"Genome analysis of NGS exome data improves clinical diagnosis yield"

Dissertation submitted in partial fulfilment for the degree of

Master of Science in Biotechnology

Submitted By

Beetihotra Sanyal

KIIT School of Biotechnology, Campus- 11

KIIT to be deemed University

Bhubaneswar, Odisha, India

Under the Supervision of

Dr. Sakthivel Murugan SM.

Associate Director

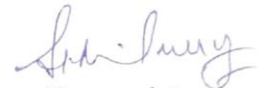
MedGenome Labs Pvt.

Bangalore- 560 099, India

MedGenome Labs Ltd.

CERTIFICATE

This is to certify the dissertation entitled "Genome analysis of NGS exome data improves clinical diagnosis yield" Submitted by ***Beetihotra Sanyal*** in partial fulfilment of the requirement for the degree of Master of Science in Biotechnology, KIIT School of Biotechnology, KIIT University, Bhubaneswar bearing Roll No.1661011 & Registration No.16674958071 is a bonafide research work carried out by him under my guidance and supervision from ***08/01/2018* to *11/01/2018.***



Name and signature

SAKTHIVEL MURUGAN SM

MedGenome Labs Ltd.

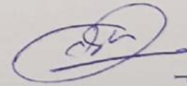
CERTIFICATE

This is to certify that the dissertation entitled "***Genome analysis of NGS exome data improves clinical diagnosis yield***" submitted by *Beetihotra Sanyal, Roll No.1661011 & Registration No.16674958071* to the KIIT School of Biotechnology, KIIT University, Bhubaneswar-751024, for the degree of Master of Science in Biotechnology is his original work, based on the results of the experiments and investigations carried out independently by him during the period from ***08/01/2018*** to ***11/05/2018*** of study under my guidance.

This is also to certify that the above said work has not previously submitted for the award of any degree, diploma, fellowship in any Indian or foreign University.

Date: 10|5|18

Place: BANGALORE



JASON K. D'SILVA

Supervisor name & signature

2 | Page

DECLARATION

I hereby declare that the dissertation entitled "***Genome analysis of NGS exome data improves clinical diagnosis yield***" submitted by me, for the degree of Master of Science to KIIT University is a record of bonafide work carried by me under the supervision of ***Dr. Sakthivel Murugan SM., Associate Director, MedGenome Labs Pvt., Bangalore- 560 099, India.***

Date: 10/05/2018
Place: Bangalore

Beetihotra Sanyal.
Name & Signature
Beetihotra Sanyal.

3 | Page

ABSTRACT

Considering the genetic heterogeneity of inherited disorders, it is becoming increasingly difficult to delineate between disorders within a similar phenotypic spectrum. Genetic analysis of NGS exome data can help clinicians reach a more conclusive diagnosis and therefore implement better treatment strategies. The prime objective of this thesis is analysis of patient samples for variant detection and interpretation responsible in causing any disease. Firstly, DNA extracted from blood is used to perform targeted gene capture using a custom capture kit. Secondly, libraries are sequenced to mean >80-100X coverage on Illumina sequencing platform. Thirdly, the sequences obtained are aligned to human reference genome and analysed using Picard and GATK version 3.6 to identify variants relevant to the clinical indication. Clinically relevant mutations were annotated using published variants in literature and a set of diseases databases – ClinVar, OMIM, GWAS, HGMD and SwissVar. Common variants are filtered based on allele frequency in 1000Genome Phase 3, ExAC, EVS, dbSNP147, 1000 Japanese Genome and our internal Indian population database. Damaging variants effect is calculated using multiple algorithms such as PolyPhen-2, SIFT, MutationTaster2, Mutation Assessor and LRT. Fourthly, based on ACMG guidelines the variants were classified as pathogenic, likely pathogenic, benign, likely benign and uncertain significance and reported. In case of reanalysis due to updated pipelines, softwares and improved knowledge the same analysis protocol mentioned above is followed and if any variant is detected associated with the newly mentioned phenotypes, it is classified and added in the report. In this project I came across 18 cases where I found 13 VUS, 4 NONE, 1 LP variants on initial analysis and upon re-analysis I got 12 VUS, 3 NONE, 1 LP and 2 variants. In one case the variant classification changed on re-analysis due to updated pipelines, databases and literature knowledge.

ACKNOWLEDGEMENTS

I take this opportunity to express my profound gratitude and deep regards to my guides Dr. Sakthivel Murugan S.M, Associated Director, MedGenome Labs Ltd., Bangalore – 560 099, India, Mr. Jason Kenneth D'Silva, Technical Manager, MedGenome Labs Ltd., Bangalore - 560 099, India, for his exemplary guidance, monitoring & constant encouragement throughout the course of the thesis. The blessings, help & guidance given by him time to time shall carry me a long way in the journey of life on which I am about to embark.

I would like to thank Miss Anshu Singh, MedGenome Labs Ltd., Bangalore for providing me the opportunity to conduct my project work successfully.

I express my deep gratitude to Miss Sukanya Ganesan (Senior Genome Analyst), Miss Nishika Sahay, Dr. Sneha P, Mr. Malaichamy Sivasankar (Senior Genome Analyst), Miss Susan Paul (Senior Genome Analyst) Mr.Rahul Kaushik for their cordial support, valuable information & guidance, which helped me in completing this work through various stages.

Above all, I want to acknowledge my parents and friends for their constant moral & economical support.

Date:

Place: *Name & Signature*

Serial No.	Contents	Page No.
1.	Abbreviations	7
2.	List of Figures	8
3.	Introduction	9
3.1	Background and Context	12
3.2	Scope and Objectives	16
3.3	Achievements	21
3.4	Overview of Dissertation	22
4.	Background and Rationale	24
5.	What did I learn?	37
5.1	Summary	37
6.	How to complement in corporate?	41
7.	References	43

1. ABBREVIATIONS

1. **ACMG**- American College of Medical Genetics and Genomics
2. **NGS**- Next generation sequencing
3. **MLPA**- Multiplex ligation-dependent probe amplification
4. **CVS**- Chorionic villus sample
5. **DNA**- Deoxyribonucleic Acid
6. **RNA**- Ribonucleic Acid
7. **OMIM**- Online Mendelian Inheritance in Man
8. **HGMD**- Human Gene Mutation Database
9. **GWAS**- Genome-wide association study
10. **ExAC**- Exome Aggregation Consortium
11. **EVS**- Exome Variant Server
12. **SIFT**- Scale-invariant feature transform
13. **LRT**- Likelihood Ratio Test
14. **VUS**- Variant of Uncertain Significance
15. **SBS**- Sequencing by synthesis
16. **VEP**- Variant effect predictor
17. **TAT**- Turn Around Time
18. **MGLIMS**- MedGenome Laboratory Information Management System
19. **FTP**- File transfer protocol
20. **HGVS**- Human Genome Variation Society

21. **cDNA**- Complementary Deoxyribonucleic Acid

22. **MODY**- Maturity Onset Diabetes of the Young

23. **IGV-** Integrative Genomics Viewer

2. LIST OF FIGURES PAGE NO.

1. **Fig** 1 - Pie chart representing variants detected across 18 cases……...20
2. **Fig 2** - Pie chart representing variants detected across 18 cases……...20
3. **Fig 3** - Graphical representation showing variants detected in analysis and re-analysis…………………………………………………..21
4. **Fig 4 -** First page of DNA test report…………………………………….32
5. **Fig 5 -** Second page of DNA test report…………………………………...33
6. **Fig 6 -** Third page of DNA test report…………………………………….34
7. **Fig 7 -** Fourth page of DNA test report…………………………………..35
8. **Fig 8 -** Last page of DNA test report……………………………………..36

3. INTRODUCTION

MedGenome Labs Ltd. is a genomic diagnostic and research-based company offering the best of health care by decoding genetic information contained in a person's genome. It has laboratories in Bangalore, Cochin, Mumbai, Singapore and California. MedGenome accepts samples like bone marrow aspirates, Peripheral blood/purified genomic DNA/chorionic villus sample (CVS)/amniotic fluid, etc for genetic tests such as Next generation sequencing (NGS), Multiplex ligation-dependent probe amplification (MLPA), Sanger sequencing, flowcytometry, etc.

Genetic testing involves examining our DNA, the chemical database that carries instructions for our body's functions. Genetic testing can reveal changes (mutations) in our genes that may cause illness or disease.

Although genetic testing can provide important information for diagnosing, treating and preventing illness, there are limitations. For example, if a healthy person, a positive result from genetic testing doesn't always mean he/she will develop a disease. On the other hand, in some situations, a negative result doesn't guarantee that someone won't have a certain disorder.

Talking to a medical geneticist or a genetic counsellor about what one should do with his/her results is an important decision in the process of genetic testing.

Everyone has a unique genome, made up of the DNA in all of a person's genes. This complex testing can help identify genetic variants that may relate to health. This testing is limited for looking at the protein-encoding regions in DNA called the exome.

I was in one of the teams involved in analysis and re-analysis of patient samples. Genetic analysis of patient genetic data is usually done to identify, interpret and report significant variants that could help clinicians for better diagnosis or improved treatment

strategy and occasionally even enable patient to live relatively better life. The analysis is carried out based on the data generated from NGS and bioinformatics analysis.

Targeted gene sequencing: Selective capture and sequencing of the protein coding regions of the genome/genes is performed. Mutations identified in the exonic regions are generally actionable compared to variations that occur in non-coding regions. Targeted sequencing represents a cost-effective approach to detect variants present in multiple/large genes in an individual.

DNA extracted from blood is used to perform targeted gene capture using a custom capture kit. The libraries are sequenced to mean >80-100X coverage on Illumina sequencing platform. The sequences obtained are aligned to human reference genome (GRCh37/hg19) using BWA program and analysed using Picard and GATK version 3.6 to identify variants relevant to the clinical indication. We follow the GATK best practices framework for identification of variants in the sample. Gene annotation of the variants is performed using VEP program against the Ensembl release 87 human gene model. Clinically relevant mutations were annotated using published variants in literature and a set of diseases databases – ClinVar, OMIM, GWAS, HGMD and SwissVar. Common variants are filtered based on allele frequency in 1000Genome Phase 3, ExAC, EVS, dbSNP147, 1000 Japanese Genome and our internal Indian population database. Non-synonymous variants effect is calculated using multiple algorithms such as PolyPhen-2, SIFT, MutationTaster2, Mutation Assessor and LRT. Only non-synonymous and splice site variants found in the panel genes are used for clinical interpretation. Silent variations that do not result in any change in amino acid in the coding region are not reported. Based on ACMG guidelines the variants are classified as pathogenic, likely pathogenic, benign, likely benign and uncertain significance and reported.

Sometimes if clinicians are not satisfied by the variations reported they ask for further re-analysis based on some new clinical synopsis. In the re-analysis we search for new

variants associated with the clinical features provided by the clinician. In some cases, we get new variants, or the previously reported variants are reclassified e.g., from variant of uncertain significance (VUS) to pathogenic or likely pathogenic.

As genetic testing becomes standard practice for the diagnosis of many “medical” diseases, such as complex diseases, who will provide adequate counselling to patients and their family members? Genetic counsellors will be required to work with physicians to help patients understand the genetic test; the risks, benefits, and accuracy of the test; and the implications for family members. Another option is to ensure that clinicians are properly trained to understand genetic testing and the importance of proper counselling. Clinical laboratories can assist in providing appropriate counselling information to physicians and patients along with the test results, but even this will not ensure that the information is properly communicated to the patient. Genetic testing also has applications for the doctor–patient relationship as these test results have implications for the family members of the patient who were not a traditional part of the doctor–patient interaction previously.

3.1 BACKGROUND AND CONTEXT

The technique of **DNA sequencing** is used to understand the nucleotide sequence of DNA (deoxyribonucleic acid). The nucleotide sequence is the most fundamental level of information of a gene or genome.

First-generation sequencing- It was initiated in the 1970s, included the Maxam-Gilbert technique, discovered by the American molecular biologists Allan M. Maxam and Walter Gilbert, and, English biochemist Frederick Sanger discovered the Sanger technique (or dideoxy method). The Sanger method became the more commonly used sequencing approaches among the two. These sequencing methods after few years were performed by using automated sequencing technologies, where the truncated DNA molecules are labelled with fluorescent tags and separated by size within thin glass capillaries and detected by laser excitation.

Next-generation or second-generation sequencing (NGS)- It had largely supplanted first-generation technologies. There are many processes which allows many DNA fragments (including millions of fragments) to be sequenced at a short interval of time and in a more cost-effective manner and much faster than first-generation technologies. The use of next-generation sequencing was improved significantly by advances in bioinformatics that allowed for increased data storage and facilitated the analysis and manipulation of very large data sets, often in the gigabases range such as 1 gigabase = 1,000,000,000 base pairs of DNA.

Knowledge of the sequence of a DNA segment has many uses such as it can be used to find any gene or segments of DNA that code for a particular protein or phenotype. When some part of a DNA is being sequenced, it can be screened for characteristic features of genes.

It has revolutionized the biological sciences. With its ultra-high throughput, scalability, and speed, NGS enables researchers to perform a wide variety of applications and

study biological systems at a level never possible. Now a days the complicated genetic research questions wants a depth of information which is beyond the capacity of old DNA sequencing techniques. Next-generation sequencing has filled that gap and become an everyday research tool to address these questions.

NGS technology has fundamentally changed the kinds of questions scientists can ask and answer. The sample is prepared with innovation and the data analysis options allows broad spectrum of applications. For example, NGS allows researchers to:

- Rapidly sequence whole genomes.
- Zoom in to deeply sequence target regions.
- Utilize RNA sequencing (RNA-Seq) to discover novel RNA variants and splice sites or quantify mRNAs for gene expression analysis.
- It can sequence cancer samples to study rare somatic variants, and more.
- Study microbial diversity in humans or in the environment.

Illumina sequencing utilizes a fundamentally different approach from the classic Sanger chain-termination method. Sequencing is done by utilizing sequencing by synthesis (SBS) technology which enables tracking of the addition of labeled nucleotides as the DNA chain is copied in a massively parallel fashion.

Next-generation sequencing produces a huge amount of sequence data which is much richer and complete than that of Sanger sequencing. The Illumina sequencing machine can deliver 300 kilobases of data up to multiple terabases in a single run, depending on instrument type and configuration. It is also cost effective and has higher reproducibility and accuracy.

Genetic analysis is done after data is generated from the illumina. These data of patients generated needs bioinformatics analysis, and the analysed data is uploaded. The main variant analysis is done by retrieving data from our database using Varminer (in house analysis tool) and the analysis is processed with respect to the supporting clinical details of particular mendelian disorders, i.e. differential diagnosis, clinical features related to the disease, laboratory findings, etc. Variants detected are interpreted and classified according to the American College of Medical Genetics and Genomics (ACMG) guidelines and are reported.

In case of re-analysis the reported diagnostic yield for clinical exome sequencing of patients with presumed Mendelian disorders is around 25%.[1–5] It remains unclear what proportion of undiagnosed patients harbour a causative mutation in the already sequenced exome that is not prioritized in the initial analysis. There are several reasons that causative mutations in the exome might be unrecognized. First, the processes both to call variants from short sequence reads and to annotate the impact of variants on genes are imperfect.[6,7] Factors at each step can result in a causative variant being unidentified. Second, exome interpretation is informed by the reported patient phenotype. If key elements of the phenotype are not available to the clinical laboratory, because they either were not recorded or have not yet clinically emerged, this decreases the likelihood that a causative variant would be prioritized. Third, the representation of known gene–disease and variant–disease associations in readily searchable structured databases is incomplete, and the gap between the primary literature and structured databases is growing[8]. Identifying relevant information in the free text primary literature relies on imperfect ad hoc heuristic searches such as with search engines, rather than the preferable systematic queries of databases. Fourth, knowledge of variant and gene function is incomplete. Over 1,600 characterized Mendelian phenotypes are not yet tied to a causative gene, some Mendelian phenotypes exhibit genetic heterogeneity and have multiple causative genes, and

many Mendelian conditions are yet to be recognized[9,10]. Furthermore, a significant amount of expert time—from 20 to 40 h—is often required to evaluate a clinical exome[11]. This introduces potential variability and bias and makes it difficult to precisely replicate exome analysis. The required manual time also may decrease the frequency of reanalysis. Some clinical laboratories reanalyse exomes upon a provider's request, limit reanalysis requests to one per year, or require a fee to reanalyse[12]. To evaluate how improvements in genome analysis methods and expanded knowledge of variant and gene function might increase the diagnostic yield of exome sequencing.

Varminer is used to annotate variants with a predicted effect on protein-coding genes from the ENSEMBL gene set, and with an allele frequency in the Exome Aggregation Consortium, 1000 Genomes Project and our internal database control human populations.[19-12] Variants are filtered to retain only rare variants (maximum allele frequency <0.1% in any 1000 Genomes Project or Exome Aggregation Consortium population or the internal database) that are predicted to be missense, truncating, in-frame insertion or deletion, stop codon loss, or splice site disrupting. Variants were classified in accordance with the guidelines of the American College of Medical Genetics and Genomics[13]. A variant is considered causative if, in the clinical judgment of the evaluating medical geneticist, it is classified as pathogenic or likely pathogenic for a disease that matches the participant's phenotype. Once a candidate variant is identified, the clinical laboratory is contacted to perform parental testing for the variant. Following parental testing, updated clinical reports are issued and participants received updated genetic counselling.

Data from the OMIM database and the Human Genome Mutation Database (HGMD) Pro 2015. are used to quantify the growth in gene-disease and variant-disease associations[23,24]. The number of HGMD variant–disease associations are calculated using the publication year tied to each variant. Only DM variants (disease-causing; demonstrated to be pathogenic) are included. The HGMD gene–disease annotation

counts use the first publication year of any variant tied to a gene. Unique gene symbols are considered distinct genes.

3.2 SCOPE AND OBJECTIVES

Scope: Genetic testing involves examination of our DNA, database of chemical which carries instructions for our body's functions. Genomic testing is important to reveal changes (mutations) in our genes that may cause illness or disease.

Though genomic testing can deliver important information that can help in diagnosing, treating and preventing illness, there are limitations. For example, if a healthy person, a positive result from genetic testing doesn't always mean he/she will develop a disease. On the other hand, in some situations, a negative result doesn't guarantee that someone won't have a certain disorder.

Talking to the doctor, a medical geneticist or a genetic counsellor about what one will do with the results is an important step in the process of genetic testing.

Genomic counselling plays an important role in determining the risk for developing certain disorders as well as their screening and sometimes medical treatment. Different types of genetic testing are done for different reasons:

- **Diagnostic testing**. If one having symptoms of some disorder that might be caused by any genetic changes, sometimes called mutated genes, genetic testing can reveal if he/she has the suspected disorder. For example, genetic testing may be used to confirm a diagnosis of cystic fibrosis or Huntington's disease.
- **Presymptomatic and predictive testing.** If one has a family history of a genetic condition, getting genetic testing before he/she has symptoms may

show if they are at risk of developing that condition. For example, this type of test may be useful for identifying your risk of certain types of colorectal cancer.

- **Carrier testing**. If someone having a family history of a genetic disorder — such as sickle cell anemia or cystic fibrosis — or in an ethnic group that has a high risk of a specific genetic disorder, he/she may choose to have genetic testing before having children. The expanded carrier screening test can also identify those genes associated with a wide variety of genetic disorders and mutations and can identify if he and his partner are carriers.

- **Pharmacogenetics**. If someone has a particular health condition or disease, this type of genetic testing may help determine what medication and dosage will be most effective and beneficial for him.

- **Prenatal testing**. If someone is pregnant, tests can detect some types of abnormalities in her baby's genes. Down syndrome and trisomy 18 syndrome are two genetic disorders that are often screened for as part of prenatal genetic testing. Traditionally this is done looking at markers in blood or by invasive testing such as amniocentesis. A new test called cell-free DNA test looks at a baby's DNA by doing a blood test of mother.

- **Newborn screening**. This is the most common type of genetic testing. In the United States, all states require that newborns be tested for certain genetic and metabolic abnormalities that cause specific conditions. This type of genetic testing is important because if results show there's a disorder such as congenital hypothyroidism, sickle cell disease or phenylketonuria, care and treatment can begin right away.

- **Preimplantation testing**. Also called preimplantation genetic diagnosis, this test may be used when one attempts to conceive a child through in vitro fertilization. The embryos are screened for genetic abnormalities. Embryos without abnormalities are implanted in the uterus in hopes of achieving pregnancy.

No matter what the results of the genetic testing, talk with the doctor, medical geneticist or genetic counsellor about questions or concerns us may have. This will help us understand what the results mean for us and our family.

Currently, the majority of genetic tests identify one or more DNA mutations in a single gene that cause a disease with a clear inheritance pattern of autosomal dominant, autosomal recessive, or X-linked. In the coming future, the potential applications for genetic testing will be more, were every known disorder having some aspect that is influenced by, if not directly caused by, mutations or polymorphisms in the genome of the patient. Lots of factors influences the translation of potential medical applications of genomic information into reality which includes our understanding of the human genome, technological advances, and the state of social, ethical, and legal issues surrounding genetic testing.

Objectives:

- Finding new variants in genes leading to pathogenic events in a patient.
- Understanding a disease, its biology and genetic cause.
- Interpretation and classification of variants based on the ACMG guidelines.
- To identify new variants on re-analysis or revision of variant significance.
- Reporting significant variants.
- Providing better diagnosis or counselling and treatment.

3.3 ACHIEVEMENTS

On **analysis** I got the following variants:

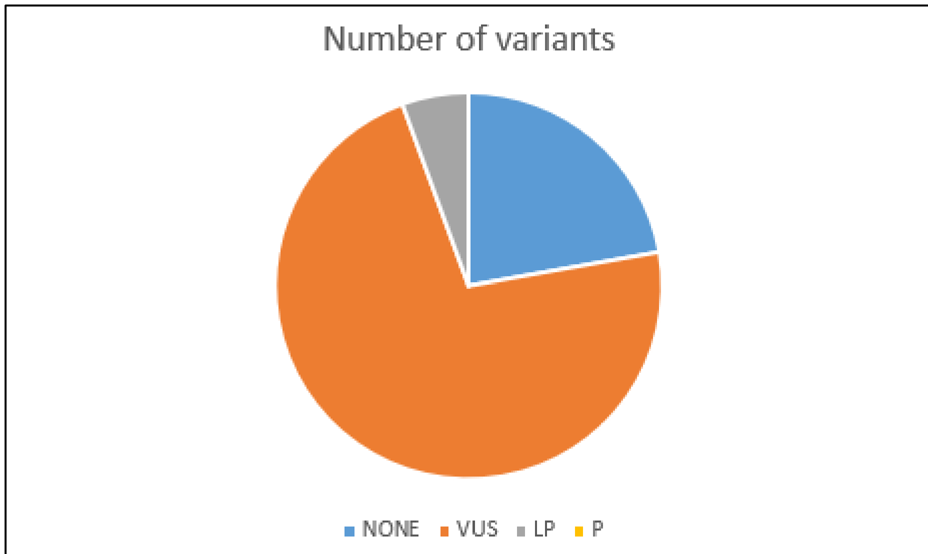


Fig 1: Pie chart representing variants detected across 18 cases.

On **Re- analysis** I got the following variants:

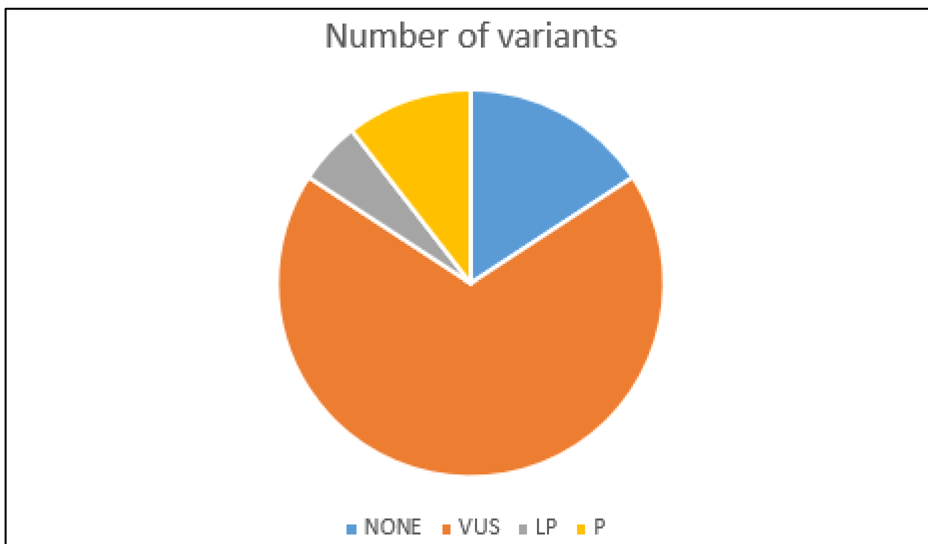


Fig 2: Pie chart representing variants detected across 18 cases.

VUS- Variant of Uncertain Significance

LP- Likely Pathogenic

P- Pathogenic

Result showing variants found in initial analysis and re-analysis.

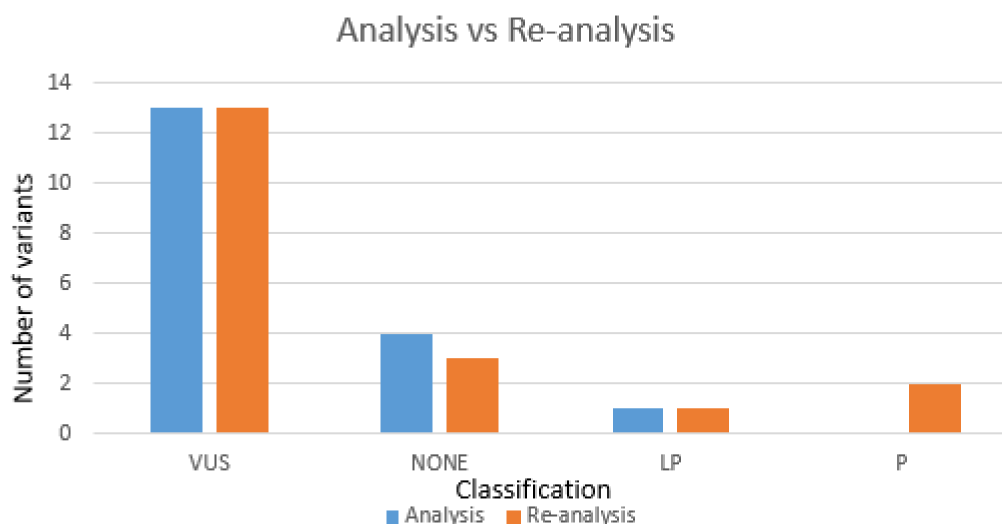


Fig 3: Graphical representation showing variants detected in analysis and re-analysis.

3.4 OVERVIEW OF DISSERTATION

I was in one of the teams involved in analysis and re-analysis of patient samples. Genetic analysis of patient genetic data is usually done to identify, interpret and report significant variants that could help clinicians for better diagnosis or improved treatment strategy and occasionally even enable patient to live relatively better life. The analysis is carried out based on the data generated from NGS and bioinformatics analysis.

Targeted gene sequencing: Selective capture and sequencing of the protein coding regions of the genome/genes is performed. Mutations identified in the exonic regions are generally actionable compared to variations that occur in non-coding regions. Targeted sequencing represents a cost-effective approach to detect variants present in multiple/large genes in an individual.

DNA extracted from blood is used to perform targeted gene capture using a custom capture kit. The libraries are sequenced to mean >80-100X coverage on Illumina

sequencing platform. The sequences obtained are aligned to human reference genome (GRCh37/hg19) using BWA program and analysed using Picard and GATK version 3.6 to identify variants relevant to the clinical indication. We follow the GATK best practices framework for identification of variants in the sample. Gene annotation of the variants is performed using VEP program against the Ensembl release 87 human gene model. Mutations which are clinically relevant are annotated using published variants in literature and set of disease databases such as ClinVar, OMIM, GWAS, HGMD and SwissVar. Common variants are filtered based on allele frequency in 1000Genome Phase 3, ExAC, EVS, dbSNP147, 1000 Japanese Genome and our internal Indian population database. Non-synonymous variants effect is calculated using multiple algorithms such as PolyPhen-2, SIFT, MutationTaster2, Mutation Assessor and LRT. Only non-synonymous and splice site variants found in the panel genes are used for clinical interpretation. Silent variations that do not result in any change in amino acid in the coding region are not reported. Based on ACMG guidelines the variants are classified as pathogenic, likely pathogenic, benign, likely benign and uncertain significance and reported.

The clinical laboratories and the clinical geneticists were largely helped by the ACMG standards and guidelines for better and effective treatments. The American College of Medical Genetics and Genomics (ACMG) previously developed guidance for the interpretation of sequence variants. In the last few years, sequencing technology has evolved rapidly with the advent of high-throughput next-generation sequencing. By adopting and leveraging next-generation sequencing, clinical laboratories are now performing an ever-increasing catalogue of genetic testing spanning genotyping, single genes, gene panels, exomes, genomes, transcriptomes, and epigenetic assays for genetic disorders. Genetic tests are done which include genotyping, single gene, panels, exomes and genomes which are recommended. The guidelines also describe about the use of specific standard term such as “pathogenic,” “likely pathogenic,”

"uncertain significance," "likely benign," and "benign"—to describe variants identified in genes that cause Mendelian disorders. Moreover, this recommendation describes a process for classifying variants into these five categories based on criteria using typical types of variant evidence (e.g., population data, computational data, functional data, segregation data). Because of the increased complexity of analysis and interpretation of clinical genetic testing described in this report, the ACMG strongly recommends that clinical molecular genetic testing should be performed in the clinical laboratory improvement amendments which should be an approved laboratory.

Varminer is used to annotate variants with a predicted effect on protein-coding genes from the ENSEMBL gene set, and with an allele frequency in the Exome Aggregation Consortium, 1000 Genomes Project and our internal database control human populations[19-22]. Variants are filtered to retain only rare variants (maximum allele frequency <0.1% in any 1000 Genomes Project or Exome Aggregation Consortium population or the internal database) that are predicted to be missense, truncating, in-frame insertion or deletion, stop codon loss, or splice site disrupting. Variants were classified in accordance with the guidelines of the American College of Medical Genetics and Genomics[13]. A variant is considered causative if, in the clinical judgment of the evaluating medical geneticist, it is classified as pathogenic or likely pathogenic for a disease that matches the participant's phenotype. Once a candidate variant is identified, the clinical laboratory is contacted to perform parental testing for the variant. Following parental testing, updated clinical reports are issued and participants received updated genetic counselling.

Sometimes if clinicians are not satisfied by the variations reported they ask for further re-analysis based on some new clinical synopsis. In the re-analysis we search for new variants associated with the clinical features provided by the clinician. In some cases, we get new variants, or the previously reported variants are reclassified e.g., from variant of uncertain significance (VUS) to pathogenic or likely pathogenic.

4. BACKGROUND AND RATIONALE

Allotment of Samples:

- The samples registered for clinical reporting in the MGLIMS will be allotted to the respective genome analysts by the person in-charge for data analysis and reporting.
- The genome analysts are required to generate and dispatch the approved reports within TAT as specified in the MGLIMS.

Retrieval of TRF:

- The test requisition form for a specific sample ID can be downloaded from the MGLIMS. This form contains details required for the analysis and reporting of the sample including:
- Sample type, consanguinity, age, gender, family history, inheritance pattern, symptoms, indications / clinical diagnosis.
- Test requested.
- Details of the requesting clinician.

Data shared from bioinformatics team:

The analysed data for the sample is shared through mail to all the genome analysts. The mail contains the path of the FTP files, raw and analysed BAM files.

Example: FTP Path: /ftp_uploads/files/12345_R123_LIB12345

This FTP folder contains

1. Quality metrics report of the sample data.
2. Variant annotated file "Variant" compatible with Varminer (in house software tool).
3. Coverage of the coding regions of the genes.
4. Low coverage regions of the genes.

5. Variant reported in HGMD.
6. Suspected large indels (annotated output from Pindel)
7. Suspected copy number variants and structural variants file.

Checking data quality for compliance:

The quality metrics shared by the bioinformatics team is thoroughly checked for compliance and sample analysis. It also has details of the library id, run id and software and version of the tools employed for data analysis.

Expectations: Data parameters including total data generated, aligned reads, passed alignment and overall duplicate ratio less than 5-10% of the expected value is accepted provided the genes of interest are well covered.

Noncompliance: When the data quality is poor the samples are reprocessed either for resequencing to generate data or sample library is regenerated based on the requirement.

Variant annotation file:

Variant annotation file is the primary annotated file used for data analysis.

The variants are annotated according to the HGVS nomenclature.

The Variant annotation file also has:

- All the mutations are annotated using published variants in literatures and set of disorder databases such as ClinVar, OMIM, GWAS, HGMD and SwissVar.
- Minor allele frequency in 1000 Genome Phase 3, ExAC, EVS and our internal Indian population database (MedVar database).
- Damaging effect of variants calculated using multiple predicton algorithms such as SIFT, Polyphen-2, MutationTaster-2 and LRT.

Varminer2:

Varminer is an in-house software tool available for variant filtration and prioritization based on the clinical phenotype and suspected mode of inheritance.

aids in better and efficient analysis and interpretation.

- The Varminer2 facilitates the viewing of “QC status”, “TRF”, “Clinical synopsis”, “Appplied filter”, “file information” and “version information”.

Analysis:

The varminar interface displays the data columns from the VariMat filr such as chromosome number, variant start position, base change, variant type, zygosity, read depth of the variant, gene in which variant is found, cDNA change, protein change, exon number, Ensemble transcript ID, RefSeq transcript ID, RefSeqProtein ID, EnsembleProtein ID, ClinVar ID, SwissVar, 1000G MAF, ExAC MAF, SIFT, PolyPhen2 HDiv, PolyPhen2_Hvar, LRT, VariantQuality, Reference base depth, Altered base depth, disease database, MedVar database, etc.

Filters:

The results obtained depends solely on the filters and their parameters used. Thus, it is important to understand the biology of the disease, mode of inheritance, commonly observed mutations and prevalence for analysis. Most of the common content in Varminer can be filtered based on the applied parameters.

Panel filter: The panel name is entered in the “Disease filter” tab which adds the gene list corresponding to the disease and can be used for filtering.

Eg. Muscular dystrophy and congenital myopathy, deafness panel, MODY panel, Charcot-Marie-Tooth disease panel, etc.

Included genes: Specific genes can be manually added to this tab.

Excluded genes: During analysis the genes can also be excluded from this tab.

OMIM filters: Disease phenotype can be added in this tab which helps in pulling all the reported genes from OMIM database.

Synopsis filters: Disease phenotype can be added in this tab which helps in pulling all the genes based on OMIM clinical synopsis.

In addition to the above filters, column filters are also available, which can be used to exclude or include specific parameters. For example, only “Homozygous” variants can be filtered or only “Missense” variants can be filtered.

Criteria for filtering and prioritization of the variants for reporting:

The goal of filtering and prioritization is to weigh the available evidence in favour of a likely role of the genetic variants detected in a patient.

- **Phenotype / clinical indications / diagnosis**: The variants prioritized for reporting is mainly based on clinical indications and diagnosis. The clinical indications given in the TRF are read and understood with respect to the genetic basis of the disorder and gene list is selected accordingly.
- **Assumed mode of inheritance based on pedigree or family history or phenotype**: Based on given clinical synopsis or diagnosis or any specific phenotype the mode of inheritance pattern of the disease is studied and included. For example- some phenotypes are very well-known mode of inheritance. Cystic fibrosis is inherited in an autosomal recessive mode. If pedigree and details of other affected and unaffected members are given, this information can be used to predict the mode of inheritance for the particular family.
- **Minor Allele Frequency**: We check how rare a variant is in the population databases (1000 genome, ExAC and internal databases). As if the variant is common in the population then the variant is unlikely to be disease causing.

Conversely, when a variant rare in the general population, the variant is likely to cause disease.

- **Reported status**: The clinically relevant variant(s) are checked whether they are reported in literature for the given phenotype, by using databases such as Clinvar, SwissVar, HGMD, etc.

  If the variant is reported in any of the databases with the suspected or similar phenotype, the variant is prioritized for reporting.
- **Prediction based on *in-silico* tools**: The current ACMG guidelines suggests that multiple lines of computational evidence support deleterious effect on the gene. Thus, in our analysis we use five different predictive tools SIFT, Polyphen2, MutationTaster2 and LRT in assigning significance and clinical correlation for reporting.
- **Genotype-Phenotype correlation**: The prioritized variants based on all the above criteria is checked for genotype-phenotype correlation to ascertain the significance of the variant(s).

PINDEL analysis: The size of the indels called in GATK tool is only around 10bp. Thus, to identify large indels PINDEL tool is used.

Integrative Genomics Viewer (IGV): The variants thus prioritized for reporting are visualized in the IGV, a high-performance visualization tool for interactive exploration of variants viewer. This aligns the BAM file of the variant against the reference sequence of NCBI. IGV helps us to identify skewed strand ratio, bad quality reads, allele change and visualize depth.

Variant Interpretation:

Variant selected after filtering based on the above criteria in a sample is evaluated and classified according to the American College of Medical Genetics and Genomics (ACMG) guidelines.

Very Strong:

- Variant reported in previous studies with substantial evidence that the variant causes the disease in question.
- Variants in a gene where loss of function is a known mechanism of disease, including: nonsense or frameshift changes including interruption of the normal start or stop codon, alteration of the splice donor or acceptor site.

Strong:

- Previously established pathogenic variant regardless of nucleotide change with the same amino acid change.
- De novo (both maternity and paternity confirmed) in a patient with the disease and no family history.
- A gene or gene product is damaging if there is supporting well established functional studies.

Moderate:

- Should be located in a mutational hotspot or critical and well-established functional domain.
- Absent from controls population.
- Detected in trans with a pathogenic variant for recessive disorders.
- Nonrepeat region or stop-loss variants occurring when protein change occurs due to inframe deletion or insertion.

- Changes in an amino acid residue which are missense in nature and where any other missense change can determine to be pathogenic has been seen before.
- It is considered as de novo variation before parental or maternal examination is done.
- Co-segregation with disease in multiple affected family members in a gene definitely known to cause the disease.
- Variants which are missense in a gene that is having a low rate of benign variation and in which missense variants are a common mechanism of the disorder.
- The computational evidence helps in understanding a deleterious effect on the gene or a gene product (conversion, evolutionary, splicing impact, etc.).
- Patient's phenotype or family history is highly specific for a disease with a single genetic etiology.

Variant Classification:

After sequence change has been reviewed it is classified into these categories:

Pathogenic: A disease causing variation in a gene which can explain the patient's symptoms has been detected. This usually means that a suspected disorder for which testing had been requested has been confirmed.

Likely Pathogenic: A novel or previously reported variants, with some evidence suggesting that this is very likely to contribute to the development of disease however, the scientific evidence is currently insufficient to prove this conclusively.

Benign: A variant that are present at a higher frequency in the general population than expected for that disease and which is known not to be responsible for disease has been detected. Generally, no further action is warranted on such variants when detected.

Likely Benign: This category includes varients in intronic region that are unlikely to affect splicing, variants in highly variable region if a gene. Variants for which multiple lines of computational evidence suggest no impact on gene/gene product, well-established functional studies show no deleterious effect and variants not segregating with the disease in a family.

Variant of uncertain significance: A variant has been detected, such as missense variant, in frame deletions/insertions, but it is difficult to classify to either pathogenic (disease causing) or benign (non-disease causing) based on current available scientific evidence. It is probable that their significance may change, subject to availability of scientific evidence.

Report Generation:

MedGenome Labs Pvt. Ltd.
3rd Floor, Narayana Nethralaya Building, Narayana Health City,
#258/A, Bommasandra, Hosur Road, Bangalore – 560 099, India.
Tel: +91 (0)80 67154932 / 933 Web: www.medgenome.com

DNA TEST REPORT – MEDGENOME LABORATORIES

Full Name / Ref No:	XXXX	Order ID/Sample ID:	XX/XX
Gender:	XX	Sample Type:	Blood
Date of Birth / Age:	XX years	Date of Sample Collection:	XX
Referring Clinician:	Dr. XXXX,	Date of Sample Receipt:	XX
	XX,	Date of Order Booking:	XX
	XX	Date of Report:	XX
Test Requested:	Comprehensive neurology panel		

CLINICAL DIAGNOSIS / SYMPTOMS / HISTORY

XXXX, born of a non-consanguineous marriage, presented with clinical indications of neuroregression followed by febrile illness. Her brain MRI showed periventricular hyper intensities region. There is a family history of psychiatric illness in father. *XXXX* is suspected to be affected with leukodystrophy and has been evaluated for pathogenic variations in the genes listed in appendix 1.

RESULTS

PATHOGENIC VARIANT CAUSATIVE OF THE REPORTED PHENOTYPE WAS IDENTIFIED

Gene (Transcript)	Location	Variant	Zygosity	Disease (OMIM)	Inheritance	Classification
MFSD8 (-) (ENST00000296468)	Exon 10	c.894T>G **(p.Tyr298Ter)**	Homozygous	Neuronal ceroid lipofuscinosis- 7	Autosomal recessive	Pathogenic

ADDITIONAL FINDINGS: NO VARIANT(S) OF UNCERTAIN SIGNIFICANCE (VUS) IDENTIFIED

No other variant that warrants to be reported was detected. Variations with high minor allele frequencies which are likely to be benign will be given upon request.

VARIANT INTERPRETATION AND CLINICAL CORRELATION

Variant description: A homozygous nonsense variation in exon 10 of the ***MFSD8*** gene **(chr4:128851942A>C; Depth: 52x)** that results in a stop codon and premature truncation of the protein at codon 298 **(p.Tyr298Ter; ENST00000296468)** was detected (Table). The observed variation has previously been reported in a patient affected with late infantile neuronal ceroid lipofuscinosis [23]. The p.Tyr298Ter variant has not been reported in the 1000 genomes and ExAC databases and has a minor allele frequency of 0.07% in our internal database. The *in silico* prediction of the variant is damaging by MutationTaster2. The reference codon is conserved across species.

Page 1 of 17
Name/Sample ID: XXXX/XX

DUMMY REPORT

Fig 4: First page of DNA test report

MedGenome Labs Pvt. Ltd.
3rd Floor, Narayana Nethralaya Building, Narayana Health City,
#258/A, Bommasandra, Hosur Road, Bangalore – 560 099, India.
Tel: +91 (0)80 67154932 / 933 Web: www.medgenome.com

MEDGENOME

OMIM phenotype: Neuronal ceroid lipofuscinosis-7 (OMIM#610951) is caused by homozygous or compound heterozygous mutations in the *MFSD8* gene (OMIM*611124). The disorder is characterized by the intracellular accumulation of autofluorescent lipopigment storage material in different patterns ultrastructurally [10].

Based on the above evidence, ***this MFSD8 variation is classified as a pathogenic variant and has to be carefully correlated with the clinical symptoms.***

RECOMMENDATIONS

Validation of the variant(s) by Sanger sequencing is recommended to rule out false positives.

Sequencing the variant(s) in the parents and the other affected and unaffected members of the family is recommended to confirm the significance.

Genetic counselling is advised for interpretation on the consequences of the variant(s).

TEST METHODOLOGY

Targeted gene sequencing: Selective capture and sequencing of the protein coding regions of the genome/genes is performed. Mutations identified in the exonic regions are generally actionable compared to variations that occur in non-coding regions. Targeted sequencing represents a cost-effective approach to detect variants present in multiple/large genes in an individual.

DNA extracted from blood was used to perform targeted gene capture using a custom capture kit. The libraries were sequenced to mean >80-100X coverage on Illumina sequencing platform. The sequences obtained are aligned to human reference genome (GRCh37/hg19) using BWA program [2, 3] and analyzed using Picard and GATK version 3.6 [4, 5] to identify variants relevant to the clinical indication. We follow the GATK best practices framework for identification of variants in the sample. Gene annotation of the variants is performed using VEP program [6] against the Ensembl release 87 human gene model [7]. Clinically relevant mutations were annotated using published variants in literature and a set of diseases databases – ClinVar, OMIM, GWAS, HGMD and SwissVar [8-15]. Common variants are filtered based on allele frequency in 1000Genome Phase 3, ExAC, EVS, dbSNP147, 1000 Japanese Genome and our internal Indian population database [16-20]. Non-synonymous variants effect is calculated using multiple algorithms such as PolyPhen-2, SIFT, MutationTaster2, Mutation Assessor and LRT. Only non-synonymous and splice site variants found in the comprehensive neurology panel genes were used for clinical interpretation. Silent variations that do not result in any change in amino acid in the coding region are not reported.

Total data generated (Gb)	5.57
Total reads aligned (%)	99.99
Reads that passed alignment (%)	96.69
Data ≥ Q30 (%)	91.36

***Genetic test results are reported based on the recommendations of American College of Medical Genetics [1], as described below:**

Variant	A change in a gene. This could be disease causing (pathogenic) or not disease causing (benign).
Pathogenic	A disease causing variation in a gene which can explain the patients' symptoms has been detected. This usually means that a suspected disorder for which testing had been requested has been confirmed.
Likely Pathogenic	A variant which is very likely to contribute to the development of disease however, the scientific evidence is currently insufficient to prove this conclusively. Additional evidence is expected to confirm this assertion of

CAP ACCREDITED

DUMMY REPORT

Fig 5: Second page of DNA test report.

MedGenome Labs Pvt. Ltd.
3rd Floor, Narayana Nethralaya Bu'lding, Narayana Health C'ty,
#258/A, Bommasandra, Hosur Road, Bangalore – 560 099, India.
Tel: +91 (0)80 67154932 / 933 Web: www.medgenome.com

	pathogenicity.
Variant of Uncertain Significance	A variant has been detected, but it is difficult to classify it as either pathogenic (disease causing) or benign (non-disease causing) based on current available scientific evidence. Further testing of the patient or family members as recommended by your clinician may be needed. It is probable that their significance can be assessed only with time, subject to availability of scientific evidence.

*The transcript used for clinical reporting generally represents the canonical transcript (according to Ensembl release 87 gene model), which is usually the longest coding transcript with strong/multiple supporting evidence. However, clinically relevant variants annotated in alternate complete coding transcripts could also be reported.

Variants annotated on incomplete and nonsense mediated decay transcripts will not be reported.

*The *in silico* predictions are based on Variant Effect Predictor, Ensembl release 87 (SIFT version - 5.2.2; PolyPhen - 2.2.2); LRT version - November, 2009 release from dbNSFPv3.1 and MutationTaster2 based on build NCBI 37 / Ensembl 69 [21].

For any further technical queries please contact techsupport@medgenome.com.

DISCLAIMER

- The classification of variants of unknown significance can change over time and MedGenome cannot be held responsible for this. Please contact MedGenome at a later date to inquire about any changes.
- Intronic variants are not assessed using this method.
- Large deletions of more than 10 bp or copy number variations /chromosomal rearrangements cannot be assessed using this method.
- Certain genes may not be covered completely and few mutations could be missed. Variants not detected by the assay that was performed may impact the phenotype.
- The mutations have not been validated by Sanger sequencing.
- Incidental or secondary findings (if any) that meet the ACMG guidelines [22] can also be given upon request.
- This is a laboratory developed test and the development and the performance characteristics of this test was determined by MedGenome.

DUMMY REPORT

Sakthivel Murugan, PhD Associate Director - Diagnostics	Vivek Gopalan, PhD Senior Bioinformatics Scientist II	V. L. Ramprasad, PhD COO/Lab Director	Dr. Udhaya H. Kotecha, MD (Paediatrics), Fellowship in Medical Genetics. Consultant - Clinical Geneticist

END OF REPORT

Page **3** of **17**
Name/Sample ID: XXXX/XX

Fig 6: Third page of DNA test report.

MedGenome Labs Pvt. Ltd.
3rd Floor, Narayana Nethralaya Building, Narayana Health City,
#258/A, Bommasandra, Hosur Road, Bangalore – 560 099, India.
Tel: +91 (0)80 67154932 / 933 Web: www.medgenome.com

MEDGENOME

REFERENCES

1. Richards, S, *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: official journal of the American College of Medical Genetics. 17.5 (2015): 405–24.
2. Li, H, and R Durbin. "Fast and Accurate Long-Read Alignment with Burrows-Wheeler Transform." Bioinformatics (Oxford, England). 26.5 (2010): 589–95.
3. Meyer, LR, *et al.* "The UCSC Genome Browser Database: Extensions and Updates 2013." Nucleic acids research. 41. (2012): n.pag.
4. McKenna, A, *et al.* "The Genome Analysis Toolkit: A MapReduce Framework for Analyzing Next-Generation DNA Sequencing Data." Genome research. 20.9 (2010): 1297–303.
5. Li, H, *et al.* "The Sequence Alignment/map Format and SAMtools." Bioinformatics (Oxford, England). 25.16 (2009): 2078–9.
6. McLaren, W, *et al.* "Deriving the Consequences of Genomic Variants with the Ensembl API and SNP Effect Predictor." Bioinformatics (Oxford, England). 26.16 (2010): 2069–70.
7. ENSEMBL: http://www.ensembl.org
8. Landrum, MJ, *et al.* "ClinVar: Public Archive of Interpretations of Clinically Relevant Variants." Nucleic acids research. 44. (2015): n.pag.
9. Hamosh, A, *et al.* "Online Mendelian Inheritance in Man (OMIM), a Knowledgebase of Human Genes and Genetic Disorders." Nucleic acids research. 33. (2004): n.pag.
10. OMIM: http://www.omim.org
11. GWAS: http://www.ebi.ac.uk/gwas/
12. Welter, D, *et al.* "The NHGRI GWAS Catalog, a Curated Resource of SNP-Trait Associations." Nucleic acids research. 42. (2013): n.pag.
13. Stenson, PD, *et al.* "Human Gene Mutation Database (HGMD): 2003 Update." Human mutation. 21.6 (2003): 577–81.
14. HGMD: http://www.biobase-international.com/product/hgmd
15. Mottaz, A, *et al.* "Easy Retrieval of Single Amino-Acid Polymorphisms and Phenotype Information Using SwissVar." Bioinformatics (Oxford, England). 26.6 (2010): 851–2.
16. The 1000 Genomes Project Consortium. "A Global Reference for Human Genetic Variation." Nature 526.7571 (2015): 68–74. PMC.
17. Lek, M, *et al.* "Analysis of Protein-Coding Genetic Variation in 60,706 Humans." Nature. 536.7616 (2016): 285–91.
18. NHLBI: https://esp.gs.washington.edu/drupal
19. Nagasaki, Masao *et al.* "Rare Variant Discovery by Deep Whole-Genome Sequencing of 1,070 Japanese Individuals." Nature Communications 6 (2015): 8018. PMC. n.pag
20. dbSNP: http://www.ncbi.nlm.nih.gov/SNP/
21. MutationTaster2: http://www.mutationtaster.org/
22. Green, RC, *et al.* "ACMG Recommendations for Reporting of Incidental Findings in Clinical Exome and Genome Sequencing." Genetics in medicine: official journal of the American College of Medical Genetics. 15.7 (2013): 565–74.
23. Siintola, E., *et al.* "The novel neuronal ceroid lipofuscinosis gene *MFSD8* encodes a putative lysosomal transporter." The American Journal of Human Genetics. 81.1 (2007): 136-46.

APPENDIX 1: COVERAGE OF ANALYZED GENES

Gene	Percentage of coding region covered	Gene	Percentage of coding region covered	Gene	Percentage of coding region covered
AARS	100.00	*FRMPD4*	100.00	*PFKM*	100.00
AARS2	100.00	*FRRS1L*	100.00	*PFN1*	100.00
ABAT	100.00	*FTCD*	100.00	*PGAM2*	100.00
ABCB7	100.00	*FTL*	100.00	*PGAP1*	100.00
ABCD1	100.00	*FTO*	100.00	*PGAP2*	100.00

Page **4** of **17**
Name/Sample ID: XXXX/XX

CAP ACCREDITED

DUMMY REPORT

Fig 7: Fourth page of DNA test report

MedGenome Labs Pvt. Ltd.
3rd Floor, Narayana Nethralaya Building, Narayana Health City,
#258/A, Bommasandra, Hosur Road, Bangalore – 560 099, India.
Tel: +91 (0)80 67154932 / 933 Web: www.medgenome.com

FBXO31	100.00	PDE6D	100.00	WDR45	100.00
FBXO38	100.00	PDGFB	100.00	WDR62	100.00
FBXO7	100.00	PDGFRB	100.00	WDR73	100.00
FGD1	100.00	PDHA1	100.00	WDR81	100.00
FGD4	100.00	PDHB	100.00	WNK1	100.00
FGF12	100.00	PDHX	100.00	WWOX	100.00
FGF14	100.00	PDK3	100.00	XRCC1	100.00
FH	100.00	PDP1	100.00	XRCC4	100.00
FHL1	100.00	PDSS1	100.00	YARS	100.00
FIBP	100.00	PDSS2	100.00	YARS2	100.00
FIG4	100.00	PDX1	100.00	YWHAE	100.00
FKRP	100.00	PDYN	100.00	ZBTB18	100.00
FKTN	100.00	PET100	100.00	ZBTB20	100.00
FLAD1	100.00	PEX1	100.00	ZBTB24	100.00
FLNA	100.00	PEX10	100.00	ZC3H14	100.00
FLNC	100.00	PEX11B	100.00	ZC4H2	100.00
FLVCR1	100.00	PEX12	100.00	ZDHHC15	100.00
FLVCR2	100.00	PEX13	100.00	ZDHHC9	100.00
FMN2	100.00	PEX14	100.00	ZEB2	100.00
FOLR1	100.00	PEX16	100.00	ZFYVE26	100.00
FOXG1	100.00	PEX19	100.00	ZFYVE27	100.00
FOXI1	100.00	PEX2	100.00	ZIC2	100.00
FOXP1	100.00	PEX26	100.00	ZMYND11	100.00
FOXRED1	100.00	PEX3	100.00	ZNF335	100.00
FRAS1	100.00	PEX5	100.00	ZNF423	100.00
FREM2	100.00	PEX6	100.00	ZNF711	100.00
FRMD4A	97.02	PEX7	91.48		

DUMMY REPORT

Page 16 of 17
Name/Sample ID: XXXX/XX

CAP ACCREDITED
COLLEGE of AMERICAN PATHOLOGISTS

Fig 8: Last page of DNA test report

5. WHAT DID I LEARN?

5.1 SUMMARY

Genomic diagnosis and testing has benefits with high potential whether the results are positive or negative for a gene mutation. Test results can provide a sense of relief from uncertainty and help people make informed decisions about managing their health care. For example, a negative result can eliminate the need for unnecessary checkup and screening tests in some cases. A positive result can direct a person toward available prevention, monitoring, and treatment options. Some test results can also help people make decisions about having children. Genetic disorders can be detected in newborn screening early in life so that treatment can be started as early as possible.

Whole genome sequencing or clinical exome sequencing and analysis results in identifying and interpreting variations which has deleterious effects and has high impact variations which leads to a disease. These null variants are classified following the ACMG guidelines and are reported. And whenever the clinicians are not satisfied with the variants reported they ask for further re-analysis based on some new clinical features. In the revised analysis we search for any new variants associated with the clinical features provided by the clinician. In some cases, we got new variants or the previously reported variant classification was changed e.g., from variant of uncertain significance (VUS) to pathogenic or likely pathogenic. This is due to over years pipelines, databases, analysis software, knowledge are updated, and this could result in the identification of new variants or revision of reported variant significance.

The Standards and Guidelines were developed primarily as an educational resource for clinical laboratory geneticists to help them provide quality clinical laboratory services. The American College of Medical Genetics and Genomics (ACMG) previously developed guidance for the interpretation of sequence variants. In the past decade, sequencing technology has evolved rapidly with the advent of high-throughput

next-generation sequencing. These recommendations primarily apply to genetic tests used in clinical laboratories, which include genotyping, single genes, panels, exomes, genomes, etc. The guidelines also describe about the use of specific standard term such as “pathogenic,” “likely pathogenic,” “uncertain significance,” “likely benign,” and “benign”—to describe variants identified in genes that cause Mendelian disorders. Moreover, this recommendation describes a process for classifying variants into these five categories based on criteria using typical types of variant evidence (e.g., population data, computational data, functional data, segregation data). The ACMG strongly recommends that clinical molecular genetic testing should be performed in the clinical laboratory improvement amendments which should be an approved laboratory.

Exome sequencing is a valuable diagnostic tool for patients with Mendelian disorders. The yield of proband-only exome sequencing is estimated at 25%. Conceptually, cases for which exome sequencing is nondiagnostic fall into two distinct categories: those where the diagnosis lies outside the data produced (e.g., the causative mutation is in a coding region not covered by the exome, or it is noncoding, or the cause is not genetic); and those where the diagnosis lies in the available genetic data, but incomplete recognition of the phenotype or limitations of current tools or knowledge hinder discovery. As tools and knowledge related to gene–disease associations improve, we will be able to diagnose more cases in the latter category. The primary reason that we find new diagnoses is the growing knowledge of gene–disease and variant–disease associations in the literature. A nondiagnostic exome could be solved by any one new publication. We show that the pace of publication is quite rapid, with around 250 new gene–disease associations and 9,200 variant–disease associations curated each year.

Frequent re-evaluation of exomes is challenging in practice because of the required labour. Clinical laboratories can devote 20 to 40 h of expert labour to issue an initial clinical exome report. Effort may accordingly be prioritized to processing new exomes

rather than reanalysing old ones, as the benefit and likelihood of identifying a new diagnosis must be balanced against the cost of reanalysis. Many clinical laboratories do reanalyse exomes upon a provider's request, but there is often a limit to the number of requests or an associated fee[12]. Increased automation could lessen the expense and shift the cost-benefit calculation toward more frequent reanalysis. Efforts should focus on automating as much of the analysis process as is feasible. In particular, standard approaches are needed to encode patient phenotypes and to measure objectively the relevance of a gene or disease to a patient's phenotype[32]. Such systems will depend on databases of well-substantiated gene–disease and variant–disease associations represented in structured ontologies. Thus, it is also important to develop systems to keep databases up to date with well-curated information from the current literature.

Both the initial analysis of exome data and its subsequent reanalysis may benefit from collaboration between the clinical laboratory and the ordering provider. Laboratories rely on providers for phenotypic data and for updates to this information to be considered at reanalysis. Providers rely on laboratories to report relevant variants and to update variant interpretation based on new information. Practical considerations suggest that benefits may follow from ordering providers initiating reanalysis. In requesting reanalysis, the provider may send an update on the patient phenotype to the laboratory. The request may also serve to confirm that the provider (i) is expecting the results of reanalysis and (ii) is in contact with the patient. Nonetheless, there are expected to be occasions when a diagnosis is found on reanalysis by a clinical laboratory without a request from an ordering provider. It may therefore be helpful for undiagnosed patients who have undergone exome sequencing to ensure providers have up-to-date contact information.

My experience illustrates that a "negative" nondiagnostic result from exome sequencing does not mean that the disease etiology lies outside of the data already

produced. For patients with a high suspicion of a Mendelian disorder, providers should periodically request a reevaluation of exome data by clinical laboratories, including in the absence of new phenotypic findings. Larger studies may be helpful to define a standard practice for the timing of reanalysis, taking into account the cost of reanalysis and the evolving rate of discovery of gene–disease relationships. Furthermore, providers should weigh policies regarding reanalysis, along with cost of testing and turnaround time, in selecting a laboratory for exome sequencing.

6. HOW TO COMPLEMENT IN CORPORATE?

Ambitious employees always find opportunities to climb up the corporate ladder. There is no doubt that the individuals who are eager to climb the professional ladder of success, but the thing matters is whether they have the discipline and the positive interest to reach the desired goal. Successful individuals always keep track of great principles, which would help them to reach the highest level.

- **Taking Initiative**- One should be more involved in any difficulty his colleague or boss has encounter. One should always be ready for any sort of work which has nothing to do with his work, and that's what makes the difference between a leader and a common employee. One should always be ready to help.
- **Being always prepared**- Preparation which is good always helps an individual to stay focused on your agenda. It also shows the managers that he is serious in his work, and he takes company meetings seriously. Avoiding being careless at the meeting, with no preparation whatsoever. Good preparations may lead to promotion.
- **Being Responsible**- A good employee is a responsible employee. Adding up responsibilities should be one's forte. He should also be ready to undertake more responsibilities if necessary. When he fails at certain task assigned to him, he should be the one standing up to admit his fault and ask for the apology. Managers are always on the lookout for such a person who is not only ready to shoulder any responsibilities, but willing to bear the consequences of his failure or irresponsibility.
- **Never Postpone**- He should finish his task on time. He should ensure that the tasks must be completed before anything. Postponing of work is looked upon as less dedicated, and not focused.
- **Effective Communication**- Proper communication is an essential tool to successful career. If someone is able to put his thoughts and opinion across in

an effective manner, he will get the most attention. His communication skills will lead him to sit among bosses and managers. He will be asked to share his views in company's meetings. He should ask politely, state clearly, say friendly, share his opinions carefully. When he is angry, he should try to say slowly.

7. REFERENCES

1. Yang Y, Muzny DM, Xia F, et al. Molecular findings among patients referred forclinical whole-exome sequencing. JAMA 2014; 312:1870–1879.
2. Yang Y, Muzny DM, Reid JG, et al. Clinical whole-exome sequencing for thediagnosis of mendelian disorders. N Engl J Med 2013; 369:1502–1511.
3. Lee H, Deignan JL, Dorrani N, et al. Clinical exome sequencing for genetic identification of rare Mendelian disorders. JAMA 2014; 312:1880–1887.
4. Iglesias A, Anyane-Yeboa K, Wynn J, et al. The usefulness of whole-exome sequencing in routine clinical practice. Genet Med 2014; 16:922–931.
5. Retterer K, Juusola J, Cho MT, et al. Clinical application of whole-exome sequencing across clinical indications. Genet Med 2016; 18:696–704
6. Cornish A, Guda C. A Comparison of Variant Calling Pipelines Using Genome in a Bottle as a Reference. Biomed Res Int 2015; 2015:456479.
7. McCarthy DJ, Humburg P, Kanapin A, et al. Choice of transcripts and software has a large effect on variant annotation. Genome Med 2014; 6:26.
8. Baumgartner WA Jr, Cohen KB, Fox LM, Acquaah-Mensah G, Hunter L. Manual curation is not sufficient for annotation of genomic databases. Bioinformatics 2007;23: i41–i48.
9. Chong JX, Buckingham KJ, Jhangiani SN, et al.; Centers for Mendelian Genomics. The Genetic Basis of Mendelian Phenotypes: Discoveries, Challenges, and Opportunities. Am J Hum Genet 2015; 97:199–215.
10. De Rubeis S, He X, Goldberg AP, et al.; DDD Study; Homozygosity Mapping Collaborative for Autism; UK10K Consortium. Synaptic, transcriptional and chromatin genes disrupted in autism. Nature 2014; 515:209–215.
11. Dewey FE, Grove ME, Pan C, et al. Clinical interpretation and implications of whole-genome sequencing. JAMA 2014; 311:1035–1045.

12. Jamal SM, Yu JH, Chong JX, et al. Practices and policies of clinical exome sequencing providers: analysis and implications. Am J Med Genet A 2013;161A:935–950.

13. Richards S, Aziz N, Bale S, et al.; ACMG Laboratory Quality Assurance Committee. 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. Genet Med 2015; 17:405–424.

14. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWAMEM. ArXiv13033997 Q-Bio March 2013. http://arxiv.org/abs/1303.3997. Accessed 14 January 2016.

15. Broad Insitute. Picard Tools. http://broadinstitute.github.io/picard/. Accessed 14 January 2016.

16. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 2011; 43:491–498.

17. Girdea M, Dumitriu S, Fiume M, et al. PhenoTips: patient phenotyping software for clinical and research use. Hum Mutat 2013; 34:1057–1065.

18. Köhler S, Doelken SC, Mungall CJ, et al. The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. Nucleic Acids Res 2014;42(Database issue): D966–D974.

19. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010;38: e164.

20. Yates A, Akanni W, Amode MR, et al. Ensembl 2016. Nucleic Acids Res 2016;44(D1): D710–D716.

21. Exome Aggregation Consortium, Lek M, Karczewski K, et al. Analysis of proteincoding genetic variation in 60,706 humans. bioRxiv 2015:030338.

22. 1000 Genomes Project Consortium, Auton A, Brooks LD, et al. A global reference for human genetic variation. Nature 2015; 526:68–74.
23. 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 Res 2015;43(Database issue): D789–D798.
24. Stenson PD, Ball EV, Mort M, Phillips AD, Shaw K, Cooper DN. The Human Gene Mutation Database (HGMD) and its exploitation in the fields of personalized genomics and molecular evolution. Curr Protoc Bioinforma Ed Board Andreas Baxevanis Al 2012; Chapter 1: Unit1.13.
25. Jones WD, Dafou D, McEntagart M, et al. De novo mutations in MLL cause Wiedemann-Steiner syndrome. Am J Hum Genet 2012; 91:358–364.
26. Bramswig NC, Lüdecke HJ, Alanay Y, et al. Exome sequencing unravels unexpected differential diagnoses in individuals with the tentative diagnosis of Coffin-Siris and Nicolaides-Baraitser syndromes. Hum Genet 2015; 134:553–568.
27. Vulto-van Silfhout AT, Rajamanickam S, Jensik PJ, et al. Mutations affecting the SAND domain of DEAF1 cause intellectual disability with severe speech impairment and behavioral problems. Am J Hum Genet 2014; 94:649–661.
28. Vissers LE, de Ligt J, Gilissen C, et al. A de novo paradigm for mental retardation. Nat Genet 2010; 42:1109–1112.
29. Rauch A, Wieczorek D, Graf E, et al. Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. Lancet 2012; 380:1674–1682.
30. Rice GI, del Toro Duany Y, Jenkinson EM, et al. Gain-of-function mutations in IFIH1 cause a spectrum of human disease phenotypes associated with upregulated type I interferon signaling. Nat Genet 2014; 46:503–509.
31. Thauvin-Robinet C, Auclair M, Duplomb L, et al. PIK3R1 mutations cause syndromic insulin resistance with lipoatrophy. Am J Hum Genet 2013;93: 141–149.

32. Akawi N, McRae J, Ansari M, et al.; DDD study. Discovery of four recessive developmental disorders using probabilistic genotype and phenotype matching among 4,125 families. Nat Genet 2015; 47:1363–1369.
33. Iossifov I, O'Roak BJ, Sanders SJ, et al. The contribution of de novo coding mutations to autism spectrum disorder. Nature 2014; 515:216–221.