

# High-Throughput Technologies Applied to the Diagnosis of Rare Diseases



Nogueira C<sup>1,2\*</sup> and Vilarinho L<sup>1,2</sup>

<sup>1</sup>Research & Development Unit, Human Genetics Department, National Institute of Health Doutor Ricardo Jorge, Portugal

<sup>2</sup>Newborn Screening, Metabolism & Genetics Unit, Human Genetics Department, National Institute of Health Doutor Ricardo Jorge, Portugal

## Abstract

Rare Diseases have an incidence of less than 5 cases per 10,000 people, with around 7,000 diseases identified worldwide, being 80% of genetic origin. It is estimated that in Portugal there are 600,000 to 800,000 patients with rare diseases. Next Generation Sequencing (NGS) technology is underlying major advances in genetic diagnosis of rare metabolic diseases, as it has the capacity to generate a huge amount of data in a short time and at an affordable cost, providing practical information to health professionals and patients within the scope of personalized medicine. The translation of NGS knowledge to the laboratorial routine is facilitating a better understanding of these rare diseases and is allowing a quicker detection of similar cases, reducing the costs in the National Health System due to prolonged and unspecific palliative treatments. This translational research is enabling our country with new technological approaches, strengthening our Center as a reference laboratory for the study of these pathologies.

**Keywords:** NGS; Rare Diseases; WES; Exome Sequencing

**\*Corresponding author:** Nogueira C, Research & Development Unit, Newborn Screening, Metabolism & Genetics Unit, Human Genetics Department, National Institute of Health Doutor Ricardo Jorge, Porto, Portugal

**Submission:**  May 25, 2022

**Published:**  July 22, 2022

Volume 4 - Issue 1

**How to cite this article:** Nogueira C, Vilarinho L. High-Throughput Technologies Applied to the Diagnosis of Rare Diseases. J Biotech Biores. 4(1). JBB. 000580. 2022.

**Copyright@** Nogueira C, This article is distributed under the terms of the Creative Commons Attribution 4.0 International License, which permits unrestricted use and redistribution provided that the original author and source are credited.

## Introduction

Rare diseases collectively are not rare, affecting millions of individuals worldwide. More than 7,000 rare diseases have been identified, being a substantial number of which life threatening or chronically debilitating [1,2]. However, the number of phenotypes that remain to be defined may be considerably higher [3].

A single rare disease affects less than 5 cases per 10,000 people, but on aggregate, an estimated 350 million people globally suffer from rare diseases (~4% of an estimated world population of 7.5 billion). The majority of rare diseases (50-75%) affect children and are responsible for 35% of deaths in the first year of life, for 30% of deaths up to 5 years of age and also for a significant number of paediatric hospital admissions with a range of phenotypes associated to these severe multisystem disorders [4]. The accurate diagnosis of a rare disease takes an average 4.8 years and involves several specialists, who may be distributed around the world [5]. About 80% of rare diseases are of genetic origin, which can be beneficial for these patients and their families, as there may be some treatment options, as well as genetic counseling and prenatal diagnosis [5]. Patients with rare diseases and their families are a major challenge for the current health system, as they carry an immense burden and psychological distress [6].

It has been a challenge to develop diagnostics, interventions and therapeutics for these patients, as each disease may affect only a few to several hundred individuals. The molecular etiology of >3,800 rare diseases has been determined, although for most of these diseases, the molecular basis and underlying cause of these diseases have not yet been determined. The total number of rare disease-causing genes is estimated of 7,000 to 15,000, based on the mutation rate of the human genome and the number of essential genes [7]. Recent advances

in Next Generation Sequencing (NGS) technologies have become an invaluable tool to rapidly analyze known disease genes and to search for pathogenic variants in new candidate genes, which is contributing for a great advance in clinical practice, as well as in basic research field. The number of rare disease-causing gene discoveries has been increasing in the last years and the genetics of many idiopathic diseases have been elucidated [8]. In this article, we describe how NGS has changed the diagnostic of rare diseases, as well as its limitations and challenges, illustrating the Portuguese reality and detail the testing technologies that are used.

### Rare Diseases in Portugal

It is estimated that in Portugal there are 600,000 to 800,000 patients with rare diseases, given the lack of an adequate single registration for these diseases, as well as the reduced number and epidemiologic studies carried out to date. However, the publication of the Integrated Strategy for Rare Diseases 2015-2020 [9] may improve the knowledge of rare diseases in our country.

This strategy is based on inter-ministerial, inter-sectoral and inter-institutional cooperation that makes use of medical, social, scientific and technological resources, focusing on the following aspects:

- A. The coordination of national health care;
- B. Access to early diagnosis;
- C. Access to treatment;
- D. The investigation (scientific) and
- E. Social inclusion and citizenship. Rare diseases don't just affect the patients, but they also have a huge impact on families, friends, caregivers and in society.

Associations of patients with rare diseases, such as "Raríssimas", were essential to support the difficulties experienced by these families. This national association of mental and rare disabilities has strategic objectives such as: promote dissemination, information and public awareness on rare diseases nationally and internationally; promote the integrated management of patients with rare disease; develop and participate in transnational research projects in the scope of rare diseases and establish national and international partnerships. Furthermore, it is responsible for the existence of the integrated resource center for rare diseases, the well-known "Casa dos Marcos", which brings together the social, health, research and training valences, among others, which aroused an international political curiosity resulting in numerous institutional and diplomatic visits. Moreover, this national association established a cooperation protocol with our Center, whose global objective was to establish scientific and pedagogical cooperation, optimizing resources between both institutions [10].

In the last years, a research project funded by European Community (DESVENDAR "DESCobrir, VENcer as Doenças rARas" - NORTE-01-0246-FEDER-000014), also allowed the implementation of Next Generation Sequencing (NGS) technology in our Center, which is giving major advances in genetic diagnosis of

rare metabolic diseases, from prenatal level to neonatal, pediatric and adult level, as it has the capacity to generate a huge amount of data in a short time and at an affordable cost, providing practical information to health professionals and patients within the scope of personalized medicine. The translation of NGS knowledge to the laboratorial routine is facilitating a better understanding of the etiology of the rare diseases and is allowing a more quickly detection of similar cases, reducing the costs in the national health system due to prolonged and unspecific palliative treatments. As a referral laboratory for the diagnosis and investigation of rare metabolic diseases in children and adults we offer this study for national public and private hospitals, mainly collaborating with pediatricians, neurologists and neuropediatricians. This translational research is enabling our country with new technological approaches, strengthening our Center as a reference laboratory for the study of these pathologies [11,12].

### The Impact of Next Generation Sequencing in Rare Diseases

A major turning point in rare diseases research came in 2009 with the publication of the first manuscript that described the use of NGS for the identification of rare diseases genes [13]. In the last 12 years, NGS technologies have enabled testing of multiple disease genes simultaneously, ranging from targeted gene panels to exome sequencing and genome sequencing [14]. In this section we will report the approach used in our Center for diagnosis of rare diseases, focusing particularly on inborn errors of metabolism and mitochondrial diseases.

#### Patients selection

NGS for rare diseases begins with selecting the subjects who will be sequenced. We study patients suspected of mitochondrial/metabolic diseases, based on their clinical, biochemical and/or neuroimaging findings but with no definitive molecular diagnosis. In agreement with the Declaration of Helsinki, informed consent for genetic studies was obtained from all investigated subjects, or their relatives.

#### Sequencing technology

**Gene panels:** Targeted gene panels, include nuclear genes selected to be already associated with inborn errors of metabolism, mitochondrial, lysosomal and other rare diseases, were designed using the software "Suredesign" from Agilent Technologies. Sequencing was performed on the Illumina MiSeq platform. and SureSelect QXT/XTHS kits from Agilent Technologies were used to capture the genes of interest.

**Exome sequencing:** Exome sequencing were performed using SureSelect Custom Constitutional Panel with 17Mb. Sequencing was performed on the Illumina MiSeq platform and SureSelect QXT/XTHS kits from Agilent Technologies were used to capture the genes of interest. For Whole Exome Sequencing (WES), a Human All Exon V7 Kit (Agilent) for exome capture on genomic DNA from the patients, who remain with no molecular diagnosis, and their healthy parents (TRIO approach) will be performed in a NextSeq Illumina sequencer.

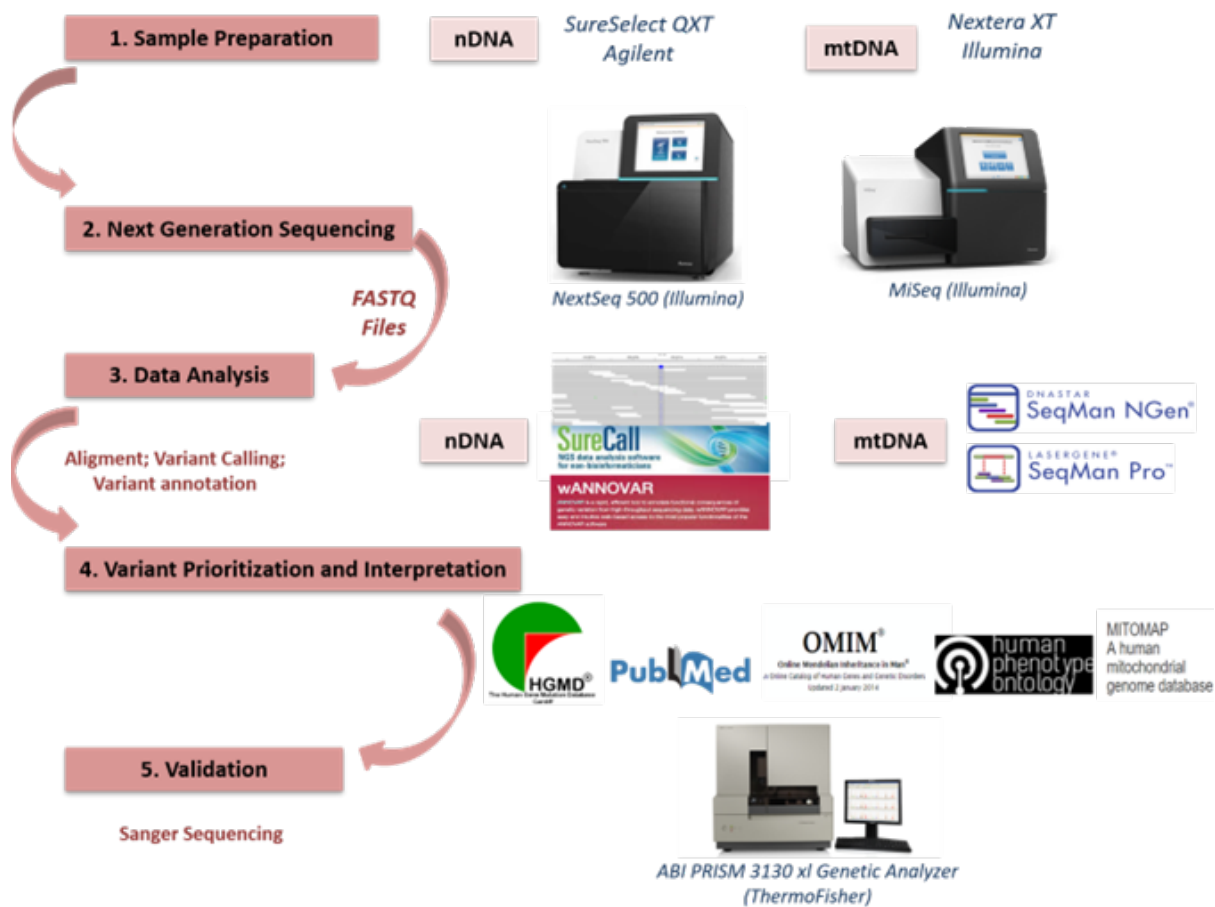
**RNA sequencing:** Transcriptome analysis using RNA Sequencing technology (RNA-seq) is used to provide functional validation of the VUS identified (all rare synonymous, splice region, untranslated region (UTR), and deep intronic variants), as well as to provide evidence of a second variant to complement a rare heterozygous mutation already identified by WES. For RNA-seq a TruSeq Stranded kit (Illumina) after poly(A) capture will be used in a NextSeq Illumina sequencer.

**Whole human mitochondrial genome:** Long-range PCR was used to amplify the entire human mtDNA in a single amplicon, using back-to-back primers [15]. Sequencing of indexed paired-end DNA libraries was performed on the Illumina MiSeq platform, according to the manufacturer instructions.

**Confirmatory sanger sequencing analysis:** Sanger sequencing, using an ABI 3130XL DNA Analyzer and BigDye Terminator Cycle Sequencing Version 3.1 (Applied Biosystems, Foster City, CA), was used to confirm all detected variants that had the potential to be disease-causing, as well as to perform co-segregation studies.

**Bioinformatics pipelines**

The available commercial programs: Surecall (Agilent Technologies) and wAnnovar (wannovar.wglab.org/), were used to perform variant calling and annotation. The Variants of Unknown Significance (VUS) were classified according to the American College of Medical Genetics and Genomics Guidelines (ACMG).



**Figure 1:** Schematic representation of NGS workflow for rare diseases diagnosis.

The filtering of variants taking into account: i) the type of mutation (missense, frame-shift, stop-gain or stop-loss, and splice-site variants), ii) in silico predictors (SIFT, PolyPhen-2, MutationTaster) [16-18] and presence in databases (dbSNP, 1000 Genomes, HGMD professional, ClinVar, ExAC, OMIM, gnomAD), and iii) the population frequency (variants with a minor allele frequency (MAF) >1% in the 1000 Genomes Project (<http://www.1000genomes.org>) and Exome Variant Server databases (<http://evs.gs.washington.edu>) were filtered out). For whole human mitochondrial genome FASTQ files were aligned to the mtDNA reference sequence with SeqMan Ngen (DNASTAR). The commercial program, SeqMan Pro

(DNASTAR), was used to perform variant calling and annotation. The filtering of variants taking into account: i) the type of mutation, ii) the population frequency, iii) heteroplasmy >5%, and iv) in silico predictors and presence in databases (Mitomap, MitImpact2, HmtVar) (Figure 1).

**Limitations and Challenges of NGS Technologies**

The search for pathogenic mutations in rare human genetic diseases during the past decade has involved huge efforts to sequence coding regions, or the entire genome, using NGS technologies.

The approximate current diagnostic rate of genetic variants in rare diseases is <50% using NGS, remaining many of them with unknown cause. Many factors contribute to the suboptimal translation of NGS technology into a rare disease diagnosis [6,19]: i) the coverage and accuracy rates <100%, results in missing variants and false positive results, ii) the filtering and interpretation of data, as more than one candidate variant is usually found. Indeed, the greater the number of kilobases sequenced, the greater the chance of finding more candidates. There are also ethical challenges related to incidental findings, adding another layer of complexity or clinical management of genetic variants [20]. To address this, the aim of genetic testing in rare disease diagnosis needs to be better defined. Although, the biggest challenge of NGS is related with functional characterization of a novel candidate variant, which take months or even years.

CRISPR/cas9 gene editing technology and the development of induced pluripotent stem cells reprogramming techniques have recently contributed to important advances in the field of functional biology, however the time scales involved are still large. The reason why it is so important to overcome this bottleneck it is because then will be possible to look at therapies that can be used to treat the patients, this may involve relieving the symptoms or ideally, preventing them from occurring in the first place [21].

## Conclusion

In this paper, we reviewed the impact of NGS on the study of rare disorders in Portugal. NGS has revolutionized the research and diagnosis of these diseases worldwide. Approximately 25% - 30% of the undiagnosed genetic cases arrive at a diagnosis, many based on previously discovered genes [12]. As more genes are discovered, a rise in the diagnostic rate is expected. Applying NGS strategy for the screening of rare diseases patients, compared to conventional approaches, represents a good balance between cost, time and amount of data for analysis and provides higher diagnostic yield. Thus, the simultaneous analysis of a large number of genes increases the discovery of new and unexpected genotype/phenotype associations, supporting its implementation in clinical practice. Although, to accelerate the NGS understanding larger databases of genomic information linked with clinical information are needed, such as the "The 100,000 Genomes Project", from UK and the "Towards access to at least 1 million Genomes", from European Union, which are enabling the access to 100,000 and one million genomes, respectively. In the near future, NGS will be an essential component in precision medicine regarding its high throughput and relatively low and further decreasing cost and the importance of individual genomic information for personalized strategies, will benefit patient care in the long term.

## References

1. Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A (2015) OMIM.org: Online Mendelian Inheritance in Man (OMIM), an online catalog of human genes and genetic disorders. *Nucleic Acids Res* 43(Database Issue): D789-D798.

2. Ayme S, Urbero B, Oziel D, Lecouturier E, Biscarat AC (1988) Information on rare diseases: The orphanet project. *Rev Med Interne* 19: 376S-377S.
3. Samuels ME (2010) Saturation of the human phenome. *Curr Genom* 11(7): 482-499.
4. Dodge JA, Chigladze T, Grossman Z, Ramos F, et al. (2011) The importance of rare diseases: From the gene to society. *Arch Dis Child* 96(9): 791-792.
5. Liu Z, Zhu L, Roberts R, Tong W (2019) Toward clinical implementation of next-generation sequencing-based genetic testing in rare diseases: Where are we? *Trends Genet* 35(11): 852-867.
6. Wright CF, Fitz Patrick DR, Firth HV (2018) Paediatric genomics: Diagnosing rare disease in children. *Nat Rev Genet* 19(5): 253-268.
7. Cooper DN, Chen JM, Ball EV, Howells K, Mort M, et al. (2010) Genes, mutations, and human inherited disease at the dawn of the age of personalized genomics. *Hum Mutat* 31(6): 631-655.
8. Sangmoon L, Murim C (2016) Ultra-rare disease and genomics-driven precision medicine. *Genomics Inform* 14(2): 42-45.
9. Ministries of Health, Education and Science and Solidarity, Employment and Social Security (2015) Approves the integrated strategy for rare diseases, Republic Dairy 41(2<sup>nd</sup> Suppl): 5190-5110.
10. Isidro G (2016) Rare diseases in Portugal. *Epidemiological Bulletin Observations* 7: 1-2.
11. Nogueira C (2019) UNLOCK discover, beat rare diseases. *Epidemiological Bulletin Observations* 11: 73-76.
12. Nogueira C, Silva L, Pereira C, Vieira L, Teles EL, et al. (2019) Targeted next generation sequencing identifies novel pathogenic variants and provides molecular diagnoses in a cohort of pediatric and adult patients with unexplained mitochondrial dysfunction. *Mitochondrion* 47: 309-317.
13. B Ng S, Turner EH, Robertson PD, Flygare SD, Bigham AW, et al. (2009) Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 461(7261): 272-276.
14. Bick D, Jones M, Taylor SL, Taft RJ, Belmont J (2019) Case for genome sequencing in infants and children with rare, undiagnosed or genetic diseases. *J Med Genet* 56(12):783-791.
15. Zhang W, Cui H, Wong LJC (2012) Comprehensive one-step molecular analyses of mitochondrial genome by massively parallel sequencing. *Clin. Chem* 58(9): 1322-1331.
16. Kumar P, Henikoff S, Ng PC (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 4(7): 1073-1081.
17. Adzhubei I, Jordan DM, Sunyaev SR (2013) Predicting functional effect of human missense mutations using polyphen-2. *Curr Protoc Hum*, Chapter 7, Unit 7.20
18. Schwarz JM, Rödelsperger C, Schuelke M, Seelow D (2010) Mutation taster evaluates disease-causing potential of sequence alterations. *Nat Methods* 7(8): 575-576.
19. Mitsuhashi S, Matsumoto N (2020) Long-read sequencing for rare human genetic diseases. *J Hum Genet* 65(1): 11-19.
20. Fernandez-Marmiesse A, Gouveia A, Couce ML (2018) NGS Technologies as a turning point in rare disease research, diagnosis and treatment. *Curr Med Chem* 25(3): 404-432.
21. Bacchelli C, Williams HJ (2016) Opportunities and technical challenges in next-generation sequencing for diagnosis of rare pediatric diseases. *Expert Rev Mol Diagn* 16(10): 1073-1082.

For possible submissions Click below:

Submit Article