Germline bi-allelic *SH2B3*/LNK alteration predisposes to a neonatal juvenile myelomonocytic leukemia-like disorder

Chloé Arfeuille,[1,2] Yoann Vial,[1,2] Margaux Cadenet,[1,2] Aurélie Caye-Eude,[1,2] Odile Fenneteau,[3] Quentin Neven,[4] Adeline A. Bonnard,[1,2] Simone Pizzi,[5] Giovanna Carpentieri,[5] Yline Capri,[6] Katia Girardi,[7] Lucia Pedace,[7] Marina Macchiaiolo,[8] Kamel Boudhar,[9] Monia ben Khaled,[10] Wadih Abou Chahla,[11] Anne Lutun,[12] Mony Fahd,[4] Séverine Drunat,[1] Elisabetta Flex,[13] Jean-Hugues Dalle,[4] Marion Strullu,[2,4] Franco Locatelli,[7,14] Marco Tartaglia[5] and Hélène Cavé[1,2]

[1]Département de Génétique, Unité de Génétique Moléculaire, Hôpital Robert Debré, Assistance Publique des Hôpitaux de Paris (AP-HP), Paris, France; [2]INSERM UMR_S1131, Institut de Recherche Saint-Louis, Université Paris-Cité, Paris, France; [3]Service d'Hématologie Biologique, Hôpital Robert Debré, Assistance Publique des Hôpitaux de Paris (AP-HP), Paris, France; [4]Service d'Onco-Hématologie Pédiatrique, Hôpital Robert Debré, Assistance Publique des Hôpitaux de Paris (AP-HP), Paris, France; [5]Molecular Genetics and Functional Genomics, Bambino Gesù Children's Hospital IRCCS, Rome, Italy; [6]Département de Génétique, Unité de Génétique Clinique, Hôpital Robert Debré, Assistance Publique des Hôpitaux de Paris (AP-HP), Paris, France; [7]Department of Hematology/Oncology and Cell and Gene Therapy, Bambino Gesù Children's Hospital IRCCS, Rome, Italy; [8]Rare Diseases and Medical Genetics, Bambino Gesù Children's Hospital IRCCS, Rome, Italy; [9]Service de Réanimation Néonatale, Hôpital Central de l'Armée, Alger, Algeria; [10]Faculty of Medicine, University of Tunis El Manar and Pediatric Immunohematology Unit, Bone Marrow Transplantation Center Tunis, Tunis, Tunisia; [11]Service d'Hématologie Pédiatrique, Centre Hospitalier Universitaire de Lille, Lille, France; [12]Service d'Hématologie Pédiatrique, Centre Hospitalier Universitaire d'Amiens, Amiens, France; [13]Department of Oncology and Molecular Medicine, Istituto Superiore di Sanità, Rome, Italy and [14]Department of Pediatrics, Catholic University of the Sacred Heart, Rome, Italy

Correspondence: Hélène Cavé
helene.cave@aphp.fr

Received: July 13, 2023.
Accepted: November 7, 2023.
Early view: November 16, 2023.

https://doi.org/10.3324/haematol.2023.283917

SUPPLEMENTARY MATERIAL

Germline bi-allelic *SH2B3* alteration predisposes to a neonatal juvenile myelomonocytic leukemia-like disorder

Chloé Arfeuille[1,2], Yoann Vial[1,2], Margaux Cadenet[1,2], Aurélie Caye-Eude[1,2], Odile Fenneteau[3], Quentin Neven[4], Adeline A Bonnard[1,2], Simone Pizzi[5], Giovanna Carpentieri[5], Yline Capri[6], Katia Girardi[7], Lucia Pedace[7], Marina Macchiaiolo[8], Kamel Boudhar[9], Monia ben Khaled[10], Wadih Abou Chahla[11], Anne Lutun[12], Mony Fahd[4], Séverine Drunat[1], Elisabetta Flex[13], Jean-Hugues Dalle[4], Marion Strullu[2,4], Franco Locatelli[11,14], Marco Tartaglia[5], Hélène Cavé[1,2]

Supplementary Material and Methods

Genomic DNA and RNA extraction

Genomic DNA was extracted using a QIAamp DNA Mini Kit (Qiagen Gmbh). DNA concentration was measured using a Varioskan LUX (Thermo Scientific). RNA was extracted using the RNeasy Mini or Micro Kit (Qiagen); cDNA was obtained by reverse transcription of 1 µg RNA with random hexamers.

Targeted sequencing

A panel of genes implicated in JMML (*AEBP2, ASXL1, BCORL1, BRAF, CBL, CDC42 .2, CDCA5, CDKN2A, CDKN2B, CDYL, CREBBP, DNMT3A, DOCK2, EED, EP300, ESCO1, ESCO2, EZH2, GATA2, HDAC8, HRAS, IKZF1, JAK3, JARID2, KRAS, LZTR1, MAP2K1, MAP2K2, MAU2, NF1, NIPBL, NPM1, NRAS, PDE8A, PDS5A, PDS5B, PLXNB2, PPP1CB, PTPN11, RAC2, RAD21, RAF1, RASA2, RBBP4, RBBP7, RIT1, RRAS, RRAS2, SETBP1, SF3B1, SH2B3, SH3BP1, SHOC2, SMC1A, SMC3, SOS1, SOS2, SPRED1, SRSF2, STAG1, STAG2, SUZ12, TP53, U2AF1, WAPAL, WT1, ZRSR2)* was analyzed by parallel sequencing (supplementary information), using capture-based target enrichment (Custom SureSelect XT-HS2, Agilent) and sequencing on a NextSeq 500 (High Output Kit v2, 2*150bp) (Illumina). Bioinformatical alignment was performed using the Pipeline Local Run Manager v.2.4.0 (Illumina). Variant calling was performed using VarScan v.2.3.5. Variant classification was conducted using Alissa Interpret (Agilent technologies). Average sequencing depth was 1500X. Variants were filtered based on >10 reads and an allelic frequency (VAF) >5%.

Whole genome sequencing analysis:

Laboratory: PCR-free libraries were prepared with NEBNext Ultra II DNA Library Prep Kit (Illumina) according to supplier recommendations and sequenced on an Illumina NovaSeq6000 as paired reads of 150 nucleotides.

Base calling was performed using Illumina Real Time Analysis with default parameters. Reads were mapped to the human genome build (hg38) using the Burrows-Wheeler Aligner (BWA) tool. Variant and copy number alteration (CNA) calling is described in supplementary data. Prediction of the functional and clinical consequences of variants was performed by CADD (GRCh38-v1.6)[34] and REVEL[35]. Previous occurrence in the ClinVar (https://www-ncbi-nlm-nih-gov/clinvar) and COSMIC (v98) (https://cancer.sanger.ac.uk/cosmic) database were checked. Minimal allele frequencies (MAF) were obtained from GnomAD version 2.1.1 (non cancer).

Variant calling: For constitutional DNA, identification of Single Nucleotide Variations (SNV) and small insertions/deletions (up to 20bp), was performed via the Broad Institute's GATK Haplotype Caller GVCF tool. Detection of germline SNV was performed on paired analysis enabling mutation enrichment determination. For somatic DNA, Broad Institute's MuTect tool (2.0, max_alt_alleles_in_normal_count=2; max_alt_allele_in_normal_fraction=0.04) was used for somatic DNA. An IntegraGen (Evry, France) post-processing, developed to filter out candidate somatic mutations which are more consistent with artifacts or germline mutations, is applied. Only somatic mutations considered as PASS, t_lod_fstar, alt_allele_in_normal, germline_risk and panel_of_normals were retained. A somatic score ranging from 1 to 30 was calculated for each mutation, when a normal sample was available, with a score of 30 translating to the highest confidence index. This score takes into account both counts and the frequencies of mutated allele in both samples in order to minimize false positive variations. Also, the mutations with a QSS score below 20 and a VAF of tumor <0.02 are removed.

CNA calling: GATK (V4.1.4.1) was used to investigate genomic CNA by comparing tumor DNA exome data to a reference sample pool and by comparing B-allele frequency (BAF) with the distribution of variant allele frequencies of SNPs reported in GnomAD. All changes were annotated with AnnotSV to provide a comprehensive summary of structural variation in the human genome.

Bioinformatics: Quality of reads was assessed for each sample using FastQC (v.0.11.4; http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). RNA-SeQC provided key measures of data quality. These metrics are shown within Reporting and include yield, alignment and duplication rates, rRNA content, regions of alignment (exon, intron and intragenic). Alignment is performed by STAR (https://github.com/alexdobin/STAR). The duplicate reads (e.g., paired-end reads in which the insert DNA molecules have identical start and end locations in the Human genome) are removed (sambamba tools). To detect fusion-genes candidates in RNA-seq data, FusionCatcher (Start with fastq files) and Star-Fusion (Start with alignment files) are used in order to achieve higher detection efficiency. These were run using defaults configurations.

Constructs

The c.1709dupA (p.Asp570Lysfs82) was introduced in an Xpress-tagged *SH2B3* cDNA (NM_005475.3) cloned in pcDNA6/HisC (Invitrogen) by site-directed mutagenesis (QuikChange XL kit, Agilent Technologies). Constructs were verified by direct sequencing.

Evaluation of protein stability

293T cell lines were cultured in Dulbecco's modified Eagle's medium supplemented with 10% heat-inactivated fetal bovine serum (FBS) (EuroClone) and 1% penicillin-streptomycin, at 37 °C, 5% CO_2. Cells were seeded in 6-well plates the day before transfection and then transfected at 70% confluency with Fugene 6 (Promega) with constructs of interest. 48 hours after transfection, cells were treated with cycloheximide (20 mg/mL) for 2, 3 and 4 hours or left untreated before lysis. Immunoblotting was performed so that Xpress-tagged *SH2B3* levels could be assessed using a monoclonal Xpress antibody (Invitrogen). Probing membranes with an anti-GAPDH antibody (Santa Cruz) allowed normalization of protein content.

Quantitative PCR (q-PCR)

Five hundred nanograms of total RNA were reverse transcribed using SuperScript-III (Invitrogen), following the manufacturer's instructions. Detection of *SH2B3* and *GAPDH* was performed using a qPCR system with SensiFAST Probe Lo-ROX Kit (Bioline), according to the manufacturer's instructions. Specific TaqMan Gene Expression Assays probes (Applied Biosystems), were used to analyze *SH2B3* mRNA expression. Samples were normalized according to *GAPDH* transcript levels. Each sample was run in triplicate, in at least three independent experiments. qPCR assays were performed in a 7500 Fast qPCR machine (Applied Biosystems). The fold change was calculated using the $2^{-\Delta\Delta Ct}$ method.

Gene expression analyses by mRNAseq

Quantification of gene expression: STAR was used to obtain the number of reads associated to each gene in the Gencode v.31 annotation (restricted to protein-coding genes, antisense and lincRNAs). Raw counts for each sample were imported into R statistical software. Extracted count matrix was normalized for library size and coding length of genes to compute FPKM expression levels.

Unsupervised analysis: The Bioconductor edgeR package was used to import raw counts into R statistical software, and compute normalized log2 CPM (counts per millions of mapped reads) using the TMM (weighted trimmed mean of M-values) as normalization procedure. The normalized expression matrix from the 1000 most variant genes (based on standard deviation) was used to classify the samples according to their gene expression patterns using principal component analysis (PCA). PCA was performed by FactoMineR::PCA function with “ncp = 10, scale.unit = FALSE” parameters.

Differential expression analysis: Differential expression analysis was performed using the Bioconductor limma package and the voom transformation. To improve the statistical power of the analysis, only genes expressed in at least one sample (FPKM ≥ 0.1) were considered. A qval threshold of ≤ 0.05 and a minimum fold change of 1.5 were used to define differentially expressed genes.

DNA methylation (capture EM-seq)

Laboratory: For the panel capture part, a total of 1500 ng from 8 libraries, 187.5 ng each, were pooled and incubated 65°C for 16h with blockers, methylation enhancer, and Human methylome panel probes in a final volume of 40 µL. Washing steps were performed following the recommendations provided by Twist. Subsequently, the pooled sample was amplified with 8 cycles of PCR, quantified using qPCR and subjected to paired-end sequencing with 100-bp reads on an Illumina Novaseq 6000 platform.

Sequence alignment and quantification of DNA methylation: Quality of reads was assessed for each sample using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). BS-Seeker2[1] was used to map capture EM-seq data to the human hg38 genome and retrieve the number of methylated and unmethylated cytosines at each covered CpG site. Methylation rates were then integrated across CpG island (CGI)-based and gene-based features. CGI-based features were defined as follows: CpG islands (from UCSC database hg38), shores (2 kb on each side of the island) and shelves (2 kb on each side of the shores). DNA methylation outside CpG islands was analyzed by grouping CpG sites not located in CGI-based features every 100kb window. Gene-based features were defined based on Ensembl *homo sapiens* hg38 genes. The methylation rate was calculated for each gene across the promoter region (TSS +/- 500bp) and the gene body. Finally, methylation was analyzed by grouping CpGs in 100 bp tiles.

Unsupervised classification: Unsupervised classifications based on the 1000 most variant 100 bp tiles (based on standard deviation) were generated using principal component analysis (PCA) and hierarchical clustering (cosine distance, Ward method) in R software. We used consensus clustering (Bioconductor ConsensusClusterPlus package[2]) to examine the stability of the clusters. We established consensus partitions of the data set in K clusters (for K = 2, 3, ..., 10), on the basis of 1,000 resampling iterations (80% of genes, 80% of sample) of hierarchical clustering, with Pearson's dissimilarity as the distance metric and Ward's method for linkage analysis. We used the cumulative distribution functions (CDFs) of the consensus matrices to determine the optimal number of clusters (K = 5), considering both the shape of the functions and the area under the CDF curves, as previously described.

Statistical analysis and graphical representation

Protein *in silico* modelisation was conducted with Uniprot database (https://www.uniprot.org/uniprotkb/Q9UQQ2/entry) and AlphaFold (https://alphafold.ebi.ac.uk/entry/Q9UQQ2)

Statistical analyses were performed using GraphPad Prism (v.8.1.2). Medians were compared using the Mann-Whitney test.

Graphical representations were performed using ggplot2 (v.3.3.5) with RStudio (2022.02.0+443). Lollipop plot was generated using cBioportal Mutationmapper tool (https://www.cbioportal.org/mutation_mapper). Fishplots were created with the chrisamiller/fishplot (v.0.5) package using RStudio (2022.02.0+443). Illustrations were created using Affinity Designer (V1.10.6.1665).

References

1. Guo W, Fiziev P, Yan W, et al. BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data. BMC Genomics 2013;14(1):774.
2. Monti S. Consensus Clustering: A Resampling-Based Method for Class Discovery and Visualization of Gene Expression Microarray Data. 28.

Supplementary Table 1

	Patient	79.1	79.2	201.1	201.2
	Gender	M	M	F	M
	Age at JMML onset	15 days	36 days	22 days	11 days
	geographic origin	Morocco	Morocco	Algeria	Algeria
	consanguinity	yes	yes	no	no
At initial presentation	**Hematological counts in peripheral blood**				
	Hemoglobin count (g/dL)	12,8	10,8	6,7	19,9
	Platelet count (×10^9/l)	21	239	37	21
	White blood cells count (×10^9/l)	57,5	65,5	46	114
	Monocytes count (×10^9/l)	14,38	7,86	2,8	11,6
	Lymphocyte count (×10^9/l)	16,7	25,5	21	40
	Myeloid precursors (%)	15	10	9	23
	Blasts in peripheral blood (%)	6	4	6	0
	Hematological counts in bone marrow				
	Blast in bone marrow (%)	11	2	4	4
	Cellularity	Rich	Rich	Moderate	Moderate
	Dysplastic features	moderate	no	no	moderate
	Erythroid lineage	decreased (7%)	Normal (22%)	decreased (4%)	decreased (13%)
	Megakaryocytic lineage	absent	fairly abundant	absent	absent
	Clinical presentation				
	Hepatomegaly	yes	no	yes	yes
	Splenomegaly	yes	yes	yes	yes
	Adenopathy	no	no	yes	no
	Pallor	yes	no	yes	no
	Fever	no	no	no	no
	Skin lesions	no	no	no	no
	Respiratory symptoms	no	no	no	no
	Cardiac symptoms	no	no	no	no
	Bleeding	no	no	no	no
	Other clinical signs	RCIU, microcephaly, cardiomegaly, tracheomalacia, short stature	no	no	no
	Familial history	Thyroiditis (mother and two maternal aunts). insulin-dependent diabetes at 25 yrs (maternal uncle). learning and language difficulties with deafness (paternal aunt)		no	
	***SH2B3* variants**				
	Nucleotidic change	c.1160G>C	c.1160G>C	c.502G>T; c.1244 G>A	c.502G>T; c.1244 G>A
	Protein change	p.Gly387Ala	p.Gly387Ala	p.Glu168*; p.Arg415His	p.Glu168*; p.Arg415His
	Allelic status	homozygous	homozygous	compound heterozygous	compound heterozygous
	Status (germline/somatic)	germline	germline	germline	germline
	inheritance	AR	AR	AR	AR
	Other JMML features				
	Karyotype	46, XY	46, XY	46, XX	ND
	Secondary genetic alteration (*SH2B3*)	no	no	no	no
	Secondary genetic alteration (*PTPN11*)	no	no	no	no
	Fetal hemoglobin elevated for age	no	no	no	no
	Spontaneous growth of myeloid progenitors	yes	yes	ND	ND
	Treatment and follow up				
	Chemotherapy	6-MP	no	6-MP	no
	Bone marrow transplantation	yes	no	no	no
	Clinical outcome	Alive. Statural delay Very large splenomegaly Abdominal distension	Alive. Slight statural delay Autoimmune hypothyroidism	Alive. Persistent splenomegaly. Neutropenia	Alive. liver transplant after cirrhosis. Persistent neutropenia
	Age (years) as of june 2023	13	7	9	5

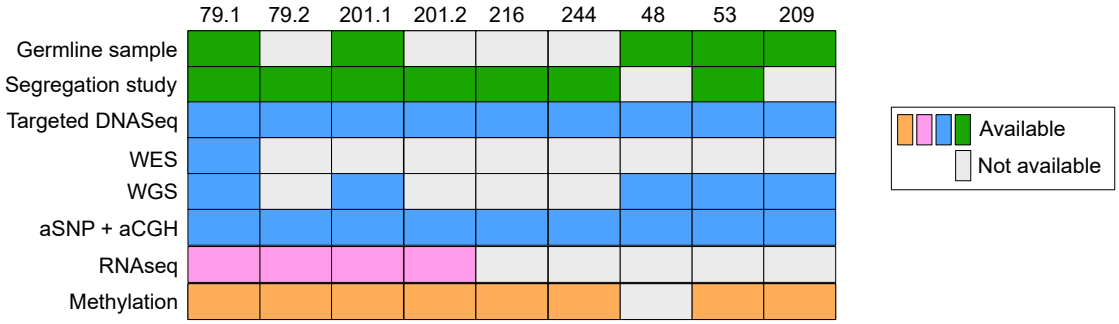
Supplementary Table 1 – continued

	Patient	216	244	OPBG_1	OPBG_2
	Gender	F	F	F	M
	Age at JMML onset	1 day	22 day	30 days	30 days
	geographic origin	Tunisia	North Africa	Romany	Romany
	consanguinity	yes	yes	yes	yes
At initial presentation	**Hematological counts in peripheral blood**				
	Hemoglobin count (g/dL)	17,8	9,4	11,8	8
	Platelet count (×10^9/l)	13	15	35	41
	White blood cells count (×10^9/l)	81	81,57	56,1	55,6
	Monocytes count (×10^9/l)	7,3	19,6	6,1	4,2
	Lymphocyte count (×10^9/l)	12,9	17,1	6,9	11,01
	Myeloid precursors (%)	9,5	11	25	3
	Blasts in peripheral blood (%)	1,5	4	5	4
	Hematological counts in bone marrow				
	Blast in bone marrow (%)	11	2	5	6,5
	Cellularity	Normal	Moderate	Moderate	Rich
	Dysplastic features	no	yes	yes	yes
	Erythroid lineage	decreased	decreased	decreased	decreased
	Megakaryocytic lineage	Present	absent	absent	dysplastic
	Clinical presentation				
	Hepatomegaly	yes	yes	yes	yes
	Splenomegaly	yes	yes	yes	yes
	Adenopathy	no	yes	yes	yes
	Pallor	yes	yes	yes	yes
	Fever	no	yes	no	no
	Skin lesions	no	no	no	no
	Respiratory symptoms	no	yes	no	no
	Cardiac symptoms	no	no	no	no
	Bleeding	Generalized petechial purpura	Meningeal haemorrhage	no	no
	Other clinical signs	no	RCIU; Facial dysmorphic features; vermis haematoma and haemoperitoneum	Facial dysmorphic features; coronary heart disease; Botallo's Duct	Facial dysmorphic features; Coronary heart disease; complex limb reduction defect
	Familial history	no	no	no	
	***SH2B3* variants**				
	Nucleotidic change	c.685_691dupGGCCCCG	c.1038dupG	c.1709dupA	c.1709dupA
	Protein change	p.Asp231Glyfs*39	p.L347Afs*38	p.Asp570Lysfs82	p.Asp570Lysfs82
	Allelic status	homozygous	homozygous	hom	hom
	Status (germline/somatic)	germline	germline	germline	germline
	inheritance	AR	AR	AR	AR
	Other JMML features				
	Karyotype	46, XX	46, XX	46 XX	46XY
	Secondary genetic alteration (*SH2B3*)	no	no	na	na
	Secondary genetic alteration (*PTPN11*)	no	no	na	na
	Fetal hemoglobin elevated for age	elevated		no	no
	Spontaneous growth of myeloid progenitors	ND	ND	Yes	Yes
	Treatment and follow up				
	Chemotherapy	no	6-MP	6-MP and cis-retinoic acid	no
	Bone marrow transplantation	no	no	no	no
	Clinical outcome	Alive. Spontaneous resolution	Alive. Persistent hepato-splenomegaly and hepatomegaly	Alive. Spontaneous resolution	Alive. Spontaneous resolution
	Age (years) as of june 2023	*5*	*2*	*12*	*8*

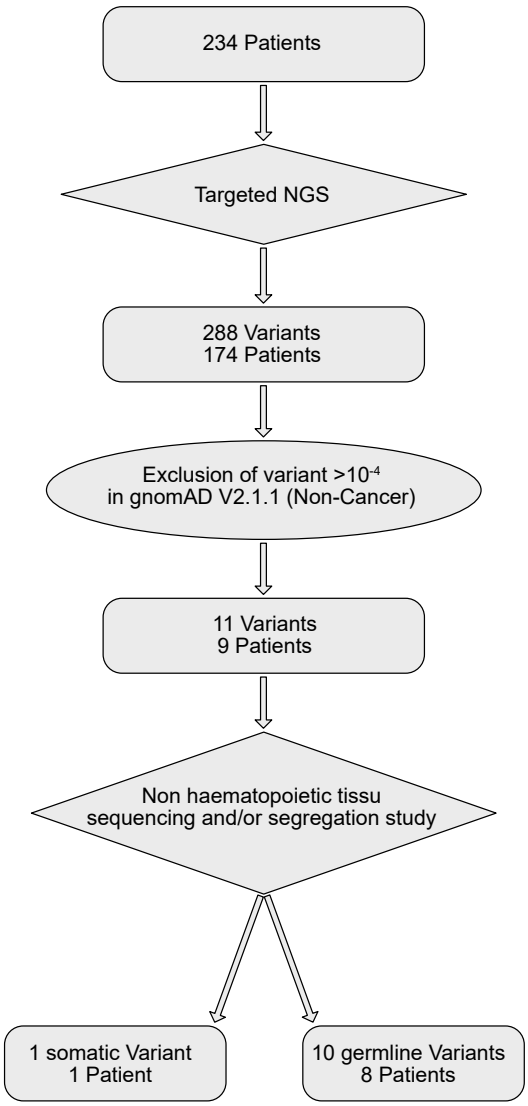
Supplementary Table 1 – continued

	Patient	48	53	209
	Gender	M	M	M
	Age at JMML onset	3,8 years	2,2 years	4,4 years
	geographic origin	unknown	unknown	unknown
	consanguinity	no	no	yes
At initial presentation	**Hematological counts in peripheral blood**			
	Hemoglobin count (g/dL)	10,1	11,2	10,5
	Platelet count (×10^9/l)	18	23	93
	White blood cells count (×10^9/l)	11,7	71,2	29,4
	Monocytes count (×10^9/l)	2,6	17	2,06
	Lymphocyte count (×10^9/l)	3,16	12,8	5
	Myeloid precursors (%)	8	14	4
	Blasts in peripheral blood (%)	17	3	1
	Hematological counts in bone marrow			
	Blast in bone marrow (%)	12	6	2
	Cellularity	Rich	Rich	Very rich
	Dysplastic features	moderate	no	no
	Erythroid lineage	decreased	normal	normal
	Megakaryocytic lineage	absent	present	decreased
	Clinical presentation			
	Hepatomegaly	yes	yes	no
	Splenomegaly	yes	yes	yes
	Adenopathy	no	yes	no
	Pallor	yes	no	no
	Fever	yes	no	no
	Skin lesions	no	xanthogranuloma	no
	Respiratory symptoms	no	no	pulmonary emphysema
	Cardiac symptoms	no	no	no
	Bleeding	yes		
	Other clinical signs	no	no	Intermittent left lameness Thyroid nodules
	Familial history		multiple sclerosis (father); Splenectomy at 23 yrs (unknown reason) (maternal grandfather)	Asthma (brother) Childhood leukemia (maternal aunt) Type 2 diabetes (paternal grandmother)
	***SH2B3* variants**			
	Nucleotidic change	c.1201T>C	c.685_691dupGGCCCCG	c.922C>T
	Protein change	p.Tyr401His	p.Asp231Glyfs*39	p.Arg308*
	Status (homozygous/heterozygous)	heterozygous	heterozygous	n/a
	Status (germline/somatic)	germline	germline	somatic
	inheritance	AD	AD	n/a
	Other JMML features			
	Karyotype	46XY	46,XY[15]/45,X,-Y[12]	46, XY with secondary +21
	Secondary genetic alteration (*SH2B3*)	no	aUPD	aUPD
	Secondary genetic alteration (*PTPN11*)	*PTPN11:* p.Glu76Gly	*PTPN11*: p.Gly503Ala	*PTPN11*: p.Asp61Tyr
	Fetal hemoglobin elevated for age	UK	yes	yes
	Spontaneous growth of myeloid progenitors	yes	yes	yes
	Treatment and follow up			
	Chemotherapy	Chemotherapy	6-MP	6-MP
	Bone marrow transplantation	yes	yes	yes
	Clinical outcome	Alive	Death following veno-occlusive disease	Alive. Post-transplant complications
	Age (years) as of june 2023	*19*	*17*	*9*

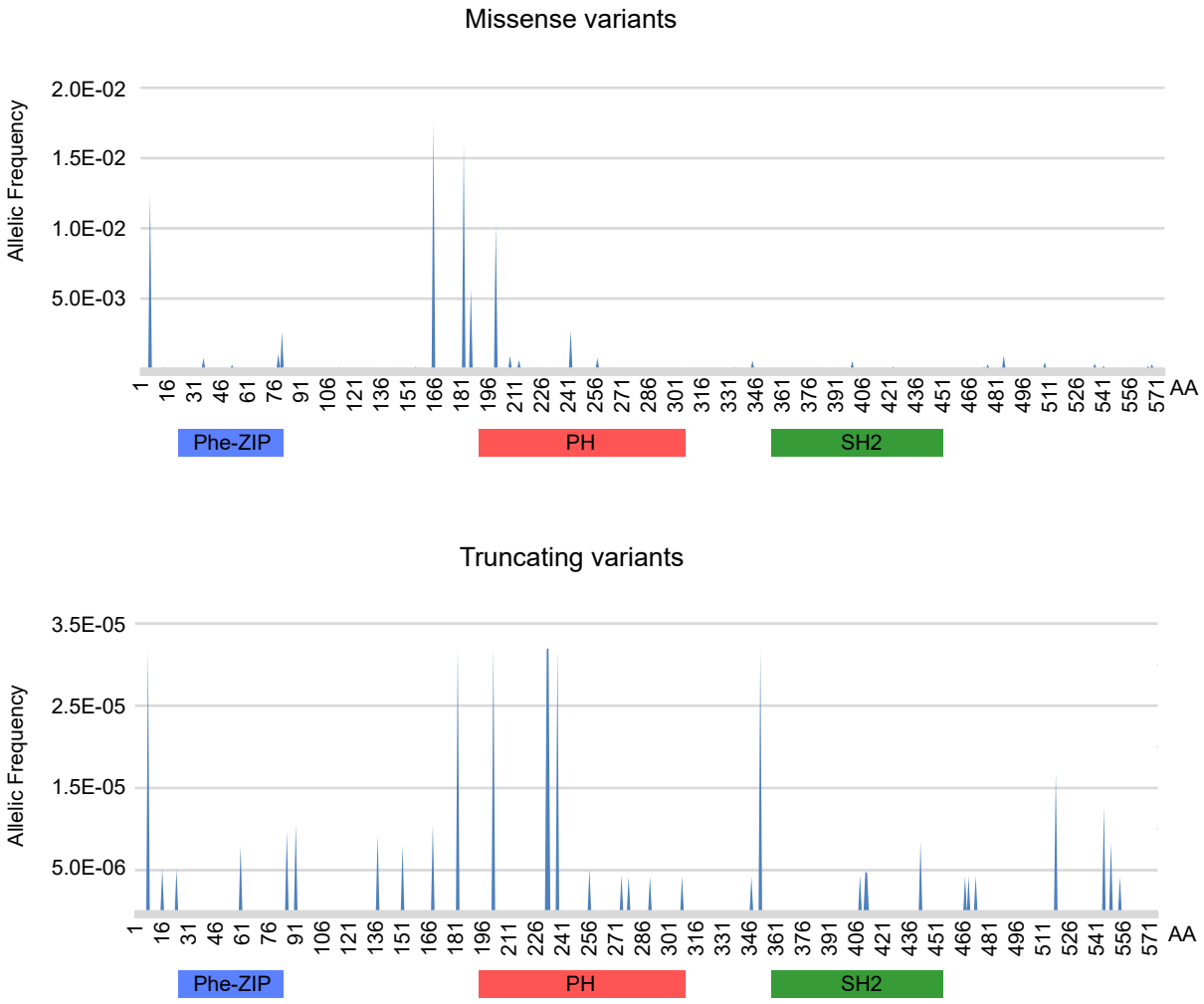
Supplementary Figure 1. Graphical representation of the analyses performed in each patient with *SH2B3* variants



Supplementary Figure 2. Flow chart of the *SH2B3* study in the French cohort of patients with JMML



Supplemental Figure 3. Frequency of *SH2B3* missense (upper panel) and truncating (bottom panel) variants in the GnomAD version 2.1.1 (non cancer) database. Notably, missense variations are mostly found in the PH domain whereas they are rare in the SH2 domain.



Supplementary Figure 4. Evolution of platelet counts with age in patients with *SH2B3* bi-allelic germline variants.