

ORIGINAL ARTICLE

Whole genome sequencing reveals the mutational landscape from disease diagnosis to relapse in patients with childhood acute myeloid leukaemia

Habsah AZIZ^{1*}, Nurul Syakima AB MUTALIB², Hamidah ALIAS^{2,3}, Rahman JAMAL²

¹Institute of Systems Biology (INBIOSIS), Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia, ²UKM Medical Molecular Biology Institute (UMBI), Universiti Kebangsaan Malaysia, Jalan Yaacob Latif, 56000 Cheras, Kuala Lumpur, Malaysia, ³Department of Paediatrics, Faculty of Medicine, Universiti Kebangsaan Malaysia, Jalan Yaacob Latif, 56000 Cheras, Kuala Lumpur, Malaysia.

Abstract

Introduction: Leukaemia is the most common cancer in children, however, there is still a big gap in knowledge about the genomic alterations in childhood acute myeloid leukaemia (AML) compared to adult AML. Relapsed AML remains as a leading cause of cancer deaths among children. This study aims to understand the molecular mechanisms of relapsed AML by elucidating the mutational landscape before and during relapse. **Materials and Methods:** Whole genome sequencing was performed on matched samples collected at diagnosis, remission and relapse from three patients of de novo childhood AML. Sanger sequencing was performed for validation in 47 patients' samples, followed by functional analysis. **Results:** Overall, we identified 312 somatic mutations including synonymous single nucleotide variants (SNVs), missense SNVs, deletions and insertion frameshifts, stopgains and splice sites. After prioritisation, only 46 variants were present at diagnosis (13-17 mutations per patient) and 49 variants at relapse (12-20 mutations per patient). Out of 81 variants, there were 35 new variants detected at relapse but not present at diagnosis. Six potential driver mutations (*KIT*, *CDC73*, *HNF1A*, *RBM10*, *ZMYM4* and *ETV6*) were identified in predicting relapse for the 3 patients, with recurrent mutations of the *ETV6* gene in 2 patients. Functional analysis of the *ETV6* mutation showed that *ETV6* lost its tumour suppressive function when both mutant *ETV6* p.P25fs and *ETV6* p.N75fs were tested in vitro. **Conclusion:** This study has uncovered the mutational landscape in three local childhood AML patients and contributes to a better understanding of the molecular mechanisms of relapsed AML.

Keywords: Acute myeloid leukaemia; relapsed AML; children; whole genome sequencing; mutation; *ETV6*; functional analysis

INTRODUCTION

Acute myeloid leukaemia (AML) is believed to be a consequence of arrested myeloid differentiation in hematopoietic system, and is characterized by abnormally differentiated and occasionally poorly differentiated cells.¹ The biology of AML is associated with its age variation, as evidenced by the significant variability of genomic aberrations in AML from infancy to adulthood. There is also a significant age-based incidence, with elevated incidence reported in both infants and older adults.²⁻⁶ AML accounts for around 25% of childhood leukaemia

cases and despite advances in therapy over the past decades, overall survival for childhood AML has not exceeded 70% compared to the overall survival of childhood ALL at 80-90%.⁷⁻¹¹ This is in part because of the disease heterogeneity. AML is the leading cause of death in childhood cancers even though the prognosis is classified in the low or middle risk group.¹²

Relapsed AML significantly contributes to the mortality of AML patients and has resulted in the death of more than half of the affected children.¹³ The median time of relapse in children with AML has been reported over the decades as 0.93 years

*Address for correspondence: Dr. Habsah Aziz, Institute of Systems Biology (INBIOSIS), Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia. Tel: +603 8921 4546, Fax: +603 8921 3398, E-mail: habsah@ukm.edu.my

(1976–1991), 0.76 years (1991–1997) and 0.8 years (2002–2008), with little improvement.¹⁴ Advances in disease management through intensification of chemotherapy, accuracy of risk classification, improvement in supportive care and the use of minimal residual disease (MRD) for monitoring therapy response have contributed to a decrease in relapse rate to 30–40%. However, new therapies are much needed because the event free survival to many sub-types of childhood AML is still at a very low level of only 50%.^{10,11} In Malaysia, the overall cure rate of childhood AML patients for 1, 3 and 5 years were 52.0%, 42.4% and 38.1% respectively from 1990 to 2010.¹⁵

Although there has been much progress in understanding the molecular heterogeneity and pathogenesis of AML, most patients still experience relapse. Identification of somatic variants is an approach to understanding the molecular mechanism of AML relapse. Nevertheless, identifying driver mutations from the list of somatic mutations is still challenging in the current cancer genome sequencing era.¹⁶ The latest techniques such as next generation sequencing (NGS) have facilitated the study of cancer genomes. Whole genome sequencing (WGS) and whole exome sequencing (WES) has successfully identified novel driver genes such as *NPM1*, *FLT3*, *CEBPA*, *DNMT3A*, *IDH1*, *IDH2*, *TET2*, *RUNX1*, *TP53*, *NRAS*, *WT1*, *PTPN11*, *KRAS*, *KIT*, *SMC3*, *PHF6*, *STAG2*, *RAD21*, *FAM5C*, *HNRNP1K* and *WT1* in adult AML.^{17–21} However, the relevance of these findings to childhood AML remains unclear given that mutations commonly found in adult AML are less discovered in childhood AML.^{12,22–24}

Therefore, understanding each patient's genome and identifying actionable mutations may help improve the survival of patients with relapsed AML. The knowledge of the childhood AML genome is still at the early stage as most genome-based studies involved adult AML patients.^{12,17,26,27,18–25} Thus, this study was carried out to fill some of the knowledge gaps in the molecular mechanisms of relapse in childhood AML.

MATERIALS AND METHODS

Study Cohort and Sample Collection

Recruitment of subjects for this study involved patients with AML from birth to 18 years of age at the Hospital Canselor Tuanku Muhriz and Institute of Paediatrics, Hospital Kuala Lumpur. This study was approved by the UKM

Research Ethics Committee (UKM 1.5.3.5/244/UMBI-2015-001) and the Medical Research Ethics Committee of the Ministry of Health Malaysia (NMRR-15-583-24799 (IIR)). This study involved the collection of bone marrow specimens from 30 paediatric AML patients at three stages: during diagnosis, after remission and after relapse. Written informed consent was obtained from all the participants.

Bone Marrow Sample Processing and Cells Sorting

Bone marrow samples of patients with paediatric AML were collected in two sodium heparin tubes through a bone marrow aspiration procedure. The mononuclear cells were isolated using the Ficoll-Paque method. As the bone marrow samples obtained did not contain 100% AML leukaemia blast cells, isolation of AML blast cells were performed using FACSAria™ II Cell Sorter flow cytometer to increase the purity of AML blast cells before DNA and RNA extraction. We used seven antibody staining markers namely BB515 Mouse Anti-Human HLA-DR, PE Mouse Anti-Human CD13 PerCP-Cy™ 5.5, Mouse Anti-Human CD56, CD117-PE-Cy7, APC Mouse Anti-Human CD34, APC-H7 Mouse anti-Human CD45 and BD Horizon™ Fixable Viability Stain 620 (BD Bioscience).

DNA and RNA Extraction

DNA and RNA were extracted from mononuclear cells using the AllPrep DNA/RNA/miRNA Extraction Kit according to the manufacturer's procedures. The NanoDrop™ 2000c spectrophotometer was used to measure the purity and concentration of nucleic acids. The Qubit® 2.0 fluorometer was used to measure DNA concentration using a Qubit® dsDNA High Sensitivity kit (Thermo Fisher Scientific). The integrity of extracted DNA was determined using 1.2% agarose gel electrophoresis.

Library Preparation and Whole Genome Sequencing

WGS was performed using the Illumina HiSeqXTen™ sequencing system. The library preparation was performed using the Truseq™ DNA Nano kit (Illumina) according to the manufacturer's protocol.

Bioinformatic Analysis

Raw data from the FASTQ files were exported from the Illumina HiSeqX sequencing system and alignment was performed against Hg19

human genome reference using Burrows-Wheeler Aligner (BWA) and variant calling was performed using Genome Analysis Toolkit (GATK). Somatic analysis of Single Nucleotide variants (SNV), Small insertions and deletions (InDels) and Structural Variants (SV) was analysed using muTect software, Strelka dan DELLY.

After variants detection was performed, subsequent variants were annotated using ANNOVAR. Gene annotation was based on the RefSeq Database and GENCODE were used to search for genome regions affected by variants and protein changes. Variants were annotated against conservation areas using alternative allele frequency data in populations as reported by large scale sequencing projects covering 1000 human genome projects and Exome Aggregation Consortium (ExAC).

The Single Nucleotide Polymorphism Database (dbSNP), COSMIC, OMIM and the Genome Wide Association Studies (GWAS) catalog were used to obtain information related to frequently encountered variants such as Single Nucleotide Polymorphism (SNP). The annotations to the function and signaling pathway were from Gene Ontology database, KEGG, Reactome and Biocarta. Protein pathogenicity prediction was performed using Sorting Intolerant From Tolerant (SIFT), PolyPhen2 HDIV and PolyPhen2 HVAR, Likelihood Ratio Test (LRT), Mutation Taster and Mutation Assessor.

All somatic variants obtained from muTect software, Strelka and DELLY were manually validated using Integrated Genomic Viewer (IGV) software to check for sequencing artifacts and alignment errors. Subsequently, somatic variants were further prioritised based on protein pathogenicity scores namely Sorting Intolerant From Tolerant (SIFT), PolyPhen2 HDIV and PolyPhen2 HVAR, Likelihood Ratio Test (LRT) and Mutation Taster. Variants were also prioritised based on sequencing depth, that is more than 10 sequencing coverage for alternative alleles.

Sanger Sequencing

Sanger sequencing was performed to validate the selected variants (*ETV6 p.P25fs*, *ETV6 p.N75fs*, *ABCG8*, *EZH1*, *EDARADD*, *RYR2*, *PKHD1*, *PLXNA1*, *PRR13*, *RBM10*, *SH2B1* and *TAS2R7*) for technical and biological validation. Technical validation involved all trio samples obtained (9 samples) that have been performed using WGS.

As for biological validation, validation was performed on 47 additional samples from another 27 patients consisting of 28 samples of AML diagnosis, 15 samples of AML remission and 4 samples of AML relapse. Sanger sequencing was performed on an automatic DNA sequencer 3130xl Genetic Analyzer. DNA chromatogram was analysed using BioEdit Sequence Alignment Editor software.

Plasmid DNA

DNA plasmids were constructed for wild-type plasmids ets variant 6 (*ETV6*) (*ETV6* wild type; WT) and mutant plasmids (*ETV6* P25fs and *ETV6* N75fs) before being transfected into cell lines. For the *ETV6* wild type Plasmid (*ETV6* wild type; WT), cDNA comprising the full-length sequence of the human *ETV6* gene (NM_001987) was cloned into the pcDNA3.1+/C- (K) DYK vector expression in the cytomegalovirus (CMV) promoter region together the DYKDDDDK tag on terminal C uses the CloneEZ cloning strategy. For *ETV6* Mutant Type Plasmids (*ETV6* P25fs and *ETV6* N75fs), the desired mutations were integrated into the *ETV6* open reading frame (ORF) on the pcDNA3.1+/C- (K) DYK vector using the CloneEZ cloning strategy.

Cell Culture

Two cell lines, Kasumi-1 (ATCC® CRL-2724™) and 293T (ATCC® CRL-3216™), were used for the functional study. Kasumi-1 cell line is a suspension cell lines and it is a myoblast cell obtained from a 7-year-old Japanese child with AML relapse (FAB M2). The Kasumi-1 cells have the translocation of t (8:21) (q22, q22) making this cell line the most suitable based on the characteristics that also exist in the trios samples of this study. While 293T cell line is an adherent type cell obtained from human kidney embryo. This 293T cells were selected for their high transfection capacity and often used as *in-vitro* models to study the function of gene mutations. Kasumi-1 cells and 293T cells were cultured in T-25 or T-75 culture flasks using RPMI and DMEM media respectively with 20% FBS and incubated at 37°C with 5% CO₂. The media for these cells were added or changed every 2 or 3 days for the cells to receive adequate nutrients. The sub-culture process was performed when the cell density reached a confluent state.

Transient Transfection

Kasumi-1 cells were transfected with wild type and mutant plasmid *ETV6* using electroporation

method and Neon™ Transfection System 10 μ l Kit (Invitrogen). The transfection protocol is set at a voltage of 1650V with 1 pulse rate (pulse) for 20 milliseconds and 3 μ g of DNA plasmid was used for 1×10^6 cells. Meanwhile, Lipofectamine™ 3000 (Thermo Fisher Scientific) transfection reagent was used for 293T cells. The transfection was carried out according to the manufacturer's protocol. After 48 hours of transfection, protein was extracted by radioimmunoprecipitation assay (RIPA) (Thermo Fisher Scientific) and Halt™ Protease Inhibitor (Pierce).

Gene Expression Analysis by Quantitative Real Time PCR

Quantitative Real-Time PCR (qRT-PCR) reaction was performed to assess the expression of wild and mutant *ETV6* gene. This analysis used extracted RNA from transfected Kasumi-1 and 293T cells with wild-type *ETV6* plasmids and mutant constructs. The cDNA synthesis process was carried out using an iScript™ cDNA synthesis kit (Bio-rad Laboratories Inc) according to the manufacturer's instruction. The synthesized cDNA was then subjected to qPCR using the SsoAdvanced™ Universal SYBR® Green Supermix kit (Bio-rad Laboratories Inc) on a CFX96™ real-time PCR machine. GAPDH is used as a reference gene for the normalization process of *ETV6*.

Western Blot

Sodium dodecyl sulfate-polyacrylamide gel with a concentration of 5% and 10% was prepared using a mixture of 30% Bis-Acrylamide, 1.5M Tris-HCL (Tris-(hydroxymethyl)-aminomethane) pH 8.8, 0.5M Tris-HCL pH 6.8 (Bio-rad Laboratories Inc), N, N, N', N'-tetramethyl-ethylenediamine and 10% ammonium persulfate (Sigma-Aldrich). After SDS-polyacrylamide gel electrophoresis and protein transfer to polyvinylidene difluoride (PVDF) membrane using a semi-dry blotter with a voltage of 15V for 15 min, immunoblots were blocked with 1% buffer 1X Tris Buffered Saline (TBS) and 1% Tween 20 containing 5% nonfat dried milk. Immunoblots were probed with 1: 1000 primary antibody Anti-DDK (Gene Script), 1: 1000 primary antibody β -actin (AC-15): clone sc-69879 (Santa Cruz Biotech), 1: 10000 secondary antibody goat anti-mouse IgG-HRP: clone sc-2005 (Santa Cruz Biotech). Proteins were detected with SuperSignal West Pico PLUS Chemiluminescent Substrate (Thermo Fisher Scientific) and Chemidoc XRS imaging system

and Chemidoc XRS Quantity One software (Bio-rad Laboratories Inc) was used to visualize the intensity of the protein.

Cell Viability Assay

Cell viability assays were conducted to study the effect of *ETV6* mutation on cell viability. Analysis was performed after the cells were transfected for 24, 48, and 72 hours. The XTT Cell Viability Assay kit (Biotium Inc) is used to assess the rate of cell proliferation based on the activity of mitochondrial enzymes in living cells. Absorption readings were taken using a Varioskan™ Flash (Thermo Fisher Scientific) microplate reader at 450 nm and 630 nm wavelengths as background absorption.

Protein Expression Assay

Protein expression assays were performed using the Proteome Profiler™ Human Apoptosis Array kit (R&D System) to assess the effect of *ETV6* mutation on apoptosis related protein. The procedures were performed according to manufacturer's protocol. The intensity of the resulting protein spots was visualized using the Chemidoc XRS imaging system and Chemidoc XRS Quantity One software (Bio-rad Laboratories Inc).

Statistical Analysis

Statistical analysis was performed using GraphPad Prism 7.0. Data were shown as mean \pm standard deviation. Two-tailed T-test was used for functional assay analysis. In all assays, $P < 0.05$ was considered significant. Experiments were performed at least twice for the assays involved in this study.

RESULTS

Demographic Data of Study Patients

We recruited 30 patients with childhood AML who received treatment at HCTM and HKL hospitals. The demographic and clinical characteristics of the patients are listed in Table 1. In total, 56 bone marrow samples from 30 patients with childhood AML aged 8 weeks to 15 years were examined in this study from the three phases of the disease that is diagnosis, remission and relapse. Three patients with the trio samples (at diagnosis, remission and relapse) were used for WGS. The clinical data for the three trio samples are shown in Table 2. We had 3 patients from whom we managed to obtain the samples at diagnosis, remission and relapse. We managed to collect 47 samples from another 27

Table 1: Demographic and clinical data of study subjects

CATEGORY	Total	Percentage %
GENDER		
Male	17	56.7
Female	13	43.3
AGE		
Infant (8 weeks-1 year old)	3	10.0
Toddler (1 - 3 years old)	12	40.0
Kid (3 - 11 years old)	11	36.7
Teenager (12 - 17 years old)	4	13.3
RACE		
Malay	25	83.3
Chinese	4	13.3
Indian	1	3.3
HOSPITAL		
HCTM	16	53.3
HKL	14	46.7
STAGE		
Diagnosis	31	55.4
Remission	18	32.1
Relapse	7	12.5
Matched diagnosis & remission	18	
Matched diagnosis, remission & relapse	3	

patients which were used for validation. The data of the 27 patients are shown in Supplementary Table A.

Whole Genome Sequencing Technical Report
WGS was performed on nine samples from the three patients (matched samples at diagnosis,

Table 2: Clinicopathological data for the three trios (matched diagnosis-remission-relapse)

Patient	Gender	Age	Race	Cytogenetic	Blast at diagnosis (%)	Blast at remission (%)	Blast at relapse (%)	Time to relapse (month)
UL8	Female	12	Malay	45,X,-X, t(8;21)(q22;q22)	60	2	33	15
UL16	Male	14	Malay	Structural abnormalities on chromosomes 2,3 & 8. There was a loss of Y seen in all cells analysed	60	3	55	10
UL27	Female	5	Indian	Female normal chromosomes, no clonal abnormalities were shown.	90	3	81	9

remission and relapse). The quality of WGS data is shown in Supplementary Table B. The quality score for the Q30 phred scale is greater than 80% for all samples, ranging from 82.97% to 92.48%. Supplementary Table C shows the mapping statistics, coverage and sequencing depth in each WGS sample.

Characterization of Somatic Single Nucleotide Variants (SNVs), Somatic Insertion and Small Deletions (InDels) and Somatic Structural Variants (SV)

From the analysis of the sequencing data from the leukaemia cells (diagnosis and relapse samples) to the remission samples, we identified the somatic mutations that have been accumulated in the Leukaemia cells. Somatic analysis of Single Nucleotide Variants (SNV), Small insertions and deletions (InDels) and Structural Variants (SV) was performed using the muTect software,²⁸ Strelka²⁹ dan DELLY³⁰ respectively. Table 3 summarises the somatic SNVs statistics observed in patients with childhood AML. The full list of somatic SNVs is listed in Supplementary Table D to I. Table 4 summarises the statistics for InDels and Table 5 shows the complete list of somatic InDels. The identified SV statistics and the distribution of somatic SV in each patient are as shown in Table 6.

Driver Mutation Identification

The driver gene identification for this study was based on *CGC513* which is the driver gene listed in the Cancer Gene Census (CGC), the Bert Vogelstein125,³¹ the SMG127,³² and Comprehensive435.³³ The driver gene must have at least one 'damaging' prediction or two 'possibly damaging' scores. The list of the driver genes found in this study is in Table 7.

Variant Prioritisation

Prioritisation of somatic SNV and InDels variants was performed by using IGV,³⁴ pathogenicity score screening based on SIFT,³⁵ PolyPhen2 HDIV and PolyPhen2 HVAR,^{36,37} LRT³⁸ Mutation Taster³⁹ and Mutation Assessor⁴⁰ as well as screening sequencing depth coverage must at least 10X for alternative alleles. After prioritisation, 312 somatic gene variants were finalised to 81 somatic gene variants, with 46 variants were present at diagnosis and 49 variants present at relapse. We then compared the variants present at diagnosis to the relapse samples using the Venn diagram. Figures 1 (A, B and C) shows the results of variant comparison on samples UL8, UL16, and UL27. The prioritisation of variants was further refined by comparing the somatic variants identified in the three patients'

Table 3: Total somatic single nucleotide variants at various genome regions

Variant type and location	UL8T	UL8T2	UL16T	UL16T2	UL27T	UL27T2
Coding sequence	58	72	49	48	43	31
Synonymous SNV	21	23	22	19	21	17
Missense SNV	33	45	25	27	17	14
Stopgain	2	1	2	1	2	0
Stoploss	0	0	0	0	0	0
Unknown	2	3	0	1	3	0
Intronic	1964	2384	1705	1742	1845	1511
3'-untranslated region	33	32	24	39	31	36
5'-untranslated region	11	13	9	6	21	16
splicing	1	2	2	4	1	1
ncRNA_exonic	32	38	19	29	30	30
ncRNA_intronic	452	629	338	476	421	375
ncRNA_UTR3	0	0	0	0	0	0
ncRNA_UTR5	0	0	0	0	0	0
ncRNA_splicing	1	0	0	0	0	0
Upstream	44	64	34	43	35	33
Downstream	47	56	43	54	53	28
Intergenic	4073	4711	3009	3648	3304	2942
Total	6716	8003	5233	6089	5787	5003

Table 4: Total somatic InDels at various genome regions

Variant type and location	UL8T	UL8T2	UL16T	UL16T2	UL27T	UL27T2
Coding sequence	3	2	0	1	0	0
Frameshift deletion	1	1	0	0	0	0
Frameshift insertion	2	1	0	1	0	0
Inframe deletion	0	0	0	0	0	0
Inframe insertion	0	0	0	0	0	0
Stopgain	0	0	0	0	0	0
Stoploss	0	0	0	0	0	0
Unknown	0	0	0	0	0	0
Intronic	31	37	11	20	10	36
3'-untranslated region	0	0	1	0	0	1
5'-untranslated region	0	0	0	1	1	1
Splicing	0	0	0	0	1	0
ncRNA_exonic	0	0	0	0	0	0
ncRNA_intronic	5	9	1	2	1	1
ncRNA_UTR3	0	0	0	0	0	0
ncRNA_UTR5	0	0	0	0	0	0
ncRNA_splicing	0	0	0	0	0	0
Upstream	0	1	0	2	0	1
Downstream	0	1	0	0	0	0
Intergenic	44	62	20	18	14	27
Total	83	112	33	44	27	67

Table 5: List of somatic InDels

Patient	Gene	InDel type	Amino acid changes	Variant Allele Frequency (VAF)
UL8T	<i>ABCG8</i>	Frameshift deletion	exon5:c.637delT;p.L213fs	38%
UL8T	<i>ETV6</i>	Frameshift insertion	exon3:c.224_225insGAGGT;p.N75fs	40%
UL8T	<i>MAGEA10</i>	Frameshift insertion	exon4:c.321_322insAG;p.G108fs, exon5:c.321_322insAG;p.G108fs, exon5:c.321_322insAG;p.G108fs	47%
UL8T2	<i>ABCG8</i>	Frameshift deletion	exon5:c.637delT;p.L213fs	41%
UL8T2	<i>ETV6</i>	Frameshift insertion	exon3:c.224_225insGAGGT;p.N75fs	42%
UL16T2	<i>ETV6</i>	Frameshift insertion	exon2:c.73_74insCCTCT;p.P25fs	39%
UL27T	<i>CTRB2</i>	Stop codon	exon8:c.631-1->T	33%

Table 6: Statistics of somatic structural variants

Patient	Tandem duplication	Translocation	Inversion	Deletion
UL8T (Diagnosis)	19	66	12	24
UL8T2 (Relapse)	105	31	27	37
UL16T (Diagnosis)	73	21	8	24
UL16T2 (Relapse)	25	93	15	88
UL27T (Diagnosis)	23	105	17	26
UL27T2 (Relapse)	18	101	17	101

Table 7: List of the driver genes

Gene	Variant classification	Amino acid change	dbSNP - RS	COSMIC	CGC	Bert Vogelstein125	Comprehensive435	SMG127	Variant Allele Frequency (VAF)	Patient
<i>KIT</i>	Nonsense mutation	KIT:NM_000222:exon17:c.T2466G;p.N822K1 KIT:NM_001093772: exon 17:c.T2454G;p.N818K	.	COSM1322	Gastrointestinal stroma tumour (GIST), AML, TGCT, Mastocytosis, Mucosal melanoma, Epithelioma	Oncogene	Driver	1.40%	38 %	UL16T
<i>RBM10</i>	Stopgain	RBM10:NM_001204466:exon16:c.C1570T;p.Q524X1 RBM10:NM_152856: exon 16:c.C1567T;p.Q523X1 RBM10:NM_001204467: exon17:c.C1798T;p.Q600X1RBM10:NM_001204468: exon17:c.C1996T;p.Q666X1RBM10:NM_005676: exon17:c.C1801T;p.Q601X	.	COSM1744247	-	-	Candidate driver	-	62%	UL8T2
<i>ZMYM4</i>	Nonsense mutation	ZMYM4:NM_005095: exon 1:c.T2G;p.M1R	.	.	-	-	Candidate driver	-	11%	UL8T2
<i>ETV6</i>	Frameshift insertion	ETV6:NM_001987: exon3:c.224_225insGAGGT;p.N75fs	.	.	Congenital fibrosarcoma, Leukaemia, Lymphoma, Breast carcinoma, MDS	-	-	-	42%	UL8T2
<i>CDC73</i>	Nonsense mutation	CDC73:NM_024529: exon 10:c.C946T;p.H316Y	.	.	Parathyroid cancer, Fibroma of the jaw	Tumour suppressor gene	Driver	-	50%	UL16T2
<i>HNF1A</i>	Nonsense mutation	HNF1A:NM_000545: exon 2:c.C391T;p.R131W1HNF1A:NM_001306179: exon 2:c.C391T;p.R131W	rs137853244	COSM24907	Liver adenoma, Hepatocellular-hepatic adenoma, Hepatocellular carcinoma	Tumour suppressor gene	-	-	50%	UL2T2
<i>ETV6</i>	Frameshift insertion	ETV6:NM_001987: exon2:c.73_74insCCTC;p.P25fs	.	.	Congenital fibrosarcoma, Leukaemia, Lymphoma, Secretory breast, MDS	-	-	-	39%	UL16T2

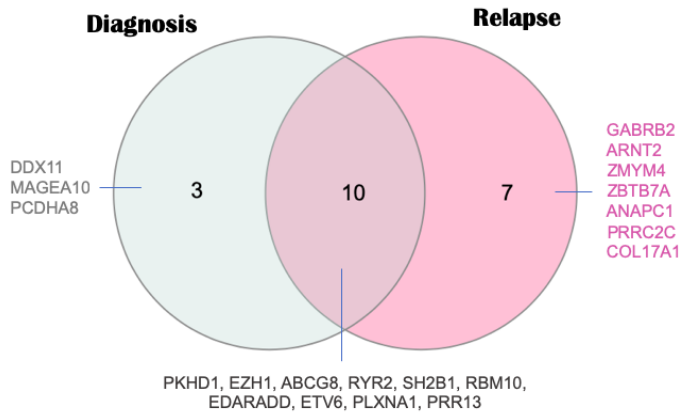


Figure 1 (A): Venn diagram of diagnosis versus relapse for UL8 patient. Ten recurrent somatic variants namely *PKHD1*, *EZH1*, *ABCG8*, *RYR2*, *SH2B1*, *RBM10*, *EDARADD*, *ETV6*, *PLXNA1* and *PRR13* in the diagnosis and relapse samples as well as having additional mutations during the relapse phase (*GABRB2*, *ARNT2*, *ZMYM4*, *ZBTB7A*, *ANAPC1*, *PRRC2C* and *COL17A1*).

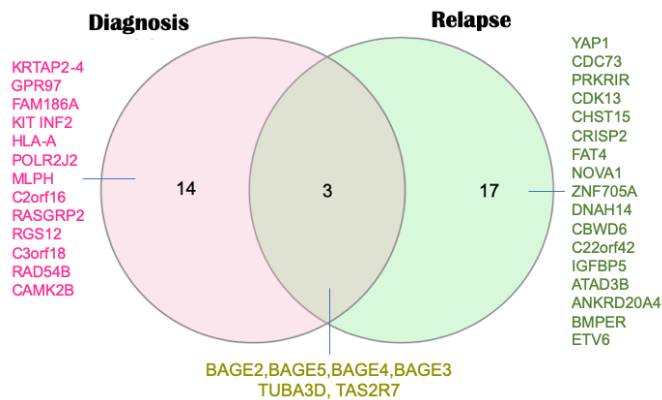


Figure 1 (B): Venn diagram of diagnosis versus relapse for UL16 patient. Three somatic variants (*BAGE2*, *BAGE5*, *BAGE4*, *BAGE3*, *TUBA3D* and *TAS2R7*) present in both sample. There are 17 new variants exist during relapse (*YAP1*, *CDC73*, *PRKRIR*, *CDK13*, *CHST15*, *CRISP2*, *FAT4*, *NOVA1*, *ZNF705A*, *DNAH14*, *CBWD6*, *C22orf42*, *IGFBP5*, *ATAD3B*, *ANKRD20A4*, *BMPER* and *ETV6*).

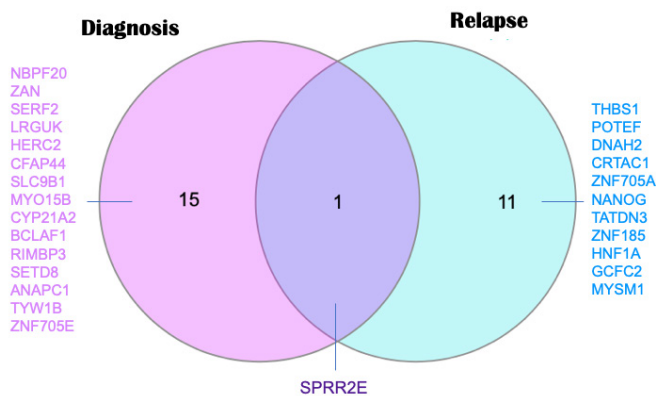


Figure 1 (C): Venn diagram of diagnosis versus relapse for UL27 patient. Only one recurrent variant *SPRR2E*. However, there are 11 new variants observed at relapse phase (*THBS1*, *POTEF*, *DNAH2*, *CRTAC1*, *ZNF705A*, *NANOG*, *TATDN3*, *ZNF185*, *HNF1A*, *GCFC2* dan *MYSM1*).

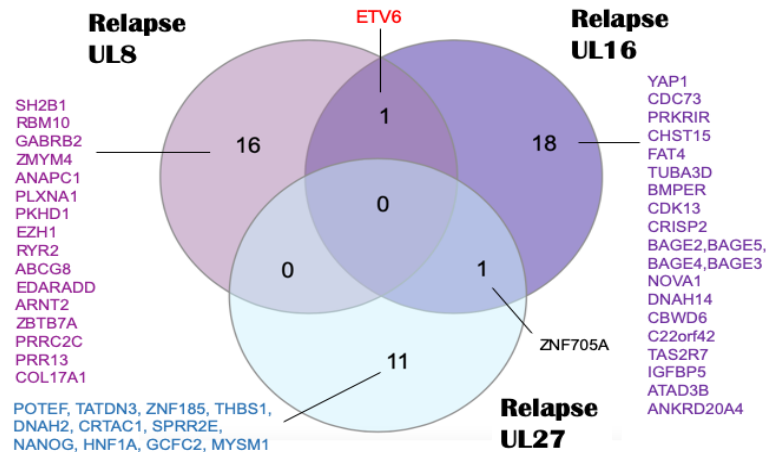


Figure 2: Venn diagram of relapse variants in three patient trios. *ETV6* is a recurrent mutation that exists in UL8 and UL16 relapse samples. The *ZNF705A* variant is a recurrent mutation in UL16 and UL27 relapse samples.

relapse samples (Figure 2). Based on the driver gene identification analysis, the *ETV6* variant is one of the variants identified as a driver gene, resulting the selection of this variant for the functional study. A lollipop plot (Figure 3) was generated to identify the distribution of the somatic variants of *ETV6* p.P25fs and p.N75fs found in this study. This study is the first to report these two somatic variants in AML.

Sanger Sequencing Validation

Technical validation involved nine samples while biological validation was performed on 47 other samples from another 27 patients. Technical validation successfully detected the presence of variants discovered using WGS. No variants were detected on the remission samples indicating all of these variants were

not germline variants. Whereas for biological validation, these 12 variants were not detected in the additional samples.

Gene Expression Analysis by Quantitative Real Time PCR

Evaluation of wild-type and mutant *ETV6* gene expression was carried out using the quantitative real time pcr (qRT-PCR). The relative expression of wild-type and mutant *ETV6* mRNAs was calculated after 48 hours transfection into Kasumi1 and 293T cells and normalized to GAPDH expression. The results were shown in and Figure 4A and 4B.

Detection of Wild-type and Mutant ETV6 Proteins Using Western Blot

Figure 5 shows the results of the western blot.

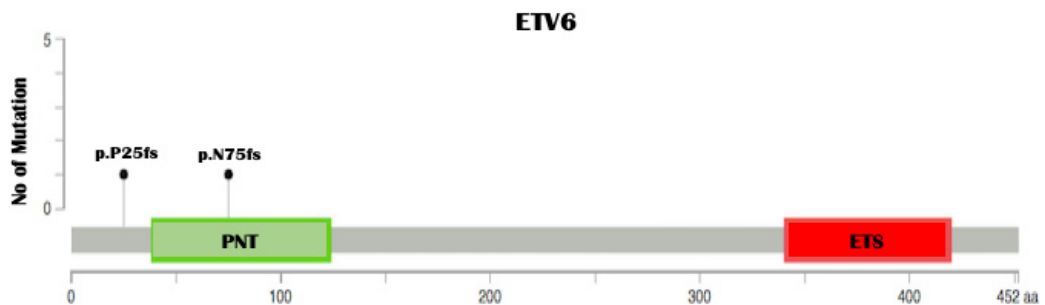


Figure 3: Distribution of *ETV6* somatic variants on *ETV6* gene domain function. The gray bars represent the whole protein with different amino acid (aa) positions and the colored boxes are specific to the domain protein function; green for the PNT domain and red for the ETS domain. *ETV6* p.N75fs is located in the PNT domain and *ETV6* p.P25fs is located before the PNT domain. Protein diagrams were generated using Mutation Mapper.

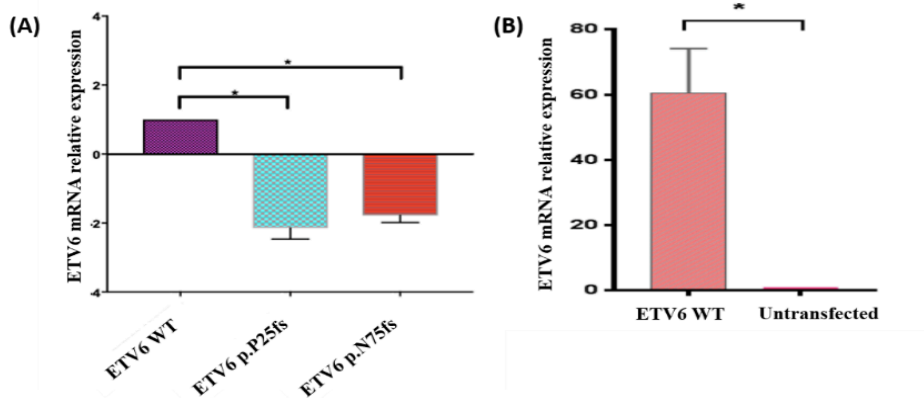


Figure 4: (A) Quantification of mRNA expression of the mutants *ETV6* p.P25fs and p.N75fs in 293T cells. *ETV6* mutants p.P25fs and p.N75fs had significantly reduced gene expression compared to wild-type *ETV6* (WT). (B) Quantification of wild-type *ETV6* mRNA expression in Kasumi1 cells compared with untransfected cells. Wild-type *ETV6* showed a high level of mRNA expression when compared to untransfected cells.

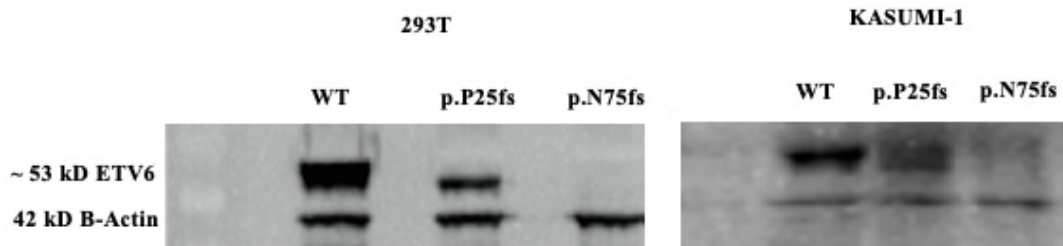


Figure 5: Western blotting analysis showed the expression of wild-type protein (WT) and *ETV6* mutants in Kasumi-1 and 293T cells after 48 hours of transfection. β -actin (42 kDa) served as a positive control of protein load.

The results indicated that transfected cells had a higher protein expression than the untransfected cells. In both cells, the concentration of *ETV6* protein in cells transfected with the mutant *ETV6* p.P25fs decreased as compared to the wild-type *ETV6*. The absence of *ETV6* protein expression in cells transfected with mutant *ETV6* p.N75fs

in both cells implies that this mutant protein truncated the *ETV6* protein as a result of a stop codon termination.

Cell Viability Assay

Figure 6 shows a graph of cell viability assay in a Kasumi-1 cell. Wild-type proteins *ETV6*

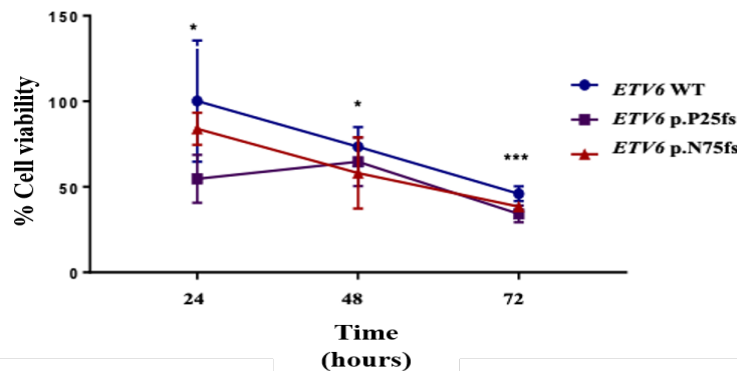


Figure 6: Cell viability assay for *ETV6* plasmid transfection. After 24, 48 and 72 hours of transfection, WT and *ETV6* mutants p.P25fs and p.N75fs, respectively, showed a significant decrease in cell viability compared to non-transfected cells in Kasumi 1 cells. Statistical significance in all cases was measured using Student's t-test (* $p < 0.05$, *** $p < 0.001$), $n = 3$. The error bar represents the average \pm S.D.

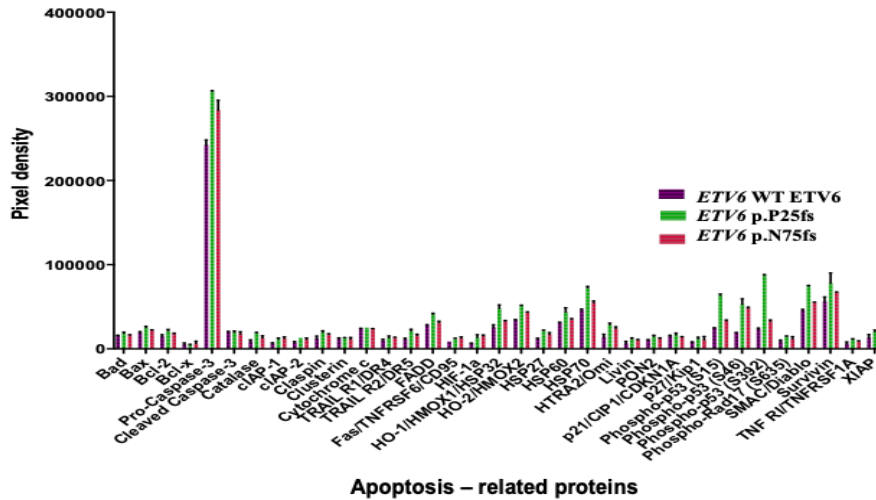


Figure 7 (A). Profile of apoptosis related proteins in 293T cells transfected with *ETV6* WT, mutant *ETV6* p.P25fs and p.N75fs.

(WT) and *ETV6* mutants p.P25fs and p.N75fs exhibited a significant decrease in cell viability after 24 hours, 48 hours and 72 hours.

apoptosis were analysed. Figure 7 (A, B and C) shows the results comparing *ETV6* WT as well as *ETV6* mutants p.P25fs and p.N75fs.

Protein Expression Assay using Proteome Profiler Human Apoptosis Array

Proteome Profiler assays were performed to analyse the expression of apoptosis-related protein profiles in 293T cells transfected with *ETV6* WT and *ETV6* mutants p.P25fs and p.N75fs. A total of 35 proteins that have roles in

Apoptosis Signaling Pathway

From the analysis of the apoptosis proteome profiler, a significant apoptosis-related protein profile in *ETV6* p.P25fs and p.N75fs mutants versus *ETV6* WT was observed in the apoptosis signaling pathway. The profile of the proteins involved is illustrated in Figure 7 (D).

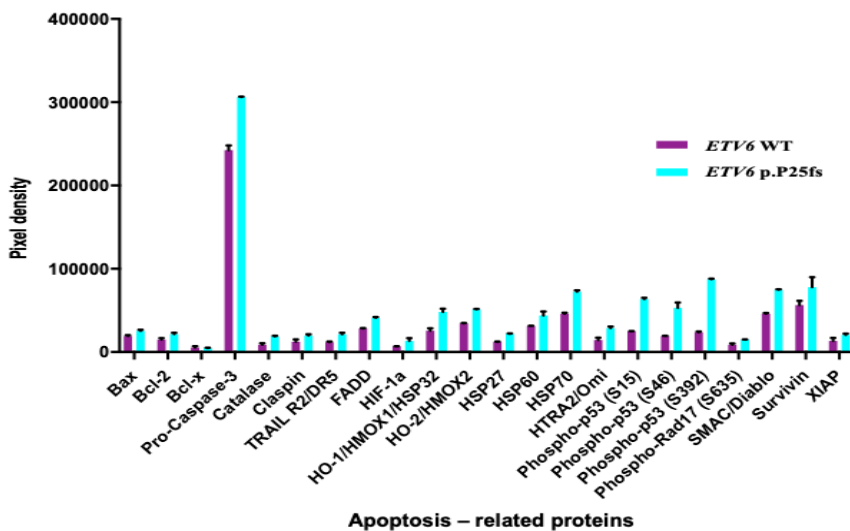


Figure 7 (B). Significantly deregulated 22 apoptosis-related protein ($p < 0.01$) by *ETV6* p.P25fs versus WT (Bax, Bcl-2, Bcl-x, Pro-Caspase-3, Catalase, Claspin, TRAIL R2/DR5, FADD, HIF-1a, HO-1/HMOX1/HSP32, HO-2/HMOX2, HSP27, HSP60, HSP70, HTRA2/Omi, Phospho-p53 (S15), Phospho-p53 (S46), Phospho-p53 (S392), Phospho-Rad17 (S635), SMAC/ Diabolo, Survivin and XIAP).

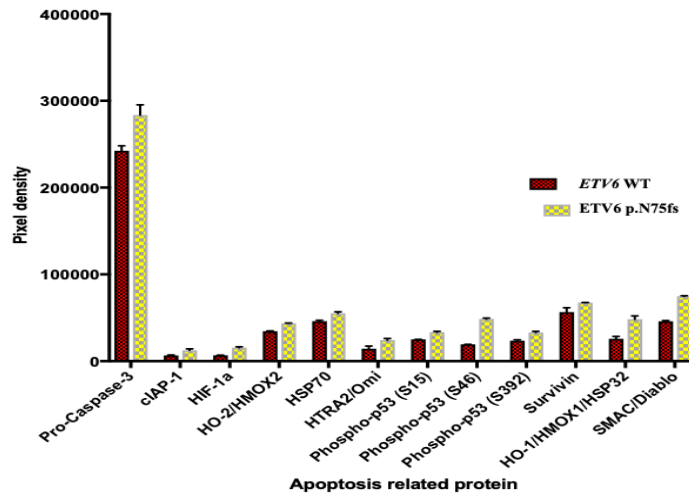


Figure 7 (C). Significantly deregulated 12 apoptosis-associated protein ($p < 0.01$) by *ETV6* p.P75fs versus WT (Pro-Caspase-3, cIAP-1, HIF-1a, HO-2/HMOX2, HSP70, HTRA2/Omi, Phospho-p53 (S15), Phospho-p53 (S46), Phospho-p53 (S392), Survivin, HO-1/HMOX1/HSP32 and SMAC/Diablo).

DISCUSSION

We identified various types of somatic mutations in the genomes of the three relapsed AML patients including synonymous single nucleotide variants (SNVs), missense SNVs, deletions and insertion frameshifts, stopgains and splice sites. A total of 312 somatic gene variants were identified: 154 variants in the diagnosis phase and 158 variants in the relapse phase. After prioritisation, the list was finalised to 81 somatic gene variants, with 46 variants were present at diagnosis (13-17 mutations per patient) and 49 variants present at relapse (12-20 mutations per patient). These findings are similar with previous studies showing that on average there are several hundred mutations in the AML genome yet only a few will have the expected translational effects.^{12,17-21,41}

Each cancer has a unique and distinctive genome profile and AML itself is so heterogenous. It is important that each case is examined separately before group observations are discussed. The AML UL8 patient was a 12-year-old Malay girl, whose cytogenetic analysis showed the t(8;21) (q22; q22). This patient had 59 somatic gene variants at the diagnosis phase and 74 somatic gene variants upon relapse 15 months later. After prioritising the variants for translational effects, we identified 10 somatic gene variants found in both the diagnosis and relapse phases, 3 variants in the diagnosis sample and 7 new somatic gene variants in the relapsed sample. Out of the 20 somatic gene variants found, 3 variants namely *RBM10*, *ZMYM4* and

ETV6 are believed to be potential driver variants responsible in the development of relapse for this patient.

The next patient (UL16) was a Malay boy, having structural abnormalities on chromosomes 2, 3 and 8. He was diagnosed with AML at the age of 14. The leukaemia cell genome of this patient had 51 somatic gene variants at the diagnosis phase and 52 somatic gene variants in the relapsed sample (10 months later). Only three somatic gene variants were found at the diagnostic phase and remained detected in the relapse sample. Interestingly, this patient has a mutation in the same gene found in the UL8 patient, that is the *ETV6* gene. However, the *ETV6* mutation in UL16 was detected at relapse whereas this mutation was found during diagnosis and relapse in UL8. Based on the AML driver gene database, the three somatic gene variants that have the potential to be the driver genes for this patient are *KIT*, *CDC73* and *ETV6*.

The AML UL27 patient was an Indian girl, aged 5 years, with no chromosomal abnormalities. She relapsed after 9 months of diagnosis. The patient's genome had 44 somatic gene variants at the diagnosis phase and only 32 somatic gene variants at relapse. Only one variant of the somatic gene has been identified as a potential driver gene in the development of relapse, the *HNF1A* gene.

The investigation of AML genomic changes from diagnosis to remission and relapse has revealed significant heterogeneity in the AML genome. We identified the variants present at diagnosis which were not detected during relapse,

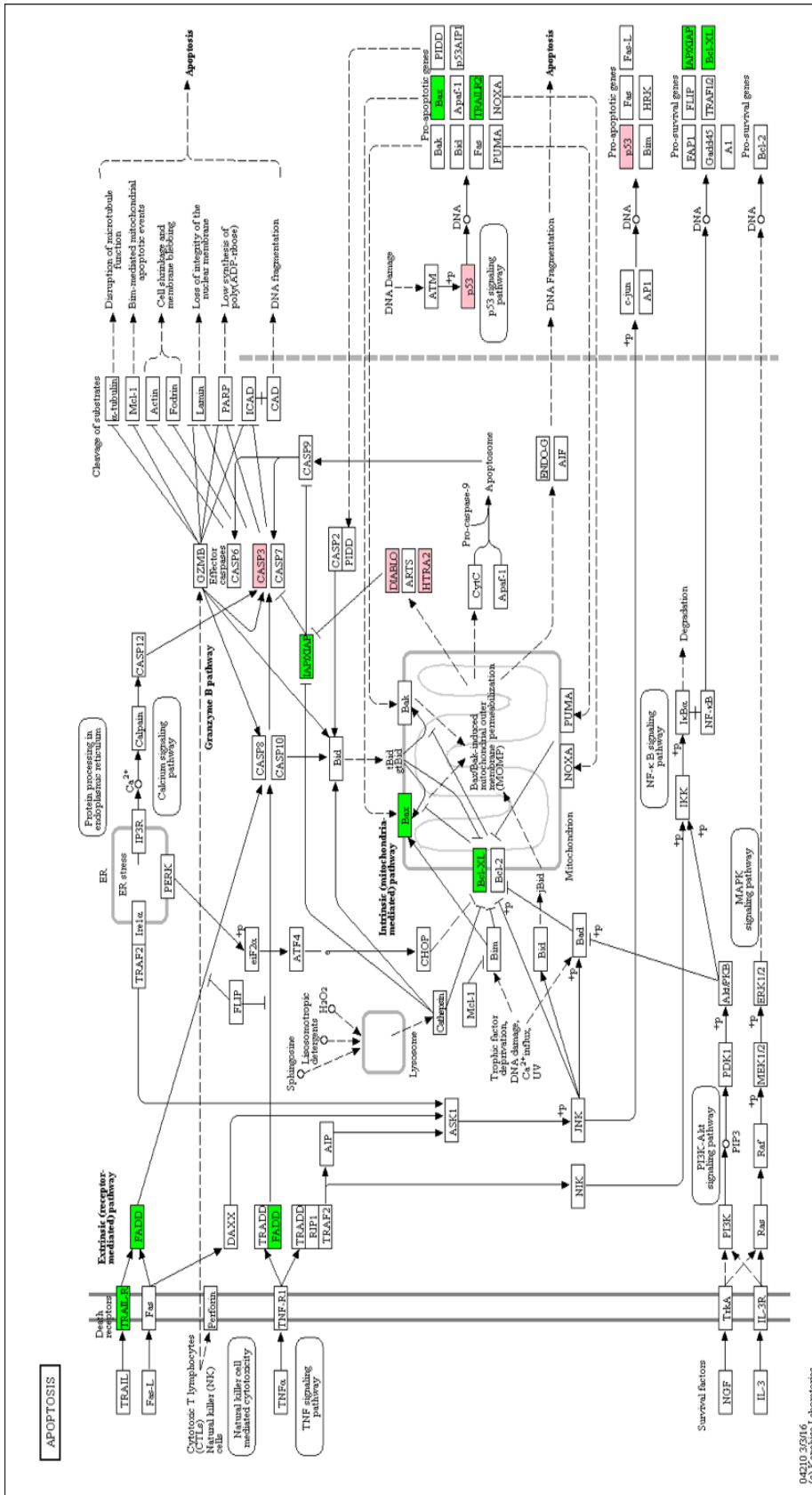


Figure 7 (D). Apoptosis signaling pathways. Significant apoptosis-related protein profiles in ETV6 p.P25fs mutants (green color) and both ETV6 p.P25fs and p.N75fs mutants (pink color)

persistent variants that remained in both phases, as well as new variants that appeared at relapse only. Out of 46 somatic gene variants in the diagnosis samples, only 14 (30%) persisted in the relapse samples. In contrast, of the 49 somatic gene variants identified in the relapse phase, 35 (71%) were newly developed variants that were not found at diagnosis. A total of 81 somatic gene mutations were found in the diagnosis and relapse phases; 32 (40%) were only found in the diagnosis sample, 35 (43%) were only identified when relapse occurred and only 14 (17%) variants were present in both samples.

The findings of this study support the evolutionary model of AML clones by Vosberg and Greif which is based on Darwin's theory of natural selection, also known as survival of the fittest.^{42,43} The population of AML cells in the relapse phase can grow from either a clone or a sub-clone of cell populations that existed at the time of diagnosis, followed by additional mutations. For linear evolution, the cell responsible for the relapse is believed to be part of the major clone at the time of diagnosis. This can be seen in UL8 patient where 10 mutations of the same gene remained from the diagnosis phase to the relapsed phase. As for branched evolution, it is believed that major clones at diagnosis were eliminated but leukaemia cells responsible for the relapse develop from sub-clones at diagnosis. The presence of many new gene mutations during relapse as in UL16 patient (17 genes) and UL27 patient (11 genes) suggests that this branching evolution occurred in them.

Our findings are almost similar to the study by Farrar and colleagues from the United States that used the whole exome sequencing on matched diagnosis, remission and relapse sample in 20 patients with childhood AML. They also found a lower number of gene mutations in the diagnosis sample than in the relapsed sample, and the number of somatic gene mutations persisting from diagnosis to relapse was higher (41%)²³ than in this study (17%). A study from Japan reported a similar pattern, using WES in 4 childhood AML cases, but with fewer number of mutations than our findings. They identified 51 gene mutations with only 5 (10%) gene mutations at diagnosis and 15 (29%) gene mutations when relapse occurred.²⁴

The findings of this study also support the belief that the genomic landscape of patients with childhood and adult AML are distinct, not only in the Caucasian population but in other ethnicities as well. The Cancer Genome Atlas (TCGA)

project involving 200 adult AML de novo patients using WGS (n = 50) and WES (n = 150) found 23 genes with recurrent mutations, including *NPM1*, *FLT3*, *CEBPA*, *DNMT3A*, *IDH1*, and *IDH2*, *TET2*, *RUNX1*, *TP53*, *NRAS*, *WT1*, *PTPN11*, *KIT*, *KRAS*, *PHF6*, *STAG2*, *RAD21*, *FAM5C*, *HNRNP* as well as newly confirmed genes have implications in leukemogenesis namely *EZH2*, *U2AF1*, *SMC3A*, and *SMC*.¹⁹ On the other hand, gene mutations commonly found in adult AML patients such as *DNMT3A* and *TP53* could not be detected in most samples of childhood AML patients, and if any at a very low rate of only 1%.^{12,22,24,44-46} Nevertheless, there are also similar recurrent gene mutations found such as *NPM1*, *FLT3*, *CEBPA*, *WT1* and *KIT*.¹² *FLT3*, *NRAS*, *WT1*, *PTPN11*, *KIT*, *KRAS* and *TET2*²³ as well. *RAD21*, *SMC3*, *STAG2*, *EZH2*, *NRAS*, *KRAS*, *KIT*, *FLT3*, *CEBPA*, *WT1*, *NPM1* and *IDH2*.²⁴ With a limited sample size, our study found recurrent mutations of *ETV6* gene in two patients out of three patients with childhood AML and all cases had no cytogenetic similarities compared to the previous three studies.

Identification of driver genes is crucial in the treatment of AML especially for the development of targeted therapies. Overall, through these three genome profiles, six driver genes that potentially play a role in leukemogenesis and can predict the occurrence of relapse have been successfully identified; *KIT*, *CDC73*, *HNFI1A*, *RBM10*, *ZMYM4* and *ETV6*. Variants that are present in both the diagnosis and persist in the relapse phase are believed to increase the risk of relapse for patients with mutations in this gene. These variants (*PKHD1*, *EZH1*, *ABCG8*, *RYR2*, *SH2B1*, *EDARADD*, *PLXNA1*, *PRR13*, *RBM10*, *TAS2R7* and *ETV6*) were validated in 4 relapsed samples, 28 diagnosis samples and 15 remission samples from other patients with childhood AML. However, all the variants were not detected hence consolidating the belief that AML is really very heterogenous. We believe that biological validation requires a larger sample size to obtain more significant and accurate results.

Our previous study using deep transcriptome sequencing with the same cohort revealed that the most enriched pathway is the transcriptional misregulation in cancer for comparison of relapse versus remission samples.⁴⁷ Therefore, based on the driver gene identification analysis and our previous finding, we performed a functional study to characterise the function of the recurrent mutation in the *ETV6* gene using an in vitro model. *ETV6* also known as Translocating E26

transforming-specific Leukaemia 1 (TEL1) is one of the family to E26-transforming specific (ETS) transcription regulators.^{48,49} The *ETV6* gene is located on chromosome 12p13 which contains 8 exons that encode a protein with 452 amino acids.⁵⁰⁻⁵³ The *ETV6* gene has been shown to play an important role in hematopoiesis as well as vascular development.⁵⁴ Each *ETV6* gene consists of key functional domains namely N-terminal pointed domain (PNT) also known as N-terminal helix-loop-helix (HLH) or sterile alpha motif (SAM), less-conserved central regulatory domain and C-terminal DNA-binding domain (ETS-domain).^{49,55-59} The PNT domain for *ETV6* consists of four α -helices responsible for homodimerization.^{57,60} This domain also contributes to the suppression of target genes.^{55,61}

We had identified mutations in the *ETV6* gene PNT domain specifically p.N75fs. Previous studies have found that mutations of the *ETV6* gene in the PNT domain in various types of blood cancers.⁶²⁻⁶⁶ To our knowledge, the *ETV6* p.N75fs mutation is a novel frameshift insertion mutation for patients with childhood AML. To date, there are two types of mutations that have been found in AML patients, nonsense mutations resulted in severely truncated *ETV6* protein production in E76X and S78X⁶⁷ as well as E115X⁶⁴ and the missense mutation involves specific amino acids such as R105P resulting in misfolding of the *ETV6* protein which in turn interferes with protein multimerization.⁶⁴ Frameshift InDels on the *ETV6* gene within PNT domain were only found in ALL patients V66fs (adult ALL patient)⁶⁸ and Q118fs (paediatric ALL patients).⁶⁵ We also discovered mutation before the PNT *ETV6* domain, p.P25fs, and this novel frameshift insertion mutation was the first to be reported in AML patients. There are missense mutations namely S16T, P25Q, V37M that located before the PNT domain have been reported in ALL patients.⁶⁹ Yet no functional analysis has ever been reported for the effect of mutations on this region of *ETV6* gene.

We demonstrated that mutations in the *ETV6* gene result in the loss of its function as a tumour suppressor gene. This is evidenced by the absence of protein bands on the western blot in the cells transfected with p.N75fs plasmids. The absence of protein bands implies that the *ETV6* protein has been truncated. Similar findings were also reported by previous researchers who examined the effects of frameshift mutations (p. R127fsX207, p. H383fsX389) and nonsense mutations (p. E76X, p.S78X) of this *ETV6* gene

in AML patients.^{67,70} Whereas the frameshift insertion of p.P25fs before the PNT domain showed the presence of lower concentration of protein bands than the wild-type *ETV6* in both transfected cells indicating the protein truncation. This frameshift insertion of p.P25fs has resulted in the loss of *ETV6* gene function as a tumour suppressor.

Evaluation of *ETV6* gene expression found that the relative mRNA expression of both p.P25fs and p.N75fs *ETV6* mutants had significantly reduced gene expression compared to wild-type *ETV6*. *ETV6* p.P25fs mutants had significantly reduced mRNA expression than *ETV6* p.N75fs mutants but it was found that protein truncation was not as complete as in *ETV6* p.N75fs mutants. Steady-state protein abundance is determined by four rates; transcription, translation, mRNA decay and protein decay. The abundance of these proteins can be obtained infinitely from various combinations of these rates.⁷¹⁻⁷³ A recent study looking at simultaneous quantification between mRNA and protein involving 86 loci in 10 genes also found that less than 20% of these loci had a similar effect between mRNA and protein, but a large number of these loci affected proteins instead of mRNA in the same gene. This suggests a complex post-transcription effect on gene expression.⁷⁴

Previous studies have also shown that high expression of the *ETV6* gene in rat erythroleukaemia cells has accelerated the cell differentiation into erythrocytes, with *ETV6* mRNA expression peaked in the first three days of cell differentiation but the presence of dominant *ETV6* negative mutants blocked this effect.⁷⁵ Other researchers examining the role of *ETV6* in the regulation of red blood cell maturation also showed that morphant *ETV6* reduces *gatal* expression but enhances erythroid cell differentiation and elevates erythroblasts in the blood.⁷⁶ Moreover, loss of *ETV6* gene expression due to deletion in wild-type *ETV6* was also found to contribute to leukamogenesis in ALL paediatric patients.⁷⁷

The hematopoietic system needs to maintain its balance between apoptosis and cell proliferation, as changes in protein expression will disrupt this balance and trigger the accumulation of malignant cells.⁷⁸⁻⁸⁰ Analysis of leukaemia cell viability under the influence of *ETV6* mutants p.P25fs and p.N75fs exhibited a significant rate of decrease in cell viability compared to non-transfected cells. These findings further strengthened the fact that the *ETV6* gene acts

as a tumour suppressor gene. Yamagata and colleagues noted that *ETV6* causes disruption to cell proliferation by blocking the cell cycle and triggering apoptosis via mitochondrial intrinsic pathways.⁸¹

Proteome profiler apoptosis assays revealed 22 and 12 apoptosis-related proteins regulated by p.P25fs and p.N75fs *ETV6* mutant respectively. There are several significant proteins regulated by p.P25fs and p.N75fs *ETV6* mutants involved in apoptosis signaling pathways namely p53, TRAIL R2, FADD, Bcl-2, Bcl-x1, Bax, Pro-Caspase -3, Diablo, HTRA2 and XIAP. Caspase (cysteine-aspartic proteases) is a key protein responsible for apoptosis function. Both *ETV6* p.P25fs and *ETV6* p.N75fs mutants significantly regulated pro-caspase 3 compared to WT *ETV6*. The activation of this executioner caspase will initiate a cascade of events that results in DNA fragmentation with endonuclease activation, destruction of nuclear and cytoskeleton proteins, protein cross-linking, ligand expression for phagocytic cells and formation of apoptotic bodies.^{82,83}

The intrinsic signaling pathway of apoptosis is also regulated by the B cell lymphoma gene 2, BCL-2, a group of proteins that inhibit cytochrome C production from mitochondria and can be classified as pro-apoptotic and anti-apoptotic proteins.^{79,84} The *ETV6* p.P25fs mutant produced significantly higher BCL-2 and BAX protein expression profiles than wild-type *ETV6*, indicating that this mutant regulates pro-apoptotic proteins. The increase in BCL-2 expression correlated with resistance to chemotherapy in AML patients.^{85,86} In addition to the BCL-2 protein, apoptosis is also inhibited in blood cancers by overexpression of inhibitors of apoptosis (IAP) family proteins that play a role in blocking the terminal caspase responsible for apoptosis activation.⁸⁷ IAP expression is correlated with poor prognosis, where cIAP expression in AML patients is able to predict poor overall survival.⁸⁸ Moreover, high XIAP expression is also associated with poor prognosis in patients with childhood AML.^{79,89,90} These findings also support the results of our study which indicated that childhood AML patients with *ETV6* mutation experienced relapsed when expressing significant level of XIAP and cIAP proteins.

The discovery of this *ETV6* gene mutation in patients with childhood AML had very different implications compared to patients with childhood ALL. Translocation of t(12;

21) (p13.1; q22) of *ETV6*-*RUNX1* is the most common genetic aberration found in patients with childhood ALL covering 25% of cases with precursor-B phenotype.⁹¹ However, patients with childhood ALL with *ETV6*-*RUNX1* have a favourable prognosis with long-term remission rate and a good survival rate.⁹² Most patients with this genetic aberration will experienced relapse between 10 and 20 years after cessation of treatment.^{93,94} In contrast, the mutation in *ETV6* gene increased the likelihood of relapse in patients with childhood AML.

CONCLUSION

This research profiled the mutational landscape of three patients with childhood AML using WGS on matched trio samples of diagnosis, remission and relapse. Our research has successfully revealed the genomic landscape from the disease diagnosis to relapse. There are genetic differences between local patients with childhood AML from other populations. Given that AML genome heterogeneity is a critical issue in the treatment of AML patients, the genetic variant profile data obtained from this study will have implications in relapse risk classification and development of genomic based therapies. The discovery of a recurrent mutation of relapse driver gene that is the *ETV6* in the genome of these patients with childhood AML will contribute to its exploitation as a future targeted therapy.

Acknowledgments: This work was funded by National University of Malaysia Dana Impak Perdana. We are grateful to all the patients for participation in this study. We also thank our colleagues at Hospital Canselor Tuanku Muhriz and Institute of Paediatrics, Hospital Kuala Lumpur for their support throughout the study.

Authors' contributions: HAZ: Conceptualization, Methodology, Investigation, Software and analysis, Writing. NSAM: Conceptualization, Methodology, Software and analysis, Review and editing. HA: Conceptualization, Review and editing. RJ: Conceptualization, Methodology, Resources & Funding, Review and editing. All authors of this paper have read and approved the final submitted version.

Conflict of interest: The authors declare no conflict of interest.

REFERENCES

1. Döhner H, Daniel J W, Clara D B. Acute Myeloid Leukemia. *N Engl J Med* [Internet]. 2015;373(12):1136–52.
2. Meshinchi S, Arceci RJ. Prognostic factors and risk-based therapy in pediatric acute myeloid leukemia. *Oncologist*. 2007;12(3):341–55.
3. Pui CH, Carroll WL, Meshinchi S, Arceci RJ. Biology, risk stratification, and therapy of pediatric acute leukemias: an update. *J Clin Oncol*. 2011;29(5):551–65.
4. Jay Y, Schiffer CA. Genetic biomarkers in acute myeloid leukemia: will the promise of improving treatment outcomes be realized? *Expert Rev Hematol*. 2012;5(4):395–407.
5. Tasian SK, Pollard JA, Aplenc R. Molecular therapeutic approaches for pediatric acute myeloid leukemia. *Front Oncol*. 2014;4:55.
6. Tarlock K, Meshinchi S. Pediatric acute myeloid leukemia: Biology and therapeutic implications of genomic variants. *Pediatr Clin North Am* [Internet]. 2015;62(1):75–93.
7. Wook Jae L, Bin C. Prognostic factors and treatment of pediatric acute lymphoblastic leukemia. *Korean J Pediatr*. 2017;60(5):129–37.
8. Tomizawa D, Kiyokawa N. Acute Lymphoblastic Leukemia. In: *Hematological Disorders in Children* [Internet]. Singapore: Springer Singapore; 2017. p. 33–60.
9. Gami AS, Alonzo TA, Perentesis JP, Meshinchi S. Children's Oncology Group's 2013 Blueprint for Research: Acute Myeloid Leukemia. *Pediatr Blood Cancer*. 2012;60(6):964–71.
10. Rubnitz JE. Current Management of Childhood Acute Myeloid Leukemia. *Pediatr Drugs*. 2017;19(1):1–10.
11. Im HJ. Current treatment for pediatric acute myeloid leukemia. *Blood Res*. 2018;53(1):1.
12. Bolouri H, Farrar JE, Triche T, Ries RE, Lim EL, Alonzo TA, *et al*. The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions. *Nat Med*. 2018;24(1):103–12.
13. Moore a S, Kearns PR, Knapper S, Pearson a DJ, Zwaan CM. Novel therapies for children with acute myeloid leukaemia. *Leukemia*. 2013;27(7):1451–60.
14. Rubnitz JE, Inaba H, Leung WH, Pounds S, Cao X, Campana D, *et al*. Definition of cure in childhood acute myeloid leukemia. *Cancer*. 2014;120(16):2490–6.
15. Nasir A, Zahari NF, Taib F, Mohamad N. Survival of Children with Acute Leukaemia: A Single Centre Experience. *Malaysian J Paediatr Child Heal*. 2020;26(1):4–13.
16. Graubert TA, Mardis ER. Genomics of acute myeloid leukemia. *Cancer J*. 2011;17(6):487–91.
17. Ley TJ, Ding L, Walter MJ, McLellan MD, Lamprecht T, Larson DE, *et al*. DNMT3A mutations in acute myeloid leukemia. *N Engl J Med*. 2010;363(25):2424–33.
18. Mardis ER, Ding L, Dooling DJ, Larson DE, McLellan MD, Chen K, *et al*. Recurring mutations found by sequencing an acute myeloid leukemia genome. *N Engl J Med*. 2009;361(11):1058–66.
19. Cancer Genome Atlas Research N. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med*. 2013;368(22):2059–74.
20. Ding L, Ley TJ, Larson DE, Miller CA, Koboldt DC, Welch JS, *et al*. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature*. 2012;481(7382):506–10.
21. Welch J, Ley T, Link D, Miller C. The origin and evolution of mutations in acute myeloid leukemia. *Cell*. 2012;150(2):264–78.
22. Ho PA, Kutny MA, Alonzo TA, Gerbing RB, Joaquin J, Raimondi SC, *et al*. Leukemic mutations in the methylation-associated genes DNMT3A and IDH2 are rare events in pediatric AML: A report from the Children's Oncology Group. *Pediatr Blood Cancer*. 2011;57(2):204–9.
23. Farrar JE, Schuback HL, Ries RE, Wai D, Hampton OA, Trevino LR, *et al*. Genomic profiling of pediatric acute myeloid leukemia reveals a changing mutational landscape from disease diagnosis to relapse. *Cancer Res*. 2016;76(8):2197–205.
24. Shiba N, Yoshida K, Shiraiishi Y, Okuno Y, Yamato G, Hara Y, *et al*. Whole-exome sequencing reveals the spectrum of gene mutations and the clonal evolution patterns in paediatric acute myeloid leukaemia. *Br J Haematol*. 2016;175(3):476–89.
25. Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P, Roberts ND, *et al*. Genomic Classification and Prognosis in Acute Myeloid Leukemia. *N Engl J Med*. 2016;374(23):2209–21.
26. Patel JP, Gönen M, Figueroa ME, Fernandez H, Sun Z, Racevskis J, *et al*. Prognostic relevance of integrated genetic profiling in acute myeloid leukemia. *N Engl J Med*. 2012;366(12):1079–89.
27. Sheng Li, Francine E Garrett-Bakelman, Stephen S Chung, Mathijs A Sanders, Todd Hricik, Franck Rapaport, *et al*. Distinct evolution and dynamics of epigenetic and genetic heterogeneity in acute myeloid leukemia. *Nat Med*. 2016;22:792–799.
28. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, *et al*. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol*. 2013;31:213–219.
29. Saunders CT, Wong WSW, Swamy S, Becq J, Murray LJ, Cheetham RK, Strelka: Accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics*. 2012;28(14):1811–7.
30. Rausch T, Zichner T, Schlattl A, Stütz AM, Benes V, Korbel JO. DELLY: Structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics*. 2012;28(18):i333–i339.
31. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW. Cancer genome landscapes. *Science* (80-). 2013;339(6127):1546–1558.
32. Kandath C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, *et al*. Mutational landscape and significance across 12 major cancer types. *Nature*. 2013;502(7471):333–9.
33. Tamborero D, Gonzalez-Perez A, Perez-Llomas C, Deu-Pons J, Kandath C, Reimand J, *et al*. Comprehensive identification of mutational cancer

- driver genes across 12 tumor types. *Sci Rep*. 2013;3:Article number: 2650.
34. Robinson PN, Tomasz Z *et al*. Integrative genomics viewer (IGV): Visualizing alignments and variants. In: *Computational Exome and Genome Analysis*. 2017. p. Chapter 17:1-13.
 35. Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: Predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res*. 2012;40(Web Server issue):W452-7.
 36. Adzhubei I, Schmidt S, Peshkin L, Ramensky V, Gerasimova A, Bork P, *et al*. PolyPhen-2 : prediction of functional effects of human nsSNPs. *Curr Protoc Hum Genet*. 2010;
 37. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet*. 2013;Chapter 7:Unit7.20.
 38. Chun S, Fay JC. Identification of deleterious mutations within three human genomes. *Genome Res*. 2009;19(9):1553–1561.
 39. Schwarz JM, Rödlersperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods*. 2010;7(8):575–6.
 40. Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: Application to cancer genomics. *Nucleic Acids Res*. 2011;39(17):37–43.
 41. Walter MJ, Shen D, Ding L, Shao J, Koboldt DC, Chen K, *et al*. Clonal Architecture of Secondary Acute Myeloid Leukemia. *N Engl J Med*. 2012;366:1090–8.
 42. Vosberg S, Greif PA. Clonal evolution of acute myeloid leukemia from diagnosis to relapse. *Genes Chromosom Cancer*. 2019;58(12):839–49.
 43. Darwin C. *On the Origin of the Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. London: John Murray. 1859.
 44. Farrar JE, Schuback HL, Ries RE, Wai D, Hampton OA, Trevino LR, *et al*. Genomic Profiling of Pediatric Acute Myeloid Leukemia Reveals a Changing Mutational Landscape from Disease Diagnosis to Relapse. 2016;
 45. Liang DC, Liu HC, Yang CP, Jaing TH, Hung IJ, Yeh TC, *et al*. Cooperating gene mutations in childhood acute myeloid leukemia with special reference on mutations of ASXL1, TET2, IDH1, IDH2, and DNMT3A. *Blood*. 2013;121(15):2988–95.
 46. Thol F, Heuser M, Damm F, Klusmann JH, Reinhardt K, Reinhardt AD. DNMT3A mutations are rare in childhood acute myeloid leukemia. *Haematologica*. 2011;96(8):1238–9.
 47. Osman SH, Abu N, Aziz H, Chow YP, Wan Mohamad Nazarie WF, Ab Mutalib NS, *et al*. Deep Transcriptome Sequencing of Pediatric Acute Myeloid Leukemia Patients at Diagnosis, Remission and Relapse: Experience in 3 Malaysian Children in a Single Center Study. *Front Genet*. 2020;11(February):1–5.
 48. Degnan BM, Degnan SM, Naganuma T, Morse DE. The ets multigene family is conserved throughout the metazoa. *Nucleic Acids Res*. 1993;21(15):3479–3484.
 49. Lopez RG, Carron C, Oury C, Gardellino P, Bernard O, Ghysdael J. TEL is a sequence-specific transcriptional repressor. *J Biol Chem*. 1999;274:30132–30138.
 50. Baens M, Peeters P, Guo C, Aerssens J, Marynen P. Genomic organization of TEL: The human ETS-variant gene 6. *Genome Res*. 1996;
 51. Hart SM, Foroni L. Core binding factor genes and human leukemia. *Haematologica*. 2002;87(12):1307–23.
 52. Bohlander SK. ETV6: A versatile player in leukemogenesis. *Semin Cancer Biol*. 2005;15(3):162–74.
 53. Zhou F, Chen B. Acute myeloid leukemia carrying ETV6 mutations: biologic and clinical features. *Hematology*. 2018;23(9):608–12.
 54. Pereira CF, Chang B, Qiu J, Niu X, Papatsenko D, Hendry CE, *et al*. Induction of a hemogenic program in mouse fibroblasts. *Cell Stem Cell*. 2013;13:205–18.
 55. Chakrabarti SR, Sood R, Nandi S, Nucifora G. Posttranslational modification of TEL and TEL/AML1 by SUMO-1 and cell-cycle-dependent assembly into nuclear bodies. *Proc Natl Acad Sci U S A*. 2000;97(24):13281–5.
 56. Gu X, Shin BH, Akbarali Y, Weiss A, Boltax J, Oettgen P, *et al*. Tel-2 is a Novel Transcriptional Repressor Related to the Ets Factor Tel/ETV-6. *J Biol Chem*. 2001;276(12):9421–36.
 57. Poirel H, Lopez RG, Lacronique V, Valle V Della, Mauchauffé M, Berger R, *et al*. Characterization of a novel ETS gene, TELB, encoding a protein structurally and functionally related to TEL. *Oncogene*. 2000;19:4802–4806.
 58. Potter MD, Buijs A, Kreider B, Van Rompaey L, Grosveld GC. Identification and characterization of a new human ETS-family transcription factor, TEL2, that is expressed in hematopoietic tissues and can associate with TEL1/ETV6. *Blood*. 2000;95(11):3341–8.
 59. Golub TR, Barker GF, Stegmaier K, Gilliland DG. The TEL gene contributes to the pathogenesis of myeloid and lymphoid leukemias by diverse molecular genetic mechanisms. *Curr Top Microbiol Immunol*. 1997;220(2):67–79.
 60. Golub TR, Goga A, Barker GF, Afar DE, McLaughlin J, Bohlander SK, *et al*. Oligomerization of the ABL tyrosine kinase by the Ets protein TEL in human leukemia. *Mol Cell Biol*. 1996;16:4107–4116.
 61. Irvin BJ, Wood LD, Wang L, Fenrick R, Sansam CG, Packham G, *et al*. TEL, a Putative Tumor Suppressor, Induces Apoptosis and Represses Transcription of Bcl-XL. *J Biol Chem*. 2003;278:46378–46386.
 62. Van Waalwijk Van Doorn-Khosrovani SB, Spensberger D, De Knecht Y, Tang M, Löwenberg B, Delwel R. Somatic heterozygous mutations in ETV6 (TEL) and frequent absence of ETV6 protein in acute myeloid leukemia. *Oncogene*. 2005;24:4129–4137.
 63. Silva FPG, Morolli B, Storlazzi CT, Zagaria A, Impera L, Klein B, *et al*. ETV6 mutations and loss in AML-M0. *Leukemia*. 2008;22(8):1639–43.

64. Wang Q, Dong S, Yao H, Wen L, Qiu H, Qin L, *et al.* ETV6 mutation in a cohort of 970 patients with hematologic malignancies. *Haematologica*. 2014;99(10):e176–8.
65. Zhang J, Mullighan CG, Harvey RC, Wu G, Chen X, Edmonson M, *et al.* Key pathways are frequently mutated in high-risk childhood acute lymphoblastic leukemia: A report from the Children's Oncology Group. *Blood*. 2011;118:3080–3087.
66. Zhang J, Walsh MF, Wu G, Edmonson MN, Gruber TA, Easton J, *et al.* Germline Mutations in Predisposition Genes in Pediatric Cancer. *N Engl J Med*. 2015;373(24):2336–46.
67. Barjesteh Van Waalwijk Van Doorn-Khosrovani S, Spensberger D, De Knecht Y, Tang M, Löwenberg B, Delwel R. Somatic heterozygous mutations in ETV6 (TEL) and frequent absence of ETV6 protein in acute myeloid leukemia. *Oncogene*. 2005;24(25):4129–37.
68. Vlierberghe P Van, Ambesi-Impiombato A, Perez-Garcia A, Haydu JE, Rigo I, Hadler M, *et al.* ETV6 mutations in early immature human T cell leukemias. *J Exp Med*. 2011;208:2571–2579.
69. Moriyama T, Metzger ML, Wu G, Nishii R, Qian M, Devidas M, *et al.* Germline genetic variation in ETV6 and risk of childhood acute lymphoblastic leukaemia: A systematic genetic study. *Lancet Oncol*. 2015;16:1659–1666.
70. Hock H, Shimamura A. ETV6 in Hematopoiesis and Leukemia Predisposition. *Semin Hematol*. 2017;54(2):98–104.
71. Hausser J, Mayo A, Keren L, Alon U. Central dogma rates and the trade-off between precision and economy in gene expression. *Nat Commun*. 2019;10(1).
72. Li GW, Burkhardt D, Gross C, Weissman JS. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell*. 2014;157(3):624–35.
73. Hargrove JL, Schmidt FH. The role of mRNA and protein stability in gene expression. *FASEB J*. 1989;3(12):2360–70.
74. Brion C, Lutz SM, Albert FW. Simultaneous quantification of mRNA and protein in single cells reveals post-transcriptional effects of genetic variation. *Elife*. 2020;9:1–34.
75. Waga K, Nakamura Y, Maki K, Arai H, Yamagata T, Sasaki K, *et al.* Leukemia-related transcription factor TEL accelerates differentiation of Friend erythroleukemia cells. *Oncogene*. 2003;22:59–68.
76. Rasighaemi P, Onnebo SMN, Liongue C, Ward AC. ETV6 (TEL1) regulates embryonic hematopoiesis in zebrafish. *Haematologica*. 2015;100:23–31.
77. Patel N, Goff LK, Clark T, Ford AM, Foot N, Lillington D, *et al.* Expression profile of wild-type ETV6 in childhood acute leukaemia. *Br J Haematol*. 2003;122(1):94–8.
78. Reed JC. Bcl-2-family proteins and hematologic malignancies: History and future prospects. *Blood*. 2008;111:3322–3330.
79. Shadia Z, Wang R, Gandhi V. Targeting the Apoptosis Pathway. *Leuk Lymphoma*. 2014;55(9):1980–92.
80. D'Arcy MS. Cell death: a review of the major forms of apoptosis, necrosis and autophagy. *Cell Biol Int*. 2019;43(6):582–92.
81. Yamagata T, Maki K, Waga K, Mitani K. TEL/ETV6 induces apoptosis in 32D cells through p53-dependent pathways. *Biochem Biophys Res Commun*. 2006;347(2):517–26.
82. Martinvalet D, Zhu P, Lieberman J. Granzyme A induces caspase-independent mitochondrial damage, a required first step for apoptosis. *Immunity*. 2005;22(3):355–70.
83. Poon IKH, Lucas CD, Rossi AG, Ravichandran KS. Apoptotic cell clearance: Basic biology and therapeutic potential. *Nat Rev Immunol*. 2014;14(3):166–80.
84. Cory S, Adams JM. The BCL2 family: Regulators of the cellular life-or-death switch. *Nat Rev Cancer*. 2002;2(9):647–56.
85. Campos L, Rouault JP, Sabido O, Oriol P, Roubi N, Vasselon C, *et al.* High expression of bcl-2 protein in acute myeloid leukemia cells is associated with poor response to chemotherapy. *Blood*. 1993;81:3091–3096.
86. Radha G, Raghavan SC. BCL2: A promising cancer therapeutic target. *Biochim Biophys Acta - Rev Cancer*. 2017;1868(1):309–14.
87. Fulda S, Vucic D. Targeting IAP proteins for therapeutic intervention in cancer. *Nat Rev Drug Discov*. 2012;11:109–124.
88. Hess CJ, Berkhof J, Denkers F, Ossenkoppele GJ, Schouten JP, Oudejans JJ, *et al.* Activated intrinsic apoptosis pathway is a key related prognostic parameter in acute myeloid leukemia. *J Clin Oncol*. 2007;25:1209–1215.
89. Sung KW, Choi J, Hwang YK, Lee SJ, Kim HJ, Kim JY, *et al.* Overexpression of X-linked inhibitor of apoptosis protein (XIAP) is an independent unfavorable prognostic factor in childhood de novo acute myeloid leukemia. *J Korean Med Sci*. 2009;24:605–613.
90. Tamm I, Richter S, Oltersdorf D, Creutzig U, Harbott J, Scholz F, *et al.* High expression levels of X-linked inhibitor of apoptosis protein and survivin correlate with poor overall survival in childhood de novo acute myeloid leukemia. *Clin Cancer Res*. 2004;10:3737–3744.
91. Pui CH, Relling M V., Downing JR. Acute lymphoblastic leukemia. *N Engl J Med*. 2004;350:1535–48.
92. Loh ML, Goldwasser MA, Silverman LB, Poon WM, Vattikuti S, Cardoso A, *et al.* Prospective analysis of TEL/AML1-positive patients treated on Dana-Farber Cancer Institute Consortium Protocol 95-01. *Blood*. 2006;107(11):4508–13.
93. Forestier E, Heyman M, Andersen MK, Autio K, Blennow E, Borgström G, *et al.* Outcome of ETV6/RUNX1-positive childhood acute lymphoblastic leukaemia in the NOPHO-ALL-1992 protocol: Frequent late relapses but good overall survival. *Br J Haematol*. 2008;140(6):665–72.
94. Sun C, Chang L, Zhu X. lymphoblastic leukemia and mechanisms underlying its relapse. *Oncotarget*. 2017;8(21):35445–59.