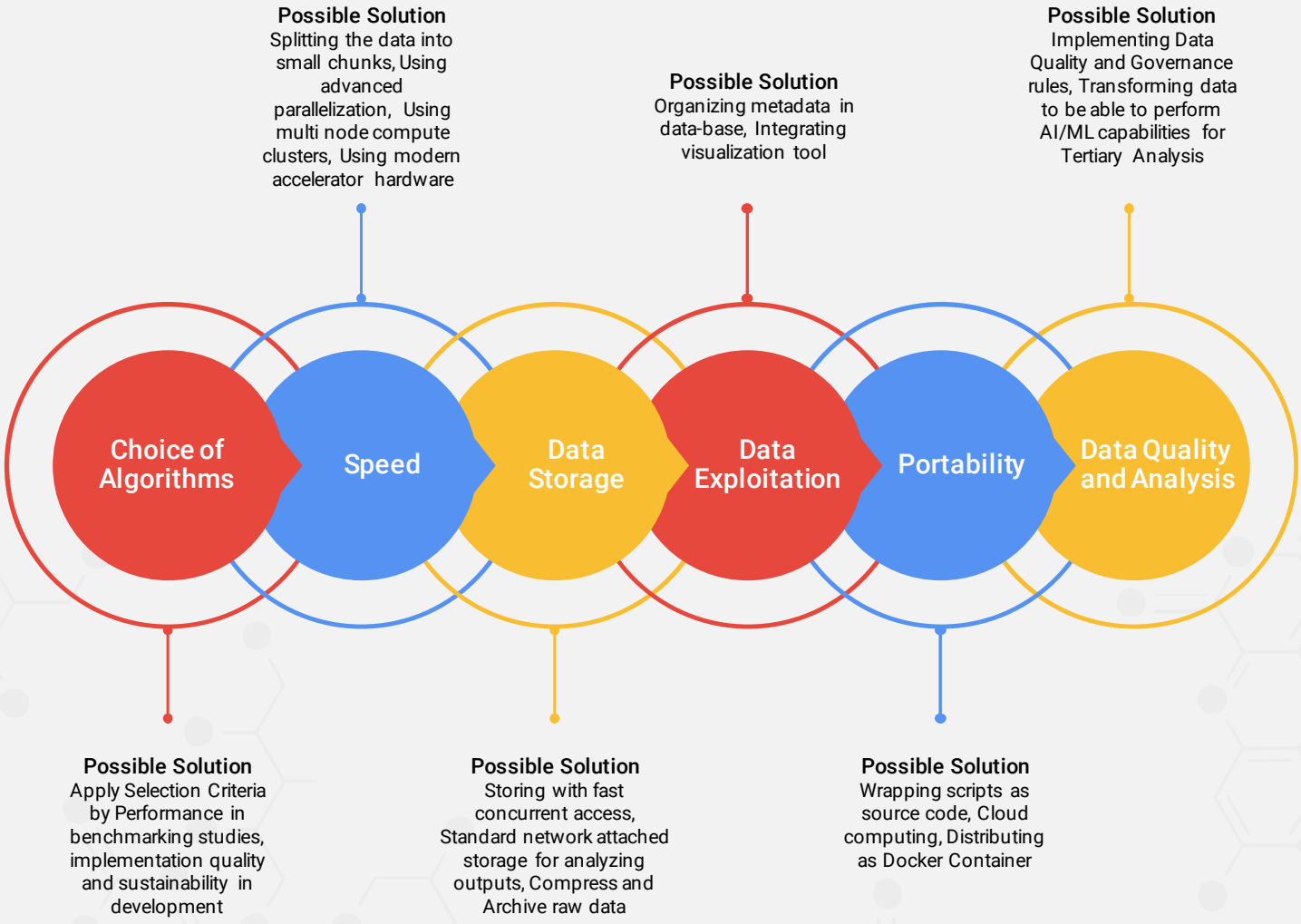

Figure 2: Challenges and possible solutions through NGS analytics

Besides this, genomic data is personal and in a clinical context, integrating big genomic data to better understand gene function and interaction creates legal/ethical issues posing a threat to this emerging field of NGS-linked Life Sciences discovery. Digital interference to manage big genomic data can only be seen as a possible solution here.

However, to be in compliance with regard to clinical genomic data privacy, publicly accessible, cloud-based read mapping services to commercial sequencing services may be seen as a roadblock. Ways to overcome this roadblock may include the establishment of non-public (protected) cloud-based solutions with strict and reliable access control, and/or the development of cloud-based read mapping that does not need a transfer of the actual read sequence to the public cloud.

NGS analysis workflow shares a number of steps which are the same for many use cases of Life Sciences Discovery projects. Therefore, it is very critical to choose the right approaches to analyse the data derived from NGS. The analysis workflow must be highly standardized, but at the same time, flexible enough to also accommodate customized analyses developed from time to time. Artificial intelligence(AI) integrated into a massive pool of data derived from NGS technology is going to be a game changer to transform the life sciences industry by faster discovery of precision medicine.[8]

The Way Forward

The advancement of NGS over the past few years has significantly narrowed down the gap between bioinformatics and clinical informatics. This, in-turn, increases the importance of NGS-based testing. However, as the technology in this field evolves continuously, big data management keeps on throwing challenges. The challenges include a new analytical pipeline, a data management system, patient privacy concerns, and laboratory regulations linked to big data. Next-generation sequencing technology will redefine the practice of medicine and Life Sciences discovery if data derived from it is managed well and smarter. To sustain the rapidly progressing healthcare and life sciences advancements, adoption of smart computing technology to NGS like genomics-linked life sciences discovery tool, is highly essential.

Endnotes

1. Park, S. T. (2016). Trends in Next-Generation Sequencing and a New Era for Whole Genome Sequencing. https://www.einj.org/articles/archive.php. Retrieved 2021, from https://www.einj.org/journal/view.php?number=618

2. Behjati, S. (2013). What is next generation sequencing? https://pubmed.ncbi.nlm.nih.gov/. Retrieved 2021, from https://pubmed.ncbi.nlm.nih.gov/23986538/

3. Samuel, G. N. (2017). The UK's 100,000 Genomes Project: manifesting policymakers' expectations. https://www.ncbi.nlm.nih.gov/. Retrieved 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5706982/

4. Thermo Fisher signs CDx agreement with Chugai Pharma. (2020, June 9). BioSpectrum Asia Edition. https://www.biospectrumasia.com/news/28/16267/thermo-fisher-signs-cdx-agreement-with-chugai-pharma.html

5. Hess, J. F. (2020). Library preparation for next generation sequencing: A review of automation strategies. https://www.sciencedirect.com/. Retrieved 2021, from https://www.sciencedirect.com/science/article/pii/S0734975020300343

6. Kulkarni, P. (2017). Challenges in the Setup of Large-scale Next-Generation Sequencing Analysis Workflows. https://www.sciencedirect.com/. Retrieved 2021, from https://www.sciencedirect.com/science/article/pii/S2001037017300776

7. Kulkarni, P. (2017). Challenges in the Setup of Large-scale Next-Generation Sequencing Analysis Workflows. https://pubmed.ncbi.nlm.nih.gov/. Retrieved 2021, from https://pubmed.ncbi.nlm.nih.gov/29158876/.

8. Wandelt, S. (2012). Data Management Challenges in Next Generation Sequencing Sebastian. https://link.springer.com/. Retrieved 2021, from https://link.springer.com/article/10.1007/s13222-012-0098-2

Authors

Dr. Ratnakar Palakodeti
VP & Practice Head- Life sciences, Tech Mahindra

Two decades of experience in pharmaceutical industry heading portfolio management, strategic marketing and planning. Ph.D in Business Management

Mr. Aniruddha Paul
Tech Manager – Healthcare & Life Sciences, Tech Mahindra

16 years of experience in Life Sciences R&D laboratory management, Molecular biology, Genomics and Microbiology applications. MSc(Microbiology), PGDM

Tech Mahindra

www.youtube.com/user/techmahindra09
www.facebook.com/techmahindra
www.twitter.com/tech_mahindra
www.linkedin.com/company/tech-mahindra
www.techmahindra.com