International Journal of Advanced Scientific Research
www.allscientificjournal.com
ISSN: 2456-0421
Received: 29-08-2021; Accepted: 14-09-2021; Published: 29-09-2021
Volume 6, Issue 5, 2021, Page No. 7-13

Whole-genome sequencing and in-silico approaches to identify the genetic basis of rare diseases

Sudheer Menon
Department of Bioinformatics, Bharathiar University, Tamil Nadu, India

Abstract
To a greater or lesser extent, genetics plays a role in all diseases. Variations in our DNA and differences in how that DNA functions, coupled with our environment, lead to disease. It investigates the genetic basis of human disease, including single-gene disorders, chromosomal imbalances, epigenetics, cancer, and complex disorders, as well as how our understanding and technological advances can be applied to the provision of appropriate patient diagnosis, management, and therapy. Mutations that alter splicing play a significant role in rare diseases but are often overlooked by diagnostic sequencing pipelines. Increasing our understanding of pathogenic splicing variants will lead to better diagnostic yields for patients and their families, and will improve treatment and care strategies. Although recent advances in sequencing technologies, predictive modeling, and understanding of splicing mechanisms have made it possible to detect and interpret splice affecting variants more accurately, several limitations still prevent their routine ascertainment in diagnostic testing.

Keywords: next-generation sequencing, rare diseases, artificial intelligence, genomics

Introduction

Genomic medicine is based on the idea that Mendelian traits have genetic changes that are both preventative and therapeutic treatments that rarely occur (Baird *et al*., 1988) [9] about 8000 distinct disease traits have been identified to date (Ayme *et al*., 1998). The Rare single nucleotide variants, copy number variants that have been many Mendelian conditions, which leads to the single gene (Harel *et al*.,2018) [26]. The study of Mendelian conditions had a substantial impact on our understanding of genomic and molecular mechanisms of rare human diseases, and many discoveries which have helped us to understand common conditions. It also highlights the amount of additional research that needs to be done, since 20% of proteins coding for human disease genes have been definitively associated with a human phenotype to date. Research from the Human Genome Research Institute (NHGRI) and National Heart, Lung, and Blood Institute (NHLBI)-funded Centers for Mendelian Genomics has revealed a steady trajectory of 263 new discoveries annually. Wider availability of ES has enabled elucidation of the molecular and genomic architecture of Mendelian conditions (Posey *et al*., 2019) [36]. With the decline in next-generation sequencing costs, genome sequencing (GS) is now widely accessible across all fields of research and medicine, as well as both research and clinical applications (wetter strand; 2019)

Rare Diseases

There are approximately 7000 rare diseases, many of which are life-threatening or chronically debilitating (Amberger *et al*., 2015). It takes an average of 4.8 years to diagnose a rare disease accurately, and more than seven physicians or specialists may be involved. Diagnosis is often a long and tedious process with rare diseases that burden the patient tremendously stresses patients and their families, and challenges the current healthcare system genetic diagnosis may benefit rare disease patients and their families. There is no direct link between genetic diagnosis and treatment options, so physicians will continue to treat symptoms according to likely prognosis, albeit more informed (Wright *et al*., 2018) [49].

Sequencing-based genetic diagnosis: Challenges and Opportunities

New genomics technologies, such as next-generation sequencing (NGS), are extensively applied in research and offer great potential in clinical settings. NGS-based genetic testing has improved the invention of genetic variants in rare diseases, but there**'s** still a translational gap between NGS-based genetic testing and clinical implementation. Many factors contribute to the suboptimal translation of NGS technology into the diagnosis of rare diseases. An NGS-based genetic test's acceptance and uptake are driven by the demonstration of a clear benefit for patients, driving physician decision-making (Shendure *et al*., 2017) [44] NGS has several challenges to overcome before it can deliver its full potential to patients, clinicians, and society. Genetic testing based on NGS can produce accurate and reproducible results that can support clinical decision-making for the diagnosis of rare diseases. A better definition of genetic testing for the diagnosis of rare diseases is necessary to address this matter. In addition, there should be a seamless and harmonized relationship among genetic testing service providers, physicians, and patients.
Figure 1.

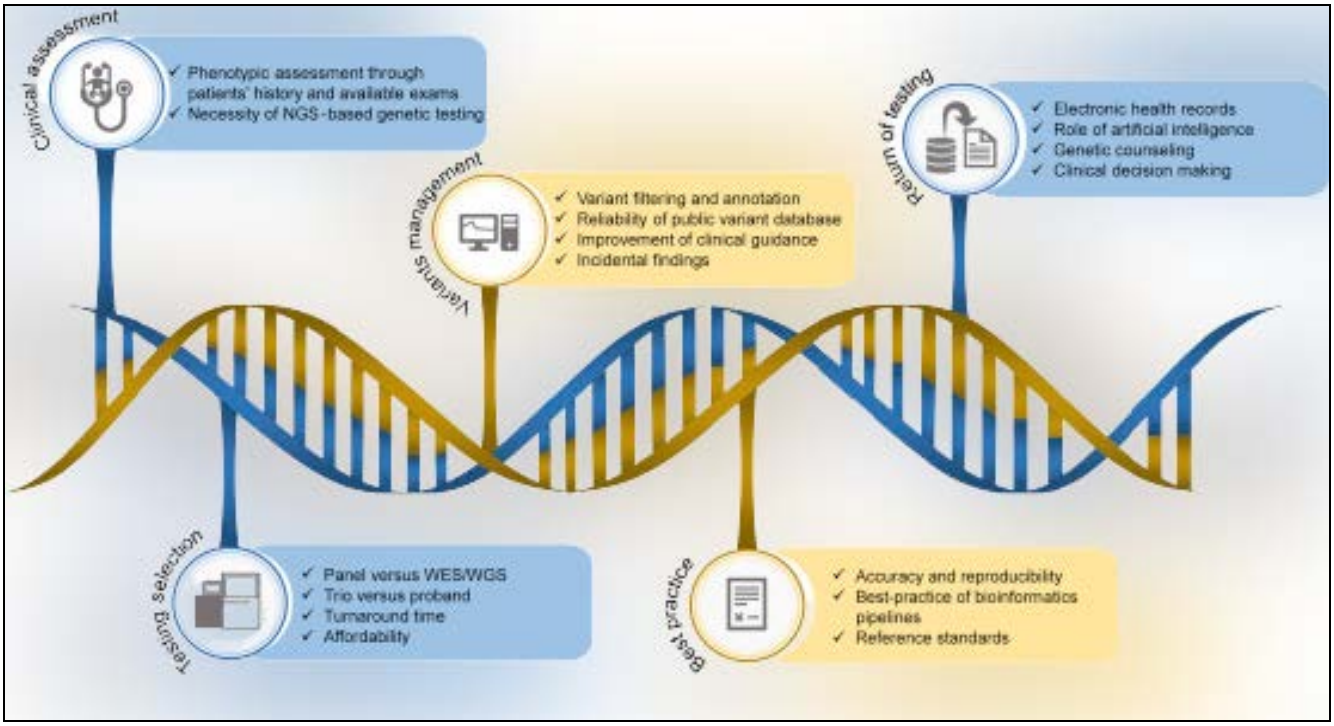

Fig 1

There are so many genetic disorders, so it is impossible to include more than a few examples in this review. These include Genetics Home Reference (http://ghr.nlm.nih.gov/edu/>), Gene Reviews (The 'Education' section at the National Human Genome Research Institute (https://www.genome.gov/education/), and Online Mendelian Inheritance in Man (http://www.omim.org/) are particularly useful on this topic.).

Mutations refer to any heritable change in the DNA sequence, where heritable refers to somatic cell division. A change in DNA may have no consequence but can sometimes cause observable differences within the individual. Furthermore, this terminology carries negative implications and brings to mind the 'mutants in science. Variants may be further classified as benign or pathogenic.

Table 1: Research Agency for Cancer Variant Classification

Variant classification	Description	Precautionary recommendations	Investigation required
5	Surely pathogenic	Complete high risk precaution as current guidance.	Genetic testing for all family members
4	Relative pathogenic	Complete high risk precaution as current guidance.	Genetic testing for all family members
3	Uncertain	Precaution base on family history and other known factors.	No genetic testing required.
2	Relative non pathogenic	Cure as if “no mutation” detected	No genetic testing required.
1	Surely not pathogenic	Cure as if “no mutation” detected	No genetic testing required.

Although increasing numbers of human DNA variants are identified. We are still unsure of the effect; these are termed variants of uncertain significance or VUS (Table 1). There are many different DNA variations in a population, referred to as alleles, and each allele represents a particular variation. Minor allele frequency has a 1% variant, so-called polymorphism.

Variants affecting only one nucleotide

Mutations affecting only one nucleotide are termed single nucleotide variants or single nucleotide polymorphisms (Figure 2) depending upon the minor or allele frequency. The human genome is estimated to contain at least 11 million SNPs.

Additions and deletions

Deletions and insertions of less than 1000 bp are also relatively common in the human genome, with the smallest indels being the most frequent.

Structural variations

These affect DNA segments longer than 1000 base pairs (1 kb). They include translocations, inversions, large deletions, and copy number variants. Copy number variations are segments of our genome with sizes ranging from 1000 to millions of base pairs and copy numbers ranging from zero to several copies in healthy individuals (Figure 2). According to the analysis of many human genomes, approximately 12% of the human genome sequence contains CNVs. CNVs may include entire genes in their most extensive forms.

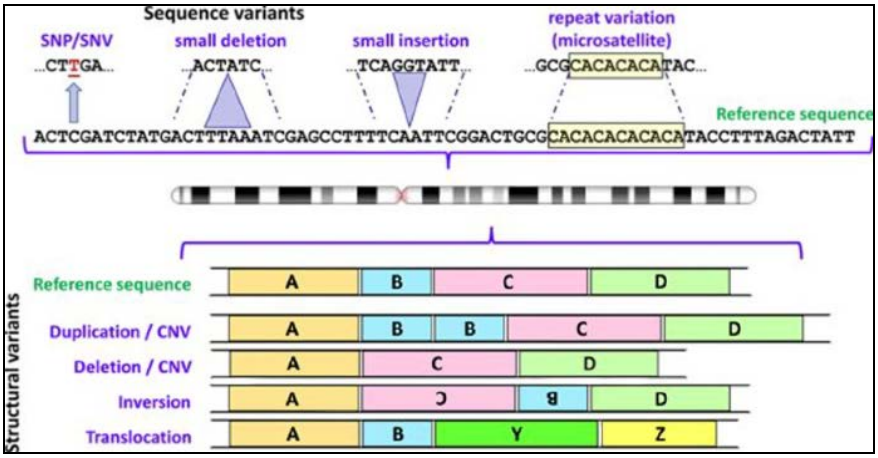

Fig 2: Different types of variations in human genomics.

Is splicing mutation-caused disease common?

The ATM and NF1 gene analysis estimate that about half of all disease-causing variants affect splicing. Scientists detected 44 changes in 52 patients with neurofibromatosis using cDNA genotyping. The majority of them disrupted splicing, including splicing variants, near-splice variants, and missense variants. The researchers examined 62 changes in ATM associated with ataxia-telangiectasia and found that half affected splicing. Both estimates rely on relatively small numbers of variants affecting single genes from cohorts of patients with certain disorders, which causes some bias in ascertainment and does not give a reliable estimate of the percentage of variants affecting splicing across the human genome (Teraoka *et al*., 1999; Ars *et al*., 2000) [46, 3].

A massively parallel approach to testing the effects of splicing can accurately quantify which variants disrupt splicing. Several examples have been published in recent years using saturation mutagenesis to examine all possible single nucleotide changes, generally focusing on short exons in disease-associated genes (Mueller *et al*., 2015) [35]. The group identified by 23% of identical variations decreased the inclusion of exon 7 to 70% of the mutant allele level (Soucek *et al*., 2019).

Randomly mutated minigenes were used to test alternative splicing of exon 11 in the gene RON (Braun *et al*., 2018) [10]. In the minigenes, which covered exons 10-12, including complete introns, there was an average of 3.6 variants in 97% of positions. To measure the impact of individual variants, a linear regression method was employed. In addition, 45% of constructs showed a change in inclusion level >10%, with a regression predicting over 90% of all positions in exon 11, and at least 50% of flanking intronic and exonic positions had at least one variant that affected isoform usage >5%. Invasive tumors are associated with transcripts lacking exon 11 (Collesi *et al*., 1996) [17]. Around 70% of all variants increased or decreased exon inclusion in some way. Around 65% of SNVs affected splicing, while around 72% of double mutations affected it (Ke *et al*., 2018).The multiplexed functional assay of splicing using sort-seq is a reporter system based on constructs of 3 exons and two introns, where skipping of the central test exon reconstitutes fluorescence. DNA sequencing is used to identify constructs, with normalized read counts being used to generating an inclusion index for that construct (Cheung *et al*., 2019) [12].

Predictions based on silicon

Many different tools are available to predict the impact of variants on pre-mRNA splicing. Still, there is no consensus on using them optimally, limiting their use in research and clinical settings. Under the American college of medical genetics and genomics guidelines, predictions from in silico tools may be used (Richards *et al*., 2015) [42]. In silico splicing, prediction tools must meet several criteria for integration into a clinical setting. First, the tool must be able to predict the functional impact of variants, which we will discuss below. Secondly, the tool should be easy to implement and the output should be easy to interpret. In addition, a single integrated platform for splicing abnormalities and other types of variation is likely to streamline analysis. Several tools for splicing prediction into a dedicated in silico splicing prediction window, including Max EntScan (Yeo and Burge, 2004) [50] and NNSPLICE (Reese *et al*., 1997) [41] for predicting disruption to, or gain of, canonical splice sites, and ESEFinder (Cartegni *et al*., 2003) [11]. Alamut enables easy comparison of wild-type and variant sequences and can generate a report on splice sites or splicing enhancers (Wai *et al*., 2020) [47].

Table 2

	Minisattellites	**Microsatellites**
	Minisattellites	**Microsatellites**
Number	Approximately 1500	Approximately 500000
Locations within our genome	Mainly near the ends of chromosomes (telomeres)	Scattered throughout the length of all chromosomes
Element repeat length1	Approximately 10 to >100 bp	(1^2) 2 to approximately 6 bp
Number of repeat units within the array	Usually from approximately 60 to >1000	Usually ~6 to ~14
	DNA fingerprinting	DNA profiling; genetic linkage studies
	Variable number tandem repeats (VNTR)	VNTR, short tandem repeats (STR), simple sequence repeats (SSR)

RNA-seq Analysis of Splicing

The major studies by using transcriptome for using different diseases and major group RNA-seq are rare in muscle disorders. In this study, muscle tissue mRNA was sequenced from 63 patients with suspected monogenic muscle disorders and compared to 184 control samples from the Genotype-Tissue Expression project (Cummings *et al*., 2017) [19]. There are different tissue types which have been developed for gene expression analysis like panel gene expression which have been introduced by different scientist (Gonorazky *et al*., 2019) [20]. With the GTEx data, it allowed comparisons of gene expression and exon splicing patterns in different tissues. Users can search for a gene, or gene panel of interest and compare across tissues to establish the most appropriate tissue to capture the RNA profile of interest (Aicher *et al*., 2020) [1].

Materials certified to be biological

The best way to address systematic errors is to use a reference standard approach to evaluate diagnostic performance and reproducibility, and to establish best practices for variant calling. (Hardwick *et al*., 2017) [21]. Several professional organizations have recommended calibrating NGS measurements routinely with reference standards.

Development of reference materials, including well-characterized cell-based genetic materials and synthetic spike-in controls (li *et al*.,2018) Several laboratories now use Genome in a Bottle Consortium (GIAB) reference samples (e.g., NA12878) as process controls to estimate detection limits, ensure repeatability and reproducibility, and calibrate their NGS workflow (zook *et al*.,2016) [27, 52].

1. The development of pan-ethnic reference materials. Genetic variations across global populations vary, resulting in regions with low complexity, low GC content, or repetitive sequence differences. Ethnic population-based reference material may not be adequate to cover genetic variants associated with rare diseases that have a wide geographical distribution and diverse epidemiology(consortium *et al*.,2015) [18].
2. Reference materials spiked to take account of matrix effects. The variant distribution within formalin-fixed, paraffin-embedded, or liquid biopsy-based samples from rare disease patients may be very different from that seen in normal tissues (Robbe *et al*.,2018) [43].
3. Rare disease-specific reference materials. Healthy donor-based reference material may only cover a tiny proportion of the causative variants of rare diseases. A solution that may be feasible is to refine the development of reference material using human induced pluripotent stem cells that harbor pathologic variants. Genetic Testing Reference Materials Coordination Program (GET-RM) efforts aim to coordinate a self-sustaining community effort to improve the availability of suitable and characterized reference materials for inherited diseases and pharmacogenomics (Kalman *et al*., 2016) [27].
4. Reference material development through genome editing. Recent advances in bioengineering, including genome editing, can be used to introduce specific genetic variants into cells. It will lead to synthetic reference material that includes causal variants for rare diseases (Zhang *et al*., 2017) [51].

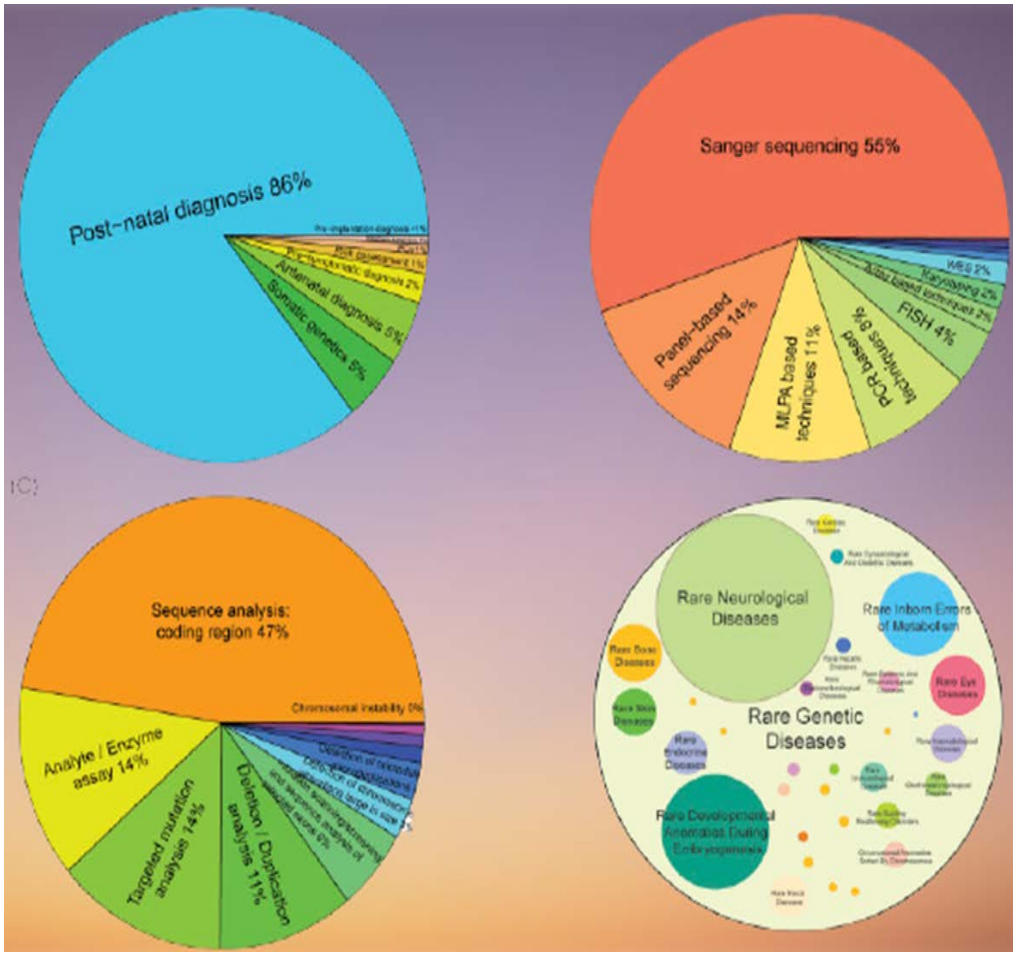

Fig 3

Conclusion

Splicing errors have severe consequences on gene function that lead to disease. By developing a multiplexed functional splicing assay using Sort-seq, we explore how genetic variation affects exon recognition. Exome Aggregation Consortium variants found within or adjacent to 2,198 human exons accounted for 3.8% (1,050) of large-effect splice disrupting variants. Importantly, we find that 83% of SDVs are situated outside of canonical splice sites, are distributed evenly between distinct exonic and intronic regions, and are difficult to predict a priori. It appears that even outside the context of disease, extant, rare genetic variants can have significant functional effects on splicing and allows their empirical assessment at scale. With the advances and improvements discussed in this review, we are closer than ever to identifying and determining the effects of variants on splicing. The discovery of these variants in the clinic could significantly impact diagnostic yields across rare diseases. The rise of splice-modifying therapies offers hope to those suffering from rare disorders. However, there is still a great deal of work to do, and the potential reward for the effort would be invaluable.

Abbreviations

Multiplexed functional assay of splicing using Sort-seq (MFASS).
Exome Aggregation Consortium (ExAC)
Splice-Disrupting variants (SDVs).
American College of Medical Genetics and Genomics (ACMG)
Genotype-Tissue Expression (GTEx)
PAGE panel gene expression

Reference

1. Aicher JK, Jewell P, Vaquero-Garcia J, Barash Y, Bhoj EJ. Mapping RNA splicing variations in clinically accessible and nonaccessible tissues to facilitate Mendelian disease diagnosis using RNA-seq. Genetics in Medicine,2020:22(7):1181-1190.
2. Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A. OMIM. org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. Nucleic acids research, 2015:43(D1), D789-D798.
3. Ars E, Serra E, García J, Kruyer H, Gaona A, Lázaro C, Estivill X. Mutations affecting mRNA splicing are the most common molecular defects in patients with neurofibromatosis type 1. Human molecular genetics,2000:9(2):237-247
4. Aymé S, Urbero B, Oziel D, Lecouturier E, Biscarat AC. Information on rare diseases: the Orphanet project. La Revue de médecine interne,1998:19:376S-377S.
5. Sudheer Menon. "Preparation and computational analysis of Bisulphite sequencing in Germfree Mice" International Journal for Science and Advance Research In Technology,2020:6(9):557-565.
6. Sudheer Menon, Shanmughavel Piramanayakam and Gopal Agarwal. "Computational identification of promoter regions in prokaryotes and Eukaryotes" EPRA International Journal of Agriculture and Rural Economic Research (ARER),2021:9(7):21-28.
7. Sudheer Menon. "Bioinformatics approaches to understand gene looping in human genome" EPRA International Journal of Research & Development (IJRD),2021:6(7):170-173.
8. Sudheer Menon. "Insilico analysis of terpenoids in Saccharomyces Cerevisiae"international Journal of Engineering Applied Sciences and Technology, 2021, 6(1). ISSN No. 2455-2143, PP(43-52).
9. Baird PA, Anderson TW, Newcombe HB, Lowry RB. Genetic disorders in children and young adults: a population study. American journal of human genetics,1988:42(5):677.
10. Braun S, Enculescu M, Setty ST, Cortés-López M, de Almeida BP, Sutandy FR, Zarnack K. Decoding a cancer-relevant splicing decision in the RON proto-oncogene using high-throughput mutagenesis. Nature communications,2018:9(1):1-18.
11. Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR. ESE finder: a web resource to identify exonic splicing enhancers. Nucleic acids research,2003:31(13):3568-3571.
12. Cheung R, Insigne KD, Yao D, Burghard CP, Wang J *et al*. A multiplexed assay for exon recognition reveals that an unappreciated fraction of rare genetic variants cause large-effect splicing disruptions. Molecular cell,2019:73(1):183-194.
13. Sudheer Menon. "Computational analysis of Histone modification and TFBs that mediates gene looping" Bioinformatics, Pharmaceutical, and Chemical Sciences (RJLBPCS),2021:7(3):53-70.
14. Sudheer Menon Shanmughavel piramanayakam, Gopal Prasad Agarwal "FPMD-Fungal promoter motif database: A database for the Promoter motifs regions in fungal genomes" EPRA International Journal of Multidisciplinary research,2021:7(7):620-623.
15. Sudheer Menon, Shanmughavel Piramanayakam, Gopal Agarwal. Computational Identification of promoter regions in fungal genomes, International Journal of Advance Research, Ideas and Innovations in Technology,2021:7(4):908-914.
16. Sudheer Menon, Vincent Chi Hang Lui, Paul Kwong Hang Tam. Bioinformatics methods for identifying hirschsprung disease genes, International Journal for Research in Applied Science & Engineering Technology (IJRASET),2021:9(VII):2974-2978.
17. Collesi C, Santoro MM, Gaudino G, Comoglio PM. A splicing variant of the RON transcript induces constitutive tyrosine kinase activity and an invasive phenotype. Molecular and cellular biology, 1996:16(10):5518-5526.
18. Consortium GP, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM. A global reference for human genetic variation. Nature, 2015:526(7571):68-74.
19. Cummings BB, Marshall JL, Tukiainen T, Lek M, Donkervoort S, Foley AR. Genotype-Tissue Expression Consortium. Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. Science translational medicine, 2017, 9(386).
20. Gonorazky HD, Naumenko S, Ramani AK, Nelakuditi V, Mashouri P, Wang P *et al*. Expanding the boundaries of RNA sequencing as a diagnostic tool for rare Mendelian disease. The American Journal of Human Genetics,2019:104(3):466-483.
21. Hardwick SA, Deveson IW, Mercer TR. Reference standards for next-generation sequencing. Nature Reviews Genetics,2017:18(8):473-484.

22. Sudheer Menon, Bioinformatics approaches to understand the role of African genetic diversity in disease, International Journal Of Multidisciplinary Research In Science, Engineering and Technology (IJMRSET),2021:4(8):1707-1713.
23. Sudheer Menon. Comparison of High-Throughput Next generation sequencing data processing pipelines, International Research Journal of Modernization in Engineering Technology and Science (IRJMETS),2021:3(8):125-136.
24. Sudheer Menon. Evolutionary analysis of SARS-CoV-2 genome and protein insights the origin of the virus, Wuhan, International Journal of Creative Research Thoughts (IJCRT),2021:9(8):b696-b704.
25. Sudheer Menon, Vincent Chi Hang Lui, Paul Kwong Hang Tam. A step-by-step work flow of Single Cell RNA sequencing data analysis, International Journal for Scientific Research and Development (IJSRD), 2021:9(6):1-13.
26. Harel T, Lupski JR. Genomic disorders 20 years on—mechanisms for clinical manifestations. Clinical genetics,2018:93(3):439-449.
27. Kalman LV, Datta V, Williams M, Zook JM, Salit M. L *et al*. Development and characterization of reference materials for genetic testing: focus on public partnerships. Annals of Laboratory Medicine, 2016:36(6):513-520.
28. Ke S, Anquetil V, Zamalloa JR, Maity A, Yang A, Arias MA, Chasin LA. Saturation mutagenesis reveals manifold determinants of exon definition. Genome research,2018:28(1):11-24.
29. Sudheer Menon. Computational characterization of Transcription End sites in Human Genome, International Journal of All Research Education and Scientific Methods (IJRESM),2021:9(8):1043-1048.
30. Sudheer Sivasankaran Menon, Shanmughavel Piramanayakam. Insilico prediction of gyr A and gyr B in Escherichia coli insights the DNA-Protein interaction in prokaryotes, International Journal of Multidisciplinary Research and Growth Evaluation, (IJMRD),2021:2(4):709-714.
31. Sudheer Menon, Vincent Chi Hang Lui and Paul Kwong Hang Tam Bioinformatics tools and methods to analyze single cell RNA sequencing data, International Journal of Innovative Science and Research Technology, (IJISRT),2021:6(8):282-288.
32. Sudheer Menon. Computational genome analysis for identifying Biliary Atresia genes, International Journal of Biotechnology and Microbiology, (IJBM), 2021:3(2):29-33.
33. Li H, Bloom JM, Farjoun Y, Fleharty M, Gauthier L, Neale B *et al*. A synthetic-diploid benchmark for accurate variant-calling evaluation. Nature methods, 2018:15(8):595-597.
34. Lionel AC, Costain G, Monfared N, Walker S, Reuter MS, Hosseini SM, Marshall CR. Improved diagnostic yield compared with targeted gene sequencing panels suggests a role for whole-genome sequencing as a first-tier genetic test. Genetics in Medicine,2018:20(4):435-443.
35. Mueller WF, Larsen LS, Garibaldi A, Hatfield GW, Hertel KJ. The silent sway of splicing by synonymous substitutions. Journal of Biological Chemistry, 2015:290(46):27700-27711.
36. Posey JE, O'Donnell-Luria AH, Chong JX, Harel T, Jhangiani SN, Akdemir ZHC *et al*. Insights into genetics, human biology and disease gleaned from family based genomic studies. Genetics in Medicine,2019:21(4):798-812.
37. Sudheer Menon. Recent Insilco advancements in genome analysis and characteristics of SARS-Cov2. International Journal of Biology Research, (IJBR),2021:6(3):50-54.
38. Sudheer Menon. Bioinformatics methods for identifying Human disease genes, International Journal of Biology Sciences, (IJBR),2021:3(2):1-5.
39. Sudheer Menon. SARS-CoV-2 Genome structure and protein interaction map, insights to drug discovery, International Journal of Recent Scientific Research, (IJRSR),2021:12(8):42659-42665.
40. Sudheer Menon. Insilico Insights to Mutational and Evolutionary aspects of SARS-Cov2, International Journal of Multidisciplinary Research and Development, (IJMRD),2021:8(8):167-172.
41. Reese MG, Eeckman FH, Kulp D, Haussler D. Improved splice site detection in Genie. Journal of computational biology,1997:4(3):311-323.
42. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J *et al*. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genetics in medicine, 2015:17(5):405-423.
43. Robbe P, Popitsch N, Knight SJ, Antoniou P, Becq JHe M, Schuh A. Clinical whole-genome sequencing from routine formalin-fixed, paraffin-embedded specimens: pilot study for the 100,000 Genomes Project. Genetics in Medicine,2018:20(10):1196-1205.
44. Shendure J, Balasubramanian S, Church GM, Gilbert W, Rogers J, Schloss JA *et al*. DNA sequencing at 40: past, present and future. Nature,2017:550(7676):345-353.
45. Souček P, Réblová K, Kramárek M, Radová L, Grymová T, Hujová P *et al*. High-throughput analysis revealed mutations' diverging effects on SMN1 exon 7 splicing. RNA biology,2019:16(10):1364-1376.
46. Teraoka SN, Telatar M, Becker-Catania S, Liang T, Önengüt S, Tolun A *et al*. Splicing defects in the ataxia-telangiectasia gene, ATM: underlying mutations and consequences. The American Journal of Human Genetics,1999:64(6):1617-1631.
47. Wai HA, Lord J, Lyon M, Gunning A, Kelly H, Cibin P *et al*. Blood RNA analysis can increase clinical diagnostic rate and resolve variants of uncertain significance. Genetics in Medicine,2020:22(6):1005-1014.
48. Wetterstrand KA. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP) Available at: www.genome.gov/sequencingcostsdata.
49. Wright CF, FitzPatrick DR, Firth HV. Paediatric genomics: diagnosing rare disease in children. Nature Reviews Genetics,2018:19(5):253-268.
50. Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. Journal of computational biology,2004:11(2-3), 377-394.
51. Zhang Z, Zhang Y, Gao F, Han S, Cheah KS, Tse HF, Lian Q. CRISPR/Cas9 genome-editing system in

human stem cells: current status and future prospects. Molecular Therapy-Nucleic Acids, 2017:9:230-241.
52. Zook JM, Catoe D, McDaniel J, Vang L, Spies N, Sidow A *et al*. Extensive sequencing of seven human genomes to characterize benchmark reference materials. Scientific data,2016:3(1):1-26.