Mutational screening of key genes in RTK/RAS signaling pathway associated with Pediatric Leukemia in Mizo population

By

ANDREW VANLALLAWMA

Registration No. and Date: MZU/M.Phil/475 of 03/08/2017

Dissertation submitted in fulfilment of the Requirements for the degree of

Master of Philosophy in Biotechnology

Under the supervision of

Dr. N. Senthil Kumar

Professor, Department of Biotechnology

School of Life Sciences, Mizoram University,

Aizawl, Mizoram-796004, India

2019

NAME OF CANDIDATE	:	ANDREW VANLALLAWMA
DEGREE	:	MASTER OF PHILOSOPHY
DEPARTMENT	:	BIOTECHNOLOGY

TITLE OF DISSERTATION **:** MUTATIONAL SCREENING OF KEY GENES IN RTK/RAS SIGNALING PATHWAY ASSOCIATED WITH PEDIATRIC LEUKEMIA IN MIZO POPULATION.

DATE OF PAYMENT OF ADMISSION : 25/08/2017
(Commencement of first semester)

COMMENCEMENT OF SECOND SEM/ DISSERTATION : FEBRUARY 2018
(From conclusion of end semester exams)

APPROVAL OF RESEARCH PROPOSAL

1.	BOS	:	26/04/2018
2.	SCHOOL BOARD	:	03/05/2018
	REGISTRATION NO. & DATE	:	MZU/M.Phil./475 of 03/08/2017
	DUE DATE OF SUBMISSION	:	31 JANUARY 2019
	EXTENSION (IF ANY)	:	28 JUNE 2019

Head
Department of Biotechnology

Mizoram University

Aizawl – 796004

June 2019

Declaration of the Candidate

I, Andrew Vanlallawma, hereby declare that the subject matter of this dissertation entitled "Mutational screening of key genes in RTK/RAS signaling pathway associated with Pediatric Leukemia in Mizo population" is a record of work done by me, that the contents of this dissertation did not form basis of the award of any previous degree to me or to the best of my knowledge to anybody else, and that the dissertation had not been submitted by me for any research degree in any other University/Institute.

This is being submitted to Mizoram University for the award of the degree of Master of Philosophy in Biotechnology.

Date:

(Andrew Vanlallawma)

Candidate

(Dr Jyotirmoy Bhattacharya)

Head

(Prof N. Senthil Kumar)

Supervisor

CERTIFICATE

This is to certify that the dissertation entitled "Mutational screening of key genes in RTK/RAS signaling pathway associated with Pediatric Leukemia in Mizo population" to Mizoram University for the award of the degree of Master of Philosophy Biotechnology by Andrew Vanlallawma Registration No. *MZU/M.Phil/475 of 03.05.2018*, Research scholar in the Department of Biotechnology, is a record of original research work, based on the results of the experiments and investigations carried out independently by his during the period from 2017-2019 of study, under my guidance and supervision and has not been previously submitted for the award of any degree in any Indian or foreign University.

It is further certified that the scholar fulfills all the requirements as laid down by the University for the purpose of submission of M.Phil dissertation.

Place:

Date:

(Prof. N. Senthil Kumar)

Supervisor

Acknowledgement

It is a pleasure to express my sincere thanks to Prof. N. Senthil Kumar, Department of Biotechnology, Mizoram University for his precious guidance, encouragement, motivation and never ending support in all aspects of life and most of all, for giving me the opportunity to work on a research topic that has been of great interest to me for years.

I am deeply grateful to DBT, New Delhi sponsored Advanced Level State Biotech Hub for the infrastructure support and financial assistance and Bioinformatics Infrastructure Facility for providing support and means for my data analysis.

I would like to express my gratitude to our collaborating clinicians Dr. Jeremy L Pautu, Dr. John Zohmingthanga, Dr. C. Lalchhandama and all the lab technicians in Mizoram State Cancer Hospital (MSCI) for their help, support and willingly allowing me to collect samples.

I am highly indebted to Dr. Sridhar Sivasubbu (Principal Scientist) and Dr. Vinod Scaria (Principal Scientist) and to all the Ph.D scholars, CSIR – IGIB, New Delhi for their kind hospitality and invaluable trainings on Whole Exome Sequencing and Data analysis. Not only did I learned techniques from them but also personally shaped me into a better researcher and sharpening my concepts on research as a whole.

I am truly grateful to Dr. Arindam Maitra (Principal Scientist) and Mr. Arindam Pallodhi, NIBMG, Kalyani for Whole Mitochondrial Sequencing and pre-processing of the raw sequencing files.

I would like to thank my lab-mates (NSK Lab) for their help in my work countless times and also am truly thankful for tolerating me for months.

To my family, I thank you for the love, unpressured support and time to achieve my goal.

Finally, I thank the Almighty God for a good health enabling me to complete this work.

Dated:

Place: Aizawl, Mizoram Andrew Vanlallawma

TABLE OF CONTENTS

CONTENTS	PAGE NO.
ACRONYMS	
LIST OF FIGURES	
LIST OF TABLES	
INTRODUCTION & REVIEW OF LITERATURES	1 – 10
MATERIALS & METHODS	11 – 18
AIM & OBJECTIVES	19
RESULTS	20 – 34
DISCUSSION	35 – 43
SUMMARY	44 – 45
REFERENCES	46 – 55
BIODATA	56 – 58

List of Figures

Figure 1	Isolated genomic DNA from peripheral blood run in 0.8% Agarose Gel Electrophoresis. L = Ladder, 1-11 = pediatric leukemia DNA samples and 12 -15 = mother's DNA sample.	Page 21
Figure 2	Samples normalized to 100ng/µl in 0.8% agarose gel electrophoresis.	Page 22
Figure 3	QC of the final library preparation using agarose gel electrophoresis. L=10kb ladder, 1-6 = the plexing id's. 1 and 2 contains the TSQ103 and 104, respectively where the samples are pooled.	Page 23
Figure 4	PCR product of 8 MSI markers for one patient and its corresponding mother. L1: BAT25, L2: BAT26, L3: D2S123, L4: D17S250, L5: D16S398, L6: 16S496, L7: D18S58, L8: D16S3057, L9: 100bp Ladder, L10: BAT25, L11: BAT26, L12: D2S123, L13: D17S250, L14: D16S398, L15: 16S496, L16: D18S58, L17: D16S3057.	Page 32

List of Tables

Table 1	List of samples obtained for the study containing sample identification, age, gender and their diagnosis. GDN: GUaRDIAN (sample identification code), M: male, F: female. ALL: Acute Lymphoblastic Leukemia, AML – M1: Acute Myeloid Leukemia of M1 subtype (according to FAB classification), CML: Chronic Myeloid Leukemia, JMML: Juvenile Myelomonocytic Leukemia. The corresponding patient (son/daughter) and mother are placed in a single row.	Page 20
Table 2	List of sample concentration quantified using nanodrop and calculations to 100ng/µl. GDN: GUaRDIAN, Abs: Absorbance, µl: microliter, RSB: Re-suspension Buffer.	Page 21
Table 3	List of samples with index and plexing ID.	Page 22
Table 4	List of sequencing reads that passed quality check and aligning to the reference genome with the percentage for each leukemic patient. QC: Quality control, %; Percentage.	Page 23
Table 5	List of total number of variants in each sample for the raw annotated files, after filter1, after filter2 and after filter3 with mean number of variants.	Page 24
Table 6	List of frequently mutated genes in leukemia	Page 24
Table 7	Variants observed in samples after applying filter 1, 2 and 3, showing the genes, amino acid change and reported status in ClinVar database. V: Valine, E: Glutamic acid, R: Arginine, C: Cystein, G: Glycine.	Page 25
Table 8	List of matched variants with 55 leukemia and ClinVar genes after filter 2. Chrom: Chromosomal number, Pos: Position, Ref: Nucleotide in the reference genome, Alt: The altered variant from the reference genome.	Page 25
Table 9	Matched Varaints with 55 Leukemia genes that are not reported in ClinVar after applying filter 1 and 2. Variants indicated in # and * are reported in dbSNP and COSMIC database respectively. Chr: Chromosome number; Pos: Position; Ref: Reference, Alt: Alternate, AA change: Amino acid change; MOI: Mode of Inheritance, Hom/Het: Homozygous or Heterozygous.	Annexure 1
Table 10	Unmatched variants to ClinVar are interpreted using OMIM. Ref: Reference, Counts: Read counts identified from IGV, Total reads: Total reads identified from read depth, AA change: Amino acid change, Hom/Het: Homozygous or heterozygous, S_P: SIF Prediction, P_P: Polyphen Prediction, MT_P: Mutation Taster Prediction, D (*): Deleterious/Pathogenic, * indicates how many are predicted as deleterious from the three predicting software	Annexure 2
Table 11	Genotypic variants confering to responsiveness of 5FU drug (a first line of treatment in Cancer) screened using DPYD gene (metabolizer of 5-FU) in Pediatric Leukemia patients	Page 26

	patients in Mizo population. * indicates the presence of a genotype. For Exon - 14 the observed genotype is the wildtype G. Genotype highlighted in RED confers highest severity, YELLOW confers mildly and GREEN confers low severity in terms of the toxic effect of the 5-FU drug.	
Table 12	List of different classes of variants in the mitochondrial genome and the patient condition. Sample highlighted with the same colour indicates patients and their corresponding mother samples. All: Total variants in each sample, Exonic: Total exonic variants in each sample, Syn: Total synonymous variants in each sample, Non – syn: Total non - synonymous variants in each sample and Condition: diagnosis of the disease	Page 27
Table 13	List of samples harbouring MT-ND2 T119A variant. * indicates patient and corresponding mother sample	Page 28
Table 14	Table 14 : List of samples harbouring T152C variant in D-loop region. * indicates patient and corresponding mother sample.	Page 28
Table 15	List of non – synonymous variants present in corresponding patient (GDN4254) and mother (GDN4263).	Page 29
Table 16	List of non – synonymous variants present in corresponding patient (GDN4255) and mother (GDN4264).	Page 29
Table 17	List of non – synonymous variants present in corresponding patient (GDN4257) and mother (GDN4265).	Page 30
Table 18	List of non – synonymous variants present in corresponding patient (GDN4258) and mother (GDN4266)	Page 30
Table 19	List of markers for MSI testing with their description.	Page 31
Table 20	List of MSI marker used and their relative fluorescence units (RFU) and their Primer annealing temperature (Tm)	Page 31
Table 21	Allele and basepair for each markers for patient (GDN4254) and corresponding mother (GDN4263) samples. a = allele, bp = base pairs	Page 33
Table 22	Allele and basepair for each markers for patient (GDN4255) and corresponding mother (GDN4264) samples. a = allele, bp = base pairs	Page 33
Table 23	Allele and basepair for each markers for patient (GDN4257) and corresponding mother (GDN4265) samples. a = allele, bp = base pairs	Page 34
Table 24	Allele and basepair for each markers for patient (GDN4258) and corresponding mother (GDN4266) samples. a = allele, bp = base pairs	Page 34

Acronyms

Acronym	Definition
5-FU	5 - Fluorouracil
AA	Amino Acid
ABL	Abelson Murine Leukemia Viral Oncogene Homolog 1
ALL	Acute Lymphoblastic Leukemia
AML	Acute Myeloid Leukemia
ATL2	A-Tailing Mix
ATM	Ataxia-Telangiesctasia mutated
ATP	Adenosine triphosphate
B-ALL	B-cell acute lymphoblastic leukemia
BCL10	B-cell lymphoma/leukemia 10
BCR	Breakpoint Cluster Region
CDH1	Cadherin 1
CEX	Coding Exome Oligos
CLL	Chronic Lymphoblastic Leukemia
CML	Chronic Myeloid Leukemia
CN-AML	Cytogenetically Normal Acute Myeloid Leukemia
CT3	Capture Target Buffer 3
CXCL12	C-X-C Motif Chemokine Ligand 12
DNA	Deoxyribonucleic Acid
DPYD	Dihydropyrimidine dehydrogenase
EE1	Enrichment Elution Buffer 1
ENHIS	Environment and Health Information System
EPM	Enhanced PCR Mix
ERK	Extracellular Sgnal - Regulated Kinases
ET2	Elute Target Buffer 2
FAB	French - American - British
FLT3	Fms Related Tyrosine Kinase 3
GATK	Genome Analysis Toolkit
GDN	GUaRDIAN
GDP	Guanosine Diphosphate
GTP	Guanosine-5'-triphosphate
IARC	International Agency for Cancer Research
IGV	Integrated Genome Viewer
ITD	Internal Tandem Duplication
JAK	JAnus Kinase
JMML	Juvenile Myelomonocytic Leukemia
KRAS	Kirsten Rat Sarcoma

LIG	Ligation
LIG2	Ligation Mix 2
LoF	Loss of Function
MAPK	Mitogen-Activated Protein Kinase
MSH2	mutS homolog 2
MUTYH	mutY homolog
ND1	NADH Dehydrogenase Subunit 1
ND2	NADH Dehydrogenase Subunit 2
ND4L	NADH dehydrogenase 4L
NEM	Enrichment Amp Mix
NF1	Neurofibromin 1
NF-κB	Nuclear Factor Kappa B
NRAS	Neuroblastoma Rat Sarcoma
NSCLC	Non-Small Cell Lung Carcinoma
OD	Optical Density
OMIM	Online Mendelian Inheritance in Man
PCR	Polymerase Chain Reaction
PI3K	phosphatidylinositol 3-kinase
PPC	PCR Primer Cocktail
PTP	Protein Tyrosine Phosphatase
PTPN11	Protein Tyrosine Phosphatase Non-Receptor Type 11
RAS	Rat Sarcoma
rCRS	revised Cambridge Reference Sequence
RFU	Relative Fluorescence Units
ROS	Reactive Oxygen Species
RSB	Resuspension Buffer
RTK	Receptor Tyrosine Kinase
SETBP1	SET binding protein 1
SHP-2	Src homology region 2 domain-containing phosphatase-2
SMB	Streptavidin Magnetic Bea
STAT	signal transducers and activators of transcription
STL	Stop Ligation Buffer
SWS	Streptavidin Wash Solution
TET2	Tet methylcytosine dioxygenase 2
WHO	World Health Organization
WSR	World Age-Standardised Rate

IEC, CIVIL HOSPITAL, AIZAWL.
COMMUNICATION TO THE PRINCIPAL INVESTIGATOR/GUIDE BY THE MEMBER SECRETARY, INSTITUTIONAL ETHICS COMMITTEE

No.B.12018/1/13-CH(A)/IEC/*70* **Dated: 18th June, 2018**

To,

Dr. N. Senthil Kumar
Professor and Head, Department of Biotechnology
Mizoram University

Subject: Ethics Committee Approval for the referenced projects.

Reference: ***"Genomic and proteomic profiling of Leukemic patients in Mizo population"***

Dear, ***Dr. N. Senthil Kumar***

With reference to submission of document for review and approval to conduct the above mentioned study. The Ethics Committee has reviewed and approved the study documents as mentioned below:

1. Curriculum Vitae of non Civil Hospital, Aizawl Investigators
2. Brief description of proposal/summary
3. Copy of the Protocol/Project and questionnaire (if any)
4. Copy of Patient information sheet & Consent form in local language
5. Copy of Clinical trial agreement
6. Copy of PI undertaking

The following members of the Ethics Committee were present at the meeting held on date 14th June, 2018 at time **1:30 pm** at Medical Superintendent Hall, Civil Hospital, Dawrpui, Aizawl, Mizoram - 796001.
The quorum met as per ICG-GCP and schedule Y guidelines as mentioned below:

S. No	Name of the member	Qualification	Designation in the Ethics Committee	Gender
1.	Dr. Lal Biakkima	Director H&ME (Rtd)	Chairman	Male
2.	Dr. T. Lalzawmliana MD	HoD, Dept. of Biochemistry	Member	Male
3.	Dr. Mary Muanpuii Ralte MD	HoD, Dept. of Anesthesiology	Basic Medical Science	Female
4.	Rev R. Lalchangliana	Pastor	Member	Male
5.	Dr. Zoengpari	Associate Professor, MZU	NGO/Social activist	Female
6.	Pu Rosangzuala Ralte	Advocate (Legal Expert)	Member	Male
7.	Dr. Saia Chenkual	HoD, Dept. of Surgery	Member	Male

Please note that this Ethics Committee is constituted as per schedule Y, ICH-GCP, applicable local laws and regulatory requirement.

We approve the project to be conducted in its present form. Ethics Committee expects to be informed about:

1. Any SAE occurring in the course of the study
2. A copy of final individual center report

We hereby confirm that neither you nor your study members have participated in the voting/decision making procedure of the Ethics Committee.

Yours sincerely,

-Sd-
(DR. LAL BIAKKIMA)
Chairman
Institutional Ethics Committee
Civil Hospital, Dawrpui, Aizawl
Mizoram – 796001

Chairman
Institutional Ethical Committee
Civil Hospital, Aizawl

T. Lalzawmliana 18/6/18

(DR T. LALZAWMLIANA)
Member Secretary
Institutional Ethics Committee
Civil Hospital, Dawrpui, Aizawl
Mizoram – 796001

Secretary
Institutional Ethical Committee
Civil Hospital, Aizawl

Note : Since Dr. C. Lalchhandama, Member Secretary is the Co-Principal Investigator for this study, this committee unanimously appointed Dr. T. Lalzawmliana (Member, IEC) as Member Secretary for this meeting.

A heterogeneous group of diverse and biologically distinct subgroups of haematopoietic malignancy or cancer of the blood is termed as leukemia. They are different from lymphoma as they do not form any solid tumour. Leukemia can be classified into their cell lineage: lymphoid (B and T - cells) and myeloid (granulocytes) as well as based on the behaviour of the cell that turns cancerous into acute or chronic. The most common cancer type in pediatric (0 – 19 years) is leukemia and in pediatric leukemia cases, lymphoid malignancy dominates other subtypes (Rudolph, 1996). Leukemia was first scientifically reported as a clinical entity by John Hughes Bennet (Thomas, 2012). Since then, different classifications and their updated versions were implemented by the World Health Organisation (WHO). Broadly they can be classified as Acute Lymphoblastic Leukemia (ALL), Acute Myeloid Leukemia (AML), Chronic Lymphoblastic Leukemia (CLL) and Chronic Myeloid Leukemia (CML), where each of them has a distinct subtype within their own types.

International Agency for Cancer Research (IARC) from 62 countries comprising 532 invited cancer registries, the incident cases for childhood cancer (age 0 – 19 years) totals to 385509 between 2001 to 2010 and the most common cancers were leukaemias with world age-standardised population incidence rate of 46.4 (WSR) (Steliarova-Foucher *et al*., 2017). In European countries, leukemia accounts for 30% of all the cancers diagnosed (less than 15 years of age), becoming the most common childhood malignancy (ENHIS – WHO, 2009). In India, cancer ranks ninth position in the cause of death in children within the age of 5 and 14 years and annually, approximately 45,000 children are diagnosed with cancer (Arora and

Kanwar, 2009). Mizoram, one of the Northeastern states of India, was being dubbed as Cancer Capital of India by Times of India in December 2016 (Times of India ; CancerIndia). A review paper describing pediatric leukemia and lymphoma incidence in India reported an average of 7 cases in Mizoram from 2012 – 2014 (Asthana *et al.*, 2018).

In the past, AML was classified according to the morphology of the cell type that turns cancerous which was called as the French - American – British (FAB) classification in 1976. This forms the basis of diagnosis for clinicians. A ground breaking discovery of fusion protein called Philadelphia chromosome (BCR-ABL) by Peter C. Nowell in 1960 (Nowell 1960) and the series of subsequent functional confirmation experiments in animal models confirmed that it was oncogenic and the causal factor for CML (Daley *et al.,* 1990; Elefanty *et al*., 1990; Heisterkamp *et al*., 1990). This formed the foundation that CML are linked with chromosomal translocation. In light of Nowell's discovery, many researchers discovered more translocations that are involved in leukemia and a signature mark for leukemia. Today, recurrent fusion genes in leukemia have also been established that have an impact on targeted diagnosis and prognosis (Wang *et al*., 2017). The findings of chromosomal translocation changed the whole scenario in leukemia classification. The earlier FAB classification of AML was also strengthened based on chromosomal translocation as the classification didn't consider the factors that have implications in the prognosis (American Cancer Society). Recently, in 2016 World Health Organisation updated a revised classification of leukemia.

Apart from chromosomal translocation positive, there is also chromosomal translocation negative leukemia such as cytologically negative AML (CN - AML) which suggested that

chromosomal translocation may not be the main causal factor for leukemiogenesis. In line with this, there are reports that chromosomal translocation is not enough to drive normal cells to malignancy (Castilla *et al*., 1999) and are even found during pregnancy (Hjalgrim *et al*., 2002; McHale *et al*., 2003; Wiemels *et al.,* 1999; Wiemels *et al*., 2002; McHale *et al*., 2003). Similar evidence also indicates that translocations are found in 100-fold more in the healthy population than leukemic patients (Eguchi-Ishimae *et al*., 2001; Mori *et al*., 2002). This strongly suggested that most translocations are not sufficient for pathogenesis but could be a predictive marker for risk of developing leukemia in those children who possess the fusions genes.

In 1971, "Knudson Hypothesis" was proposed stating the requirement of two hits for cancer to develop (Knudson 1971). Following the same hypothesis, a two hit model of leukemiogenesis was suggested for haematopoietic malignancy. One hit confers a proliferative advantage and the other hit confers differentiation arrest and evasion of programed cell death ie. apoptosis (Gilliland and Griffin, 2002). This could also mean that the above stated translocation could be the first hit during pregnancy where the translocated region would be the genes that confer proliferative advantage and the second hit could be a somatic mutation in the genes that controls the differentiation of the haematopoietic progenitor cells or may play a role in signalling the cell to die.

Leukemic pathogenesis is one of the most extensively studied in cancer that give rise to targeted therapies leading to a successful treatment depending on the sub type of leukemia with cure rate of about 80%, thus reducing the mortality (Wiemels, 2008). This is also in - part because of the identification of co-operating gene mutations that involves in its

pathogenesis as stated from the two hit model. Identifying a specific gene mutation for a specific subtypes of leukemia plays a critical role in its diagnosis, prognosis and also to the level of predicting its disease free survival rate and its relapse. Similar to translocations, these gene mutations plays a key event in diagnosis and prognosis of leukemia (Boissel *et al*., 2008).

With the advancement of technology in DNA sequencing, Massively Parallel Sequencing now holds the power for mass screening of these gene mutations either through Whole Genome Sequenicng, Whole Exome Sequencing or targeting a panel of the desired genes. These platforms produce high throughput data as well as large scale genomic discovery of variants in the DNA with a much lower cost (Mardis, 2011). In parallel with the development of the technology, the bioinformatics tools also grows rapidly thus facilitating the ease of variant detection. The first Next Generation Sequencing platform was developed in 2005 by 454 Life Sciences in 2005 (http://www.my454.com) which laid the foundation for discovery genomics in identifying variants that have a clinical significance. Since then, it is widely being used for whole exome study that specifically targets the coding regions suggesting a direct link to the disease phenotype. From such studies, mutational background of pediatric leukemia has also been established in different ethnicity revealing recurrent mutational hotspots, driver genes and variants effecting in different pathways: RTK/RAS signalling pathway and its downstream MAPK/ERK signalling pathway, PI3K/AKT pathway and MTOR pathway, JAK/STAT signalling pathway Notch signalling pathways, WNT/β-catenin Pathway, CXCL12 Pathway, NF-κB pathway, Metabolic pathway and other pathways including p53 pathways (Bonaccorso *et al*., 2015; Farrar *et al*., 2016; Zhang *et al*., 2019; Mirabilii *et al*., 2012). The class of genes that are frequently mutated includes

lymphoid/myeloid differentiation, transcription factors, epigenetic regulators, signal transduction, apoptotic regulators etc. (Grimwade *et al*., 2015; Ding *et al*., 2018).

RAS signaling pathway is one of the well-studied pathways relating with childhood leukemia. Mutations in RAS genes are frequent in pediatric AML (Goemans *et al*., 2005). RTKs activates the mitogen-activated protein (MAP) kinase cascade which is one of the most common signal transduction pathways within the cell. The pathway is initiated by activating RAS, which is attached to the plasma membrane after triggering by a signal induced - RTK (Receptor Tyrosine Kinase). The end signal in the pathway activates transcription factors, expressing genes responsible for proliferation and angiogenesis. Mutations affecting RTK/Ras/MAPK signaling causes many human diseases, one of which is leukemia (Malumbres and Barbacid, 2003).

Members of the RAS GTPase family are one of the crucial molecules in many signaling pathways, involving in cell cycle regulation, proliferation and subsequent programmed cell death (Medarde and Santos, 2011). Mutations in RTKs and RAS were identified in 90% of patient samples and were the predominantly altered pathways during AML leukemogenesis (Farrar *et al*., 2016). Similarly, N-RAS, K-RAS, FLT3, NF1 and PTPN11 genes are frequently mutated in pediatric acute lymphoblastic leukemias (ALLs), which plays a role in RTK/RAS signaling pathway (Paulsson *et al*., 2008).

RAS oncogenes binds to many receptors present on the membrane there by regulating the signals among which is FLT3, and functions in cell growth, differentiation as well as leading the cell to apoptosis. Single nucleotide mutations are in N-RAS are found in approximately 15% in AML patients, where in K-RAS it was found to be about 5% (Renneville *et al*., 2008).

In many of the cancer sites, mutations in N-RAS or K-RAS have been identified especially in lung, colon as well as in leukemia (Goemans *et al*., 2005). Mutational hotspots of the RAS genes falls in the amino acid positions 12, 13, and 61 in exons 2 and 3 in the N-RAS and K-RAS (Perentesis *et al*., 2004; Shu *et al*., 2004).

The FMS-like tyrosine (FLT3) gene functions in controlling hematopoiesis. Expression of this gene at high level has been reported in haematological malignancy as in the case of acute myeloid leukemias (AMLs) and B-cell acute lymphoblastic leukemia (B-ALL) (Testa and Pelosi, 2013). FLT3 is the commonly mutated gene in AML, occurring in about 25–45% of all AML patients. Internal tandem duplication (FLT3 - ITD) in exons 11 and 12 were seen in 12% of pediatric AML constituting majority of the FLT3 mutations (Iwai *et al*., 1999; Kondo *et al*., 1999). The second is a single nucleotide mutation in exon 17, of tyrosine kinase domain (TKD) within activation loop involving the amino acid position 835 and/or 836 and infrequently also observed in N841 or Y842 which are reported approximately in 5 - 10% of AML patients (Yamamoto *et al*., 2001).

The Protein Tyrosine Standard Phosphatase Nonreceptor 11 (PTPN11) gene translates to a protein tyrosine phosphatase (PTP) and is localized in the cytoplasm called SHP-2, which actively plays a role in signal transduction via. the RTK/RAS/MAP kinase cascade (Tartaglia *et al*., 2004; Goemans *et al.*, 2005). PTPN11 mutations are reported to constitute about 35% in patients with juvenile myelomonocytic leukemia (JMML). JMML patients with PTPN11 mutation negative either harbours a homozygous NF1 deletion/inactivation or a RAS mutation highlighting the crucial role of activated RAS in leukemia (Tartaglia *et al.*,

2004). Common mutations on exons 3 and 13 were identified (9%) which play a role in RTK/RAS signaling pathway (Paulsson *et al*., 2008).

RAS signaling pathway regulates the normal cell growth and differentiation by a cyclic interchange of conformation from a GTP bound the active conformation to a GDP bound form the inactive conformation (Hall *et al*., 1992). NF1 gene encodes neurofibromin protein which interacts with RAS with high affinity, and increases the cyclic conversion of RAS to form the active to inactive complexes, to and fro (Boguski and McCormick, 1993). This suggested that NF1 negatively regulates RAS-GTP and acts as a tumour suppressor (Basu *et al*., 1992). Infants and pediatrics with NF1 polymorphisms have a higher chance of risk in developing malignant myeloid disorder (Shannon *et al*., 1992). Patients with these NF1 microdeletions display a more severe NF1 phenotype, characterized by increased risk for developing leukemia. One of the most mutated hotspots are exons 22 (R1276) (Klose *et al*., 1998), 28 and 29 (Peters *et al*., 1999).

Apart from the nuclear genome, mitochondrial genome also plays a critical role in disease phenotype as the 16,569 bp mitochondrial genome comprises of 2 rRNAs, 22 tRNAs and 37 genes coding for 13 polypeptides which are involved in respiratory mechanism of the cell (DiMauro *et al*., 2003). Mitochondria is devoid of histone proteins and an inadequate repairing of DNA, on top of that the high amount of reactive oxygen species (ROS) as a result of oxidative process in the mitochondrial membrane makes mtDNA highly susceptible to aberration (Yakes *et al*., 1997; Marcelino *et al*., 1999). The mtDNA has about ten to twenty - fold higher mutation frequency when compared with the nuclear DNA (Johns, 1995; Grossman and Shoubridge, 1996). Events of mutations in mtDNA in tumour cells are

found to be true with the reports that tumour cells are subjected to constitutive oxidative stress (Toyokuni *et al*., 1995), which is in line with the function of reactive oxygen species (ROS) that initiates and promotes carcinogenesis (Wei, 1998). The mtDNA polymerase has no proofreading activity as compared with its nuclear counterpart suggesting that mutations in the mtDNA could accumulate more in the mitochondrial genome than the nuclear genome (Shadel *et al*., 1997). The environment (high ROS), function (respiration: ATP synthesis), as well as its machinery (DNA polymerase) acting on mitochondrial genome favours higher rate of mutations facilitating carcinogenesis.

mtDNA variants have been characterized for different malignancy. In colorectal cancer, mtDNA mutations suggest that it might confer proliferative advantage than the neighboring cells and homoplasmic mutant cells could have provided a replicative advantage (Polyak *et al*., 1998). However, a causative link between mtDNA mutations and malignancy has not yet been firmly established (Chial and Craig, 2008). In hepatocellular carcinoma, the D-loop polymorphism 150 C than T is associated with early recurrence and 16263 T than C with tumour-free survival time, but not for overall survival time (Shilai *et al*., 2016). In NSCLC (Non-Small Cell Lung Carcinoma), mtDNA polymorphisms were found to be correlated with pathogenesis and prognosis (Hu *et al*., 2015, Lam *et al*., 2012). mtDNA polymorphism of D310 is known globally as a biomarker for lung cancer diagnosis (Chen *et al*., 2016) as well as Tcell-ALL risk (Kwok *et al*., 2011).

Several studies on the mtDNA mutations in adult leukemia has been reported with interesting findings like 40% of the patients harboured mtDNA mutations and that A15296G is a leukemia specific marker (He *et al*., 2003). Another group also conducted a study and

found that 21% of the studied 56 AML patients have T16311C which is associated with chromosomal aberrations with better prognosis (Silkjaer *et al*., 2013). The study on the mtDNA mutations in paediatrics is much lesser that adult (Kang *et al*., 2016). Regions in mtDNA like D-loop, ND1, region has also been studied in adult leukemic patients (Grist *et al*., 2004; Yao *et al*., 2007; Han *et al*., 2013). However, in 44 cases of paediatric AML, three variations each of which a T>C change in hyper variable region-I within the D – Loop region positions at 16126, 16224 and 16311 confers a poor event free survival (Sharawat *et al*., 2009). From the studied 39 pediatric ALL patients, eight mutations (missense or nonsense) were observed which are in heteroplasmic condition as well as novel synonymous variants were also reported in genes that encodes subunits of complex1 such as ND1,ND2,ND4L,ND6 and ATP6 and 8 respectively (Järviaho *et al*., 2018). Loss of function (LoF) mutation was found in 6.3% in ALL and 10% in AML among the studied pediatric haematological malignancies (Triska *et al*., 2019).

Genomic screening in cancer for the identification of biomarkers that has a diagnostic and prognostic potential has been carried out for many decades, among which identification of High – MSI (MSI-H) and Low – MSI (MSI-L) is one of the approach. Microsatellites are di-, tri-, or tetra nucleotide tandem repeats present in thousands of regions within the genome (Richard *et al*., 2008) and have higher chance of mutation then other regions (Brikmann *et al*., 1998). Mechanism of this aberration by repetition of segments was proposed to be DNA polymerase slippage (Tautz *et al*., 1994). Cancer cells that have large numbers of microsatellite instability may have defects in the ability to correct mistakes that occur when DNA is copied in the cell. These defects are corrected by a class of proteins called DNA mismatched repair enzymes (Boland and Goel. 2010). Therefore, presence of MSI-H could

potentially relate to the loss of function or inactivation of the Mismatched Repair genes affected by mutations with thin the gene.

MSI is also frequently found in cancer cells and also might confer benefits from clinical immunotherapies (Bonneville *et al.*, 2017). The first case of pediatric neoplasia related with genomic instabilities suggested that defects in DNA repair machinery might involve in pediatric T cell ALL (Baccichet *et al*., 1997). Several studies have reported MSI in leukemic patients (Webb *et al*., 1999; Kendall *et al*., 2004; Mao *et al*., 2008; Patel *et al*., 2017) but some authors also reported no MSI in their study group (Walker *et al*., 2017).

This pilot study to understand the germline mutational background in both the nuclear and mitochondrial genomes of pediatric leukemia cases is the first scientific report to be conducted in Mizo population. This study could be a platform in understanding the genetic predisposition for high incidence of cancer within the Mizo population.

Sample collection

A total of 11 children between 2- 16 years (mean age = 8.5) who are diagnosed with leukemia (5-ALL, 4-AML, 1-CML and 1-JCML) and underwent treatment at Mizoram State Cancer Institute, Aizawl from January – July 2018 were included in this study (Table 1). After obtaining consent from the mother or father, 2 ml of peripheral blood was drawn from the children and from four mothers. Peripheral blood was collected in EDTA coated vials and stored in -20^{0}C for DNA isolation. Ethical clearance was obtained from Institutional Ethics Committee, Civil Hospital Aizawl.

DNA isolation

DNA was isolated from whole blood by using QIAamp DNA Mini Kit (CA, USA) with some modifications as per the manufacturer's protocol. The quality of isolated DNA was checked using Nanodrop (NanoDrop™ 1000 Spectrophotometer, Thermofisher) at optical density (OD) 260 nm. The purity of the isolated DNA was checked by measuring OD at 260/280 for protein contamination as well as 260/230 for RNA contamination (Table 2). The quality of the isolated DNA was also checked by 0.8% Agarose Gel Electrophoresis (Fig.1).

DNA library preparation

After the required concentration of 100ng for library preparation was obtained (Fig. 2), DNA library was prepared by using Illumina v4 TruSeq Exome library prep as per the manufacturer's protocol.

DNA fragmentation by Sonication

From all the samples, 50 µl was taken for DNA fragmentation using Covaris S220, to shear the DNA to a desired product of 150 basepair (bp) genomic DNA fragment using the following settings as Duty factor: 10%, Peak Power: 175W, Cycle burst: 200 and Duration: 280 secs and Temperature: 7^{o}C.

End Repair and Purification by size selection using beads

Shearing of DNA by sonication leads to an uneven breakage of the DNA, and this was converted to blunt ends by End Repair Mix (ERP) buffer which removes the 3' – 5' overhangs and polymerizes the 5' – 3' ends of the sheared DNA by the content of this buffer. Magnetic bead which are a component of Sample Purification Buffer (SPB) was used to select only the desired 150 bp fragment by changing the concentration of SPB ratio with the fragment DNA to be purified. The repaired, blunt ends of a fragment DNA was adenylated for the ease of ligating the adapters to both of the fragmented DNA. This was done by using A-Tailing Mix (ATL2) which ligates the nucleotide adenine to the 3' ends. This felicitates the adapter ligation as thymine is present at the 5'- ends of the adapters as well as ensures a low rate of producing chimeras (concatenated template).

Adapter ligation

The indexing adapters (Table 3) were ligated to the adenylated 3' – ends using Ligation Mix 2 (LIG2) for hybridization in the downstream process. After adding LIG2, it was incubated at 30^{0}C for 10 mins after which it was kept in thermal cycler and the Ligation program (LIG

program) was run. To stop the ligation process, Stop Ligation Buffer (STL) was added and was incubated in a shaker at 1200 rpm for 2 mins and the clean-up was performed.

Enrich DNA Fragment

To enrich the DNA fragment, PCR was performed using Enhanced PCR Mix (EPM) and PCR Primer Cocktail (PPC) as 95°Cfor 3 mins, 8 cycles of 98°C for 20 secs 60°C for 20 secs 72°C for 30 secs, 72°C for 5 mins and held at 4°C with a 50 µl final reaction volume. The quantity and quality of the enriched library was measured using Nanodrop.

Probe Hybridization

The enriched DNA fragments were then subjected to hybridization with a capture probes that will only target the exonic region of the genome using Coding Exome Oligos (CEX) and Capture Target Buffer 3 (CT3) after pooling of the samples for 9-plex (Table 4) as 95°C for 10 mins, 18 cycles of 1 min each, starting at 94°C, then decreasing 2°C per cycle, 58°C for 90 mins and hold at 58°C. Each well contained 100 µl of hybridized captured fragments.

Using Enrichment Elution Buffer 1 (EE1), Elute Target Buffer 2 (ET2), HP3 (2 N NaOH) Streptavidin Magnetic Beads (SMB) and Streptavidin Wash Solution (SWS) the hybridized probes were captured. After two heated wash to remove non-specific binding was performed it was ready for a second capture. A second Capture was performed to ensure high specificity of the target regions followed by the clean-up process.

Amplification of Enriched Library

An 8- cycle PCR was performed using Enrichment Amp Mix (NEM) and PCR Primer Cocktail (PPC) to amplify the enriched library giving conditions as 98°C for 30 seconds, 8 cycles of: 98°C for 10 secs, 60°C for 35 secs, 72°C for 30 secs, 72°Cfor 5 mins and held at 4°C. Each well contained 50 µl and a clean-up was performed. Agarose Gel Electrophoresis (2% gel) was ran to check the quality of the amplified enriched library (Fig. 3).

Cluster Generation

A bridge PCR was performed to generate cluster of the DNA fragments using C – Bot, illumina and after which it was loaded to illumine HiSeq2500 for sequencing.

WES data analysis

Pre-processing of the sequence file

The quality of the raw read fastq files were checked twice before and after trimming the adapter sequence and the low quality reads by Trimmomatic software (Bolger, 2014) (AVG = Q20, MINLENGHT = 50, SLIDINGWINDOW = 5:20) using fastQC (Andrews, 2010).

Aligning to the reference genome and Variant Calling

Processed fastq files were mapped on human reference genome sequence (hg19) using BWA (Li *et al.*, 2009). Then variant calling was done using GATK haplotype caller (McKenna *et al.*, 2010). The results of alignment are given in Table 4.

Annotating variants

Annotation of the variants was performed with Annovar (Wang *et al*., 2010) using databases such as refGene (http://varianttools.sourceforge.net/Annotation/RefGene), ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/), LJB (https://sites.google.com/site/jpopgen/dbNSFP), dbSNP (https://www.ncbi.nlm.nih.gov/snp/), ExAc (exac.broadinstitute.org/), esp6500 (https://evs.gs.washington.edu/EVS/), 1000 Genome (Gibbs *et al*., 2015) for testing the reported variants as well as their allele frequency and predicting the functional effect of the variant using different functional predicting software (LJB database).

Prioritization of variants

The list of genes present in the annotated file for each sample was subjected to VarElect (Stelzer *et al*., 2016) to find the genes which are only related to leukemia. The genes which show the association were selected for further analysis. From the selected genes, the first filter for the variants was performed by filtering only to the exonic region and discarding the synonymous variants. Then, the second filter was used to remove those variants that have a higher allele frequency (> 0.05). Gene function prediction tools like SIFT (https://sift.bii.a-star.edu.sg/), Polyphen2 (http://genetics.bwh.harvard.edu/pph2/), Provean (http://provean.jcvi.org/index.php) and Mutation Taster (http://www.mutationtaster.org/) to identify deleterious/pathogenic variants were used as filter three to predict the effect of the variants to their protein functions. List of variants after each filter steps are given in Table 5.

Leukemia related genes

Gene variants which are reported in leukemia patients were listed out after performing data mining through literature survey as well as which are catalogued in databases. The list of genes (Table5) was then matched with the outputs of both filters 2 and 3. The variants which were left after matching with filter 2 were then matched with CIViC database (https://civicdb.org/ - which shows the clinically actionable variants) and also with ClinVar for reported disease specific variants, avSNP dataset (reformatted dbSNP datasets) to check the known variants and OMIM (https://www.omim.org/) for the hereditary conditions. List of the 55 leukemia related genes are given in Table 6.

Clinically actionable variants

All the samples were screened for the four genotypes in exons 6, 13, 14 and 22 in Dihydropyrimidine dehydrogenase (DPYD) gene for the responsiveness of the chemotherapy drug 5-Fluorouracil, a first line of treatment for cancer patients (Hariprakash *et al*., 2017) from the exome data using Integrated Genome Viewer (IGV) (Robinson *et al*., 2011).

Whole Mitochondrial Genome Sequencing

Sample collection and DNA isolation

Four patients with their corresponding mother blood samples were taken for this study (Table 12). DNA isolation from their peripheral blood was performed using QIAamp DNA Mini Kit (CA, USA) with some modifications as per the manufacturer's protocol. The quality of the isolated DNA was analysed in 0.8% Agarose Gel Electrophoresis (1X TE buffer and

Ethidium Bromide dye) at 80V for 30mins and the resulting gel images were obtained in GBOX gel documentation system, UK (Fig. 3).

Library preparation of the whole mitochondrial genome

The whole mitochondrial genome (16.6 kb) was subjected to PCR amplification to give rise to a two overlapping amplified products (9.3 + 7.6 kb) using Long PCR Kit (Sequal Prep™) (Sha *et al.,* 2010) in ABI 9700 thermal cycler (Thermo Scientific). The two overlapping amplicons was run on 0.8% Agarose Gel for quality checking and purification was performed using QIAquick Gel Extraction Kit (Qiagen). The purified amplicons were then pooled together for library preparation which was performed using Nextera XT DNA Library Kit (illumina, USA) as per the manufacturer's protocol. The libraries prepared was subjected to a pair-end sequencing in illumine HiSeq 2500 (illumiina, USA).

Whole mitochondrial genome sequence analysis

Pre-processing of the raw fastQ files were analysed using FastQC and BWA – MEM was used to align it to the reference mitochondrial sequence which is the revised Cambridge Reference Sequence (rCRS). The aligned .sam file was ensured to possess a read quality score of 40 (Q40) and < Q40 was discarded. The file was subsequently converted into .bam file using Samtools (Li, 2009). QualiMap was used to ensure enough coverage of the BAM file to rCRS. VarScan2 was used for variant calling and the called variants were annotated using ANNOVAR.

Variant prioritization

Variants that have read depth lesser than 3% of the total reads or have a coverage lesser than 10% of the total reads from either the directions were discarded. The variants obtained were then analysed using MitoMap and mtBrowse for reported variants or their functional implications to the disease.

Microsatellite Instability Test

Four patients with their corresponding mothers were screened for instability in their microsatellite region. The description of the screened region and the marker is given in Tables 19 and 20, respectively. For each of the markers used, PCR amplification was carried out for 35 cycles each at 95°C for 40 secs, varying annealing temperatures, 72°C for 40secs, followed by final extension at 72°C for 5mins and hold at 4°C (Table 8). The amplified PCR products were analysed in 2% Agarose Gel Electrophoresis in 1X TE buffer stained with Ethidium Bromide at 80V for 45 mins and the resulting gel images were obtained in GBOX gel documentation system (UK) (Fig. 4).

Sample preparation for MSI

Calculating for each of their relative fluorescence units (RFU) value to 100 units, each of the markers was pooled to their corresponding samples. From the pooled markers for each samples, 1 µl was taken and mixed with 9 µl of the cocktail reaction (Liz dye: 0.3 µl + 8.7 µl) to make the 10 µl final volume. Then, it was subjected to ABI 3500 automated DNA sequencer (Applied Bio system, USA) to check the mobility shifts of the interested repeated segments to confirm microsatellite instable.

Understanding the Germline Genomic Mutational background of Pediatric Leukemia patients in Mizo population.

The collected samples and their description is depicted in table 1

Sample ID	Age	Gender	Condition
GDN4252	13	F	ALL
GDN4253	11	M	ALL
GDN4254	14	M	ALL
GDN4263	38	F	Mother
GDN4255	16	M	AML – M1
GDN4264	42	**F**	Mother
GDN4256	12	M	ALL
GDN4257	11	M	AML
GDN4265	32	**F**	Mother
GDN4258	2	F	AML – M1
GDN4266	36	F	Mother
GDN4259	5	M	JCML
GDN4260	4	M	ALL
GDN4261	2	F	AML - M1
GDN4262	4	M	CML

Table1: List of samples obtained for the study containing sample identification, age, gender and their diagnosis. GDN: GUaRDIAN (sample identification code), M: male, F: female. ALL: Acute Lymphoblastic Leukemia, AML – M1: Acute Myeloid Leukemia of M1 subtype (according to FAB classification), CML: Chronic Myeloid Leukemia, JMML: Juvenile Myelomonocytic Leukemia. The corresponding patient (son/daughter) and mother are placed in a single row.

Isolation of DNA from pediatric leukemia patients and mother's sample (Figure 1)

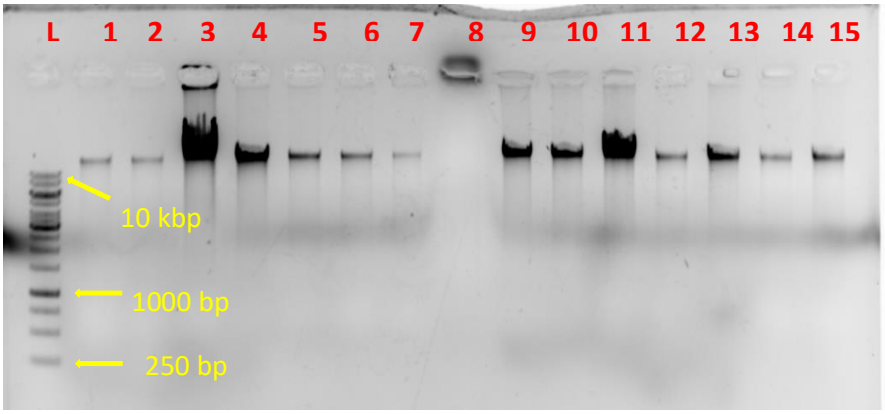

Figure1: Isolated genomic DNA from peripheral blood run in 0.8% Agarose Gel Electrophoresis. L = Ladder, 1-11 = pediatric leukemia DNA samples and 12 -15 = mother's DNA sample.

After normalising the samples to 100ng, quality control was perform in 0.8% agarose gel electrophoresis given in table2

GDN ID	Abs ng/ul	For 100 ng (µl)	RSB (µl)
GDN4252	15.71	31.82	18.17
GDN4253	12.09	41.35	8.64
GDN4254	405.49	1.23	48.76
GDN4255	51.89	9.63	40.36
GDN4256	19.84	25.20	24.79
GDN4257	15.01	33.31	16.68
GDN4258	14.4	34.72	15.27
GDN4259	139.54	3.58	46.41
GDN4260	43.65	11.45	38.54
GDN4261	42.28	11.82	38.17
GDN4262	95.63	5.22	44.77
GDN4263	19.54	25.58	24.41
GDN4264	40.46	12.35	37.64
GDN4265	20.38	24.53	25.46

Table 2: List of sample concentration quantified using nanodrop and calculations to 100ng/µl. GDN: GUaRDIAN, Abs: Absorbance, µl: microliter, RSB: Re-suspension Buffer.

Quality control was performed using 0.8% agarose gel electrophoresis (Fig2). Indexing used to differentiate each specific sample is given in Table 3

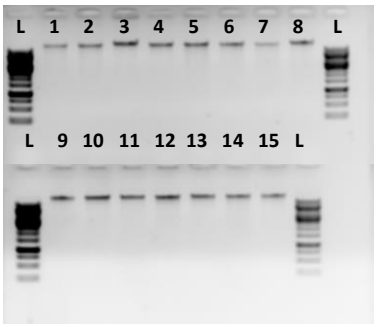

Fig 2: Samples normalized to 100 ng/µl in 0.8% agarose gel electrophoresis. L: 10kilobase pair ladder, 1-15: samples.

Samples	Index used	Plexing ID
GDN4252	D701 – D501	TSQ103
GDN4253	D701 – D502	
GDN4254	D701 – D503	
GDN4255	D701 – D504	
GDN4256	D701 – D505	
GDN4257	D701 – D506	
GDN4258	D701 – D507	
GDN4259	D701 – D508	
GDN2460	D702 - D501	TSQ104
GDN4261	D702 - D502	
GDN4262	D702 - D503	
GDN4263	D702 - D504	
GDN4264	D702 - D505	
GDN4265	D702 - D506	

Table 3: List of samples with index and plexing ID.

Final QC was performed after library preparation was done using agarose gel electrophoresis (Fig. 3). Number of reads aligned to the reference genome (hg19) for all sample is listed in Table 4.

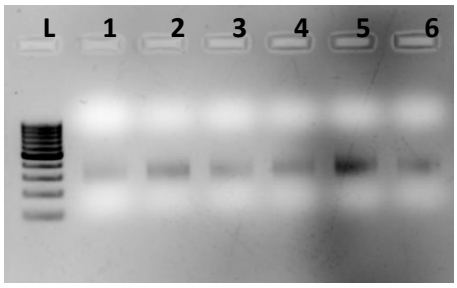

Fig 3: QC of the final library preparation using Agarose gel electrophoresis. L=10kb ladder, 1-6 = the plexing ID's. 1 and 2 contains the TSQ103 and 104, respectively where the samples are pooled.

Samples	QC pass reads	Aligned reads	% of alignment
GDN4252	64265254	64261094	99.99
GDN4253	35548941	35546525	99.99
GDN4254	84911078	84907234	99.99
GDN4255	75146989	75143138	99.9
GDN4256	30669406	30658577	99.96
GDN4257	10760797	10755716	99.95
GDN4258	33683692	33668629	99.96
GDN4259	64428370	64404387	99.96
GDN4260	49560787	49557628	99.99
GDN4261	59679153	59675442	99.99
GDN4262	65613087	65608945	99.99
MEAN	**52206141**	**52198847**	**99.97**

Table 4: List of sequencing reads that passed quality check and aligning to the reference genome with the percentage for each leukemic patients. QC: Quality control, %; Percentage.

From the aligned reads, variant calling was performed using GATK – haplotype caller and the variants obtained are annotated using ANNOVAR. The variants were then filtered using the aforementioned filters, the number of variants observed after each filters are given in Table 5. For ease of prioritizing the variants, frequently mutated genes associated with leukemia are listed which will be matched with the sample variants (Table 6)

Samples	**Annotated file**	**After Filter1**	**After filter 2**	**After filter 3**
GDN4252	512926	12980	1699	129
GDN4253	348029	12476	1732	148
GDN4254	683227	13022	1884	162
GDN4255	743429	12798	1823	147
GDN4256	337397	12723	1902	154
GDN4257	129832	11783	2120	234
GDN4258	331558	12754	1865	168
GDN4259	557585	12808	1739	136
GDN4260	534159	12811	1848	165
GDN4261	668326	12914	1767	160
GDN4262	656041	13015	1850	144
Mean	**500228**	**12735**	**1839**	**159**

Table 5: List of total number of variants in each sample from the raw annotated files.

NRAS	NF1	CBLB	CRLF2	CEBPA	IDH2
KRAS	NPM1	KMT2A	SETBP1	PDGFRB	CTCF
FLT3	BCR	KMT2D	BCL2L11	U2AF1	PTPN11
ABL1	JAK2	CREBBP	CSF1R	PDGFRA	GATA2
ATM	BCL10	EP300	BTK	IKZF1	CDNK2A
KIT	BIRC3	ARID1A	GNAS	MTOR	
JAK3	TP53	ASXL1	MYD88	NT5C2	
CSF3R	JAK1	TET2	BIRC3	DNMT3A	
RUNX1	NPM1	PAX5	POT1	IKZF3	
ETV6	NOTCH1	IKZF2	IDH1	WT1	

Table 6: List of frequently mutated genes in leukemia.

Each sample had a total mean variant of 500228, after discarding the variant that lies in the intronic, variant resulting in a synonymous change, stringent allele frequency cut off and using 3 prediction tools showing pathogenic or deleterious effect of the variant, three patients harboring one variant each are observed (Table 7).

Sample	Gene	AA change	ClinVar
GDN4255	NOTCH1	V1699E	Not Reported
GDN4257	RUNX1	R376C	Not Reported
GDN4258	MUTYH	G143E	Polyposis, Hereditary Cancer Predisposing Syndrome (rs730881833)

Table7: 3 Variants observed in samples after applying filter 1, 2 and 3. V: Valine, E: Glutamic acid, R: Arginine, C: Cystein, G: Glycine.

After applying filter 2, variants observed were then matched with the aforementioned 55 Leukemia related genes and then to ClinVar database to check for known variants with their clinical significance and the associated disease phenotypes (Table 8).

Chr	Pos	Ref	Alt	Gene	Clinical Significance	Disease associated
11	108098555	A	G	ATM	Conflicting interpretations of Pathogenicity	Ataxia-telangiectasia syndrome, Hereditary cancer-predisposing syndrome
11	108159732	C	T	ATM	Benign / Likely Benign	Ataxia-telangiectasia syndrome, Hereditary cancer-predisposing syndrome
11	119156193	C	T	CBL	Benign / Likely Benign	Rasopathy, Noonan-Like Syndrome Disorder
12	49434409	G	A	KMT2D	Benign	Kabuki syndrome
1	45797401	G	A	MUTYH	Conflicting interpretations of Pathogenicity	MYH-associated polypopsis, Hereditary cancer-predisposing syndrome
1	**45797914**	**C**	**T**	**MUTYH**	**Pathogenic / Likely Pathogenic**	**MYH-associated polypopsis, Hereditary cancer-predisposing syndrome**
1	45800146	C	T	MUTYH	Benign, Uncertain Significance	MYH-associated polypopsis, Hereditary cancer-predisposing syndrome
1	45800167	G	A	MUTYH	Benign, Uncertain Significance	MYH-associated polypopsis, Hereditary cancer-predisposing syndrome
18	42643270	G	T	SETBP1	likely Benign	Schinzel-Giedion syndrome
1	85742023	C	A	BCL10	Benign	Immunodeficiency 37
20	31022469	G	A	ASXL1	Benign	C-like syndrome
22	23654017	G	A	BCR	Uncertain Significance	ALL and AML
4	106158550	G	T	TET2	Not provided	
4	55589830	A	G	KIT	Uncertain Significance	Gastrointestinal stroma tumor
9	139401375	C	T	NOTCH1	Uncertain Significance	Adams-Oliver syndrome 5, Cardiovascular phenotype
9	139410139	T	C	NOTCH1	Uncertain Significance	Adams-Oliver syndrome 5

Table 8: List of matched variants with 55 leukemia and ClinVar genes after filter 2. Chrom: Chromosomal number, Pos: Position, Ref: Nucleotide in the reference genome, Alt: The altered variant from the reference genome.

A total of 46 variants that are not found in ClinVar are listed in Table 9 (Annexure I).

The 46 variants that are unmatched with ClinVar were then interpreted using OMIM for identifying the inheritance pattern as well as its prediction score using SIFT, Polyphen2 and Provean as given in Table 10 (Annexure II). The genotype of Dihydropyrimidine dehydrogenase (DPYD) for clinical interpretation for responsiveness of 5-FU drug is also screened from the bam files of each sample as listed in Table 11.

	Exon - 6, Chr1:98165019 (hg19) WT Genotype (T)			**Exon- 13, Chr1:97981343 (hg19) WT Genotype (A)**			**Exon - 14, Chr1:97915631 (hg19) WT Genotype (G)**			**Exon - 22, Chr1:97547947 (hg19) WT Genotype (T)**		
Sample	**C/C**	**C/T**	**T/T**	**C/C**	**C/A**	**A/A**	**T/T**	**T/C**	**C/C**	**A/A**	**A/T**	**T/T**
GDN4252			*			*						*
GDN4253			*			*						*
GDN4254			*			*						*
GDN4255			*			*						*
GDN4256			*			*						*
GDN4257			*			*						*
GDN4258			*			*						*
GDN4259			*			*						*
GDN4260			*			*						*
GDN4261			*			*						*
GDN4262			*			*						*

Table 11: Genotypic variants confering to responsiveness of 5FU drug (a first line of treatment in Cancer) screened using DPYD gene (metabolizer of 5-FU) in Pediatric Leukemia patients patients in Mizo population. * indicates the presence of a genotype. For Exon - 14 the observed genotype is the wildtype G. Genotype highlighted in RED confers highest severity, YELLOW confers mildly and GREEN confers low severity in terms of the toxic effect of the 5-FU drug.

Whole mitochondria genome sequencing reveals a list of variants in each sample as given in Table 12.

Sample	All	Exonic	Syn	Non - Syn	Condition
GDN4252	38	24	17	7	ALL
GDN4253	49	31	20	11	ALL
GDN4254	44	26	19	6	ALL
GDN4263	40	18	11	6	Mother
GDN4255	31	12	7	5	AML – M1
GDN4264	33	12	7	5	Mother
GDN4256	39	24	17	7	ALL
GDN4257	41	25	19	6	AML
GDN4265	46	28	20	7	Mother
GDN4258	46	28	20	7	AML – M1
GDN4266	45	27	20	6	Mother
GDN4259	45	27	20	7	JCML
GDN4260	40	20	14	6	ALL
GDN4261	33	12	6	6	AML
GDN4262	29	10	7	3	CML
MEAN	**39.6**	**21.6**	**14.9**	**6.3**	

Table 12: List of different classes of variants in the mitochondrial genome and the patient condition.

The corresponding patient (son/daughter) and mother are placed in a single row.

All: Total variants in each sample, Exonic: Total exonic variants in each sample, Syn: Total synonymous variants in each sample, Non – syn: Total non - synonymous variants in each sample and Condition: diagnosis of the disease.

Two novel unreported variants T119A (MT-ND2 gene) and Y165 (MT-ND6 gene) which encodes subunits of complex I of respiratory machinery were found (Table 13). D-loop variant T152C was also found in 7 samples (Table 14).

Variant	Gene	Sample	Condition
T119A	MT-ND2	GDN4255*	AML – M1
		GDN4261	Mother
		GDN4264*	Mother
Y165C	MT-ND6	GDN4253	ALL

Table 13: List of samples harbouring MT-ND2 T119A variant.

* indicates patient and corresponding mother sample.

A total of 7 samples out of the studied 15 harbours T152C variant in D-loop region which was suggested to be involved in AML pathogenesis in previous report. Among the 7 samples harbouring the variant one corresponding patient and mother was identified which could be linked to predisposition.

Sample	Condition
GDN4255*	ALL
GDN4259	ALL
GDN4261	ALL
GDN4262	AML – M1
GDN4263	Mother
GDN4264*	Mother
GDN4266	Mother

Table 14: List of samples harbouring T152C variant in D-loop region.

* indicates patient and corresponding mother sample.

Comparison of non-synonymous variants present in patients and mother samples favours a change from A>G (70.2%) with less C>T (20.8%), T>A (6.25%) and G>A (4.1%) while in the amino acids, a T>A (70.2%) change is highest followed by T>I (16.6%), A>T (4.1%), H>Y (4.1%), S>P (4.1%) and M>T (2.08%) (Tables 15 – 18).

Samples	Gene	Pos	cDNA pos	AA change
GDN4254 (Patient)	ND2	4883	c.A364G	p.T122A
	ATP6	8701	c.A175G	p.T59A
	ATP6	8860	c.A334G	p.T112A
	ND3	10398	c.A340G	p.T114A
	CYTB	14766	c.C20T	p.T7I
	CYTB	15326	c.A580G	p.T194A
GDN4263 (Mother)	ATP6	8860	c.A334G	p.T112A
	ND3	10398	c.A340G	p.T114A
	ND4L	10609	c.T140C	p.M47T
	CYTB	14766	c.C20T	p.T7I
	CYTB	15119	c.G373A	p.A125T
	CYTB	15326	c.A580G	p.T194A

Table 15: List of non-synonymous variants present in corresponding patient (GDN4254) and mother (GDN4263) samples.

Samples	Gene	Pos	cDNA pos	AA change
GDN4255 (Patient)	ND2	4824	c.A355G	p.T119A
	ATP6	8794	c.C268T	p.H90Y
	ATP6	8860	c.A334G	p.T112A
	ND3	14766	c.C20T	p.T7I
	CTYB	15326	c.A580G	p.T194A
GDN4264 (Mother)	ND2	4824	c.A355G	p.T119A
	ATP6	8794	c.C268T	p.H90Y
	ATP6	8860	c.A334G	p.T112A
	CTYB	14766	c.C20T	p.T7I
	CTYB	15326	c.A580G	p.T194A

Table 16: List of non-synonymous variants present in corresponding patient (GDN4255) and mother (GDN4264) samples.

Samples	Gene	Pos	cDNA pos	AA change
GDN4257 (Patient)	ND2	4833	c.A364G	p.T122A
	ATP6	8701	c.A175G	p.T59A
	ATP6	8860	c.A334G	p.T112A
	ND3	10398	c.A340G	p.T114A
	CTYB	14766	c.C20T	p.T7I
	CTYB	15326	c.A580G	p.T194A
GDN4265 (Mother)	ND2	4833	c.A364G	p.T122A
	ATP6	8686	c.T160C	p.S54P
	ATP6	8701	c.A175G	p.T59A
	ATP6	8860	c.A334G	p.T112A
	ND3	10398	c.A340G	p.T114A
	CTYB	14766	c.C20T	p.T7I
	CTYB	15326	c.A580G	p.T194A

Table 17 : List of non-synonymous variants present in corresponding patient (GDN4257) and mother (GDN4265) samples.

Samples	Gene	Pos	cDNA pos	AA change
GDN4258 (Patient)	ND2	4833	c.A364G	p.T122A
	ATP6	8701	c.A175G	p.T59A
	ATP6	8860	c.A334G	p.T112A
	ATP6	8686	c.T160C	p.S54P
	ND3	10398	c.A340G	p.T114A
	CTYB	14766	c.C20T	p.T7I
	CTYB	15326	c.A580G	p.T194A
GDN4266 (Mother)	ATP6	8548	c.G58A	p. A20T
	ATP6	8701	c.A175G	p.T59A
	ATP6	8860	c.A334G	p.T112A
	ND3	10398	c.A340G	p.T114A
	CTYB	14766	c.C20T	p.T7I
	CTYB	15326	c.A580G	p.T194A

Table 18: List of non – synonymous variants present in corresponding patient (GDN4258) and mother (GDN4266).

MSI was performed using various markers relating to different chromosomes (Table 19).

Microsatellite Marker Name	Sequence	Dye	Marker Size	Repeat	Gene name and Chromosome number
BAT25	F: 5′ -TCG CCT CCA AGA ATG TAA GT - 3′ R: 5′ - TCT GCA TTT TAA CTA TGG CTC - 3′	PET	110 - 125	(T)25	KIT proto-oncogene receptor tyrosine kinase (KIT) , chromosome 4
BAT26	F: 5′ - TGA CTA CTT TTG ACT TCA GCC -3′ R: 5′ - AAC CAT TCA ACA TTT TTA ACC C -3′	NED	100 - 120	(A)26	mutS homolog 2 (MSH2), chromosome 2
D2S123	F: 5′- AAA CAG GAT GCC TGC CTT TA - 3′ R: 5′ - GGA CTT TCC ACC TAT GGG AC -3′	NED	200 - 230	(CA)29	DNA segment containing (CA) repeat, chromosome 2
D17S250	F: 5′ - GGA AGA ATC AAA TAG ACA AT – 3′ R: 5′ - GCT GGC CAT ATA TAT ATT TAA ACC – 3′	VIC	140 - 170	(CA)19	DNA segment containing (CA) repeat, chromosome 17
D16S398	F: 5'-CTTGCTCTTTCTAAACTCCA-3' R: 5'-GAAACCAAGTGGGT TAGGTC-3'	PET	175 - 195	(CA)23	CDH1, chromosome 16
D16S496	F: 5′- GAAAGGCTACTTCATAGATGGCAAT-3′ R: 5′- ATAAGCCACTGCGCCCAT-3′	VIC	200 - 230	(T)13 and (CA)21	CDH1, chromosome 16
D18S58	F: 5′-GCTCCCGGCTGGTTTT-3′ R: 5′- GCAGGAAATCGCAGGAACTT -3′	6-FAM	140 - 155	(CA)18	DNA segment containing (CA) repeat, chromosome 18
D16S3057	F: 5′-CCTGTGTGTATAACTATGTCAAAAT-3′ R: 5′-GCCCTTGAAACTAGGCAATA-3′	6-FAM	190 - 205	(CG)19	DNA segment containing (CG) repeat, chromosome 17

Table 19: List of markers for MSI with their description.

Name	RFU	Tm (˚C)
BAT25	25	56
BAT26	40	56
2S123	40	59
D17S250	100	52
D16S398	25	55.5
D16S496	100	61
D18S58	100	60
D16S3057	100	57

Table 20: List of MSI marker used and their relative fluorescence units (RFU) and their Primer annealing temperature (Tm).

PCR was performed for the MSI markers in the corresponding 4 patient and mother samples using all the eight markers (Fig 5).

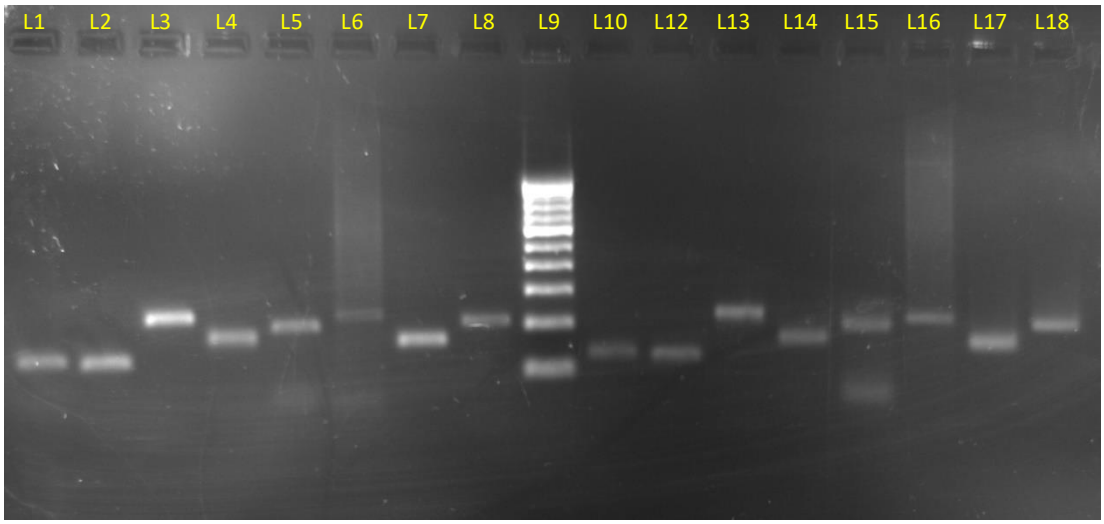

Fig 4: PCR product of 8 MSI markers for one patient and its corresponding mother. L1: BAT25, L2: BAT26, L3: D2S123, L4: D17S250, L5: D16S398, L6: 16S496, L7: D18S58, L8: D16S3057, L9: 100bp Ladder, L10: BAT25, L11: BAT26, L12: D2S123, L13: D17S250, L14: D16S398, L15: 16S496, L16: D18S58, L17: D16S3057.

From the total 8 MSI marker used, only one marker D16S398 showed a significant change in three of the tested corresponding patient and mother samples. The 23 CA repeats are absent in all the three mothers but is present in all the 3 patients.

Microsatellite instability was tested for 8 markers and results are listed in table 20 for each corresponding patients and mother samples separately (Tables 21 – 24).

GDN4254 (Patient) vs GDN4263 (mother)		
Markers	**Patient**	**Mother**
D18S58	a1 :140bp, a4:140bp	a5:140bp, a4:140bp
D16S3057	a4:200bp	a3:200bp, a5:200bp
D17S250	a3:150bp, a5:160bp, a5:220bp	a2:150bp, a4:160bp, a4:220bp
D16S496	a5:225bp	a4:225bp
BAT26	a1,2,5:120bp	a1,2:120bp
BAT25	a?,29,30,31,?:130bp	a?,29,30,31,?:130bp
D2S123	a1,4:210bp	a2:210bp
D16S398	a3:190bp	a1,6:180bp

Table 21 : Allele and basepair for each markers for patient (GDN4254) and corresponding mother (GDN4263) samples. a = allele, bp = base pairs.

GDN4255 vs GDN4264		
Markers	**Patients**	**Mothers**
D18S58	a3,11 :140bp	a1,3:140bp
D16S3057	a6:200bp	a4,6:200bp
D17S250	a2,3:150bp, a4:160bp, a5,14:165bp	a10:150bp, a5:160bp
D16S496	a4:225bp,	a4,5:225bp
BAT26	a2,5:120bp	a1,2:120bp
BAT25	a?,29,30,31,?:130bp	a?,29,30,31,?:130bp
D2S123	a1:210bp, a5:230bp	a4,11:220bp
D16S398	a4,5:190bp	

Table 22: Allele and basepair for each markers for patient (GDN4255) and corresponding mother (GDN4264) samples. a = allele, bp = base pairs.

GDN4257 vs GDN4265		
Markers	**Patients**	**Mothers**
D18S58	a1,4 :150bp	a5,3:150bp
D16S3057	a6:200bp	a2,6:200bp
D17S250	a3,4:150bp, a5:160bp	a2,3:150bp
D16S496	a5:225bp	a4,7:225bp
BAT26	a1,2,5:120bp	a1,2,5:120bp
BAT25	a?,29,30,31,?:130bp	a?,29,30,31,?:130bp
D2S123	a1,4:210bp	a1,2:210bp
D16S398	a7,?:180bp	

Table 23: Allele and basepair for each markers for patient (GDN4257) and corresponding mother (GDN4265) samples. a = allele, bp = base pairs.

GDN4258 vs GDN4266		
Markers	**Patients**	**Mothers**
D18S58	a2,3 :150bp	a1,3:150bp
D16S3057	a5,6:200bp	a2,6:200bp
D17S250	a2,3:150bp	a2,3:150bp
D16S496	a4,7:225bp	a4,7:225bp
BAT26	a2,5:120bp	a1,2,5:120bp
BAT25	a?,29,30,31,?:130bp	a?,29,30,31,?:130bp
D2S123	a1:210bp	a1,2:210bp
D16S398	A4,5:190bp	

Table 24 : Allele and basepair for each markers for patient (GDN4258) and corresponding mother (GDN4266) samples. a = allele, bp = base pairs.

Discussion

The present study involves 11 pediatric leukemia patients, along with 4 corresponding mother samples. Whole exome analysis performed in the germline genomic mutational screening shows indecisive results where the important heterozygous variants observed are found only in the patients and not in the corresponding mother samples, suggesting that it could be a *de novo* germline mutation or is inherited from the father. The exception was for two homozygous variants, BCL10: A5S and ASXL: G652 which were reported as benign in ClinVar for immunodeficiency syndrome and C-like syndrome, respectively. Unreported variants were observed in this study which could be population specific variants, and a possible biomarker. However, more number of samples will be required to confirm their role and functions. Mizo population is a homogenized population due to the practise of endogamy which is reflected in the observed mitochondrial genomic screening (Tables 15 -18) as well as could be true for the nuclear genome, if both the parents of patients were studied. Nonetheless, the observed variants from the exome data could also be linked to predisposition as a result of endogamy since the genes observed are frequently found in different cancer types (*https://cancer.sanger.ac.uk/cosmic*).

MUTYH encodes an enzyme DNA glycosylase that functions in base excision repair when there is DNA damage from oxidation. The enzyme repairs a point mismatch introduced at a nucleotide complementary to adenine. Normally adenine pairs with thymine, but sometimes this Chargaff's rule fails and adenine pairs with guanine or

cytosine. Similarly when there is oxidation in the nucleotide guanine it leads to the formation of its base analogues called oxoguanine and pairs with adenine leading to a change in A-T pair to G-C pair in the subsequent cell replication. From a series of mismatch repair genes, for this type is incorrect or inappropriate base pairing, MUTYH gene product repairs the damage. Mutations in MUTYH gene is known for an autosomal recessive phenotype MUTYH-associated polyposis (MAP) which causes adenomatous polyposis of the colon and rectum also is associated with high risk of colorectal cancer with a 28 fold as DNA repair genes plays a huge role in cancer (Sieber *et al*., 2003; Sampson *et al*., 2003; Lubbe *et al* ., 2009).

MUYTH gene variants are also found in different type of cancers. MUYTH gene mutations in two pediatric high grade midline gliomas patients both harbouring mutation leading to inactivation of the gene was also reported (Kline *et al*., 2016) as well as in gastric cancer (Kim *et al*., 2004). It is also reported in pediatric leukemia (Stanczyk *et al*., 2011; Akyerli *et al*., 2003). However, an unreported variant G143E was found on a two year old female with AML-M1 subtype with a history of gastric cancer, but the mother sample did not carry the same mutation. Nonetheless, as the variant was predicted as pathogenic by three predicting softwares, as well as categorized as MAP and Hereditary Cancer Predisposing Syndrome in ClinVar, the variant might confers loss of the protein function.

NOTCH1 encodes a transmembrane receptor protein that is required in differentiation and maturation process and is activated during early embryo or in haematopoiesis (Kojika and Griffin, 2001; Schroeder *et al*., 2013). Mutations in the PEST and heterodimer domains within NOTCH1 are found in 50% of T-cell-ALL patients (Weng *et*

al., 2004). Mutations in the gene are subjected likely in ALL patients where its role is poorly understood in myeloid malignancies. This maybe because activation of Notch pathway varies between cell types (Baldi *et al*., 2004). Lu Fu *et al*. (2006) reported the first NOTCH1 mutation in AML patients and even suggest that NOTCH1 mutations are rare events in AML patients. The expression of NOTCH1 is significantly high in AML patients, however the activation is low indicating that it is not constitutively activated. Study reported that in vivo activation of NOTCH by its ligands arrest AML growth while inhibition confers proliferation (Kannan *et al*., 2013). This suggested that NOTCH plays a role as tumour suppressor in AML.

A novel pathway that activates NOTCH for inhibiting cell growth was identified (Lobry *et al*., 2013). The mutation observed in this study as predicted by the prediction softwares (SIFT, PolyPhen2 and Provean) was deleterious, suggesting that NOTCH1 V1699E mutation might confer loss of function and its ability to suppress tumour might be lost. From the aforementioned studies, inactivation or loss of function aids in cell proliferation, suggesting that the patient in this study with AML-M1 subtype might have a proliferative advantage as extensive expression of NOTCH1 especially in M1 and M0 – AML patients with simultaneous expression of CD7 which is a marker for immaturity was observed that reflects in a poor overall survival rate (Sliwa *et al*., 2014).

SHP-2 is a non-receptor tyrosine phosphatase protein which is encoded by PTPN11 gene that functions in transducing signals to RAS after receptors confined to growth factors are activated (Hof *et al*., 1998) like c-KIT, erythropoietin receptors, IL-3 and granulocyte-macrophage colony-stimulating factor receptor (Vactor *et al*., 1998; Feng, 1999; Itoh *et al*., 1998). It functions majorly in relaying signals in RTK/RAS/RAF/ERK pathways. Mutations in

PTPN11 are found commonly in JMML patients without RAS and NF1 mutation and are involved in leukemiogenesis by negative regulation of the RAS pathway by conferring growth advantage (Tartaglia *et al*., 2003; Loh *et al*., 2004). Loss or inactivation of SHP–2 function has tremendous effects in haematopoiesis (Qu *et al*., 1997; 2001). Most of the mutations reported in PTPN11 are within the domain N-terminal src-homology-2 (N-SH2) and protein tyrosine phosphatase (PTP) domain.

In consistent with other findings, GND4261 has a mutation in PTP domain (S502P) with no RAS mutation but positive for FLT3-ITD as well as PTPN11 mutation was found to be seen more among boys (Loh *et al*., 2004), but in the present study, the sample in which this mutation was found is a female. In contrast to adult AML patients, where there is no association observed between the two gene mutations, PTPN11 and FLT3-ITD (Hou *et al*., 2008), it was observed in the patient involved in this study. Amino acid change from serine position 502 to alanine, threonine and leucine within the PTP domain has been reported for Noonan-syndrome as well as JMML (Kratz *et al*., 2005). This study observed serine to proline substitution in the same position at 502 where pathogenic prediction of serine to threonine change was considered as likely oncogenic, the change from serine to proline is also predicted as likely oncogenic (OncoKB).

FLT3 is a proto-oncogene involved in haematopoiesis to exert its functions in signalling a cell to proliferate, differentiate and survive. The expression of FLT3 in acute leukemia was studied using RT-PCR where they found a surprising larger amplicon than the expected size. This was the first report of FLT3-ITD in hematological malignancies (Nakao *et al*., 1996) and from then onwards, FLT3 has been a crucial biomarker in several leukemia

especially in AML, where mutations in FLT3-ITD have been correlated with the clinical diagnosis and prognosis in leukemic patients (Swerdlow *et al*., 2008).

Different type of FLT3-ITD has been reported (Blau *et al*., 2013) and also by many other groups but the ITD found in this study has not been reported. However, the site of duplication observed in this study is consistent with other duplication site which is in the juxtamembrane domain, amino acid 591 -599 (Meshinchi *et al*., 2008). Family of type III receptor tyrosine kinases involves receptors like FLT3, c-KIT and platelet-derived growth factor receptors α and β (PDGFRα/β) where amino acid sequences of them shows a high homology and conserved regions. Mutations in amino acid positions 572, 576, 578, 579, 589, 591, 597, 599 in binding, switch and zipper motif within the juxtamembrane domain in PDGFRβ confers constitutive activation of the gene (Irusta *et al*., 2002). In line with this findings, our study observed two tyrosine amino acid positioned in 589 and 591 to be deleted and in its place an in-frame insertion consistent with ITD (Levis and Small 2003) tryptophan (W), alanine (A), glycine (G), aspartic acid (D) are inserted which also suggest that this indel might lead to a constitutive activation of FLT3 as it was observed with its homolog PDGFRβ. A study has indicated that FLT3-ITD is a driver mutation in AML patients (Smith *et al*., 2013) and also the size of insertion, region of insertion as well as the allelic ratio of mutant and wildtype has a clinical significance (Thiede *et al*., 2002; Kayser *et al*., 2008). The patient in this study was found to be harbouring this insertion that confers a worst prognosis and low overall survival rate.

From the list of variants from all patients after filter 2 and matched with 55 Leukemia genes and ClinVar, only two from 46 variants namely BCR gene (D1106N) and

TET2 gene (E1151Ter) shows an interesting result. The BCR variant D1106N was found to be associated with ALL and AML as per phenotypic condition in ClinVar but the significance was shown to be uncertained which was reported by Division of Genomic Diagnostics (The Children's Hospital of Philadelphia) but no supporting evidence was given. The TET2 variant E1151Ter indicates that a mutation confers a stop gain. Similarly as in the case of BCR, TET2 variant also has no other information regarding the phenotypic condition and its significance. Nonetheless, these two variants observed in this study are in heterozygous condition.

All the variants after filter2 and matched with 55 leukemia associated genes were interpreted using OMIM database for their mode of inheritance and their phenotypes (Table 8). In this study, two variants ASXL1 (G652S) and BCL10 (A5S) were found in homozygous condition in two patients, respectively. The G652S variant was reported earlier as not a true mutation but as a rare germline polymorphism (Schnittger *et al*., 2011; Wen-Chien *et al*., 2010). However, a study from Turkey found that the variant as a second most common variant in case of Philadelphia chromosome – negative myeloproliferative neoplasms (Yonal-Hindilerdene *et al*., 2015). The A5S variant was reported as benign in ClinVar.

The first report of mutations in mitochondria in association with a disease was observed in mitochondrial myopathy patients (Holt *et al*., 1988). Ever since, the search for mitochondrial variants that have implications in disease manifestation has been studied extensively and mutations in mtDNA have been identified in various diseases (Wallace, 1992; Brown and Wallace, 1994). Mutations in the genes coding for enzyme complex

subunits that functions in oxidative phosphorylation are one of the causal factor for mitochondria related disease (Taylor and Turnbull, 2005).

mtDNA mutation was observed in several types of cancers like colorectal, ovarian, pancreatic, prostate, lung, breast, hepatocellular and gastric cancer (Polyak *et al*., 1998; Liu *et al*., 2001; Jones *et al*., 2001; Jessie *et al*., 2001; Sanchez – Cespedes *et al*., 2001; Tan *et al*., 2002; Nomoto *et al*., 2002; Zhao *et al*., 2005). Mutations in mtDNA in leukemic patients were also reported (He *et al*., 2003; Carew *et al*., 2003; Yao *et al*.,2007).

mtDNA variant T152C in the control region of D-loop was identified to be involved mostly in AML – M3 pathogenesis and that it could be a diagnostic biomarker but was not found to be associated with clinical prognosis (Zhou *et al*., 2005). The present study identified T152C variant in 4 patients (2 AML – M1, 1 JCML and 1 CML) and in 3 mother's sample. Contrasting to the previous study, this study observed the T152C variant in AML – M1 patients. Amongst the 3 mothers who are positive for this variant, only one corresponding mother and patient harbor's this variant. This indicates that the variant alone could not initiate leukemiogenesis. However, the variant lies in the transcriptional control region-OH which might confer a proliferative advantage as the same variant was also reported in ALL patients (Kwok *et al*., 2011). Contrasting to Zhou *et al.* (2005), Kwok reported that the variant was identified in a better prognostic group among ALL patients, where he suggested that due to rapid replication (proliferative advantage) the cells are subjected to oxidative environment with high levels of ROS which facilitates apoptosis after chemotherapy treatment.

The same corresponding mother and patient (with T152C variant) and 1 patient also harbored an unreported variant (T119A) in the MT-ND2 gene. The gene is responsible for encoding one of the subunits of complex I that functions in the respiratory mechanism. The variant was predicted as pathogenic/deleterious by three functional predicting softwares (SIFT, PolyPhen2 and Provean). It was suggested that in carcinogenesis, a crucial lesion in the respiratory complexes results in high glycolytic ATP production, which change the energy capacities in cancerous cells with increase ROS generation and subsequent DNA damage/instability (Warburg, 1956). The increased glycolytic production could be used by the rapidly growing malignant cells as a source of food to exert its advantage in proliferation and resist to apoptosis. Kwok *et al*. (2011) suggested that malignant cells that resist apoptosis may lead to constant proliferation and oncogenesis.

Similar to the T119A, an unreported variant (Y165C) in MT-ND6 gene was identified in one patient which was also predicted to be pathogenic/ deleterious by the same three functional predicting softwares. Since the encoded protein of this gene also forms a subunit of complex I, variation in this gene is likely to affect the respiratory machinery resulting in increased production of glycolytic ATP.

For detecting the presence of MSI across mother and corresponding patients, 8 markers which are frequent to show instability in cancers were used (Table 19). This approach was performed in this population for gastric cancer and MSI was observed (unpublished data). In line with previous studies (Faulkner *et al*., 2003), BAT26 and BAT 26 was found to be not good as MSI marker for leukemia as all eight samples tested did not show any changes in the present study as well. Of the tested 8 markers, only one marker

D16S398 showed a change for the three corresponding mother and patient samples. This was also in consistent with earlier report where the marker was suggested to be a crucial biomarker; however it was reported for adult T-cell leukaemia patients (Miyashita *et al*., 2017). The MSI result shows there is an introduction either by insertion or duplication of CA repeats in all the three patients where no such allele was present in their corresponding mothers.

Earlier report suggest that MSI might not be a frequent event or hardly plays a crucial role in myeloid malignancies as no evidence of MSI was reported from the screened 1371 AML patients (Walker *et al*., 2017). Also, another study compares the MSI and defect in Mismatched Repair genes on 50 ALL patients with 40 healthy controls revealing High-frequency MSI (MSI-H) in 2 cases and healthy control, suggesting a generally low MSI in ALL cases (Kendell *et al*., 2004). However, a pilot study in India concluded that for CML patient with two markers- D17S261 and D3S643 shows instability and that they could be involved in the pathogenicity and evolution of the disease (Patel *et al*., 2017). It is difficult to draw conclusion based on the present study as: the test was done not from the patient's tumour and matched normal sample, but was compared with their mothers. Also only four matched patient-mother were tested for their microsatellite instability indicating less sample size.

Summary

- Whole exome analysis of Pediatric leukemia samples identified an unreported variant PTPN11 (S502P) in 1 patient which was found to be predicted as oncogenic by OncoKB database.
- From all Leukemic patients after filters 1 and 2, 16 reported and 27 unreported variants were found in ClinVar, suggesting few novel variants which might either have a diagnostic or prognostic value.
- Further, prioritizing the variants using all the three filters, only 2 variants were found: NOTCH1: V1699E (Not reported) and MUTYH: G143E (reported- rs730881833, Polyposis, Hereditary Cancer Predisposing Syndrome).
- An unreported 9ntd's insertion (Y589delinsWAGD) in the FLT3-ITD gene was found in one patient.
- A prognostic marker for responsiveness of chemotherapy drug 5-FU through DPYD genotyping for Clinically significant variant were found to confer a good prognosis by minimal toxicity.
- Whole mitochondrial DNA (mtDNA) sequencing revealed the mean variants in all samples to be 39.6, in Exonic regions to be 21.6. The average number of Synonymous variants within the exonic region was 14.9 and Non-Synonymous was 6.7.
- Variation signature A>G (70.2%) in mtDNA was observed to be associated to T>A (70.2%) amino acid change.
- In the D-loop region of the mtDNA, a T152C change was found in 3 AML patients and 1 JCML patient which has been reported to be possibly involved in the pathogenesis of AML, especially M3 subtype.

- Variants T119A in MT-ND2 and Y165C in MT-ND6 were identified in 3 and 1 patient respectively, and are predicted as pathogenic/deleterious by functional predicting tools which might confers proliferative advantage.
- From the eight MSI markers tested, BAT25 and BAT26 markers were not found to be suitable for leukemia as previously reported. D16S398 was found to be present in both mother and son, but further studies are required to validate the usefulness of this MSI marker in Mizo population.

Akyerli CB, Özbek U, Aydın-Sayitoğlu M, *et al*. (2003) Analysis of MYH Tyr165Cys and Gly382Asp variants in childhood leukemias. J Cancer Res Clin Oncol. 129:604.

Andrews S. (2010). FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Arora B and Kanwar V. (2009) Childhood cancers in India: Burden, barriers, and breakthroughs. Indian Journal of Cancer. 46(4):257–9.

Asthana S, Labani S, Mehrana S and Bakhshi S. (2018) Incidence of childhood leukemia and lymphoma in India. Pediatric Hematology Oncology Journal. 3(4):115-20.

Baccichet A, Benachenhou N, Couture F, Leclerc JM and Sinnett D. (1997). Microsatellite instability in childhood T cell acute lymphoblastic leukemia. Leukemia: official journal of the Leukemia Society of America. Leukemia Research Fund. 11:797-802.

Baldi A, De Falco M, De Luca L, *et al*. (2004). Characterization of tissue specific expression of Notch-1 in human tissues. Biol. Cell. 96:303–11.

Basu TN, Gutmann DH, Fletcher JA, *et al*. (1992). Aberrant regulation of ras proteins in malignant tumour cells from type 1 neurofibromatosis patients. Nature. 356:713-5.

Blau O, Berenstein R, Sindram A & Blau IW. (2013). Molecular analysis of different *FLT3*-ITD mutations in acute myeloid leukemia. Leuk. Lymphoma. 54(1):145-52.

Boguski M and McCormick F. (1993). Proteins regulating Ras and its relatives. Nature. 366:643-53.

Boland CR and Goel A. (2010) Microsatellite instability in colorectal cancer. Gastroenterology. 138:2073–87.

Bonaccorso P, Nellina A, Valeria I, *et al*. (2016). Molecular Pathways in Childhood Acute Lymphoblastic Leukemia: From the Bench to the Bedside. Journal of Pediatric Biochemistry. 5(4):146-56.

Bonneville R, Krook MA, Kautto EA, *et al*. (2017). Landscape of Microsatellite Instability Across 39 Cancer Types. JCO Precision Oncology. 1:1-15.

Cancer India: http://cancerindia.org.in/

Castilla LH, Garrett L, Adya N, *et al*. (1999). The fusion gene Cbfβ blocks myeloid differentiation and predisposes mice to acute myelomonocytic leukemia. Nat Genet. 23:144-6.

Daley GQ, Van Etten RA and Baltimore D. (1990). Induction of chronic myelogenous leukemia in mice by the P210bcr/abl gene of the Philadel-phia chromosome. Science. 247:824–30.

Ding LW, Sun QY, Tan KT, *et al*. (2017). Mutational landscape of pediatric acute lymphoblastic leukemia. Cancer Research. 77(2):390–400.

Eguchi-Ishimae M, Eguchi M, Ishii E, *et al*. (2001). Breakage and fusion of the TEL (ETV6) gene in immature B lymphocytes induced by apoptogenic signals. Blood. 97(3):737–43.

Elefanty AG, Hariharan IK and Cory S. (1990). bcr-abl, the hallmark of chronic myeloid leukaemia in man, induces multiple haemopoietic neoplasms in mice. EMBO J. 9(4):1069–78.

ENHIS – WHO, (2009) http://www.euro.who.int/__data/assets/pdf_file/0005/97016/4.1.-Incidence-of-childhood-leukaemia-EDITED_layouted.pdf?ua=1

Farrar JE, Schuback HL, Ries RE, *et al*. (2016). Genomic profiling of pediatric acute myeloid leukemia reveals a changing mutational landscape from disease diagnosis to relapse. Cancer Res. 76:2197–205.

Faulkner RD, Seedhouse CH, Das-Gupta EP and Russell NH. (2004). BAT-25 and BAT-26, two mononucleotide microsatellites, are not sensitive markers of microsatellite instability in acute myeloid leukaemia. Br J Haematol. 124(2):160-5.

Feng G S. (1999). Shp-2 tyrosine phosphatase: signalling one cell or many. Exp Cell Res. 253:47-54.

Fu L, Kogoshi H, Nara N and Tohda S. (2006) *NOTCH1* mutations are rare in acute myeloid leukemia. Leuk. Lymphoma. 47(11):2400-3.

García-Alcalde F, Okonechnikov K, Carbonell J, Cruz LM, Götz S, Tarazona S, Dopazo J, Meyer TF, Conesa A. Qualimap: evaluating next-generation sequencing alignment data. Bioinformatics. 2012 Oct 15;28(20):2678-9.

Gilliland DG and Griffin JD. (2002). The roles of FLT3 in hematopoiesis and leukemia. Blood. 100(5):1532-42.

Goemans BF, Zwaan CM, Miller M, *et al*. (2005). Mutations in KIT and RAS are frequent events in pediatric core-binding factor acute myeloid leukemia. Leukemia. 19:1536–42.

Grimwade D, Ivey A and Huntly BJ. (2016) Molecular landscape of acute myeloid leukemia in younger adults and its clinical relevance. Blood. 127(1):29-41.

Grist SA, Lu XJ and Morley AA. (2004). Mitochondrial mutations in acute leukaemia. Leukemia. 18:1313 6.

Hall A. (1992). Signal transduction through a small GTPases – A tale of two GAP's. Cell. 69:389-91.

Hariprakash JM, Vellarikkal SK, Keechilat P, *et al*. (2018). Pharmacogenetic landscape of DPYD variants in south Asian populations by integration of genome-scale data. Pharmacogenomics. 19(3):227-241.

Heisterkamp N, Jenster G, ten Hoeveet J, *et al*. (1990). Acute leukaemia in bcr/abl transgenic mice. Nature. 344:251–3.

Hjalgrim LL, Madsen HO, Melbye M, *et al.* (2002). Presence of clone-specifi c markers at birth in children with acute lymphoblastic leukaemia. Br J Cancer. 87(9):994–9 .

Hof P, Pluskey S, Dhe-Paganon S, Eck MJ and Shoelson SE. (1998). Crystal structure of the tyrosine phosphatase SHP-2. Cell. 92:441–50.

Hou HA, Chou WC, Lin LI, *et al*. (2008). Characterization of acute myeloid leukemia with PTPN11 mutation: the mutation is closely associated with NPM1 mutation but inversely related to FLT3/ITD. Leukemia. 22:1075–8.

Irusta PM, Luo Y, Bakht O, *et al*. (2002). Definition of an inhibitory juxtamembrane WW-like domain in the platelet-derived growth factor β receptor. J Biol Chem. 277:38627-34.

Itoh T, Liu R, Yokota T, Arai KI and Watanabe S. (1998). Definition of the role of tyrosine residues of the com-mon beta subunit regulating multiple signalling pathways of granulocyte-macrophage colony-stimulating factor receptor. Mol Cell Biol. 18:742-52.

Iwai T, Yokota S, Nakao M, *et al*. (1999). Internal tandem duplication of the FLT3 gene and clinical evaluation in childhood acute myeloid leukemia. The Children's Cancer and Leukemia Study Group, Japan. Leukemia. 13:38-43.

Kannan S, Sutphin RM, Hall MG, *et al*. (2013). Notch activation inhibits AML growth and survival: a potential therapeutic approach. J. Exp. Med. 210(2):321-37.

Kayser S, Schlenk R, Breitenbucher F, *et al*. (2008). Prognostic implication of insertion of FLT3 internal tandem duplication in the BETA-1 sheet of the Tyrosine Kinase Domain-1. *Blood.* 112:2514.

Kelliher, M.A., McLaughlin, J., Witte, O.N., and Rosenberg, N. 1990. Induction of a chronic myelogenous leukemia-like syndrome in mice with v-abl and BCR/ABL.Proc. Natl. Acad. Sci. U. S. A.87:6649–6653.

Kendall HE, Vacek PM and Finette BA. (2004). Analysis of microsatellite instability in children treated for acute lymphocytic leukemia with elevated HPRT mutant frequencies. Mutagenesis. 19(5):409–12.

Kim CJ, Cho YG, Park CH, *et al*. (2006) Genetic alterations of the MYH gene in gastric cancer. Oncogene. 23(40):6820-2.

Kline CN, Joseph NM, Grenert JP, *et al*. (2016). Inactivating MUTYH germline mutations in pediatric patients with high-grade midline gliomas. Neuro Oncol. 18(5):752–3.

Klose A, Nurnberg P, Tinschert S and Peters H. (1998b). Recurrence of the neurofibromatosis type 1 (NF1) mutation R1276X probably due to interchromosomal gene conversion. Hum Mutat. 12:291.

Knudson A. (1971). Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci USA. 68(4):820–3.

Koboldt, D., Zhang, Q., Larson, D., Shen, D., McLellan, M., Lin, L., Miller, C., Mardis, E., Ding, L., & Wilson, R. (2012). VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing Genome Research DOI: 10.1101/gr.129684.111 URL: http://varscan.sourceforge.net

Kojika S and Griffin JD. (2001). Notch receptors in hematopoiesis. Exp. Hematol. 29:1041-52.

Kondo M, Horibe K, Takahashi Y, *et al*. (1999). Prognostic value of internal tandem duplication of the FLT3 gene in childhood acute myelogenous leukemia. Med Pediatr Oncol. 33:525-9.

Kratz CP, Niemeyer CM, Castleberry RP, *et al.* (2005). The mutational spectrum of PTPN11 in juvenile myelomonocytic leukemia and Noonan syndrome/myeloproliferative disease. Blood. 106(6):2183-5.

Kwok CS, Quah TC, Ariffin H, Tay SK and Yeoh AE. (2011). Mitochondrial D-loop polymorphisms and mitochondrial DNA content in childhood acute lymphoblastic leukemia. J Pediatr Hematol Oncol. 33:e239 4.

Levis M and Small D. (2003). FLT3: ITDoes matter in leukemia. Leukemia. 17(9):1738–52.

Li H, Handsaker B, Wysoker A, Fennell T, *et al.* (2009). 1000 Genome Project Data Processing Subgroup..The Sequence Alignment/Map format and SAMtools. Bioinformatics. 25(16):2078-9.

Li H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v1 [q-bio.GN]

Li H. and Durbin R. (2009.) Fast and accurate short read alignment with Burrows-Wheeler Transform. Bioinformatics. 25:1754-60.

Lobry C, Ntziachristos P, Ndiaye-Lobry D, *et al*. (2013). Notch pathway activation targets AML-initiating cell homeostasis and differentiation. J. Exp. Med. 210(2):301-19.

Loh ML, Reynolds MG, Vattikuti S, *et al*. (2004). PTPN11 mutations in pediatric patients with acute myeloid leukemia: results from the Children's Cancer Group. Leukemia. 18:1831–4.

Loh ML, Vattikuti S, Schubbert S, *et al*. (2004) Mutations in PTPN11 implicate the SHP-2 phosphatase in leukemogenesis. Blood. 103:2325–31.

Lubbe SJ, Di Bernardo MC, Chandler IP and Houlston RS. (2009). Clinical implications of the colorectal cancer risk associated with MUTYH mutation. J. Clin. Oncol. 24:3975-80.

Malumbres M and Barbacid M. (2003). RAS oncogenes: the first 30 years. Nature Reviews Cancer. 3:459-65.

Mao G, Yuan F, Absher K, *et al*. (2008). Preferential loss of mismatch repair function in refractory and relapsed acute myeloid leukemia: Potential contribution to AML progression. Cell research. 18:281-9.

Mardis ER. (2011). A decade's perspective on DNA sequencing technology. Nature. 470:198-203.

McHale CM, Wiemels JL, Zhang L, *et al*. (2003). Prenatal origin of TEL-AML1-positive acute lymphoblastic leukemia in children born in California. Genes Chromosomes Cancer. 37(1):36–43.

McHale CM, Wiemels JL, Zhang L, *et al*. (2003). Prenatal origin of childhood acute myeloid leukemias harboring chromosomal rearrangements t (15;17) and inv (16). Blood. 101(11):4640–1.

Medarde FA and Santos E. (2011). Ras in cancer developmental disease. Genes Cancer. 2:344-58.

Meshinchi S, Stirewalt DL, Alonzo TA, *et al*. (2008). Structural and numerical variation of FLT3/ITD in pediatric AML. Blood. 111(10):4930-3.

Mirabilii S, Ricciardi MR, Allegretti M, *et al*. (2012). Targeting Metabolic Pathways for Leukemia Treatment. Blood. 120(21):1371.

Miyashita K, Fujii K, Taguchi K, *et al*. (2017). A specific mode of microsatellite instability is a crucial biomarker in adult T-cell leukaemia/lymphoma patients. J Cancer Res Clin Oncol. 143:399.

Mori H, Colman SM, Xiao Z, *et al*. (2002). Chromosome translocations and covert leukemic clones are generated during normal fetal development. Proc Natl Acad Sci USA. 99(12):8242–7.

Nakao M, Yokota S, Iwai T, *et al*. (1996). Internal tandem duplication of the flt3 gene found in acute myeloid leukemia. Leukemia. 10(12):1911–8.

Nowell PC and Hungerford DA. (1960). A minute chromosome in chronic granulocytic leukemia. Science. 132(3438):1488–501.

Patel TN, Chakraborty M and Bhattacharya P. (2017). Microsatellite Instability in Chronic Myeloid Leukemia using D17S261 and D3S643 markers: A Pilot Study in Gujarat Population. Indian J Cancer. 54:426-9.

Paulsson K, Horvat A, Strombeck B. *et al*. (2008). Mutations of FLT3, NRAS, KRAS, and PTPN11 are frequent and possibly mutually exclusive in high hyperdiploid childhoodacute lymphoblastic leukemia. Genes Chromosomes Cancer. 47:26–33.

Perentesis JP, Bhatia S, Boyle E, *et al*. (2004). RAS oncogene mutations and outcome of therapy for childhood acute lymphoblastic leukemia. Leukemia. 18:685–92.

Peters H, Luder A, Harder A, Schuelke M and Tinschert S. (1999). Mutation screening of neurofibromatosis type 1 (NF1) exons 28 and 29 with single strand conformation polymorphism (SSCP): five novel mutations, one recurrent transition and two polymorphisms in a panel of 118 unrelated NF1 patients. Hum Mutat. 13:258

Qu CK, Nguyen S, Chen J and Feng GS. (2001). Requirement of Shp-2 tyrosine phosphatase in lymphoid and hematopoietic cell development. Blood. 97:911-91.

Qu CK, Shi ZQ, Shen R, *et al*. (1997). A deletion mutation in the SH2-N domain ofShp-2 severely suppresses hematopoietic cell development. Mol Cell Biol. 17:5499-507.

Renneville A, Roumier C, Biggio V, *et al*. (2008). Cooperating gene mutations in acute myeloid leukemia: a review of the literature. Leukemia. 22:915-31.

Robinson JT, Thorvaldsdóttir H, Winckler W, *et al*. (2011). Integrative Genomics Viewer. Nature Biotechnology. 29:24–26.

Rudolph's Pediatrics. (1998) Rudolph EAM, Hoffman JIE, and Rudolph CD. Leukemia. Appleton and Lange. Stamford, CT. 20th edition. 18:1269-75.

Sampson JR, Dolwani S, Jones S, *et al*. (2003). Autosomal recessive colorectal adenomatous polyposis due to inherited mutations of MYH. Lancet. 362:39-41.

Schnittger S, Eder C, Fasan A, *et al*. (2011). Somatic Mutations and Inborn Variants in Exon 12 of ASXL1 in Different Myeloid Neoplasms. Blood. 118(21):1394.

Schroeder T, Kohlhof H, Rieber N and Just U, (2003) Notch signalling induces multilineage myeloid differentiation and up-regulates PU.1 expression. J Immunol. 170:5538–48.

Shannon KM, Watterson J, Johnson P, *et al*. (1992). Monosomy 7 myeloproliferative disease in children with neurofibromatosis, type 1: epidemiology and molecular analysis. Blood. 79:1311-8.

Sharawat SK, Bakhshi R, Vishnubhatla S and Bakhshi S. (2010). Mitochondrial D-loop variations in paediatric acute myeloid leukaemia: A potential prognostic marker. Br J Haematol. 149:391 8.

Shu XO, Perentesis JP, Wen W. *et al*. (2004). Parental exposure to medications and hydrocarbons and ras mutations in children with acute lymphoblastic leukemia: A report from the Children's Oncology Group. Cancer Epidemiol Biomarkers Prev. 13:1230–5.

Sieber OM, Lipton L, Crabtree M, *et al*. (2003). Multiple colorectal adenomas, classic adenomatous polyposis, and germ-line mutations in MYHN. Engl J Med. 348:791-799.

Silkjaer T, Nørgaard JM, Aggerholm A, *et al*. (2013) Characterization and prognostic significance of mitochondrial DNA variations in acute myeloid leukemia. Eur J Haematol. 90:385 6.

Sliwa T, Awsa S, Vesely M, *et al*. (2014). Hyperexpression of NOTCH-1 is found in immature acute myeloid leukemia. Int J Clin Exp Patho. 7(3):882–89.

Smith CC, Wang Q, Chin CS, *et al*. (2012) Validation of ITD mutations in FLT3 as a therapeutic target in human acute myeloid leukemia. *Nature*. 485:260–3.

Stanczyk M, Sliwinski T, Cuchra M, *et al*. (2011). The association of polymorphisms in DNA base excision repair genes XRCC1, OGG1 and MUTYH with the risk of childhood acute lymphoblastic leukemia. Molecular Biology Reports. 38(1):445-51.

Steliarova-Foucher E, Colombet M, Ries LAG, *et al*. (2017) International incidence of childhood cancer, 2001-10: a population-based registry study. Lancet Oncol. 18(6):719–31.

Swerdlow SH, Campo E, Harris NL, *et al*. (2008). WHO classification of tumours of haematopoietic and lymphoid tissues. 4th Edition. Lyon: International Agency for Research on Cancer (IARC).

Tartaglia M, Martinelli S, Cazzaniga G, *et al*. (2004). Genetic evidence for lineage-related and differentiation stage-related contribution of somatic PTPN11 mutations to leukemogenesis in childhood acute leukemia. Blood. 104:307–13.

Tartaglia M, Niemeyer CM, Fragale A, *et al*. (2003). Somatic mutations in PTPN11 in juvenile myelomonocytic leukemia, myelodysplastic syndromes and acute myeloid leukemia.Nat Genet. 34:148–50.

Testa U and Pelosi E. (2013). The Impact of FLT3 Mutations on the Development of Acute Myeloid Leukemias. Leukemia research and treatment. 2013:1-14.

Thiede C, Stendel C, Mohz B, *et al*. (2002). Analysis of FLT3 activating mutations in 979 patients with acute myelogenouc leukemia: association with FAB subtypes and identification of subgroups with poor prognosis. Blood. 99(12):4326–35.

Thomas X. (2013) First contributors in the history of leukemia. World J Hematol. 2(3):62-70

Times of India: https://www.indiatimes.com/news/india/cancer-cases-on-the-rise-in-mizoram-state-being-called-cancer-capital-of-india-263190.html

Van Vactor D, O'Reilly AM, Neel BG. (1998). Genetic analysis of protein tyrosine phosphatases. Curr Opin Genet Dev. 8:112-26.

Walker CJ, Eisfeld AK, Genutis LK, *et al*. (2017). No evidence for microsatellite instability in acute myeloid leukemia. Leukemia. 31(6):1474–6.

Wang K, Li M, Hakonarson H. (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. (16):e164.

Wang Y, Wu N, Liu D and Jin Y. (2015). Recurrent Fusion Genes in Leukemia: An Attractive Target for Diagnosis and Treatment. Curr Genomics. 18(5):378–84.

Webb JC, Golovleva I, Simpkins AH, *et al*. (1999). Loss of Heterozygosity and Microsatellite Instability at the *MLL* Locus Are Common in Childhood Acute Leukemia, but not in Infant Acute Leukemia. Blood. 94(1):283-90.

Wen-Chien C, Huai-Hsuan H, Hsin-An H, *et al* (2010). Distinct clinical and biological features of de novo acute myeloid Leukemia with additional sex comb-like 1 (ASXL1) mutations. Blood. 116:4086-94.

Weng AP, Ferrando AA, Lee W, *et al*. (2004). Activating mutations of NOTCH1 inhuman T cell acute lymphoblastic leukemia. Science. 306:269–71.

Wiemels JL, Cazzaniga G, Daniotti M, *et al*. (1999) Prenatal origin of acute lymphoblastic leukaemia in children. Lancet. 354(9189):1499–503.

Wiemels JL, Xiao Z, Buffler PA, *et al*. (2002). In utero origin of t (8;21) AML1-ETO translocations in childhood acute myeloid leukemia. Blood. 99(10):3801–5.

Yamamoto Y, Kiyoi H, Nakano Y, *et al*. (2001). Activating mutation of D835 within the activation loop of FLT3 in human hematologic malignancies. Blood. 97:2434–9.

Yonal-Hindilerden I, Daglar-Aday A, Akadam-Teker B, *et al*. (2015). Prognostic significance of ASXL1, JAK2V617F mutations and JAK2V617F allele burden in Philadelphia-negative myeloproliferative neoplasms. J Blood Med. 6:157-75.

Zhang J, Mullighan CG, Harvey RC, *et al*. (2011). Key pathways are frequently mutated in high-risk childhood acute lymphoblastic leukemia: A report from the Children's Oncology Group. Blood. 118(11):3080–87.

BIODATA

ANDREW VANLALLAWMA
M. Phil Research Student
Department of Biotechnology
Mizoram University
Aizawl -796 004, Mizoram, India
E-mail: andrew13johnhelen@gmail.com
Mobile: +91 - 8794647992

Summary

I have completed my Bachelor's and Master's degree in Biotechnology. I am doing my M. Phil in Department of Biotechnology, Mizoram University, under the guidance of Prof. N. Senthil Kumar. I am currently working on mutational screening of key genes in RTK/RAS signalling pathway associated with Pediatric Leukemia with a focus on Mizo populations.

Educational Qualification

Degree	Board	Div	Month and Year of Joining	Month and Year of Passing
HSLC (Class 10)	Mizoram Board of School Education	I	March 2007	April 2008
HSSLC (Class 12)	Mizoram Board of School Education	I	March 2010	April 2011
B.Sc. Biotechnology	North Eastern Hill University	I	June 2011	March 2014
M.Sc. Biotechnology	Mizoram University	I	Aug 2014	July 2016
M.Phil. Course Work	Mizoram University	I	Aug 2017	February 2018

Awards

1. Cleared the Council of Scientific and Industrial Research, CSIR – NET, Lectureship in the year 2015.
2. Best Oral presentation in International workshop on "Introduction to basic and advanced biomedical approaches for enhancing QOL in aging societies" held at Biomedical Research Institute, AIST, Tsukuba Science City, Japan inder DIACENTER – SUNRISE PROGRAM supported by AIST & JST (Govt. of Japan) and DBT (Govt. of India) from 14 – 21 Oct 2018.

Skills developed

DNA/RNA Isolation, Gel electrophoresis, PCR, Sanger Sequencing, Bioinformatics tools and methods. Whole Exome Sequencing and Whole Mitochondria Sequencing data analysis.

Conference attended

- Poster presentation on International Conference on "Biodiversity, Environment and Human Health : innovations and Emerging trends (BEHIET 2018), organized by School of Life Sciences, Mizoram University and Association of Biotechnology and Pharmacy (ABAP), India from 12 – 14 Nov 2018.

Workshop / Trainings attended

1. Participant in workshop on "Techniques in Molecular Biology", organized by DBT – State Biotech Hub, Dept. of Biotechnology, Mizoram University, from 14 – 18 Oct 2015.
2. Participant in workshop on "Understanding basic principles in Human Molecular Genetics", organized by DBT – BIF, Dept. of Biotechnology, Mizoram University, from 07 – 11 Sept 2016.
3. Participant in International workshop on "Cancer epidemiology", organized by DBT – State Biotech Hub, Dept. of Biotechnology, Mizoram University, from 29 – 30 Oct 2016.
4. Participant in workshop on "Hands on training on DNA barcoding and Phylogenetics", organized by DBT – State Biotech Hub, Dept. of Biotechnology, Mizoram University, from 20 – 25 March 2017.
5. Participant in workshop on "Homology modelling and Molecular Docking", organized by DBT – State Biotech Hub, Dept. of Biotechnology, Mizoram University, from 18 – 22 April 2017.
6. Participant in Science Communication Workshop (*SciComm 101*) held on 6 June, 2017 at Mizoram University.
7. Participant in International Workshop on "Molecular Entomology", organized Global initiative for Academic Networks, from 19 – 28 June 2017.
8. Participate in workshop on "Antibiotic awareness and Infection Control Program" organized by DBT – State Biotech Hub, Dept. of Biotechnology, Mizoram University, from 14 Nov. 2017.
9. Participate in 3rd NER, Research training Workshop on "Understanding Human Disease and Improving Human Health Using Genomics – Driven Approaches" co-organized by National Institute of Biomedical Genomics, Kalyani and DBT – State Biotech Hub, Dept. of Biotechnology, Mizoram University, from 19 - 24 Nov. 2017.
10. Acted as a Resource Person in a workshop on "Hands –on training on Molecular research technique" organized by DBT – Institutional Biotech Hub, GZRSC, funded by Rashtriya Uchchatar Shiksha Abhiyan (RUSA), GZRSC from 31 May – 6 June, 2018.
11. Participate in 3rd NER, ADVANCED Research training Workshop on "Understanding Human Disease and Improving Human Health Using Genomics – Driven Approaches" co-organized by National Institute of Biomedical Genomics, Kalyani and DBT – State

Biotech Hub, Dept. of Biotechnology, Mizoram University, from 23 July - 31 July, 2018.

12. Participant in CME course/workshop on "The concept and Application of genomics in clinical Medicine", jointly organized by Civil Hospital Aizawl and Mizoram University, conducted by CSIR – Institute of Genomics and Integrative Biology, New Delhi in 11 August 2018.
13. Participant in International workshop on "Introduction to basic and advanced biomedical approaches for enhancing QOL in aging societies" held at Biomedical Research Institute, AIST, Tsukuba Science City, Japan inder DIACENTER – SUNRISE PROGRAM supported by AIST & JST (Govt. of Japan) and DBT (Govt. of India) from 14 – 21 Oct 2018.
14. Poster presentation in "42nd Annual IMA Pre-Conference Continuing Medical Education" organised by The Academic Wing of IMA, MSB on 13 December 2018 at Aijal Club, Mizoram
15. Participant in workshop on "A brief introduction to bioinformatics and System Biology" organized by DBT – BIF, Dept. of Biotechnology, Mizoram University from 13 – 14 Dec 2018.
16. Participate in 3rd NER Hands – on training workshop on "Gene Cloning, Protein Biochemistry, Structure Biology and Bioinformatics" organised by DBT – NER Biotechnology / Bioinformatics Training Centre, Advanced Centre for treatment, Research & Education in Cancer (ACTREC), Kharghar, Navi Mumbai. India from 4 – 15 feb 2019.

BEHIET-2018

The 12th Annual Convention of Association of Biotechnology and Pharmacy (ABAP) &

International Conference on Biodiversity, Environment and Human Health: Innovations and Emerging Trends (BEHIET 2018)

Organized by

School of Life Sciences,
Mizoram University,
Aizawl, Mizoram,
India

Association of Biotechnology
and Pharmacy (ABAP),
India

CERTIFICATE OF PARTICIPATION

This is to certify that ~~Prof.~~ / ~~Dr.~~ / Mr. / ~~Ms.~~ Andrew Vanlallawma has ~~participated~~ / presented / ~~oral~~ / poster in the 12th Annual Convention of Association of Biotechnology and Pharmacy (ABAP) & International Conference on Biodiversity, Environment and Human Health: Innovations and Emerging Trends (BEHIET 2018) organized at the School of Life Sciences, Mizoram University, Aizawl, Mizoram 796 004 during November 12 to 14, 2018.

Title: Pathogenicity of gene mutation in RTK/RAS signaling pathway contributing to leukemiogenesis.

Vice Chancellor,
Mizoram University

President,
ABAP

Dean, School of Life Sciences &
Organizing Secratrary

Variants which are not matched to ClinVar database and are likely to be novel

Chr	Pos	Ref	Alt	Gene	Nature of mutation	AA Change	MOI	Hom/Het
Chr10	104849436	TCC	-	NT5C2	Nonframeshift deletion	368_369del	-	Het
Chr11	102196019	A	G	BIRC3	Nonsynonymous SNV	K260R #	-	Het
Chr11	108163354	G	T	ATM	Nonsynonymous SNV	C1482F		Het
Chr11	108170524	A	C	ATM	Nonsynonymous SNV	T1697P		Het
Chr11	118352586	G	A	KMT2A	Nonsynonymous SNV	R1264Q	AD	Het
Chr12	112926884	T	C	PTPN11	Nonsynonymous SNV	S502P *	-	Het
Chr12	49426730	GCT	-	KMT2D	Nonframeshift deletion	3919_3920del	-	Het
Chr12	49447307	T	C	KMT2D	Nonsynonymous SNV	K264R	-	Het
Chr13	28608283	GTAGAA	-	FLT3	Nonframeshift deletion	590_591del	-	Het
Chr13	28608290	-	CCCCCGCCC	FLT3	Nonframeshift insertion	Y589delinsWAGD	-	Het
Chr15	90645557	CGC	-	IDH2	Nonframeshift deletion	22_22del	-	Het
Chr19	33792755	GGC	-	CEBPA	Nonframeshift deletion	69_70del	AD	Het
Chr19	49458971	G	-	BAX	Frameshift deletion	M1fs	-	Het
Chr20	31022784	C	T	ASXL1	Stopgain	Q757X	-	Het
Chr20	31024002	G	A	AXL2	Nonsynonymous SNV	D1163N	-	Het
Chr20	57429754	G	A	GNAS	Nonsynonymous SNV	R416Q	-	Het
Chr21	36164749	G	A	RUNX1	Nonsynonymous SNV	R376C	AD	Het
Chr2	198269877	T	C	SF3B1	Nonsynonymous SNV	S488G	-	Het
Chr22	23523586	C	-	BCR	Frameshift deletion	P147fs	-	Het
Chr22	23653975	-	CCGG	BCR	Frameshift insertion	S1092fs	-	Het
Chr2	25505431	C	-	DNMT3A	Frameshift deletion	G109fs	-	Het
Chr5	149457680	TGT	-	CSF1R	Nonframeshift deletion	241_242DEL	-	Het
Chr9	133759373	C	-	ABL1	Frameshift deletion	P566fs	-	Het
Chr9	133759490	AAG	-	ABL1	Nonframeshift deletion	605_605del	-	Het
Chr9	133759703	T	C	ABL1	Nonsynonymous SNV	S676P	-	Het
Chr9	133759794	G	T	ABL1	Nonsynonymous SNV	G706V	-	Het
Chr9	139397705	A	T	NOTCH1	Nonsynonymous SNV	V1699E	-	Het

Table 9: Matched Variants with 55 Leukemia genes that are not reported in ClinVar after applying filter 1 and 2.

Variants indicated in # and * are reported in dbSNP and COSMIC database, respectively. Chr: Chromosome number; Pos: Position; Ref: Reference, Alt: Alternate, AA change: Amino acid change; MOI: Mode of Inheritance, Hom/Het: Homozygous or Heterozygous.

Variants which are not matched to ClinVar are then interpreted using OMIM database for their gene function to its phenotype condition as well as the mode of inheritance are listed in Table 10.

Sample	Gene	Ref	Counts (%)	Alt	Counts (%)	Total reads	AA change	Hom/ Het	OMIM phenotype and Mode of inheritance	S_P	P_P	MT_P	D (*)
GDN4252	BCL10	C	49 (43%)	A	64 (56%)	114	A5S	Het	Male germ cell tumor, somatic	T	B	N	
	CBL	C	61 (48%)	T	67 (52%)	128	L620F	Het	Juvenile myelomonocytic leukemia (SMu, AD)	T	D	D	**
	ASXL1	G	10 (63)	A	6 (38)	16	G652S	Het	Myelodysplastic syndrome, somatic	T	B	N	
	BCR	C	2 (50%)	-	2 (50%)	4	P147fs	Het	ALL, CML somatic	0	0	0	
	BCR	-	107 (100)	CCGG	27 (25%)	107	S1092fs	Het	ALL, CML somatic	0	0	0	
GDN4253	BCL10	C	28 (55%)	A	23 (45%)	51	A5S	Het	Male germ cell tumor, somatic	T	B	N	
	ABL1	G	3 (27%)	T	8 (73%)	11	G725V	Het	Philadelphia chromosome-positive, resistant to imatinib	T	B	N	
	NT5C2	TCC	8 (80%)	-	2 (20%)	10		Het	-	0	0	0	
	BIRC3	A	28 (52%)	G	26 (45%)	54	K260R	Het	-	T	B	N	
	CBL	C	36 (45%)	T	44 (55%)	80	L620F	Het	Juvenile myelomonocytic leukemia (Smu, AD)	T	D	D	**
	ASXL1	G	0 (0%)	A	5 (100%)	5	G652S	Hom	Myelodysplastic syndrome, somatic	T	B	N	
	BCR	-	36 (58%)	CCGG	26 (42%)	62	S1092fs	Het	ALL, CML somatic	0	0	0	
GDN4254	SF3B1	T	107 (57%)	C	81 (42%)	188	S488G	Het	Myelodysplastic syndrome, somatic	D	B	D	**
	TET2	G	183 (52%)	T	169 (48%)	352	E1151X	Het	Myelodysplastic syndrome, somatic	D	0	D	**
	ABL1	T	13 (46%)	C	15 (54%)	28	S695P	Het	Philadelphia chromosome-positive, resistant to imatinib	T	B	N	
	NOTCH1	T	16 (38%)	C	26 (62%)	42	I567V	Het	-	T	B	D	**
	ATM	G	112 (53%)	T	99 (47%)	211	C1482F	Het	T-cell prolymphocytic leukemia, somatic	T	B	N	
	BCR	-	133 (67%)	CCGG	66 (33%)	199	S1092fs	Het	ALL, CML somatic	0	0	0	
	BCR	G	59 (71%)	A	24 (29%)	83	D1106N	Het	ALL, CML somatic	T	D	D	**
GDN4255	BCL10	C	58 (48%)	A	62 (51%)	121	A5S	Het	Male germ cell tumor, somatic	T	B	N	
	NOTCH1	A	104 (75%)	T	35 (25%)	139	V1699E	Het	-	D	D	D	***
	BIRC3	A	57 (55%)	G	46 (45%)	103	K260R	Het	-	T	B	N	
	ASXL1	G	90 (49%)	A	95 (51%)	185	D1163N	Het	Myelodysplastic syndrome, somatic	T	B	N	
	BCR	-	88 (71%)	CCGG	36 (29%)	124	S1092fs	Het	ALL, CML somatic	0	0	0	
	BCR	G	35 (71%)	A	14 (29%)	49	D1106N	Het	ALL, CML somatic	T	D	D	**
GDN4256	ABL1	C	6 (75%)	-	2 (25%)	8	P585fs	Het	Philadelphia chromosome-positive, resistant to imatinib	0	0	0	
	BIRC3	A	14 (39%)	G	22 (61%)	36	K260R	Het	-	T	B	N	
	KMT2A	G	17 (35%)	A	31 (65%)	48	R1264Q	Het	Leukemia, myeloid/lymphoid or mixed-lineage (AD)	T	D	D	**
	BCR	C	23 (55%)	A	19 (45%)	42	S1092fs	Het	ALL, CML somatic	0	0	0	
GDN4257	BCL10	C	8 (42%)	A	11 (58%)	19	A5S	Het	Male germ cell tumor, somatic	T	B	N	

	DNMT2A	C	9 (69%)	-	4 (30%)	13	A5S	Het	-	0	0	0	
	CSF1R	TGT	3 (38%)	-	5 (62%)	8	G109fs	Het	Leukoencephalopathy, diffuse hereditary, with spheroids (AD)	0	0	0	
	NOTCH1	T	3 (38%)	C	5 (63%)	8	I567V	Het	-	T	B	D	**
	KMT2D	T	8 (53%)	C	7 (47%)	15	K264R	Het	Leukemia, myeloid/lymphoid or mixed-lineage	D	P	N	**
	RUNX1	G	0 (0%)	A	2 (100%)	2	R376C	Het	AML (Smu, AD)	D	D	D	***
GDN4258	MUTYH	G	35 (49%)	A	37 (51%)	72	A230V	Het	-	T	P	D	*
	MUTYH	C	69 (47%)	T	77 (53%)	146	G143E	Het	-	D	D	D	***
	KIT	A	47 (44%)	G	59 (56%)	106	I438V	Het	Germ cell tumors, somatic ,Leukemia, acute myeloid (Smu,AD)	T	B	D	*
	ABL1	AAG	43 (87%)	-	6 (12%)	49	K642del	Het	Philadelphia chromosome-positive, resistant to imatinib	0	0	0	
	ATM	A	68 (46%)	G	80 (54%)	149	H24R	Het	T-cell prolymphocytic leukemia, somatic	T	B	N	
	ATM	C	49 (49%)	T	50 (51%)	99	H1380Y	Het	T-cell prolymphocytic leukemia, somatic	T	B	N	
	KMT2D	GCT	13 (87%)	-	2 (13%)	15	Q3919del	Het	Leukemia, myeloid/lymphoid or mixed-lineage	0	0	0	
	SETBP1	G	14 (47%)	T	16 (53%)	30	E1466D	Het	-	T	B	N	
	CEBPA	GGC	2 (50%)	-	2 (50%)	4	Pdel	Het	AML (Smu,AD)	0	0	0	
	BCR	-	49 (91%)	CCGG	5 (9%)	54	S1092fs	Het	ALL, CML somatic	0	0	0	
GDN4259	ABL1	T	8 (67%)	C	4 (33%)	12	S695P	Het	Philadelphia chromosome-positive, resistant to imatinib	T	B	N	
	NOTCH1	C	72 (43%)	T	94 (56%)	167	V1232M	Het	-	T	P	N	
	ASXL1	G	12 (52%)	A	11 (48%)	23	G652S	Het	Myelodysplastic syndrome, somatic	T	B	N	
	ASXL1	C	51 (49%)	T	53 (51%)	104	Q757X	Het	Myelodysplastic syndrome, somatic	T	0	D	*
	GNAS	G	8 (50%)	A	8 (50%)	16	R416Q	Het	-	T	0	N	
	BCR	-	70 (75%)	CCGG	23 (25%)	93	S1092fs	Het	ALL, CML somatic	0	0	0	
	BCR	G	45 (87%)	A	7 (13%)	52	D1106N	Het	ALL, CML somatic	T	D	D	**
GDN4260	MUTYH	C	66 (55%)	T	54 (45%)	121	G25D	Het	-	T	P	N	
	MUTYH	G	55 (51%)	A	52 (49%)	107	P18L	Het	-	T	B	D	*
	BCL10	C	0 (0%)	A	57 (97%)	59	A5S	Hom	Male germ cell tumor, somatic	T	B	N	
	BCR	-	74 (100%)	CCGG	24(32%)	74	S1092fs	Het	ALL, CML somatic	0	0	0	
GDN4261	PTPN11	T	156 (83%)	C	32 (17 %)	188	S502P	Het	Leukemia, juvenile myelomonocytic, somatic	T	P	D	*
	FLT3		11 (12%)	del and ins	71 (78%)	85	YFY589-91delWAGDins	Het	ALL , AML	0	0	0	
	IDH2	CGC	6 (66%)	CGC del	2 (33%)	6	22 del	Het	-	0	0	0	

	BAX	G	3 (23)	Gdel	10 (76%)	13	M1fs	Het	T-cell ALL, somatic	0	0	0	
	ASXL1	G	11 (61)	A	7 (39%)	18	G652S	Het	Myelodysplastic syndrome, somatic	T	B	N	
	BCR	-	44 (72)	CCGGins	17 (27)	61	S1092fs	Het	ALL, CML somatic	0	0	0	
GDN4262	ABL1	T	8 (42%)	C	11 (58%)	19	S676P	Het	Philadelphia chromosome-positive, resistant to imatinib	T	B	N	
	ATM	A	107 (54%)	C	91 (46%)	198	T1697P	Het	T-cell prolymphocytic leukemia, somatic	T	B	N	
	KMT2D	G	21 (42%)	A	29 (58)	50	P2382S	Het	Leukemia, myeloid/lymphoid or mixed-lineage	D	P	D	**
	ASXL1	G	15 (65%)	A	8 (35%)	23	G652S	Het	Myelodysplastic syndrome, somatic	T	B	N	

Table 10: Unmatched variants to ClinVar are interpreted using OMIM.

Ref: Reference, Counts: Read counts identified from IGV, Total reads: Total reads identified from read depth, AA change: Amino acid change, Hom/Het: Homozygous or heterozygous, S_P: SIF Prediction, P_P: Polyphen Prediction, MT_P: Mutation Taster Prediction, D (*): Deleterious/Pathogenic, * indicates how many are predicted as deleterious from the three predicting software.