

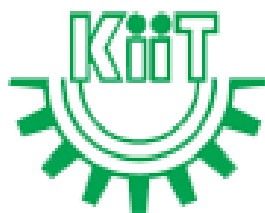
TARGETED GENE SEQUENCING AND ANALYSIS OF HEMATOLOGICAL MALIGNANCIES BY NGS

Dissertation submitted in partial fulfilment for the degree of

Master of Science in Biotechnology

Submitted By

Sukanya Sarkar



School of Biotechnology, Campus- 11

KIIT Deemed to be University

Bhubaneswar, Odisha, India

Under the Supervision of

Dr. Vidya H. Veldore

MedGenome Labs Pvt. Ltd.

3rd Floor, Narayana Netralaya Building, Narayana Health City,
#258/A, Bommasandra, Hosur Road, Bangalore – 560 099, India.

CERTIFICATE

This is to certify the dissertation entitled "**Targeted Gene Sequencing and analysis of hematological malignancies by NGS**" submitted by **Sukanya Sarkar** in partial fulfilment of the requirement for the degree of Master of Science in Biotechnology, KIIT Deemed to be University, Bhubaneswar bearing Roll No. **1661042** & Registration No. 16647151468 is a bonafide research work carried out by her under my guidance and supervision from 8th January 2018 to 11th January 2018.

Name and signature

DECLARATION

I hereby declare that the dissertation entitled "*Targeted Gene Sequencing and analysis of hematological malignancies by NGS*" submitted by me, for the degree of Master of Science to KIIT Deemed to be University is a record of bonafide work carried by me under the supervision of, **Dr. Vidya H. Veldore** Principal Scientist, Oncology department, MedGenome Labs Pvt. Ltd., Bangalore, India.

Date:

Name & Signature

Place:

ABSTRACT

Recent studies in genomic analysis have rapidly advanced the diagnostics of hematological malignancies. The combination of genomic single nucleotide variations copy number variants, genomic methylation pattern and novel fusion transcripts specific for cancer, helps in achieving high accuracy in predicting clinical outcomes of hematological cancers. Some of these genetic variations trigger new pathways that are essential for carcinogenesis in multiple human cancers and create novel functions that are not present in healthy individuals. Some of these novel variants are highly targetable. Analyses of these mutated genes helps in the diagnosis and prognosis that suggest targeting therapy for some unique protein changes of genes which can be an effective in treating for human cancers, thus improving the survival rate of the effected patients. The NGS based multi-gene analysis, allows to sequence and identify variants associated with multiple genes with diagnostic, prognostic and therapeutic implications in different cancer types; the focus of this project is based on panel designed to screen 57 cancer related genes associated with tumorigenesis, prognostication and predictive value in chemotherapy and targeted drug therapy in different hematological malignancies. Targeted sequencing represents a cost-effective approach with the ability to detect specific variants causing protein-coding changes in individual human genomes. These multi-gene, affordable tests will enable personalized treatment by matching the patient's tumor with the appropriate drug based on the mutational findings.

ACKNOWLEDGEMENT

This dissertation though an individual work, has benefitted in various ways from several people. I am thankful to every one of them.

The enthusiastic guidance from my manager Dr. Vidya H. Veldore inspired me to stretch beyond my limits and learn every new aspect of the project. Her profound insights have guided my thinking and improve the final product. She has been a great mentor for me in all aspects. I am grateful to her.

I would like to thank Dr. Harendra Modak and Tejaswi Routhu to teach me the analysis of genome and the other aspects of it and devotedly helping me in learning and for the support. My sincere thanks to Dr. Bagavathi Kausalya and Dr. Abdul Mueed Bidchol for teaching me the reporting and variant filtration and table formation.

I am grateful to Dr. Vivek Gopalan, Praveen Raj and the Clinical Bioinformatics team for the constant helps and teaching the use of IGV and giving the detailed information about the bioinformatics pipelines. I would like to thank Dr. Pawan Upadhyay for teaching me the basics of NGS and the pipelines.

My humble acknowledgement to the COO of MedGenome Labs Dr. Ramprasad V.L and the Associate Director Dr. Sakthiveil M.S. for giving me the opportunity to work with the Oncology team A very special thanks to the fellow teammate Somi Srivastava and Manali Kamath for their sincere help also the batchmates working in different teams to keep me motivated.

A very special thanks to my parents and the Institute and my college and Dr.Mrityunjaya Suar for trusting me and helping me to reach where I am today.

CONTENTS

Certificate.....	ii
Declaration.....	iii
Abstract.....	1
Acknowledgement.....	2
Contents.....	3-4
Abbreviations.....	5-6
List of Figures.....	7-8
1. Introduction	
1.1 Background and Context.....	
1.2 Scope and Objective.....	
1.3 Achievement.....	
1.4 Overview of Dissertation.....	
2. Background and Rationale	
2.1 Usefulness of NGS.....	
2.2 Hematological Malignancies.....	
2.3 Types of Hematological Malignancies.....	
2.3.1 Other types of Leukemia.....	
2.3.2 2.3.2 Different levels of genetic information captured based on NGS technology.....	
2.3.3 Bioinformatic analysis for the NGS raw data.....	
3. Things Learnt	
3.1 Summary	

3.1.1 Aims and Objective.....

3.1.2 Materials and Methods.....

3.1.3 Laboratory Methods.....

3.1.4 NGS.....

3.1.5 Bioinformatics pipelines.....

3.1.6 Genome Analysis.....

3.1.7 IGV Verification.....

3.1.8 Genome Analysis Method.....

3.1.9 OncoMiner.....

3.1.10 Variant Classification.....

3.1.11 Clinical Report

3.1.12 Result.....

3.2 Evaluation of Result.....

3. How to Complement in Corporate.....

References.....

ABBREVIATION

DNA	De-oxyribo Nucleic Acid
RBC	Red Blood Cell
WBC	White Blood Cell
NGS	Next Generation Sequencing
IGV	Integrated Genome Viewer
SOP	Standard Operating Procedure
AML	Acute Myeloid Leukemia
ALL	Acute Lymphoblastic Leukemia
CML	Chronic Myeloid Leukemia
CLL	Chronic Lymphoblastic Leukemia
MDS	Myelodysplastic Syndrome
MPN	Myeloid Neoplasms
CMML	Chronic Myelo Monocytic Leukemia
JMML	Juvenile Myelo Monocytic leukemia
RNA	Ribonucleic Acid
chIp	Chromatin Immuno-Precipitation
TCR	T-Cell Receptor
CNV	Copy Number Variants
WGS	Whole Genome Sequencing
WES	Whole Exome Sequencing
GiAB	Genome in A Bottle
SNP	Single Nucleotide Polymorphism
Indel	Insertion -Deletion
GWAS	Genome Wide Association Study
TCGA	The Cancer genome Atlas
SAM	Sequence Alignment/ Map
BAM	Binary Alignment Map
BWA	Burrow-Wheeler Algorithm
VCF	Variant Calling Format

LIMS	Laboratory Information Management System
SS-ACR	Splice Site Acceptor
SS-DNR	Splice Site Donor
SS-ACR-PRX	Splice Site Acceptor Proximal
SS-DNR-PRX	Splice Site Donor Proximal
FTP	File Transfer Protocol
TRF	Test Requisition Form
DOCM	Database of Curated Mutation
FDA	Food and Drug Administration(U.S)

LIST OF FIGURES

Fig 1: List of Blood related diseases.....	
Fig 2 : Evolution of NGS in the field of genomic studies.....	
Fig 3 : Step-wise representation of Clinical report generation.....	
Fig 4 : DNA Library Preparation.....	
Fig 5 : Sample preparation for Library generation.....	
Fig 6 : Cluster Generation.....	
Fig 7 : Sequencing of the generated templates.....	
Fig 8 : Reading of aligned sequence.....	
Fig 9 : Flowchart of Bioinformatics pipeline.....	
Fig 10 : Flowchart of BAM file to VCF file conversion.....	
Fig 11 : IGV snapshot showing TRUE variants.....	
Fig 12 : IGV snapshot showing INSERTION.....	
Fig 13 : IGV snapshot showing DELETION.....	
Fig 14 : Snapshot of Oncominer.....	
Fig 15 : Databases to rely on for clinical reporting.....	
Fig 16 : Snapshots showing FLT3 variants reported in JAX and Cbioportal.....	
Fig 17 : Snapshots showing <i>NPM1</i> variants reported in databases like Cbioportal and JAX.....	
Fig 18 : Snapshots showing <i>JAK2</i> variant reported in databases like My Cancer and DoCM.....	

Fig 19 : Snapshot from JAX showing *TET2* variant.....

Fig 20 : Snapshot of Cbioportal showing NRAS variant.....

Fig 21 : Snapshot showing *U2AF1* variant in JAX.....

Fig 22 : Snapshot showing *PTPN11* variant in Cbioportal.....

1. INTRODUCTION

Cancer remains one of the most lethal diseases for human. It has become the second leading cause of death in the world. Cancer incidence rates are growing rapidly globally. Most of the increase is mainly expected to be concentrated in the developing countries due to heavy smoking, poor diet, physical inactivity and environmental pollution. In principle, cancers arise from a group of cells that has gained mutation and cannot be changed by the DNA repair system, further mutation occurs on tumor suppressor genes, oncogenes gaining functional mutation. DNA mismatch repair will further accumulate mutations in cell populations^[1]. However, all cancers are characterized by uncontrolled growth of abnormal cells, invading surrounding tissues, metastasize (spread to distant sites). With early diagnosis with regular genetic screening and removal of precancerous growth can provide total cure in many cancers. Cancer mortality rate has decreased to 1.8% in males and 1.5% in females in recent 5 years due to development in the field of diagnostic instrumentation and progress in therapeutics ^[2].

1.1 BACKGROUND AND CONTEXT:

Malignant disorder of blood is known as hematological malignancy. Among the different types of cancers, hematological malignancies account for a substantial percentage in worldwide^[4]. Hematological cancers are heterogenous group of cancer of Blood, Bone marrow, lymph nodes. This type of cancer is derived from either of the two blood lineages: **Myeloid** and **Lymphoid** cell lines^[5].

DISORDERS	PATHOLOGY	DISEASES
Erythrocyte	Increased RBC	Polycythemia
	Decreased RBC	Anemia
Leukocyte	Increased WBC (non-malignant)	Eosinophilia Infectious Mononucleosis Sepsis
	Decreased WBC (non-malignant)	Leukopenia

	Malignant WBC disorders	Leukemia Lymphomas
Homeostatic	Quantitative Platelet Disorder	Primary Thrombocythemia Allergic purpura
	Coagulation Disorder	Hemophilia
	Vascular Disorder	Purpura Simplex

Fig 1. List of Blood related diseases

Even though leukemias are very common in children, in India, 90% of all the Leukemia are diagnosed into adults. The high mortality rate of Leukemia is mainly due to late diagnosis and mostly because the symptoms of Leukemia mimics those of the other diseases. Leukemia is the most common Hematological malignancies in India³¹.

➤ **NGS:**

Next-generation sequencing (NGS), also known as high-throughput sequencing, is a recent technology that allows us to sequence DNA and RNA much more quickly and cheaply than the previously used Sanger sequencing, and as such have revolutionized the study of genomics and molecular biology.

➤ **IMPORTANCE OF RISK STRATIFICATION OF SOMATIC MUTATION IN HEMATOLOGICAL MALIGNANCIES:**

Prognosis of patients with Hematological malignancies can be calculated using several scoring systems, which includes analysis of peripheral cytopenias, percentage of blasts in the bone marrow and cytogenetic characteristics, somatic mutations thus can help define prognosis and should be considered as new prognostic factors. Current prognostic models are mainly based on clinical and hematologic parameters, but innovative models that include genetic data are being developed for both clinical and trial settings.

Therapy is selected based on risk, transfusion needs, percent of bone marrow blasts and cytogenetic and mutational profiles. Goals of therapy are different in lower risk patients than in higher risk. In lower risk, the goal is to decrease transfusion needs and transformation to higher risk disease, as well as to improve survival. In

higher risk, the goal is to prolong survival. Current available therapies include growth factor support, hypomethylating agents, intensive chemotherapy and allogeneic stem cell transplantation⁶¹.

1.2 SCOPE AND OBJECTIVES:

Taking Genome Analysis as a Dissertation project has taught me about the important outcomes of NGS and Bioinformatics in the field of Hemato-oncology, as it is a constantly evaluating path; in every step there is something new to learn. As a student we generally restricted ourselves within the circumference of the molecular background of cancers. Genomic Analysis is totally a new thing for me that came up with the molecular diagnostics of the cancer, variant filtration, variant calling, checking the variants in IGV and making clinical reports and sending to the clinician for immediate treatment based on the mutational findings.

1.3 ACHIEVEMENTS:

During this project work I have learnt to identify true variants for different HEME samples and use the link between NGS and Bioinformatics to verify the variants and use those in Clinical reports.

1.4 OVERVIEW OF DISSERTATION

Rest of the thesis is as follows----

CONTEXT OF THESIS this part describes the usefulness of NGS and different types of Hematological Malignancies; types of Leukemias.

MATERIALS AND METHODS this part deals with how the project was carried off by learning to go through the patient's Clinical details minutely and by the help of the Bioinformatics tools, downloading the files into FileZilla format ; filtering the variants and choose the true mutation using the OncoPrint and checking the gene coverage, depth and visualizing the gene in IGV (Integrated Genome Viewer), and making clinical reports according to that within the TAT (Turn Around Time). Following the SOP (Standard Operating Procedure).

RESULT & DISCUSSION in this the focus is on the cases I have gone through following the methods and the outcome I got from the project and what

contribution I gave and what kind of limitation I have faced during the project is mentioned; how much professional I have become working in a corporate company.

COMPLEMENTATION IN CORPORATE WORLD this part fully talks about my experience and views about a corporate world.

CONCLUSION the chapter provides the concluding remarks with a stress on the total proposed schemes.

The contributions made in each chapter are discussed in sequel, which include proposed schemes, their simulation results and performance analysis.

2. BACKGROUND AND RATIONALE:

NGS IN ONCOLOGY:

The fundamental premise of cancer genomics is that cancer is caused by somatically acquired mutations, and consequently it is a disease of the genome. Although capillary-based cancer sequencing has been ongoing for over a decade, these investigations were limited to relatively few samples and small numbers of candidate genes. With the advent of NGS, cancer genomes can now be systemically studied in their entirety, an endeavor ongoing via several large-scale cancer genome projects around the world, including a dedicated cancer genome project. This may provide many benefits including a more precise diagnosis and classification of the disease, more accurate prognosis, and potentially the identification of ‘drug-able’ causal mutations. Individual cancer sequencing may, therefore, provide the basis of personalized cancer management^[7]. Currently pilot projects are underway using NGS of cancer genomes in clinical practice, mainly aiming to identify mutations in tumors that can be targeted by mutation-specific drugs^[9].

2.1 Usefulness Of NGS:

1. Next-Generation Sequencing has changed the way of carrying out molecular biology and genomic studies.
2. It has allowed to sequence and annotate genomes at a much faster rate.
3. It has allowed us to study variation, expression and DNA binding at a genome-wide level.
4. Clearly detects single nucleotide variant, deletions, insertions and structural variation.^[8]

2.2 Hematological Malignancy:

In hematological malignancies, the normal blood cell development process is interrupted by uncontrolled growth of an abnormal type of blood cell. This abnormal or cancerous blood cells prevents the normal blood functions.

- There are three main types of blood cancers----

- **Leukemia:** this type of blood cancer is generally found in blood and bone marrow, caused by rapid production of abnormal WBC.
Leukemia can be **Chronic** or **Acute**. Chronic Leukemia progresses slowly but in acute leukemia immediate treatment is required.
Leukemia can be differentiated according to the cell types---
- **Lymphatic Leukemia:** this a type of blood cancer where the lymphatic system gets affected, removes excess fluids from the body and produces immune cells. The abnormal lymphocytes become lymphoma cells, which multiply and blocks the lymph nodes resulting in impairing the immune system.
- **Myeloid Leukemia:** this is a cancer of plasma cells, leading towards weak immune system and susceptibility to infection.^[10]

2.3 Types of Hematological Malignancies:

- There are four main types of leukemia:
- **AML (Acute Myeloid Leukemia):** Acute Myeloid Leukemia specifically occurs in blood and bone marrow.
- Tiredness, weight loss, bleeding gums, flu and fatigue is normal symptom of AML.
- Average age of a person having AML is 67 years.
- **ALL (Acute Lymphocytic leukemia):** Acute Lymphocytic Leukemia is a very common childhood cancer, due to overgrowth of Lymphocytes.
- In children the disease if treated can be a chance of remission, but there is no chance for the adults for remission of the disease.
- The symptoms are--- hepatomegaly, splenomegaly, lymphadenopathy, weakness, petechiae, bruises and paleness etc.
- **CML (Chronic Myeloid Leukemia):** this type of cancer starts in the blood forming tissue.
- This is slower but as life threatening as AML.
- Symptoms are: fatigue, fever, swollen spleen.
- **CLL (Chronic Lymphocytic Leukemia):** This type of cancer generally affects people over 55 years.
- This cancer affects the B- Lymphocytes.

2.3.1 Some other types of Leukemias are----

- **MDS (Myelodysplastic Syndrome):**
- In this disease bone marrow cells fail to produce healthy RBC (Red Blood Cell), WBC (White Blood Cell) or platelets.
- Bone marrow produces under developed cells, with abnormal shape and size; these cells are known as Blast Cells.
- In high risk condition MDS can become AML (Acute Myeloid Leukemia).
- Symptoms are--- paleness, weight-loss, loss of appetite, anemia.^[11]
- **MPN (Myeloproliferative Neoplasms):**
- In this disorder, the bone marrow stem cells get affected.
- The affected stem cells overproduce any one kind of cells among RBC, WBC or platelets.
- MPN can progress gradually.
- Symptoms are—headache, dizziness, pruritus, hypertension, fatigue etc.
- Types of MPN are---Polycythemia Vera(PV), Essential Thrombocythemia(ET), Myelofibrosis.^[12]
- **CMML (Chronic Myelo Monocytic Leukemia):**
- It is a type of cancer where the blood forming cells of bone marrow invades the blood.
- It generally affects the adult population.
- Symptoms are---thrombocytopenia, anemia, leucopenia, fatigue, petechiae etc.^[13]
- **JMML (Juvenile Myelo Monocytic Leukemia):**
- It is a mixture of MDS/MPN leukemias, generally affecting children below 2 years.
- More common in males.
- It takes place due to overproduction of Myelomonocytes.
- Symptoms are—polycythemia, thrombocytopenia, anemia, neutropenia etc.^[14]

2.3.2 Different levels of genetic information captured based on NGS technology:

WHOLE GENOME SEQUENCING: Whole Genome Sequencing (WGS) provides complete information on genetic makeup of organisms, helps to study variations within the species and compare them with others which can help in solving unanswered puzzles in the genomics and providing information for research for biomarker discovery and tailor-made medicine. WGS can reveal germline variation, somatic variation, copy number variation, changes in transposable elements, and structural variants.

WHOLE EXOME SEQUENCING: Whole Exome Sequencing sequences the protein coding part of the genome (exomes) rather than the complete genome because the protein coding regions of the human genome constitute about 85% of the disease-causing mutations. Exome sequencing is done to identify the functional variation that is responsible for both Mendelian and common diseases.

RNA SEQUENCING: RNA sequencing (RNA-Seq) is a powerful technique that allows researchers to study the complete transcriptome profiling which gives information about the functional components of the genome, molecular components of cell, tissues and knowledge on disease development.

miRNA SEQUENCING: miRNA-seq helps in the study of tissue-specific expression patterns, isoforms of miRNAs, disease relationships. With unprecedented sensitivity and dynamic range, small RNA sequencing methods allow for the most accurate detection and quantification of rare small RNA sequences.

ChIP-SEQ: ChIP-seq is a powerful method to map transcription factor binding sites and histone modification status on a genome-wide scale, including include cross-linking a protein to chromatin, shearing the chromatin, using a specific antibody to precipitate the protein of interest with its associated DNA, reverse cross linking and finally purifying the associated DNA fragments.

METHYL SEQUENCING: Methyl sequencing is an approach developed to study the cysteine methylation patterns across the genome.

TCR SEQUENCING: Productive T-cell activation results in the clonal expansion of a specific T-cell and this expansion can be accurately determined by TCR sequencing. It can also help in sequencing the CDR regions and can have a potential therapeutic outcome in the field of cancer.

SINGLE CELL SEQUENCING: Single cell sequencing enables the study of sequence information from discrete cells with next generation sequencing (NGS) platforms such as Illumina's HiSeq, and NovaSeq. For samples that have a heterogeneous microenvironment, this yields a higher resolution understanding of the

inherent cellular differences even at low quantity of target DNA/RNA. Single cell techniques have found various applications that includes detection of DNA mutations, copy-number variants (CNVs), DNA-protein binding, RNA splicing, and the measurement of mRNA expression^[15].

2.3.3 Bioinformatic analysis for the NGS raw data:

After sequencing, automated bioinformatics pipelines and programs analyze large-scale biological data sets with special emphasis on high volume NGS data. All optimized parameters to ensure fast on-time delivery without compromising data accuracy or quality.

The bioinformatics pipeline for WGS and WES strictly adheres to the gold standard set by Genome in a Bottle Consortium (GIAB), who have created reference genotypes with truth values for single nucleotide polymorphisms (SNPs) and insertions/ deletions (Indels). The pipeline uses GATK best practices for germline variant calling and captures several quality metrics during the analysis to deliver the best quality results. With the help of in-house tools (VariMat, Varminer) and databases (OncoMD, MedVarDb) along with available databases like ExAc, COSMIC, OMIM, 1000Genome, ClinVar, SwissVar and GWAS ^[16].

3. THINGS LEARNT:

3.1 SUMMARY

3.1.1 AIMS AND OBJECTIVE:

Due to the remarkable effort of bioinformatics to identify the basis of hematological malignancies, there is exponential growth in high-throughput sequencing and thus providing access to affordable genetic testing. NGS encompasses different aspects of genomics, transcriptomics and epigenomics. By the help of NGS and Bioinformatics the copy number variations, deletions, single nucleotide variants can be easily identified. Moreover, the amount of DNA required is very less for library preparation for NGS. With the help of pair end sequencing, the position of structural changes, duplications can be easily identified and that too with smaller insert sizes^[18]. Low coverage of the genome can be enough for the detection of structural abnormalities of the genome along with filtering out the mutated site that can be a possible area for further targeted therapy or treatment. The WGS (Whole Genome Sequencing) has brought a huge evolution in the field of Genome Analysis of Hematological malignancies that can lead to the correct interpretation of mutated genes, even with the evolution we can easily compare and identify unique mutations in complex pathways, pathology associated with the specific genetic variation as well as the heterogeneity of the specific malignancy^[19]. This NGS method is also a time and cost effective method.

Some of the main challenges with NGS based testing are:, different analysis pipelines can generate different kind of results for a same sample that can be an issue for clinical malignancy as well as other patient samples. There is no standard guideline based detailed information about the bioinformatics pipelines and gene alignment tools in other than manuscripts. During the generation of data for analysis some of the data: the variant calling and variant filtration is done manually, that leads to a greater chance of mistakes.

Although having some negative sides NGS not only helps in identifying mutations but also helps in differential diagnosis of malignancies. From sequencing results, we may easily find the links between common/uncommon phenotype correlation

within hematological malignancies. It also tells us about the disease progression and even the response of drug in the diseases.

The main goal is to use the NGS, Bioinformatics in the diagnostics, prognostication and treatment of hematological malignancies in high through deep sequencing approach to identify the specific variant or marker for the given disease; arrive at a correlation between the observed variants and their relevance in any given leukemia subtype in these patients thus guiding personalized treatment and management. [19].

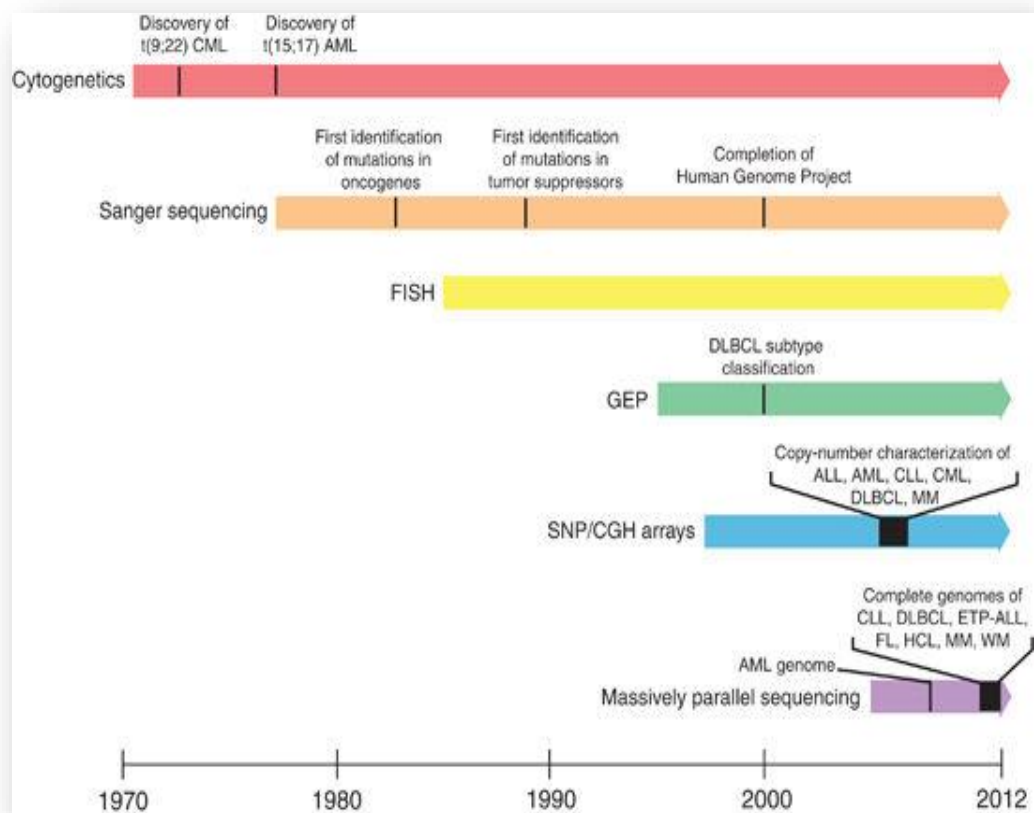


Fig 2: Evolution of NGS in the field of genomic studies

3.1.2 MATERIALS AND METHODS:

The targeted gene sequencing and analysis follows some key procedures, through which from a sample collection to NGS data generation. It is divided in some categories ----

- **Pre- analytical:** it includes sample labeling to the registration of the sample inside the laboratory.
- **Analytical:** this includes the whole lab work-flow to NGS sequencing.
- **Post-analytical:** this part covers the generation of raw data from bioinformatics to validation of the results and deriving clinically meaningful information.

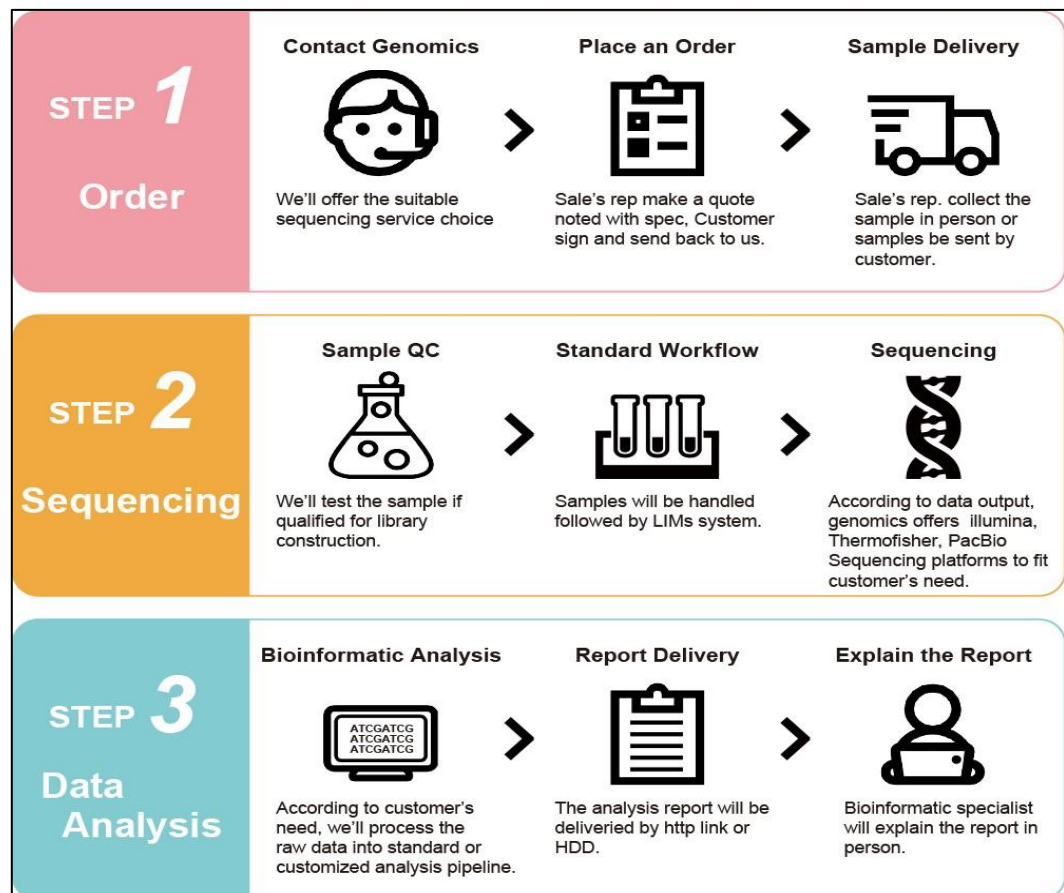


Fig 3: Step-wise representation of Clinical report generation

❖ **TARGETED GENE SEQUENCING:**

A subset of genes or regions of the genome are isolated and sequenced, which can analyze the data of a specific region of interest and with a high range of coverage value. It helps to identify rare variants, variants that would be too rare/low mutant allele depth. Targeted sequencing is enriched with probes that are involved in areas of **Cancers** and other diseases. The studies can also help in the follow-up of the disease from GWAS or WGS. With the help of target enrichment and amplicon-based process in illumina, can enable a wide range of application in discovery, screening and validation or screening of genetic variants., which can also help in finding out rare somatic mutations from complex disease conditions^[20].

3.1.3 LABORATORY METHODS:

DNA extraction and QC checking:

- Generally, the samples are **Blood in streck/EDTA tubes** or **Bone Marrow aspirates**.
- The DNA extraction and QC is verified

Library preparation:

- The isolated DNA is fragmented, and 5' adapter ligation is done.
- Reverse transcription of the DNA strand is performed.
- After that 5' adapter-oligo primer is added, and second strand synthesis is performed.
- The synthesized sequences are amplified using PCR.
- The QC stats of the sample is checked by Qubit Value.
- If the sample is failed in QC checking, the further work-flow will be stopped.
- The library preparation generally takes 7 days to complete.
- After the library is made it is subject to sequencing by NGS on Illumina platform.

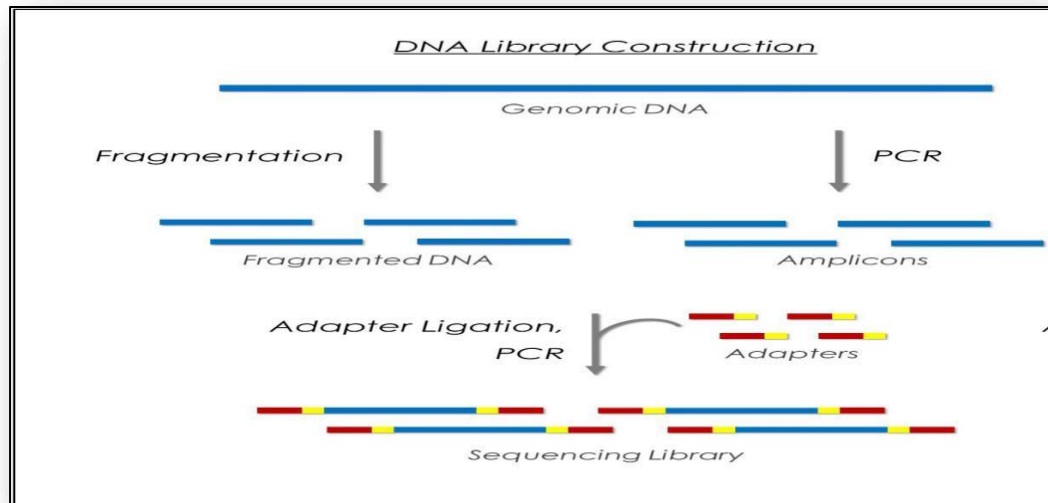


Fig 4: DNA Library Preparation

3.1.4 NGS:

- the concept behind NGS technology is, DNA polymerase catalyzes the incorporation of fluorescently labeled deoxyribonucleotide triphosphates (dNTPs) into a DNA template strand during sequential cycles of DNA synthesis. During each cycle, at the point of incorporation, the nucleotides are identified by fluorophore excitation. The critical difference is that, instead of sequencing a single DNA fragment, NGS extends this process across millions of fragments in a massively parallel fashion. More than 90% of the world's sequencing data are generated by Illumina sequencing^[15].

- **NGS illumina sequencing follows four basic steps:**

- **SAMPLE PREPARATION:**

The library is prepared by random fragmentation of genomic DNA, followed by 5' or 3' adapter ligation; if this fragmentation and ligation is done in a single step this increases the efficiency of the library preparation process. The adapter ligated sequences are both PCR amplified, and gel purified.

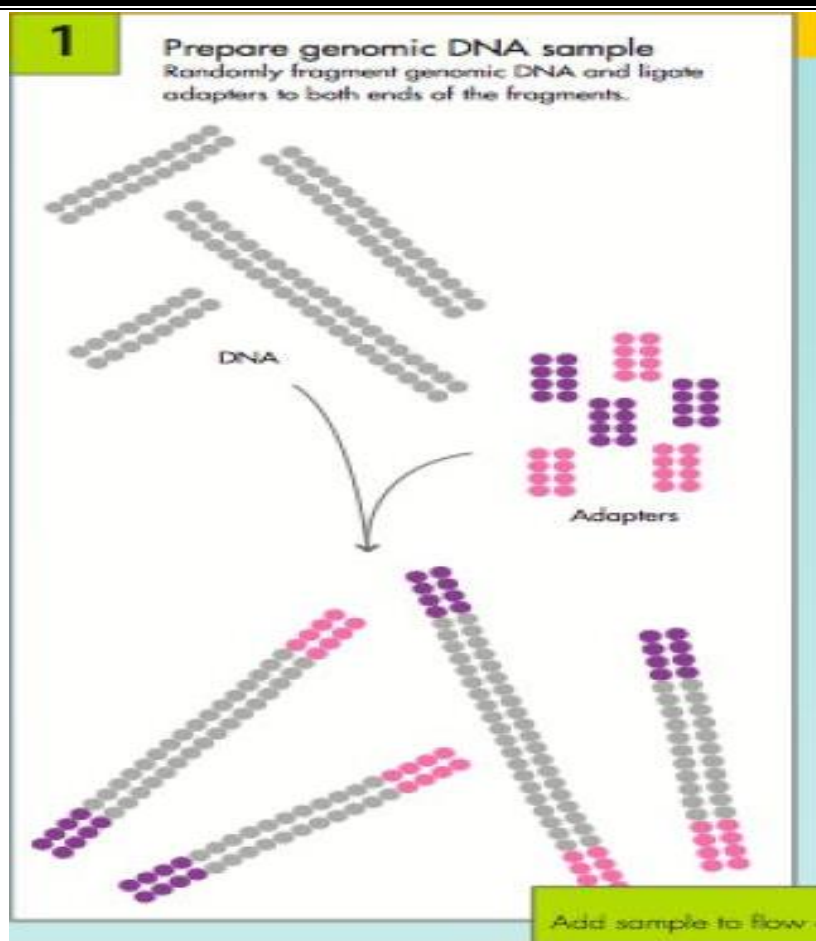


Fig 5 : Sample preparation for Library generation

▪ **CLUSTER GENERATION:**

In this step, the library is loaded into a flow cell consisting a lawn of surface bound oligos complimentary to the library adapters. After that, each cluster is amplified by bridge amplification into distinct, clonal clusters. After the cluster formation, the sequences are ready for sequencing.

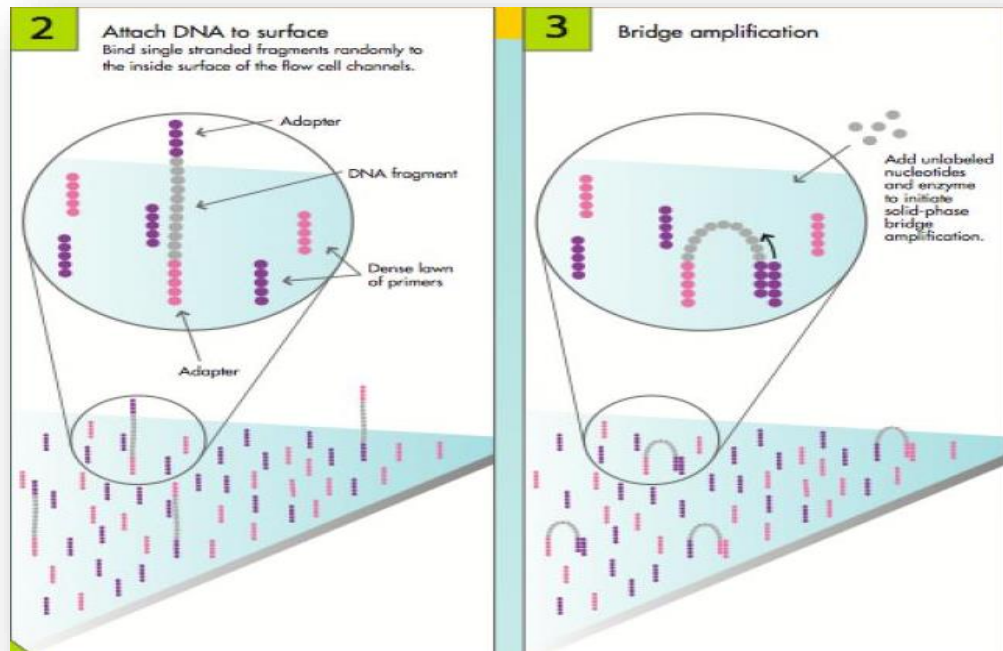


Fig 6: Cluster Generation

▪ **SEQUENCING:**

In sequencing, NGS illumina platform uses reversible terminator-based method that detects every single base as they keep on incorporate into the DNA template strand. Using of all four terminator-based nucleotides ensure that there should be no bias of incorporation and error of raw data. It also very accurate for base-by-base sequencing and eliminates the context specific data for repetitive regions and homopolymers.

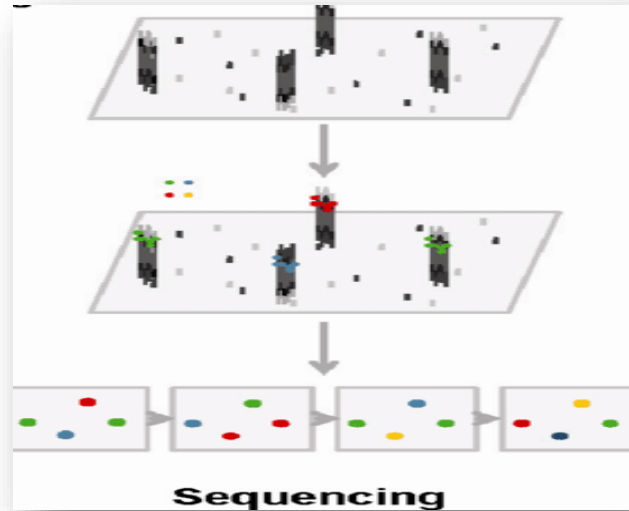


Fig 7 : Sequencing of the generated templates

▪ **DATA ANALYSIS:**

The sequenced raw data is aligned to reference genome, to interpret the presence of SNP (Single Nucleotide Polymorphism), INDEL (Insertion-Deletion).

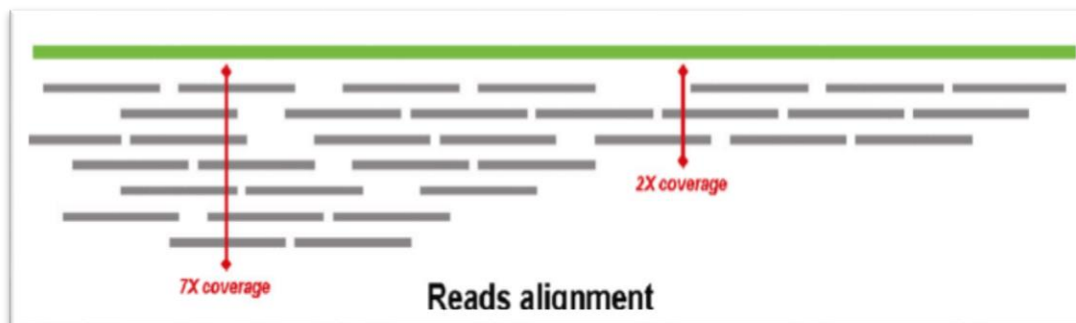


Fig 8 : Reading of aligned sequence

The generated data is studied well, and again checked for QC, in this stage if less data is generated or the depth is not good, the sample is rejected, and the further processes are stopped^[7].

3.1.5 BIOINFORMATICS PIPELINES:

- After sequencing, the generated data are kept with base call per cycle along with the quality of each call.
- Raw data are generated in **.bcl** format.
- File generated by Illumina DNA sequencing instruments, including HiSeq and MiSeq; saves base call and quality score information for each cycle that was processed.
- **.bcl** files help in demultiplexing the data and present as it was sequenced by the machine earlier.
- **Demultiplexing** is the way to change **.bcl** format to **fastq** format.
- **fastq** is universally known sequencing data file, which helps to align the sequence to the reference genome.
- The sequences obtained are aligned with the human reference sequence (GRCh37/hg19) using BWA program^[21-22] and analyzed using Picard and GATK-toolkit^[23-24] to identify variants relevant to the clinical indication.
- Gene annotation of variants are done by VeP program^[25] against the Ensembl release 87 human gene model^[26].
- Somatic variants are identified and annotated using in-house database VariMat.
- Clinically relevant mutations were annotated using published literature, 1000japanese, dbSNP, UK10K, MedVarDb (in-house database) and filtered from reporting^[27-32].
- The curated in-house database OncoMD includes somatic variants from TCGA,^[33,34]
- Fastq files are transferred into SAM(sequence Alignment/Map) files, which is the universal file format for mapped sequence reads.
- SAM files are then compressed and made into BAM (Binary Alignment/Map) files.
- BAM files are supported by BAI files.

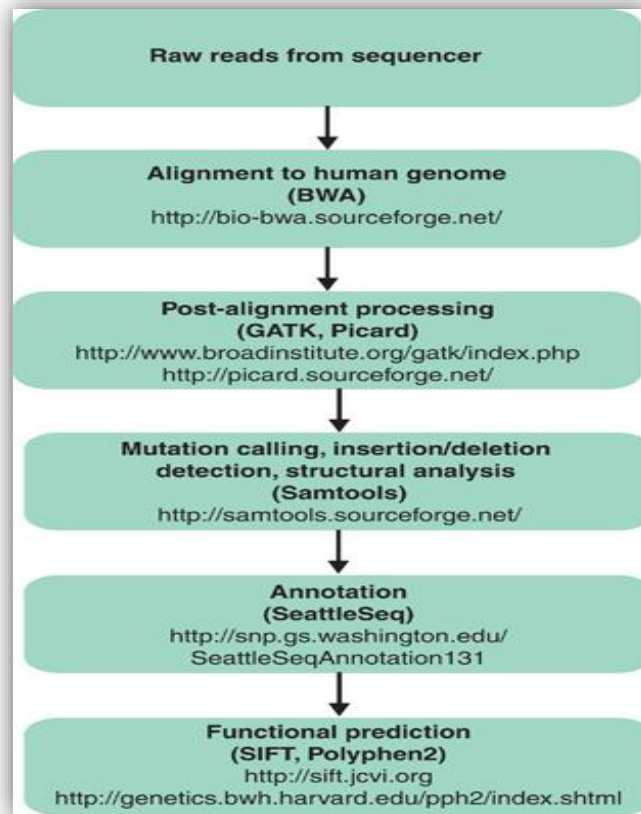


Fig 9 : Flowchart of Bioinformatics pipeline

- The illumina sequencing produces millions of reads in a short single run and mapping these large number of short reads into a genome becomes huge challenge; but using the concept of Burrows-Wheeler transform, has lead the bioinformatics to use Bowtie and BWA (Burrows-Wheeler Alignment tool) with which doing the backward process easily can mimic the top down transversal prefix trie of the genome with a relatively small footprint and counting the exact number of hits within given time.
- Using the technique, the exact repeats are collapsed on one path on the prefix trie.
- It helps to achieve the variant calling or analyses with open sources like SAM tools software package.
- By this approach ambiguous reads are identified, paired end mapping can be done, the exact number of hits can be determined, and map quality scores can be calculated.^{[bw'94]35}

- BAM file is further changed into VCF (Variant Calling Format) follows some steps---

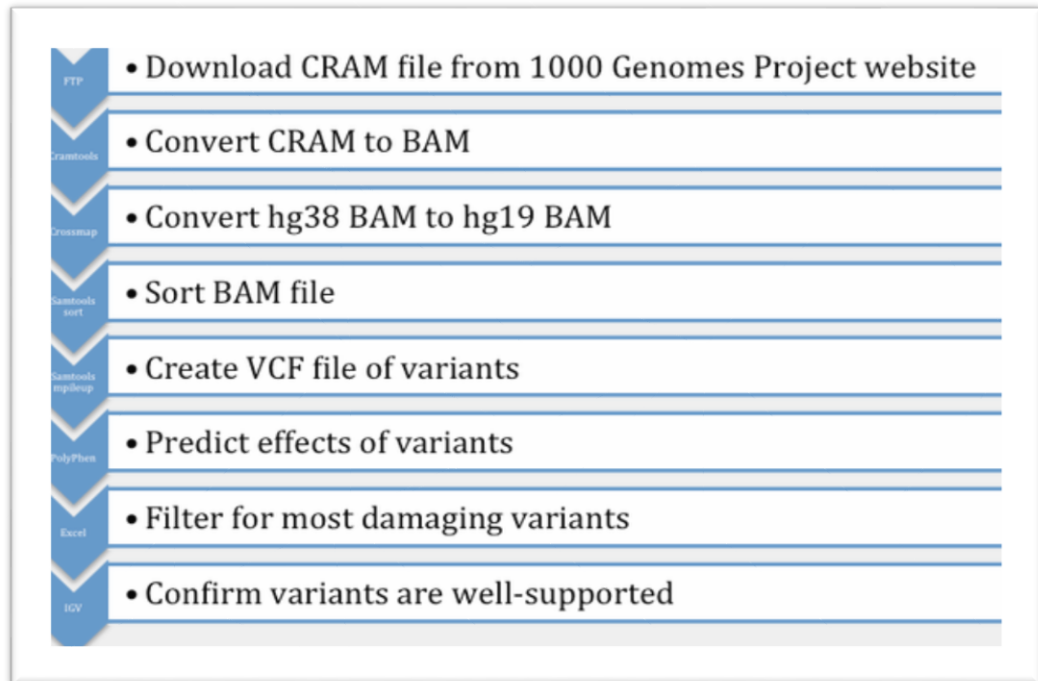


Fig 10 : Flowchart of BAM file to VCF file conversion

- 1000Genome have advances in sequence data generation, achieving and analysis, which helps in improving primary sequence data generation by increasing the read length and depth, reducing base-pair errors and the introduction of base pair errors also the pair end sequencing.

3.1.6 GENOME ANALYSIS:

🚦 Bioinformatics Data & Verification

- The file contains – Analysis QC Report, Vcf files, VariMAT files, Gene Coverage details, Somatic trim BAM files supported by BAI files, Locus coverage depth summary, Low coverage region summary.
- Generally, for HEME samples the **expected data** that should be generated is **>350 Mb**, **panel coverage** should be **≥ 90-100%**, **Panel depth** should be **≥ 350X**, **Phred score Q30 ≥ 70%** and **ontarget reads** should be **≥30%**.

- There are different kind of variant calling files for somatic variant pipeline analysis, among these, GATK is the best framework for identification of variants.
- The variants are filtered to exclude the polymorphisms and the germline variants from the VariMAT files.
- There are some specific criteria to check a variant that should be followed.

LIMS:

- LIMS (Laboratory Information Management System) is a software that effectively manage samples and associated data.
- It enables workflow automation.
- Centralize access and storage of quality control data
- Support compliance efforts.
- Track reagents and lots.
- Perform instrument run monitoring.
- Initiate downstream data analysis.
- Integrate with instruments or other in-lab systems to improve lab efficiency.

3.1.7 IGV Verification:

- **IGV** is Integrative Genome Viewer is a high-performance visualization tool for interactive exploration of large, integrative datasets.
- For analysis purpose somatic BAM trim files are given for each gene.
- There is a procedure for calling a variant as a true variant:
- At first the somatic BAM trim files are downloaded.
- There should be ≥ 10 base nucleotide changes in the gene and the Phred score should be good then the variant is considered as a good SNP.

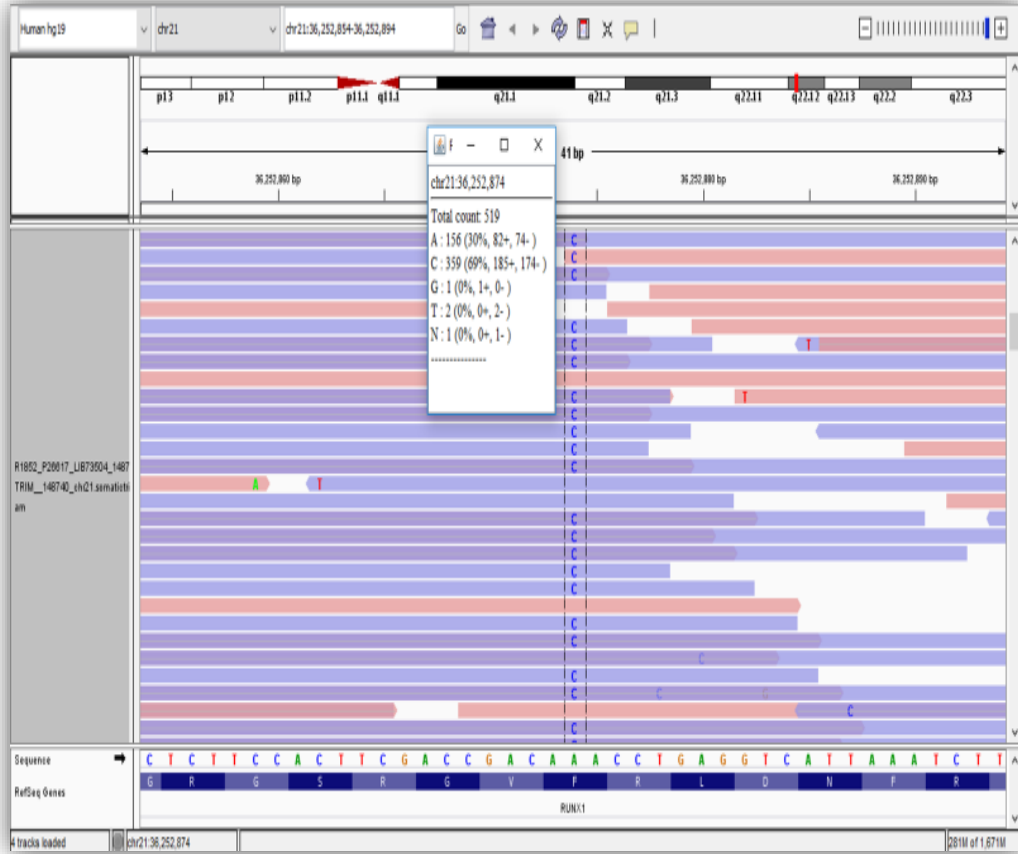


Fig 11 :IGV snapshot showing TRUE variants

- In case of Insertions, it is denoted by **I**, and always view as pairs for the quality check and confirm, the minimum of 10 reads without any repeat region, involving any soft clips, edge effects and low complexity region.
- <20 bp length indels can be taken as a confident call.
- The number of indels are checked by clicking the very next position.

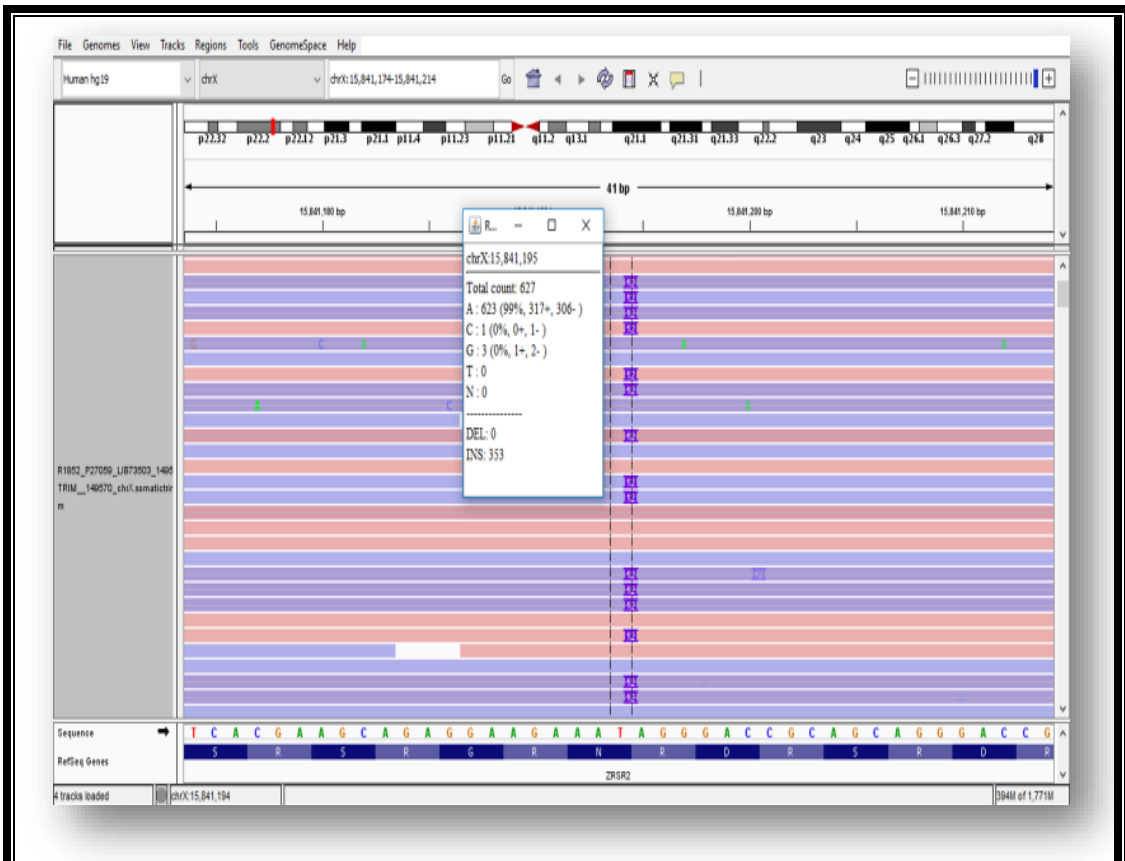


Fig 12 : IGV snapshot showing INSERTION

- For Deletions, there should be no edge effects and no soft clips. And obviously the reads should be more than 10.
- There should not be a single or more than repeat region for a deletion.
- The splice sites are considered when they are +2 or -2 not more than that, also the varclasses that are only considered are Missense, Nonsense, Inframe-del, Inframe-ins, Frameshift-del, Frameshift-Ins, Intronic-SS-ACR, Intronic-SS-DNR, Startloss and Stoploss.
- The variants are also needing to be protein-coding Transcripts, Canonical type, and Ontargets only.
- Intronic-SS-ACR-PRX and intronic-SS-DNR-PRX are ignored for each of the missense, nonsense or indels.

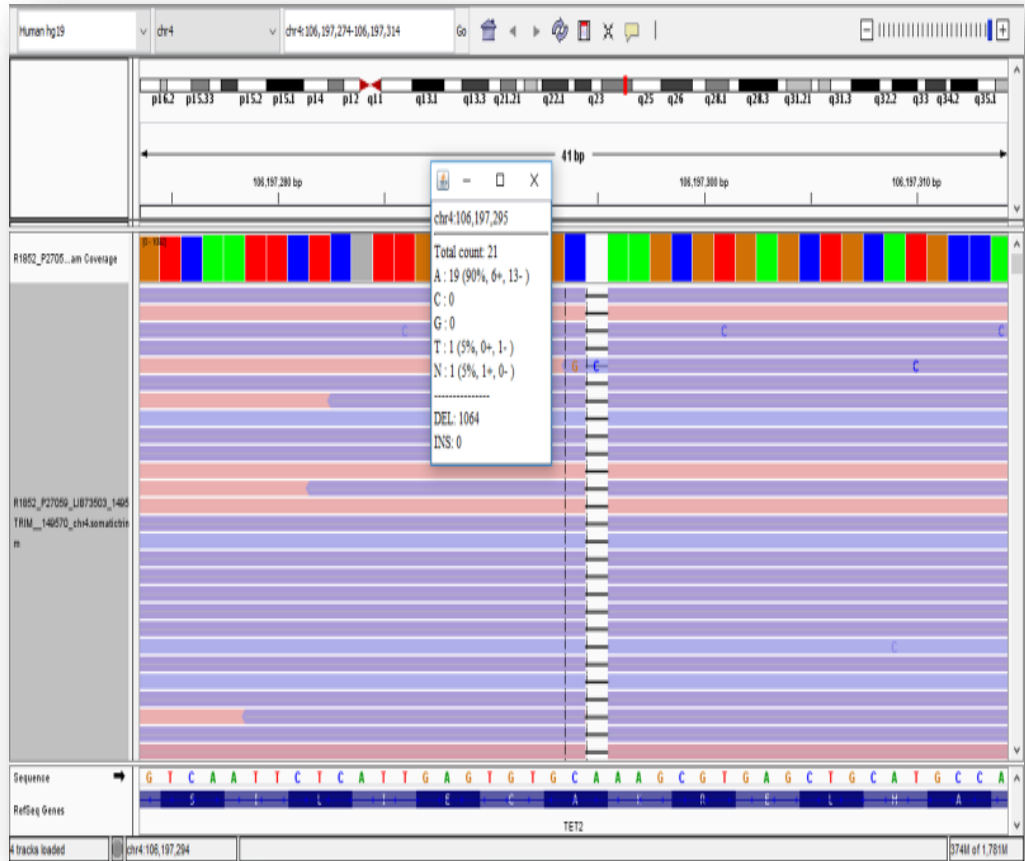


Fig 13 : IGV snapshot showing DELETION

3.1.8 GENOME ANALYSIS METHOD :

- At first the sample id is taken from LIMS and checked for either the bioinformatics analysis is completed.
- As soon as the analysis completes, an FTP link of analysis mail is shared by the Bioinformatics team.
- The files regarding the sample id is downloaded using FileZilla client software.

3.1.9 Oncominer:

- The sample id is given and searched in Oncominer, a modified software to filter out variants for analysis using specific input.

- After uploading the sample id, all the genes will be showing in columns with every annotation like—gene name, variants, cDNA change, Amino acid change, reported in disease databases, alt allele percentage, varclass etc.

Home > Samples > Sample ID: 159666 (283 variants in 5 genes) > QC Status | TRF | Clinical Synopsis | Applied Filter | File Information | Version Information

Variants Coverage Tables (1) Report

Total 223 variants in 4 genes History

	Gene	Variant	cDNA Change	AA Change	Mut Allele %	Disease DB	VarClass	Transcript/Exon
	NRAS	chr1:115252180:T>G	c.450+10A...	NA	NA	●●●●●●	INTRONIC...	ENST00000369
	NRAS	chr1:115252200:T>G	c.440A>C	p.Lys147Thr	NA	●●●●●●	MISSENSE	ENST00000369
	NRAS	chr1:115252216:T>G	c.424A>C	p.Ile142Leu	0.84	●●●●●●	MISSENSE	ENST00000369
	NRAS	chr1:115252225:T>G	c.415A>C	p.Ile139Leu	0.79	●●●●●●	MISSENSE	ENST00000369
	NRAS	chr1:115252230:T>G	c.410A>C	p.Tyr137Ser	NA	●●●●●●	MISSENSE	ENST00000369
	NRAS	chr1:115256493:C>A	c.218G>T	p.Arg73Met	NA	●●●●●●	MISSENSE	ENST00000369
	NRAS	chr1:115256505:T>C	c.206A>G	p.Asp69Gly	NA	●●●●●●	MISSENSE	ENST00000369
	NRAS	chr1:115256506:C>T	c.205G>A	p.Asp69Asn	NA	●●●●●●	MISSENSE	ENST00000369
	NRAS	chr1:115256511:A>C	c.200T>G	p.Met67Arg	0.54	●●●●●●	MISSENSE	ENST00000369
	NRAS	chr1:115256526:T>G	c.185A>C	p.Glu62Ala	NA	●●●●●●	MISSENSE	ENST00000369
	NRAS	chr1:115256531:TC>T	c.179del	p.Gly60As...	NA	●●●●●●	FRAMESHI...	ENST00000369
	NRAS	chr1:115256545:G>C	c.166C>G	p.Leu56Val	NA	●●●●●●	MISSENSE	ENST00000369

Copyright © 2015, Medgenome. All rights reserved.

Fig 14: Snapshot of OncoPrint

- From the OncoPrint, gene coverage, disease association, depth, mRNA details, GWAS studies for each gene can be known.
- Filters according to the SOP (Standard Operation Procedure) is put and the filtered-out genes are analyzed for checking the disease-causing changes.
- Each gene is checked in IGV.

3.1.10 VARIANT CLASSIFICATION:

- If true variant comes from IGV verification following the criteria for calling a variant as a true one, those variants are taken screen shot and pasted in ppt(power point Presentation).

- databases like JAX, DoCM, CbioPortal, My Cancer Genome, clinvar and even PubMed search is done for checking either the variant is reported for the cancer type.

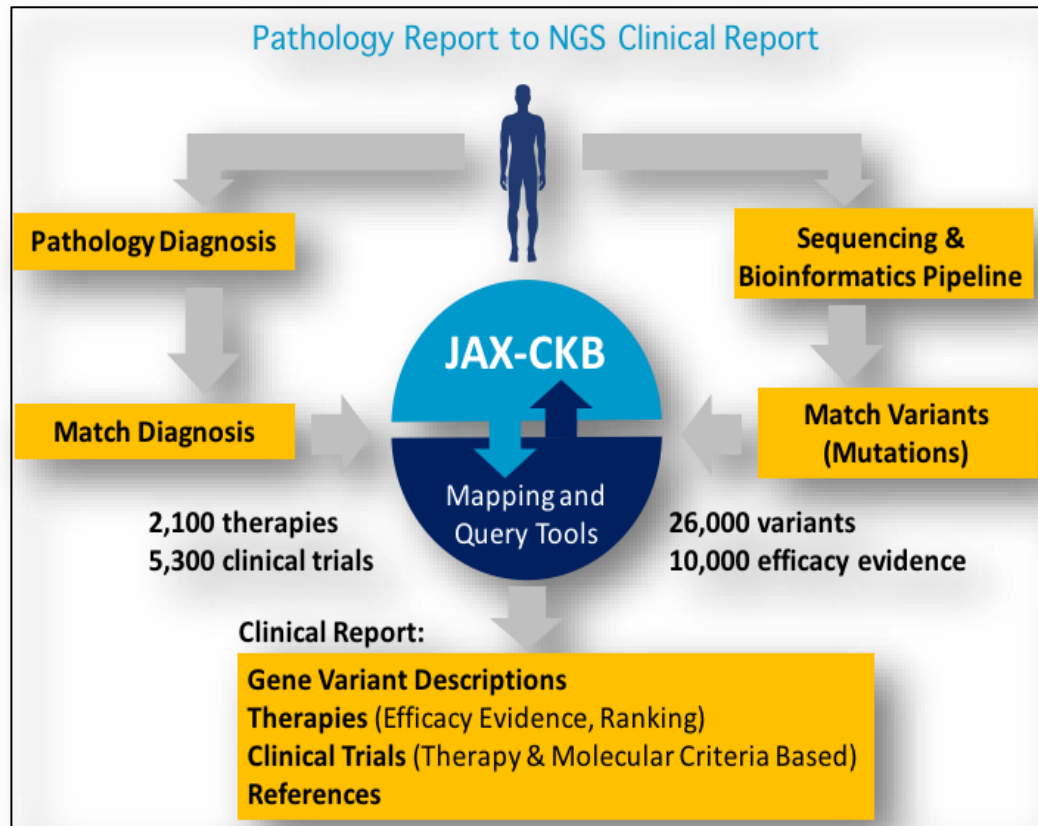


Fig 15 : Databases to rely on for clinical reporting

- If that variant is reported in the specific type of cancer, and it has any reported FDA approved drug for the same, the variant is classified as **TIER 1** or **TABLE 1** for that cancer type.
- If the variant is not reported in the cancer type, but present in any other cancer type, and does not have any FDA approved drug, the variant will be in **TIER 2** or **TABLE 2**.
- If the variant is not reported in any cancer type, then the variant will be in **TIER 3** or **TABLE 3** or **VUS**.

3.1.11 CLINICAL REPORT:

- In case of clinical reporting all the basic details about the patient, like age, gender, name is provided.
- Details about the gene and the role of the specific variant in the cancer type is given in the report.
- Coverage of all the genes that are verified or covered by the test panel are also given with the report.
- Although the TABLE 1 and TABLE 2 variants are most important in the report, but the VUS (Variants of Unknown Significance) are given in the report as further information.
- Once a variant is put in the report format, every detail of the variant is given in a table format like, transcript id, exon number, alt allele percentage.

3.1.12 RESULT:

❖ CASE 1:

AML RISK SRATIFICATION

- **SYMPTOMS:**

The case was suspected to be **AML RISK** by the clinician.

- **SAMPLE TYPE:**

Blood in EDTA tubes (2)

- Analysis was done using the filters, and variants were identified.

- Variants were verified by IGV and put into tables.

- **VARIANTS IDENTIFIED:**

FLT3, NPM1

-

GENE	ROLE IN CANCER
<i>FLT3</i> (p. Asp835Glu)	Oncogenic, gain of function mutation in AML
<i>NPM1</i> (p. Trp288CysfsTer12)	Associated with poor prognosis in presence of FLT3-ITD mutations

- The variants were put in **TABLE 1**.
- *DNMT3A, ASXL1, SMC1A* genes were found to be VUS, because these variants did not have any role in AML of the specific case study.
- The report was released mentioning “**clinically significant mutations were found in the *FLT3, NPM1* genes of the subject**”.

❖ **CASE 2:**

MPN PROGNOSTIC AND STARTIFICATION

- **SYMPTOMS:**
Subject was having splenomegaly; the clinician guessed and sent **MPN panel**.
- **SAMPLE TYPE:**
Blood in EDTA tubes (2)
 - Analysis was done using the filters, and variants were identified.
 - Variants were verified by IGV and put into tables.
 - **VARIANT IDENTIFIED:**
JAK2

GENE	ROLE IN CANCER
<i>JAK2</i> (p. Val617Phe)	Diagnostic marker in MPN, 95% of PV patients carry this mutation, involved in leukemogenesis.

- Variant was put into **TABLE 1**.
- No other variant was identified as **Table 2** or **VUS**.
- The report was released mentioning “**clinically significant mutation was found in the *JAK2* gene of the subject**”.

❖ **CASE 3:**

MDS PROGNOSTIC AND STRATIFICATION

○ **SYMPTOMS:**

The FISH result of the subject depicted that the subject was having monosomy 5, monosomy 7 and trisomy 8 and *TP53* gene deletion.

The clinician guessed as **MDS**.

○ **SAMPLE TYPE:**

Blood in EDTA tubes.

○ Analysis was done using the filters, and variants were identified.

○ Variants were verified by IGV and put into tables.

○ **VARIANTS IDENTIFIED:**

TET2, RUNX1, ZRSR2, PHF6

GENE	ROLE IN CANCER
<i>TET2</i> (p. Gln1389Ter)	Poor overall survival in MDS
<i>RUNX1</i> (p. Tyr377LeufsTer223)	Poor overall survival in MDS
<i>RUNX1</i> (p. Phe336LeufsTer264)	Poor overall survival in MDS
<i>RUNX1</i> (p. Gln286LeufsTer315)	Poor overall survival in MDS
<i>ZRSR2</i> (p. Ser447_Arg448dup)	Poor prognosis in MDS

<i>PHF6</i> (p. Tyr103Ter)	Involved in leukemogenesis and associated with high risk MDS.
<i>PHF6</i> (p. Arg225Ter)	Involved in leukemogenesis and associated with high risk MDS.

- Variants were put into **TABLE 1**.
- A splice variant of *TET2* gene was put into VUS.
- The report was released mentioning “ **clinically significant mutations were found in the *TET2*, *RUNX1*, *ZRSR2*, *PHF6* genes of the subject**”.

❖ **CASE 4:**

CHRONIC MYELOMONOCYTIC LEUKEMIA

○ **SYMPTOMS:**

The patient was diagnosed with **CMML**

○ **SAMPLE TYPE:**

Blood in EDTA tubes.

○ Analysis was done using the filters, and variants were identified.

○ Variants were verified by IGV and put into tables.

○ **VARIANTS IDENTIFIED:**

TET2, NRAS, U2AF1, ZRSR2

GENE	ROLE IN CANCER
<i>TET2</i> (p.Lys1877SerfsTer10)	Loss of function mutation predicts overall less survival in CMML patients.
<i>NRAS</i> (p.Gly12Val)	Inferior outcome like MPN like CMML phenotype.
<i>U2AF1</i> (p.Ser34Phe)	Predicts poor prognosis in MDS.
<i>ZRSR2</i> (p.Asp432_Arg435dup)	Associated with unfavorable outcomes in MDS.

- *TET2*, *NRAS* variants were put in **TABLE 1**.
- *U2AF1*, *ZRSR2* variants were put in **TABLE 2**.
- A *NRAS* gene variant was put in **VUS**.
- The report was released mentioning “ **clinically significant mutations were found in the *TET2*, *NRAS*, *U2AF1*, *ZRSR2* genes of the subject**”.

❖ **CASE 5:**

JMML RISK STRATIFICATION

○ **SYMPTOMS:**

the subject was having JMML (Juvenile Myelomonocytic Leukemia).

○ **SAMPLE TYPE:**

1 EDTA tube and 1 Heparin tube of blood.

○ Analysis was done using the filters, and variants were identified.

○ Variants were verified by IGV and put into tables.

○ **VARIANTS IDENTIFIED:**

PTPN11

GENE	ROLE IN CANCER
<i>PTPN11</i> (p.Ser502Leu)	Pathogenic variant associated with onset of JMML.

- No other variants were identified as **TABLE 2** or **VUS**.
- Clinical report was dispatched mentioning “ **clinically significant mutations were found in the *PTPN11* gene of the subject**”.

3.2 EVALUATION OF RESULT:

❑ CASE 1:

In this case, the subject was diagnosed with **AML Risk Stratification**, after genome analysis, *FLT3* and *NPM1* gene variants were filtered out.

- *FLT3* (p.Asp835Glu) variant was searched in Jax, DoCM, cbiportal to check the variant is reported or not and it has any oncogenic role in AML.

Gene Description: FLT3, receptor-type tyrosine-protein kinase FLT3, is involved in differentiation, proliferation, and survival of hematopoietic progenitor cells. Activating mutations of FLT3 are common in hematologic tumors (PMID: 19467916).

Gene Variants: 166 | Molecular Profiles: 223 | Gene Level Evidence: 183

Showing 1 to 1 of 1 entries (filtered from 166 total entries) | Filter rows: D835E

Variant	Impact	Protein Effect	Variant Description
D835E	missense	gain of function	FLT3 D835E lies within the protein kinase domain activation loop of the Flt3 protein (UniProt.org, PMID: 11290608). D835E confers a gain of function to the Flt3 protein resulting in constitutive phosphorylation and transformation of cultured cells (PMID: 11290608).

Showing 1 to 1 of 1 entries (filtered from 166 total entries) | Filter rows: D835E

©2017 THE JACKSON LABORATORY
 Website content is for educational and research purposes only and is not intended to be used for medical advice, diagnosis or treatment
 This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License

Current view shows filtered results. Click [here](#) to reset all filters.

9 Mutations: includes 6 duplicate mutations in patients with multiple samples (page 1)

Study	Sample ID	Cancer Type	Protein Change	Annotation
Acute Myeloid Le...	TCGA-AB-2814-03	Acute Myeloid Leukemia	D835E	🔍
Acute Myeloid Le...	TCGA-AB-2814-03	Acute Myeloid Leukemia	D835E	🔍
Acute Myeloid Le...	TCGA-AB-2814-03	Acute Myeloid Leukemia	D835E	🔍
Acute Myeloid Le...	TCGA-AB-2834-03	Acute Myeloid Leukemia	D835E	🔍
Acute Myeloid Le...	TCGA-AB-2945-03	Acute Myeloid Leukemia	D835E	🔍
Acute Myeloid Le...	TCGA-AB-2945-03	Acute Myeloid Leukemia	D835E	🔍
Acute Myeloid Le...	TCGA-AB-2834-03	Acute Myeloid Leukemia	D835E	🔍
Acute Myeloid Le...	TCGA-AB-2834-03	Acute Myeloid Leukemia	D835E	🔍
Acute Myeloid Le...	TCGA-AB-2945-03	Acute Myeloid Leukemia	D835E	🔍

Showing 1-9 of 9 Mutations: includes 6 duplicate mutations

FLT3 D835E in acute myeloid leukemia

CLINICAL IMPLICATIONS	BIOLOGICAL EFFECT
Oncogenic	Gain-of-function

FLT3, a receptor tyrosine kinase, is recurrently altered in acute myeloid leukemia and other hematologic malignancies.

The FLT3 D835E mutation is known to be oncogenic.

While there is promising clinical data in patients with FLT3 Y630ins altered acute myeloid leukemia treated with the FDA-approved RAF-inhibitor sorafenib, its clinical utility in patients with FLT3 D835E mutant acute myeloid leukemia is unknown.

The information above is intended for research purposes only and should not be used as a substitute for professional diagnosis and treatment.

Levels ▼

OncoKB Feedback

Fig 16 : Snapshots showing *FLT3* variants reported in databases like JAX and Cbioportal

- Checking the databases, the variant was put in **TABLE 1** for **AML**.
- **NPM1** gene was checked in the same databases—

Current view shows filtered results. Click [here](#) to reset all filters.

149 Mutations: includes 98 duplicate mutations in patients with multiple samples (page 1 of 6)

Study	Sample ID	Protein Change	Annotation	Mutation Type	Copy #	COSMIC	Allele Freq (T)	# Mut in Sample
Acute Myeloid Le...	TCGA-AB-2809-03	W288Cfs*12						4
Acute Myeloid Le...	TCGA-AB-2816-03	W288Cfs*12						6
Acute Myeloid Le...	TCGA-AB-2818-03	W288Cfs*12						8
Acute Myeloid Le...	TCGA-AB-2826-03	W288Cfs*12						5
Acute Myeloid Le...	TCGA-AB-2837-03	W288Cfs*12						1
Acute Myeloid Le...	TCGA-AB-2853-03	W288Cfs*12						7
Acute Myeloid Le...	TCGA-AB-2871-03	W288Cfs*12						15
Acute Myeloid Le...	TCGA-AB-2877-03	W288Cfs*12						21
Acute Myeloid Le...	TCGA-AB-2884-03	W288Cfs*12						6
Acute Myeloid Le...	TCGA-AB-2903-03	W288Cfs*12						1
Acute Myeloid Le...	TCGA-AB-2919-03	W288Cfs*12						10
Acute Myeloid Le...	TCGA-AB-2925-03	W288Cfs*12						16
Acute Myeloid Le...	TCGA-AB-2931-03	W288Cfs*12						11
Acute Myeloid Le...	TCGA-AB-2932-03	W288Cfs*12						6
Acute Myeloid Le...	TCGA-AB-2969-03	W288Cfs*12						8
Acute Myeloid Le...	TCGA-AB-2973-03	W288Cfs*12		FS ins	Diploid	2261		4

NPM1 W288Cfs*12 in acute myeloid leukemia

CLINICAL IMPLICATIONS
Likely Oncogenic

BIOLOGICAL EFFECT
Likely Loss-of-function

NPM1 encodes for a gene involved in regulating multiple cellular activities including ribosome biogenesis and nucleolus function. NPM1 mutations are found in hematologic malignancies.

The NPM1 W288Cfs*12 mutation is likely oncogenic.

There are no FDA-approved or NCCN-compendium listed treatments specifically for patients with NPM1 W288Cfs*12 mutant acute myeloid leukemia.

The information above is intended for research purposes only and should not be used as a substitute for professional diagnosis and treatment.

Levels ▼

OncoKB Feedback

Gene Description: NPM1, nucleophosmin, is a phosphoprotein that shuttles between the nucleus and cytoplasm, and is involved in several cellular processes including ribosomal processing and export, chromatin remodeling, DNA replication and repair, and cell cycle control (PMID: 29157973, PMID: 23436734). NPM1 is frequently mutated in acute myeloid leukemia (PMID: 29157973, PMID: 23436734).

Gene Variants 31 | Molecular Profiles 60 | Gene Level Evidence 175

Showing 1 to 1 of 1 entries (filtered from 31 total entries) | Filter rows: W288Cfs 12

Variant	Impact	Protein Effect	Variant Description
W288Cfs*12	duplication	loss of function	NPM1 W288Cfs*12 is a hotspot mutation that results from a 4 bp-duplication causing a frameshift aa 288 of 294, followed by 12 amino acids creating a novel nuclear export sequence (PMID: 16501600, PMID: 15659725). W288Cfs*12 confers a loss of function to the Npm1 protein as demonstrated by mislocalization of the Npm1 protein (PMID: 15659725).

Showing 1 to 1 of 1 entries (filtered from 31 total entries) | Filter rows: W288Cfs 12

The Jackson Laboratory | ©2017 THE JACKSON LABORATORY

Fig 17 : Snapshots showing *NPM1* variants reported in databases like Cbioportal and JAX.

- *NPM1* was also put as TABLE 1 for AML, because the variant was reported for AML, and has FDA approved drug.

❑ CASE 2:

- The subject was diagnosed with MPN.
- The variant was found to be *JAK2* (p.Val617Phe), after filtering and IGV verification.
- The variant was checked in JAX, DoCM, Cbioportal, My Cancer Genome etc.

The screenshot shows the My Cancer Genome website interface. At the top, there is a search bar and navigation links for Home, DIRECT, and About Us. The main content area is titled "JAK2 c.1849G>T (V617F) Mutation in Myeloproliferative Neoplasms". On the left, a sidebar lists various mutations, with "JAK2 c.1849G>T (V617F)" selected. The main content includes a "Properties" table, "Implications for Targeted Therapeutics" table, and a paragraph of text.

What is JAK2? **JAK2 in Myeloproliferative Neoplasms** **JAK2 c.1849G>T (V617F)** **Clinical Trials**

JAK2 c.1849G>T (V617F) Mutation in Myeloproliferative Neoplasms

Properties	
Location of mutation	Pseudokinase domain (exon 14); James et al. 2005
Frequency of JAK2 mutations in MPNs	97–98% in PV (Baxter et al. 2005 ; James et al. 2005) 50–60% in ET (Baxter et al. 2005 ; Kralovics et al. 2005 ; Levine et al. 2005) 55–65% in PMF (Baxter et al. 2005 ; Kralovics et al. 2005 ; Levine et al. 2005 ; Rumi et al. 2014)
Frequency of JAK2 V617F mutations in JAK2-mutated MPNs	95% in JAK2-mutated PV (Levine et al. 2005 ; Pardananani et al. 2007 ; Nangalia et al. 2013) ~100% in JAK2-mutated ET (Cross et al. 2005) ~100% in JAK2-mutated PMF (Cross et al. 2005)
Implications for Targeted Therapeutics	
Response to JAK inhibitors	Unknown at this time ^a

The [JAK2 V617F mutation](#) is the most frequent [mutation](#) in MPNs and results in an [amino acid](#) substitution at position 617 in exon 14, from a valine (V) to a phenylalanine (F). Crystallography studies have shown that V617F is located within the JAK2 JH2 domain, which exerts an inhibitory effect on the JH1 [kinase](#) domain ([Bandaranayake et al. 2012](#))

DoCM
DATABASE OF CURATED MUTATIONS

About News Contact Sources API Documentation FAQ

Version: Current (3.2) Submit New Variants

Variants

Show 25 entries Search: jak2

HGVS	CHR	Gene	Amino Acid	Mutation Type	Diseases	Citations
ENST00000381652.c.1691G>T	9	JAK2	p.R564L	missense	bone marrow cancer	Scott et al., 2007, N. Engl. J. Med.
ENST00000381652.c.1849G>T	9	JAK2	p.V617F	missense	acute myeloid leukemia, bone marrow cancer, chronic myeloid leukemia, lymphoid leukemia, myelofibrosis, and 1 more.	Baxter et al., Lancet, Levine et al., 2005, Blood, Campbell et al., 2006, Blood, Kitadjan et al., 2006, Blood, Wernig et al., 2008, Cancer Cell, and 14 more.

- *JAK2* was put in **TABLE 1** for **MPN**, because the variant is reported and there is FDA approved drugs for it.

Fig 18 : Snapshots showing *JAK2* variant reported in databases like My Cancer Genome and DoCM

□ CASE 3:

- The subject was diagnosed with MDS.
- Different variants of *TET2*, *RUNX1*, *ZRSR2*, *PHF6* genes were filtered and verified in IGV.

Gene TET2, tet methylcytosine dioxygenase 2, plays a role in DNA methylation and is involved in hematopoiesis (PMID: 26099018, PMID: 25510268). TET2 has been identified as a tumor suppressor in hematological malignancies, and mutations in TET2 are common in hematological cancers including myelodysplastic syndrome and chronic myelomonocytic leukemia (PMID: 26099018, PMID: 22240200).

Gene Variants 75 Molecular Profiles 77 Gene Level Evidence 4

Filtering and Sorting ⓘ

Showing 1 to 1 of 1 entries (filtered from 75 total entries) Filter rows: Q1389

Variant	Impact	Protein Effect	Variant Description
Q1389*	nonsense	loss of function - predicted	TET2 Q1389* results in a premature truncation of the Tet2 protein at amino acid 1389 of 2002 (UniProt.org). Due to the loss of the substrate binding domain (UniProt.org), Q1389* is predicted to result in a loss of Tet2 function.

Showing 1 to 1 of 1 entries (filtered from 75 total entries) Filter rows: Q1389

The Jackson Laboratory ©2017 THE JACKSON LABORATORY

Fig 19 : Snapshot from JAX showing *TET2* variant

- *TET2* was reported in JAX so, it was in TABLE 1 for MDS.
- In the same way, *RUNX1*, *ZRSR2*, *PHF6* genes were searched for those variants in JAX, DoCM, and cbiportal, My Cancer Genome and PubMed search too.
- Each variant was reported and had FDA approved drugs against them, so every identified variant from this case was put in TABLE 1 for MDS.

❑ **CASE 4:**

- In this case the subject was diagnosed with CMML.
- After the OncoPrint and IGV verification, *TET2*, *NRAS*, *U2AF1*, *ZRSR2* gene variants were found.

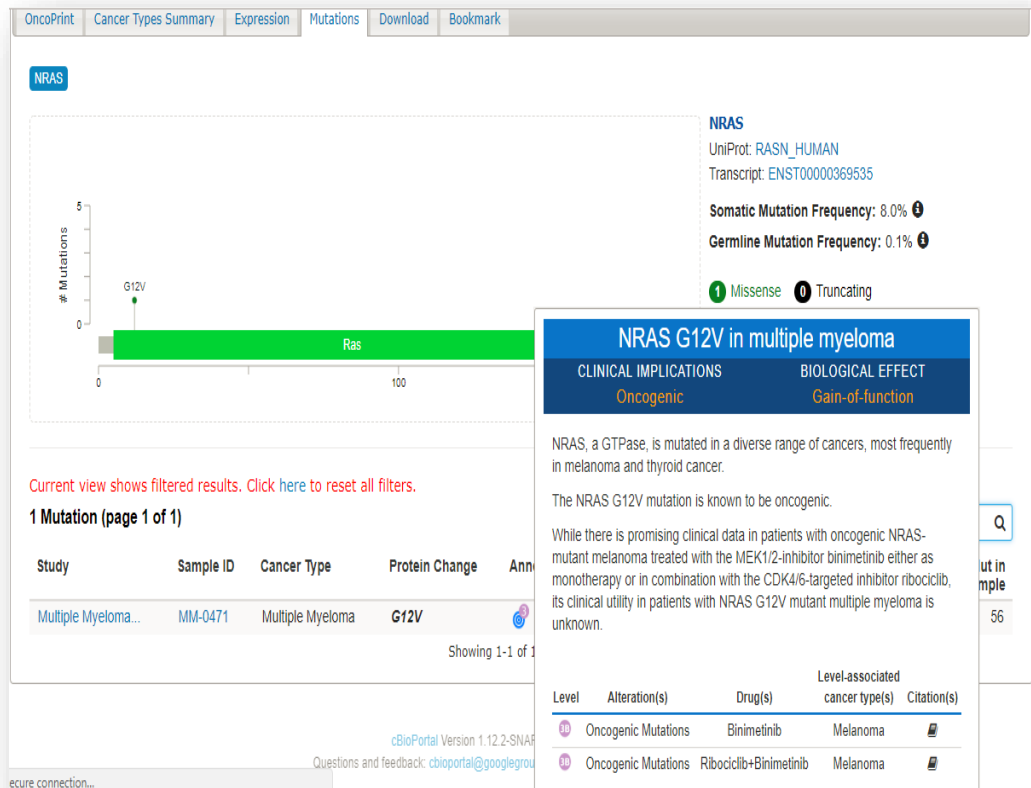


Fig 20 : Snapshot of Cbioportal showing *NRAS* variant

- *NRAS* (p.Gly12Val) was reported in cBioportal and the mutation was oncogenic for CMML, so NRAS was put into TABLE 1 for CMML.

Gene U2AF1, U2 small nuclear RNA auxiliary factor 1, functions in 3' splice site recognition therefore regulates the expression of genes involved in hematological and solid tumor development (PMID: 2963698, PMID: 28986033). U2AF1 is frequently mutated in myelodysplastic syndrome, acute myeloid leukemia, and lung adenocarcinoma (PMID: 23645665, PMID: 22678168, PMID: 24498085).

Gene Variants 34 Molecular Profiles 34 Gene Level Evidence 6

Filtering and Sorting

Showing 1 to 1 of 1 entries (filtered from 34 total entries) Filter rows: S34F

Variant	Impact	Protein Effect	Variant Description
S34F	missense	loss of function	U2AF1 S34F lies within the C3H1-type 1 Zinc finger region of the U2af1 protein (UniProt.org). S34F confers a loss of function to the U2af1 protein as indicated by aberrant splicing of U2af1 target genes in cell culture (PMID: 25267526, PMID: 28436936).

Showing 1 to 1 of 1 entries (filtered from 34 total entries) Filter rows: S34F

The Jackson Laboratory ©2017 THE JACKSON LABORATORY


 This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike license.

Fig 21 : Snapshot showing *U2AF1* variant in JAX

- *U2AF1* gene was reported for MDS, but not in CMML so, the gene was put in TABLE 2 for CMML.
- One *NRAS* variant was put in VUS because the variant was not reported in any of the databases.

❑ **CASE 5:**

- The subject was suspected to have JMML.
- After filtering and verifying the variants *PTPN11* (p.Ser502Pro) was put in TABLE 1.

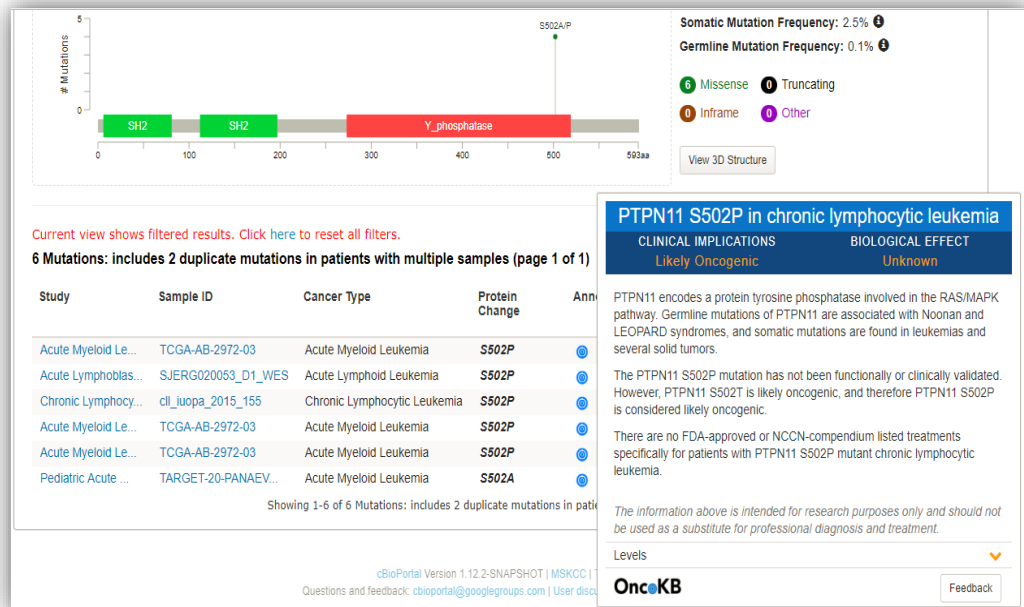


Fig 22 : Snapshot showing *PTPN11* variant in cBioportal

- The variant was reported in **cBioportal**, and associated with NOONAN SYNDROME, which is one of the symptoms of JMML.
- Hence, the variant was put into TABLE 1 for JMML.

HOW TO COMPLEMENT IN CORPORATE

Coming out from college life, and adjusting own-self to a new environment of corporate world is difficult. To cope up with the situations and endeavours one should follow some basic rules and it will not only help them learn to survive in the corporate world but also to excel in field of the same.

Some basic rules to be followed in a corporate world to complement own-selves are:

- A good communication skill is the key to success.
- An open mind and eagerness to learn without holding any prejudice helps a lot.
- As a corporate employee one should minimize their expectation and maximize their commitments.
- One should be confident enough to face the tasks coming on the way.
- One should not be shy enough to shoulder all responsibilities.
- One should be humble, invariable of their designation and qualification.
- Emotional outbreaks in the office premises will not help in the long run.
- One should promise what they can deliver to the betterment of the company.
- One should go along with the team and be a team player.
- Rather than own-self, one should think of the team always.
- Self-study won't help until that is not shared and discussed with the team.
- Going with the flow will not give anyone recognition but the hard work and dedication will.

REFERENCE

1. *Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012*. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F.
2. K.Park. *Textbook of preventive and social medicine*. Bhanot, 18th edition, 2005.
3. National cancer Institute, Bethesda, Maryland, *SEER cancer statistics review, 1975-2009*, 2011.
4. American Cancer Society, Atlanta. *Cancer Facts and Figures 2013*, 2013.
5. G. E. Xueling and X. Wang. *Role of wnt canonical pathway in hematological malignancies*. Journal of hematology and oncology,3(33):1756-8722,2010.
6. *Risk of hematological malignancies*. Kanti R. Rai; Steven E. Coutre, 2009.
7. *What is next generation sequencing?* Sam Behjati and Patrick S Tarpey, 2013.
8. *Common applications of next-generation sequencing technologies in genomic research*. Chien-Yueh Lee^{1*}, Yu-Chiao Chiu^{1*}, Liang-Bo Wang^{2*}, Yu-Lun Kuo^{3*}, Eric Y. Chuang^{1,2,4}, Liang-Chuan Lai⁵, Mong-Hsun Tsai⁶.(2013)
9. *Diagnosis, risk stratification, and response evaluation in classical myeloproliferative neoplasms*; Elisa Rumi^{1,2} and Mario Cazzola^{1,2}.(2016)
10. <http://www.hematology.org/Patients/Cancers>.
11. <http://www.aamds.org/diseases/mds>
12. <https://www.leukaemia.org.au/disease-information/myeloproliferative-disorders/>
13. <https://www.cancer.org/cancer/chronic-myelomonocytic-leukemia.html>
14. <https://www.lls.org/leukemia/juvenile-myelomonocytic-leukemia>
15. https://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf
16. <https://www.medgenome.com/ngs-bioinformatics-molecular-biology-services/>

17. <https://www.illumina.com/informatics/sample-experiment-management/lims.html>
18. *Lessons from next-generation sequencing analysis in hematological malignancies*. E Braggio, J B Egan, R Fonseca & A K Stewart.(2013)
19. <https://www.illumina.com/techniques/sequencing/dna-sequencing/targeted-resequencing/targeted-panels.html>.
20. Li, H. and R. Durbin. *Fast and accurate long-read alignment with Burrows-Wheeler transform*. *Bioinformatics*, 2010. 26(5): p. 589-95.
21. Meyer, L.R., et al., *The UCSC Genome Browser database: extensions and updates 2013*. *Nucleic Acids Res*, 2013. 41(D1): p. D64-9.
22. Li, H., et al., *The Sequence Alignment/Map format and SAMtools*. *Bioinformatics*, 2009. 25(16): p. 2078-9
23. McLaren, W., et al., *Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor*. *Bioinformatics*, 2010. 26(16): p. 2069-70.
24. <http://www.ensembl.org>
25. *The 1000 Genomes Project Consortium*. "A Global Reference for Human Genetic Variation." *Nature* 526.7571 (2015): 68–74. PMC.
26. Lek, M, et al. "Analysis of Protein-Coding Genetic Variation in 60,706 Humans." *Nature*. 536.7616 (2016): 285–91.
27. NHLBI: <https://esp.gs.washington.edu/drupal>
28. Nagasaki, M, et al. "Rare Variant Discovery by Deep Whole-Genome Sequencing of 1,070 Japanese Individuals." *Nature communications*. Vol 6. (Article 8018) (2015):
29. Moayyeri A, Hammond CJ, Hart DJ, Spector TD. The UK Adult Twin Registry (TwinsUK Resource). *Twin research and human genetics: the official journal of the International Society for Twin Studies*. 2013;16(1):144-149.
30. dbSNP: <http://www.ncbi.nlm.nih.gov/SNP/>
31. Raphael Bueno et al *Comprehensive genomic analysis of malignant pleural mesothelioma identifies recurrent mutations, gene fusions and splicing alterations:* *Nature Genetics*: Vol 48(4); 407-416 (oncomd.medgenome.com)
32. Li MM, Datto M, Duncavage EJ, Kulkarni S, Lindeman NI, Roy S, Tsimberidou AM, Vnencak-Jones CL, Wolff DJ, Younes A, Nikiforova

MN. *Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer: A Joint Consensus Recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists.* J Mol Diagn. 2017 Jan;19 (1):4-23.

34. www.cancer.sanger.ac.uk/cosmic/