A somatic *UBA2* variant preceded *ETV6-RUNX1* in the concordant BCP-ALL of monozygotic twins

Benedicte Bang,[1] Jesper Eisfeldt,[1,2] Gisela Barbany,[1,2] Arja Harila-Saari,[3] Mats Heyman,[4] Vasilios Zachariadis,[5] Fulya Taylan,[1,2,*] and Ann Nordgren[1,2,*]

[1]Department of Molecular Medicine and Surgery, Center for Molecular Medicine, Karolinska Institutet, Stockholm, Sweden; [2]Department of Clinical Genetics, Karolinska University Hospital, Stockholm, Sweden; [3]Department of Women's and Children's Health, Uppsala University Hospital, Uppsala, Sweden; [4]Department of Women's and Children's Health, Karolinska University Hospital Solna, Stockholm, Sweden; and [5]Department of Oncology-Pathology, Cancer Centre Karolinska (CCK), Karolinska Institutet, and Karolinska University Hospital, Stockholm, Sweden

---

Key Points

- A somatic *UBA2* deletion preceded the well-established leukemia initiating event *ETV6-RUNX1* fusion in monozygotic twins with BCP-ALL.

- A shared complex rearrangement created an *ETV6-RUNX1* fusion and provided evidence of a common clonal in utero origin.

Genetic analysis of leukemic clones in monozygotic twins with concordant acute lymphoblastic leukemia (ALL) has proved a unique opportunity to gain insight into the molecular phylogenetics of leukemogenesis. Using whole-genome sequencing, we characterized constitutional and somatic single nucleotide variants/insertion-deletions (indels) and structural variants in a monozygotic twin pair with concordant *ETV6-RUNX1*⁺ B-cell precursor ALL (BCP-ALL). In addition, digital PCR (dPCR) was applied to evaluate the presence of and quantify selected somatic variants at birth, diagnosis, and remission. A shared somatic complex rearrangement involving chromosomes 11, 12, and 21 with identical fusion sequences in leukemias of both twins offered direct proof of a common clonal origin. The *ETV6-RUNX1* fusion detected at diagnosis was found to originate from this complex rearrangement. A shared somatic frameshift deletion in *UBA2* was also identified in diagnostic samples. In addition, each leukemia independently acquired analogous deletions of 3 genes recurrently targeted in BCP-ALLs (*ETV6*, *ATF7IP*, and *RAG1/RAG2*), providing evidence of a convergent clonal evolution only explained by a strong concurrent selective pressure. Quantification of the *UBA2* deletion by dPCR surprisingly indicated it persisted in remission. This, for the first time to our knowledge, provided evidence of a *UBA2* variant preceding the well-established initiating event *ETV6-RUNX1*. Further, we suggest the *UBA2* deletion exerted a leukemia predisposing effect and that its essential role in Small Ubiquitin-like Modifier (SUMO) attachment (SUMOylation), regulating nearly all physiological and pathological cellular processes such as DNA-repair by nonhomologous end joining, may hold a mechanistic explanation for the predisposition.

Introduction

Childhood acute lymphoblastic leukemia (ALL) is a genetically heterogenous disease,[1-3] largely affecting B-lymphoid cells (85%).[4,5] A variety of recurrent genetic aberrations, predominantly chromosomal translocations or nonrandom loss or gain of entire chromosomes, are currently considered the initiating events and drivers of the disease.[6] These aberrations are also the basis for subtype classification, the 2 most common being high hyperdiploidy and t(12;21)(p13;q22)/ *ETV6-RUNX1*.[6] The etiology of childhood ALL remains unknown in the vast majority of cases. Nevertheless, compelling evidence of constitutional predisposition to ALL has emerged in recent years.[7]

Monozygotic (mz) twins with concordant B-cell precursor ALL (BCP-ALL) have played a central role in exploring the timing of disease initiation. Molecular studies of initiating translocations have shown identical breakpoints and chimeric fusion sequences in leukemias of twins, providing solid evidence of a common clonal origin.[8-12] In addition, fusion sequences from *ETV6-RUNX1*–carrying preleukemic clones have been identified in archived neonatal dried blood spots and cord blood, also in nontwin cases.[10-22] Preleukemic clones arise in 1 twin in utero and spread to the sibling by vascular transfusion through anastomoses in the shared placenta.[10,11,23] In due time, secondary genetic alterations accumulate in genes crucial for B-cell development and eventually render leukemic transformation.[19,24,25]

Further, studies of latency (time from initiation to diagnosis), mutation rates, and concordance rates of BCP-ALL in mz twins have informed us on the nature of different initiating and disease-driving genetic events. In cases where latency is short (<1 year), differs little within a twin pair (months), and concordance rates are high (close to 100%), the genetic event is suggested to have stronger oncogenicity.[23] Such events are either alone sufficient to cause leukemia or efficient in causing the additional driving events required for overt leukemia.[26,27] In contrast, initiating events such as t(12;21)(p13;q22)/*ETV6-RUNX1* associated with longer latency (median 4 years, range 1-12 years[28-31]), larger differences in latency (up to 9 years[11]) within a twin pair, and lower concordance rates (10% to 15%[23]) are considered weaker oncogenic drivers. The need for additional disease-driving genetic events is also greater as illustrated by higher mutation rates and recurrent secondary genetic events.[32,33] Also, more recent studies have used concordant ALL to closer explore the molecular phylogenetics of clonal evolution.[34,35]

Despite the above efforts, our knowledge and understanding of molecular phylogenetics and causes of mutational processes during leukemogenesis remain limited. Molecular phylogenetics of leukemia is currently restricted to distinguishing shared (early in utero) from unique (late in utero or postnatal) somatic variants. In this study, we performed a comprehensive genetic analysis of a pair of mz twins with concordant *ETV6-RUNX1*⁺ BCP-ALL. We evaluate constitutional variants for predisposing effects, provide molecular proof of a common clonal origin of leukemia, characterize the nature of shared (early) variants, and explore the genetic divergence of concordant BCP-ALL in each twin. Unexpectedly, we were also able to elucidate the temporal order of 2 significant shared genetic events.

Methods

Ethical approval and consent

The ethics review board at the Karolinska Institutet approved this study (ethics number 2015-293-31/4 and 04-638/4), and informed consent from the parents were obtained according to the Declaration of Helsinki.

Samples

Genomic DNA was extracted from samples collected from both twins at 4 different timepoints: (1) neonatal dried blood spots (and their adjacent control) from the Swedish National Phenylketonuria Register at the Karolinska University Hospital, (2) bone marrow at leukemia diagnosis, (3) peripheral blood at clinical remission, and (4) saliva samples 5 years after clinical remission (Figure 1A). We retrieved 5 punches from each neonatal dried blood spot card, each 3 mm in diameter and from different locations of the spots. We also collected saliva samples from the parents.

DNA extraction

Genomic DNA extraction from dried blood spots was performed using QIAamp DNA Micro Kit (Qiagen, Hilden, Germany), following the manufacturer's recommendations. Genomic DNA from bone marrow at diagnosis and blood at clinical remission had been performed prior to this study at the Department of Clinical Genetics at Karolinska University Hospital, following standard procedures. Saliva samples were collected using Oragene DNA saliva collection kit (DNA Genotek, Ontario, Canada). Genomic DNA extraction from saliva samples was performed using prepIT-L2P kit (DNA Genotek, Ontario, Canada), following the manufacturer's recommendations. The concentrations were determined by Qubit dsDNA HS Assay Kit in Qubit 3.0 fluorometer (Life Technologies, Carlsbad, CA, USA).

Clinical genetic analysis of leukemias

Karyotyping with G-banding (Giemsa staining), interphase fluorescence in situ hybridization (FISH), and array comparative genomic hybridization had been performed prior to this study, following standard protocols at the Department of Clinical Genetics at Karolinska University Hospital. We obtained clinical and laboratory data for each twin, summarized in Table 1, from medical records.

Whole-genome sequencing and bioinformatic analysis

Whole-genome sequencing (WGS) was performed on genomic DNA from diagnostic bone marrow and matched peripheral blood from remission. Libraries for sequencing on Illumina HiSeq X (Illumina Inc, San Diego, CA, USA) were prepared from genomic DNA using Illumina TruSeq polymerase chain reaction (PCR)-free kit with a mean insert size of >350 base pairs, resulting in over 700 million (range 729-915M) mapped unique sequences per sample with mean read depth 37× (range 34-39×). Alignment of reads to human reference genome (GRCh37/hg19) and variant calling was performed by Science for Life Laboratory (SciLifeLab, Stockholm, Sweden) (supplemental Table 1).

Somatic single nucleotide variants (SNVs)/indels were identified using MuTect2.[36] Constitutional variants were identified and processed using best practices of the Genome Analysis Toolkit.[37] Variants were functionally annotated using Variant Effect Predictor (version 89)[38]
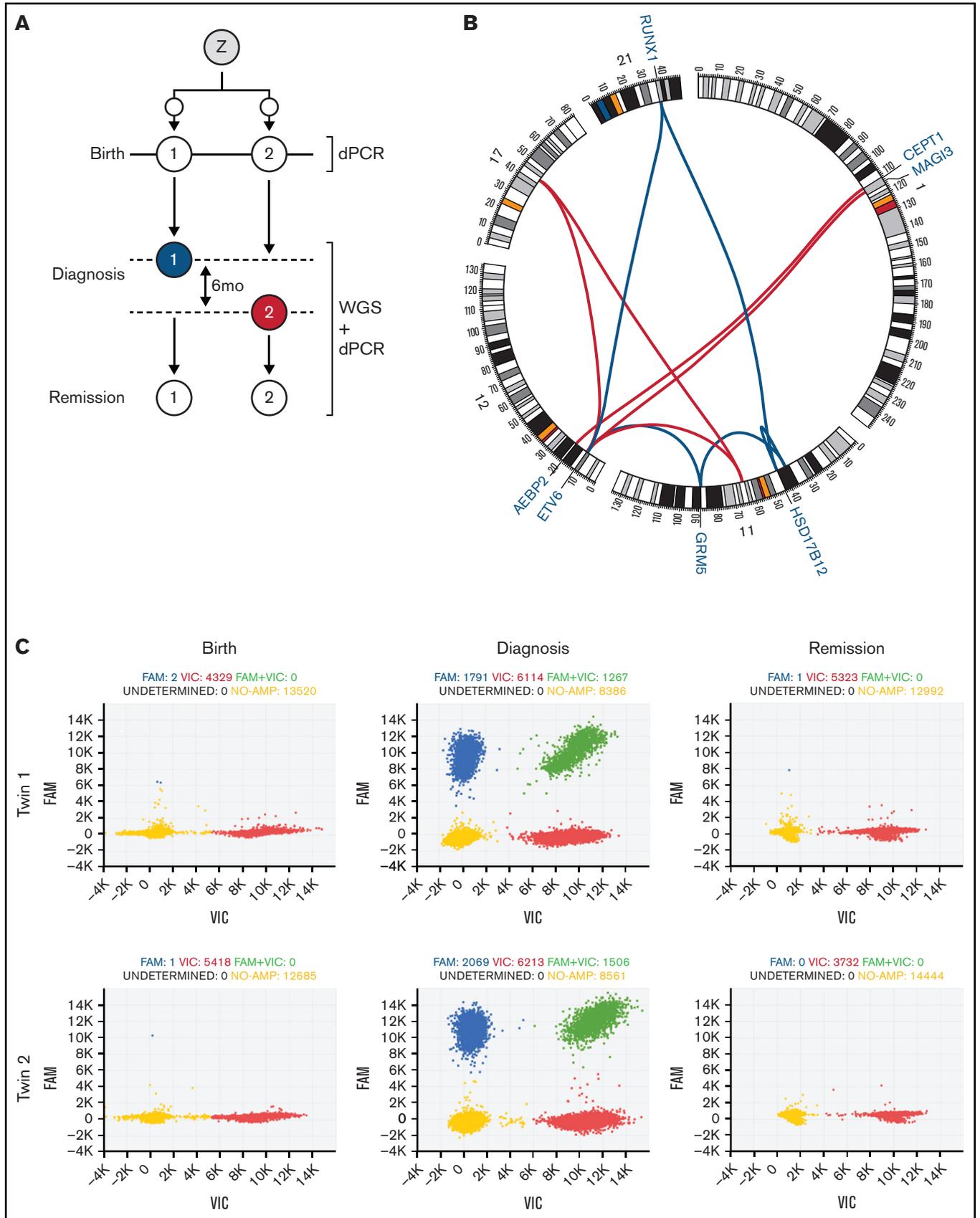
Table 1. Clinical characteristics

	Twin 1 (Tw1)	Twin 2 (Tw2)
Diagnosis	BCP-ALL	BCP-ALL
Age at diagnosis	3y 4m	3y 10m
Full blood count at diagnosis (peripheral blood)		
Hb (g/L)	41	69
WBC ($\times 10^9$/L)	2.8	2.8
Platelets ($\times 10^9$/L)	128	163
Immunophenotype (bone marrow)	CD45dim, CD19$^+$, CD10^{++}, CD20$^-$, TdT$^+$, CD22$^+$, CD cyt79a$^+$, CD38$^+$, HLADR$^+$, CD123dim, CD58$^+$, CD66c$^-$, cytIgM$^-$, no myeloid markers, subpopulation (36%) CD34$^+$	CD45dim/neg, CD19$^+$, CD10^{++}, CD20$^-$, Tdt$^+$, CD22$^+$, CD cyt79a$^+$, CD38$^+$, HLADR$^+$, CD123dim/neg, CD58$^+$, CD66c$^-$, cytIgM$^-$, no myeloid markers, CD34hetero/$^+$, no T-cell markers, CD99^{++}
Blast count at diagnosis (bone marrow)	69%	74%
CNS engagement of leukemia	No	No
Cytogenetics (bone marrow at diagnosis)		
Karyotyping with G-banding	46,XX,t(11;12)(q21;p13)[6]/46,XX[19]	46,XX,t(11;12)(q23;p13)[5]/47,XX,sl, t(1;12)(p13;p13),+der(1)t(1;12)(p13;p13)[9]/ 46,XX[3]
FISH	nuc ish(ETV6x2,RUNX1x3)(ETV6 con RUNX1x1)[145/ 206],(ETV6x1,RUNX1x3)(ETV6 con RUNX1x1)[12/206]	nuc ish(ETV6x2,RUNX1x3)(ETV6 con RUNX1x1)[98/ 222],(ETV6x1,RUNX1x3)(ETV6 con RUNX1x1)[98/ 222]
ArrayCGH	No detectable copy number changes	No additional finding
Interpretation	Main clone with t(11;12) and *ETV6-RUNX1* (70%), subclone with additional del*ETV6* (6%)	Main clone with t(11;12) and *ETV6-RUNX1* (88%), subclone with additional del*ETV6* and t(1;12) with extra der(1)t(1;12) (44%)
Image cytometric DNA analysis (ICDA)	Diploid (DNA index: 0.5, S-phase 8%)	Diploid (DNA index: 1, S-phase 8%)
Treatment protocol	NOPHO-ALL 2008 standard risk arm	NOPHO-ALL 2008 standard risk arm
Stem cell transplantation	No	No
Minimal residual disease		
Day 15	<0.1%	<0.1%
Day 29	<0.1%	<0.01%
Day 79	<0.01%	<0.01%
Complications and treatment related toxicities	Gastroenteral chlostridium difficile infection. Bilateral purulent heamophilus influenzae conjunctivitis. HSV keratitis. Severe varicella infection with concurrent hepatitis of unknown etiology	Vincristine neuropathy with remaining muscular weakness in lower limbs. Acute ITP after end of intense treatment phase, spontaneous regression after steroid treatment
Adjustments to treatment protocol	Dose reduction of high-dose methotrexate due to high top-concentrations and delayed excretion with renal toxicity	Dose reduction of Vincristine due to toxicity (peripheral neuropathy)
Relapse	No. Currently in complete remission 5.5 y after diagnosis	No. Currently in complete remission 5 y after diagnosis
Current growth and developmental parameters		
Height	−1.64 SD	−1.72 SD
Weight	−0.94 SD	−1.3 SD
Head size	Not available	Not available
Psychomotor development	Normal	Normal

CGH, comparative genomic hybridization; HSV, Herpes simplex virus; ITP, immune thrombocytopenic purpura; NOPHO, Nordic Society of Paediatric Haematology and Oncology.

and loaded into a database using GEMINI (GEnome MINIng) (v0.20.0).[39] Variants were explored in GEMINI using built-in tools and visualized in Integrative Genomics Viewer.[40] Leukemia predisposition–centered constitutional variant analysis was performed applying the gene panel "Hematological malignancies cancer susceptibility," curated by experts in Genomics England PanelApp[41] (supplemental Table 2), to whole-genome data from remission.

Structural variants were detected using FindSV pipeline (https://github.com/J35P312/FindSV) merging calls from CNVnator v0.3.2 and TIDDIT.[42,43] Immunoglobulin heavy chain (IgH) and T-cell receptor rearrangements were excluded from this analysis. Immunoglobulin rearrangements were instead identified using IgCaller.[44] Somatic structural variants were analyzed as previously described.[45,46] Complex translocations were plotted using circos.[47] To detect presence of any copy number neutral structural aberrations, loss-of-heterozygosity (LOH) analysis was performed. Regions of homozygosity were called using GEMINI built-in regions-of-homozygosity function with window sizes ranging between 100 Kb and 300 Kb in both tumor-normal pairs.

Figure 1. Schematic overview and SV analysis. (A) Monozygotic twins developed BCP-ALL with a 6-month difference in latency. Sampling took place at birth (dried neonatal blood spots), diagnosis (bone marrow), and in remission (peripheral blood, 2 years 11 months [Tw1] and 2 years 5 months [Tw2] after diagnosis). WGS and dPCR analysis were applied to samples as illustrated to uncovered constitutional and somatic variants and quantify key mutations, respectively. Blue and red circles

Sanger sequencing

Sanger sequencing was performed using ABI3730xl DNA Analyzer (Applied Biosystems, Foster City, CA, USA) and BigDye Terminator v3.1 (Applied Biosystems, Foster City, CA, USA) to validate translocation breakpoints and for segregation analysis. Primer sequences are available upon request.

Genome amplification

Ten nanograms DNA extracted from dried blood spots generated up to 12 to 20 micrograms of DNA using illustra Ready-To-Go GenomiPhi V3 DNA Amplification Kit (GE Healthcare, Buckinghamshire, UK) according to the manufacturer's protocol.

Chip-based digital PCR

Genomic DNA (15-50 ng) was amplified with 1X QuantStudio 3D Digital PCR Master mix, TaqMan assay for reference gene, and hydrolysis probes for target region according to manufacturer's instructions. Data were analyzed using QuantStudio 3D Analysis Suite Cloud Software, version 3.1.6-PRC-build2 with default parameters (confidence level of 95%, desired precision of 10%, and Poisson plus quantification algorithm).

We have previously demonstrated the versatility of digital PCR to detect and quantify somatic structural variants and its applicability to DNA extracted from dried blood spots.[48] TaqMan probe for the shared complex rearrangement was designed manually, targeting the fusion sequence of *GRM5-ETV6* (chromosome 11q to 12p). Probes for *NSD2* p.E1099K and *UBA2* deletion were designed by ThermoFisher Scientific's internal bioinformatics platform. TaqMan Copy Number Reference Assay (Applied Biosystems), human, ribonuclease P RNA component H1 (*RPPH1*; chromosome 14(GRCh37): 20811565), labeled with VIC was used as internal control.

Droplet digital PCR

Genomic DNA (67 ng) from saliva was amplified in triplicates using the droplet digital PCR (ddPCR) Supermix for Probes (No dUTP) kit (BioRad, Hercules, CA, USA). The TaqMan assay for *UBA2* used for chip-based dPCR was also used in this experiment. Droplets were generated on the Automated Droplet Generator (BioRad, Hercules, CA, USA), and PCR was performed according to the manufacturer's instructions. After completion of PCR, the droplets were read QX200 on the Droplet Reader (BioRad, Hercules, CA, USA). The QuantaSoft Analysis Pro v.1.0 was used to analyze the data. The threshold for true signal positivity was adjusted based on the signal in the control samples. The ratio of the number of positive droplets to the total number of droplets was calculated for each sample using Poisson 95% confidence intervals.

Results

Clinical findings

The twins studied here were monozygotic, monochorionic, and diamniotic females born by acute cesarean section at gestational week 36 plus 3. Both twins developed BCP-ALL at ages 3 years 4 months (Tw1) and 3 years 10 months (Tw2), a 6-month difference in latency. All clinical data have been summarized in Table 1. Immunophenotypes displayed the classical composition of BCP-ALL and were by and large identical, differing only in expression of CD99 and CD34. Cytogenetic analysis at diagnosis identified a t(11;12)(q21;p13) as well as the classical t(12;21)(p13;q22) rearrangement in the main clone of both leukemias. In addition, Tw1 carried a subclonal deletion of *ETV6* in 6% of cells, whereas Tw2 carried a subclonal deletion of *ETV6* and a t(1;12)(p13;p13),+der(1)t(1;12) in 44% of cells.

Both twins were treated according to the NOPHO-ALL 2008 protocol, standard risk arm, and remain in full remission 5.5 (Tw1) and 5 (Tw2) years after diagnosis. Both twins have a normal psychomotor development, growth parameters within the normal range, and no malformations, dysmorphic features, or clinical signs of neurofibromatosis.

Constitutional variant analysis

The increasing awareness of the contribution of constitutional predisposition to childhood leukemia urged our inclusion of a constitutional variant analysis. Both twins were found heterozygous for a constitutional missense variant in tumor suppressor gene *NF1* (chr17:g.29557883C>A, NM_000267.3:c.3137C>A) NP_001035 957.1:p.(Thr1046Lys). The detected variant was not present in gnomAD,[49] COSMIC,[50] or LOVD[51] and had a CADD[52] (v1.4) c-score of 27.1, indicative of a novel, seemingly damaging, variant, further supported by a majority of in silico predictions (supplemental Table 3). Pathogenic variants in *NF1* are known to cause neurofibromatosis type 1 (NF1).[53] However, the variant detected here is not previously described in NF1, neither did the twins display any clinical signs obligate for NF1 diagnosis. Segregation analysis showed that this variant is inherited from the healthy mother, who is also heterozygous. Hence, we classified this aberration as a variant of unknown significance. The potential impact of this variant is discussed below. No other pathogenic constitutional variants were detected.

Somatic variant analysis

Shared variants. From WGS data, we characterized a shared complex rearrangement t(11;12;21)(q23;p13;q22) (Figure 1B; Table 2) with identical fusion sequences in both twins' leukemias (supplemental Figure 1), providing proof of a common clonal origin. The rearrangement gave rise to the classical *ETV6-RUNX1* fusion gene recurrent in BCP-ALL, which was detected by FISH analysis at diagnosis, although its origin in a complex rearrangement was

---

Figure 1 (continued) represent Tw1's and Tw2's leukemia, respectively. (B) Circos plot showing somatic SVs in leukemias. Only chromosomes involved in rearrangements are displayed (1, 11, 12, 17, and 21). Colored lines illustrate how breakpoints have fused. Genes disrupted by or in close proximity of the breakpoints are indicated. Blue lines represent the shared complex rearrangement t(11;12;21)(q23;p13;q22), involving 2 inversions on chromosome 11 and generating a *ETV6-RUNX1* fusion. Redlines represent SVs unique to Tw2: 2 subclonal translocations t(1;12) and 1 t(11;12;17). (C) dPCR detection and quantification of shared complex rearrangement t(11;12;21)(q23;p13;q22) at birth, diagnosis, and in remission of both twins. TaqMan assay targeted chromosome 11;12 junction sequence. Clusters of dPCR chip wells positive for internal reference control *RPPH1* (red), target region (blue), reference and target (green), and with no amplification (yellow). Complex rearrangement readily detected at diagnosis but beyond detection at birth and in remission. Detection limit: 1 in 1000 copies. Images acquired from QuantStudio 3D Analysis Suite Cloud Software, version 3.1.6-PRC-build2 with default parameters. Z, zygote.

Table 2. Somatic structural variants, shared and unique, in twins' BCP-ALLs and supporting reads at diagnosis

	Chromosome A	Position A	Orientation A	Chromosome B	Position B	Orientation B	Length (bp)	Variant	Genes	Present in	Supporting reads	Detection
Shared complex rearrangement	11	88738784	Reverse	12	12032793	Reverse	NA	Translocation	GRM5 - ETV6	Twin 1 and 2	13 and 11	Pipeline
	12	12032592	Forward	21	36263860	Forward	NA	Translocation	ETV6 - RUNX1	Twin 1 and 2	9 and 12	Pipeline
	11	43665521	Forward	21	36264195	Reverse	NA	Translocation	HSD17B12 - RUNX1	Twin 1 and 2	10 and 5	Pipeline
	11	38994730	Forward	11	43665507	Forward	4670777	Inversion	HSD17B12|AC104387.1||MIR670|RN7SKP287|RP11-148I19.1|AC027806.1|RP11-810F22.1|AC021749.1|RP11-375D13.4|RP11-375D13.3|RP11-375D13.2|RP11-375D13.1|RP11-124G5.2|RP11-124G5.3|RP11-124G5.1|CTBP2P6|RNU6-99P|RNU6-365P|RP11-484D2.3|RP11-484D2.2|RP11-484D2.5|RP11-484D2.4|Y_RNA|MIR129-2|RP11-40H19.1|RP11-407P18.1|RP11-63C8.1|HNRNPKP3|CTD-2572N17.1|API5|LRRC4C|AC090720.1|MIR670HG|TTC17|CTD-2537L20.1|RP11-111A24.1|RP11-111A24.2	Twin 1 and 2	16 and 12	Pipeline
	11	38995567	Reverse	11	88738383	Forward	49742816	Inversion	GRM5, More than 200 genes!	Twin 1 and 2	7 and 9	Pipeline
Unique analogous deletions	2	89132165	Forward	2	89521183	Reverse	389018	Deletion	IGK locus	Twin 1	16	Pipeline
	2	89130687	Forward	2	89568151	Reverse	437464	Deletion	IGK locus	Twin 2	15	Manual
	2	114164337	Forward	2	114195636	Reverse	31299	Deletion	RP11-480C16.1|CBWD2|AC016745.3|IGKV1OR2-108	Twin 1	17	Pipeline
	2	114164336	Forward	2	114195596	Reverse	31260	Deletion	IGKV1OR2-108,AC016745.3, RP11-480C16.1, CBWD2	Twin 2	9	Pipeline
	11	36600026	Forward	11	36638041	Reverse	38015	Deletion	RAG1, RAG2, C11orf74	Twin 1	12	Pipeline
	11	36598835	Forward	11	36638100	Reverse	39265	Deletion	RAG1, RAG2, C11orf74	Twin 1	2	Manual
	11	36598831	Forward	11	36638246	Reverse	39415	Deletion	RAG1, RAG2, C11orf74	Twin 2	2	Manual
	11	36619725	Forward	11	36637900	Reverse	18175	Deletion	RAG2, C11orf74	Twin 2	1	Manual
	12	14522437	Forward	12	14652349	Reverse	129912	Deletion	ATF7IP	Twin 1	16	Pipeline
	12	14520351	Forward	12	14652520	Reverse	132169	Deletion	ATF7IP	Twin 2	4	Pipeline
	12	11809132	Forward	12	11933621	Reverse	124489	Deletion	ETV6	Twin 1	1	Manual
	12	11804005	Forward	12	11998522	Reverse	194517	Deletion	ETV6	Twin 2	3	Manual
Unique to twin 1	1	45985006	Forward	1	46024845	Reverse	39839	Deletion	PRDX1,HMGB1P48,AKR1A1	Twin 1	13	Pipeline
	9	21975708	Forward	9	22009692	Reverse	33984	Deletion	CDKN2A/CDKN2B	Twin 1	4	Manual
	19	36980100	Forward	19	37019980	Reverse	39880	Deletion	ZNF566,CTD-2630F21.1, CTBP2P7,ZNF260, AC092295.4	Twin 1	11	Pipeline
	X	48433300	Forward	X	48456163	Reverse	22863	Deletion	RBM3,RP11-1148L6.5,MRPL32P1,WDR13	Twin 1	11	Pipeline

bp, base pairs.

Table 2. (continued)

| Chromosome A	Position A	Orientation A	Chromosome B	Position B	Orientation B	Length (bp)	Variant	Genes	Present in	Supporting reads	Detection
Unique to twin 2											
1	111714901	Forward	12	19743172	Reverse	NA	Translocation	CEPT1 - AEBP2	Twin 2	6	Pipeline
1	114007276	Reverse	12	11666791	Forward	NA	Translocation	MAGI3 - RP11-434C1.3	Twin 2	14	Pipeline
11	64654367	Reverse	12	11742071	Forward	NA	Translocation		Twin 2	5	Manual
11	64653556	Forward	17	43457999	Reverse	NA	Translocation		Twin 2	6	Manual
12	11668138	Reverse	17	43457943	Reverse	NA	Translocation	RP11-434C1.3	Twin 2	6	Pipeline
1	154580133	Forward	1	154615230	Reverse	35097	Deletion	ADAR, AL606500.1	Twin 2	34	Pipeline
7	142334794	Forward	7	142495145	Reverse	160351	Deletion	TRBV20-1,TRBV21-1,TRBV22-1,TRBV23-1,TRBV24-1,MTRNR2L6,TRBV25-1,TRBVA,TRBV26,TRBVB,TRBV27,TRBV28,PGBD4P1,TRBV29-1,PRSS1,PRSS3P1,PRSS3P2,WBP1LP1,TRBJ2-1,TRBJ2-2,TRBJ2-2P,TRBJ2-3,TRBJ2-4,TRBJ2-5,TRBJ2-6,TRBJ2-7	Twin 2	18	Pipeline
14	22907933	Forward	14	22982929	Reverse	74996	Deletion	AE000661.37,TRDD2,TRDD3,TRDJ1,TRDJ4,TRDJ2,TRDJ3,TRDC,TRDV3,TRAJ61,TRAJ60,TRAJ59,TRAJ58,TRAJ57,TRAJ56,TRAJ55,TRAJ54,TRAJ53,TRAJ52,TRAJ51,TRAJ50,TRAJ49,TRAJ48,TRAJ47,TRAJ46,TRAJ45,TRAJ44,TRAJ43,TRAJ42,TRAJ41,TRAJ40,TRAJ39,TRAJ38,TRAJ37,TRAJ36,TRAJ35,TRAJ34,TRAJ33,TRAJ32,TRAJ31,TRAJ30,TRAJ29	Twin 2	7	Pipeline
22	22569566	Forward	22	22599665	Reverse	30099	Deletion	IGLV10-54,IGLVIV-53,TOP3BP1,LL22NC03-123E1.5,VPREB1	Twin 2	16	Pipeline

bp, base pairs.
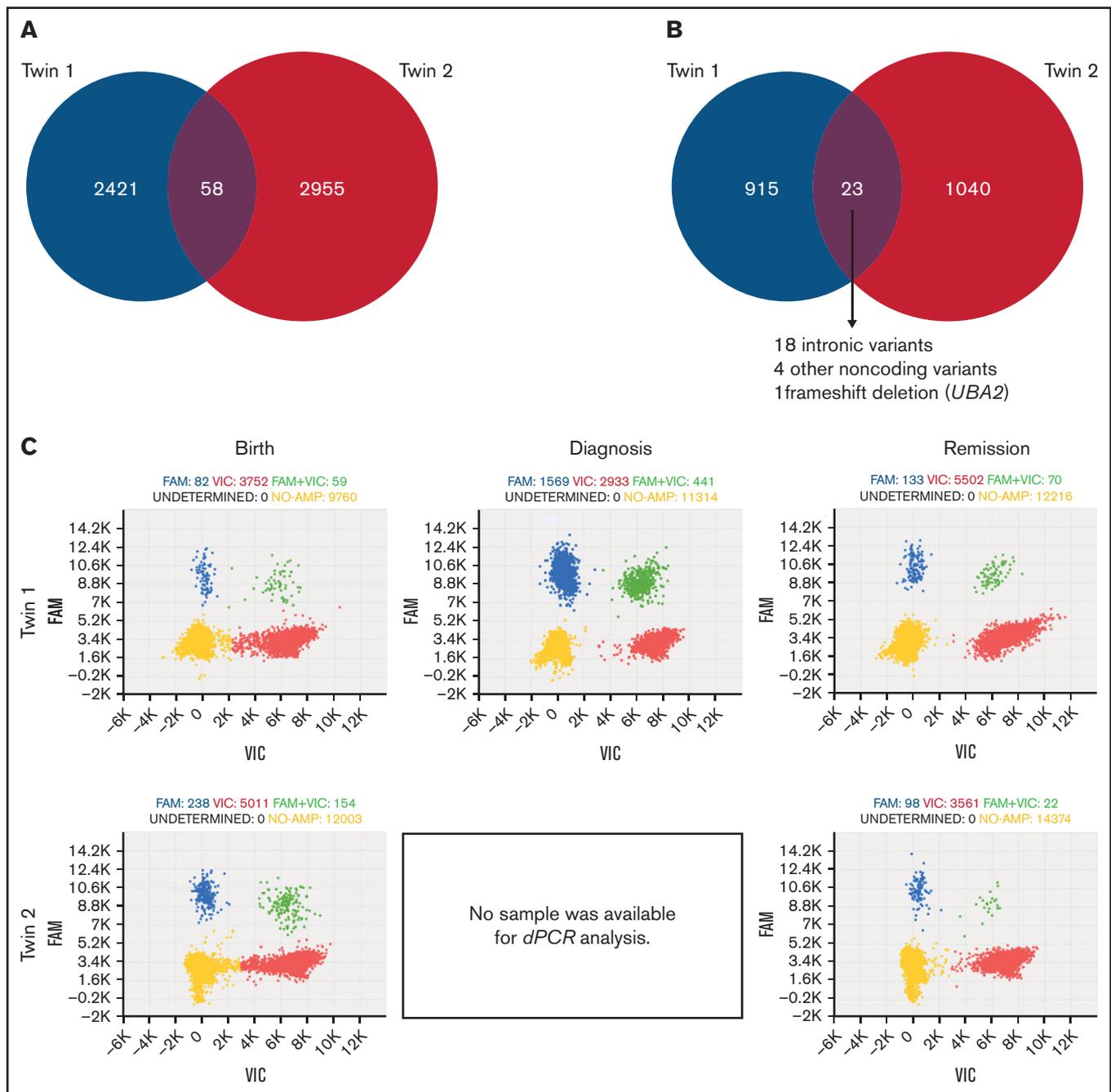
Figure 2. Shared and unique somatic SNVs/indels in twins' leukemias. (A) All somatic SNVs/indels across the genome. Two thousand, four hundred twenty-one and 2955 unique to Tw1 and Tw2, respectively, and 58 shared. (B) SNVs/indels in protein coding genes. Nine hundred fifteen and 1040 unique to Tw1 and 2, respectively: 23 shared. (C) dPCR detection and quantification of *UBA2* deletion (NM_005499.2: c.463_470del; NP_005490.1: p.(Thr156Leufs*2) at birth, diagnosis, and in remission. Diagnostic sample of Tw2 not available for analysis. Clusters of dPCR chip wells positive for wildtype allele (red), mutant allele (blue), wildtype and mutant allele (green), and with no amplification (yellow). *UBA2* deletion was detected at birth (both twins), diagnosis (Tw1; Tw2 lacked sample for analysis), and, unexpectedly, also in remission (both twins). Detection limit: 1 in 1000 copies. Images acquired from QuantStudio 3D Analysis Suite Cloud Software, version 3.1.6-PRC-build2 with default parameters.

not. It also created an additional fusion gene, *GRM5-ETV6,* by the in-frame fusion of *GRM5* exon 2 (chr 11) to *ETV6* exon 6 (chr 12), previously unreported in Mitelman[54] and St. Jude Cloud (https://www.stjude.cloud)[55] (accessed December 2020). Sequence coding for erythroblast transformation specific-domain in ETV6 but only part of atrial natriuretic factor-receptor domain in GRM5 was retained in the fusion gene. Nevertheless, under the promoter of *GRM5*, expression of *GRM5-ETV6* fusion in lymphocytes is likely low to nonexisting (https://www.gtexportal.org/home/) (accessed December 2020). No other shared structural variants were found.
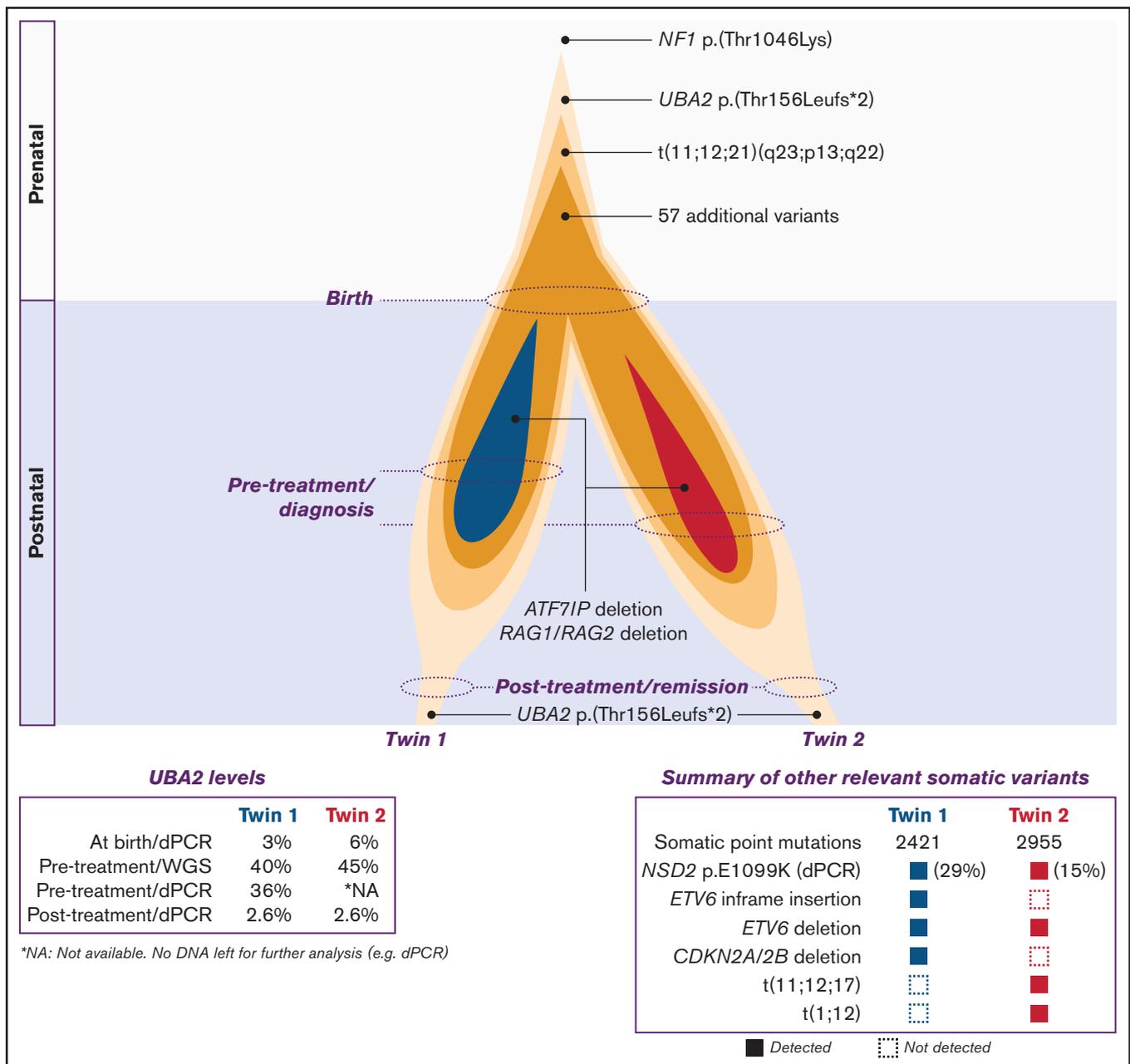
Figure 3. Illustration of clonal evolution from in utero leukemia initiation (prenatal) to remission (postnatal). Strikingly, *UBA2* deletion proceeded *ETV6-RUNX1* fusion generated by the shared complex rearrangement in utero. Fifty-seven additional shared SNVs/indels were acquired during the prenatal period. Clonal evolution of preleukemic clones established prenatally in both twins continued separately, mainly postnatally, acquiring SVs and SNVs/indels unique to each twins' leukemia. Genes known to be recurrent targets of secondary events in BCP-ALL, *ATF7IP*, *RAG1/RAG2*, and *ETV6*, were targeted by unique analogous deletions in both twins. The *UBA2* deletion persisted subclonally in remission of both twins. dPCR results stated in percent refers to the fraction of mutant target DNA in the analyzed sample.

With the aim to quantify the complex rearrangement (ie, a preleukemic clone) at birth, the unique fusion sequence of *GRM5-ETV6* was used for chip-based dPCR (Figure 1C). Results stated in percent (%) refers to the fraction of mutant target DNA in the analyzed sample. As expected, the complex rearrangement was readily detected at diagnosis in Tw1 (26%) and Tw2 (29%) but not at remission. However, the rearrangement was beyond detection in both twins also at birth, indicating its copy number was beyond our detection limit (1 in 1000 copies).

In addition, 58 somatic SNVs/indels were found shared by both leukemias (Figure 2A; supplemental Table 4). Our analysis highlighted a somatic frameshift deletion in *UBA2* (NM_005499.2: c.463_470del; NP_005490.1: p.(Thr156Leufs*2)), which has recently been implicated in childhood BCP-ALL as a novel driver gene (Figure 2B).[56] Variant allele frequencies from WGS at diagnosis was 40% (Tw1) and 44% (Tw2). The deletion, affecting the *UBA2* exon 6 of 17, caused a frameshift, introducing a proximate premature stop.

To quantify the presence of *UBA2* deletion at diagnosis and, simultaneously, repeat our effort to quantify the presence of preleukemic clones at birth, we again applied chip-based dPCR analysis (Figure 2C). As expected, the *UBA2* deletion was readily detected at diagnosis in Tw1 (36%) and at birth in Tw1 (3%) and Tw2 (6%). Unfortunately, lack of further sample prevented analysis of Tw2 at diagnosis. Unexpectedly, the *UBA2* deletion was detected at remission (sampling 2 years 11 months [Tw1] and 2 years 5 months [Tw2] after diagnosis) in both twins (2.6%). When reviewed manually, the variant was found in 1 out of 45 WGS reads (2.2%) at remission of Tw1. Five years after the clinical remission, the *UBA2* deletion in DNA extracted from the twins' saliva was detected 1.14% in Tw1 and 0.78% in Tw2 (data not shown). In children, cellular content of saliva consists of ~13.5% lymphocytes and 70% epithelial cells.[57] Therefore, our finding is most likely due to the presence of lymphocytes in the saliva rather than presence of mosaicism in epithelial cells.

Unique variants. We identified 10 deletions unique to Tw1's leukemia and 10 deletions, 2 translocations, and 1 complex rearrangement unique to Tw2's leukemia from WGS data (Table 2). Six deletions were analogous to one another (ie, affecting the same chromosomal regions but with different breakpoints). Two analogous deletions encompassed the *ETV6* and *ATF7IP* locus, respectively, whereas *RAG1/RAG2* locus was affected by 2 deletions in each twin (supplemental Figures 2, 3, and 4, respectively). The *ETV6* deletions were also seen at diagnosis by FISH analysis and detected here by targeted manual inspection of WGS data.

In Tw1, the 4 remaining deletions, sized 22 to 40 kb, affected chromosomes 1, 9, 19, and X (Table 2). The 34 kb deletion of *CDKN2A/CDKN2B* (supplemental Figure 5) on chromosome 9 was detected through manual investigation of recurrent second-hit regions in BCP-ALL. In Tw2, the 4 remaining deletions, sized 30 to 160 kb, affected chromosomes 1, 7, 14, and 22 (Table 2). The 2 translocations fused proximate regions of chromosome 1p and 12p, whereas the complex rearrangement involved chromosomes 11, 12, and 17 (Table 2; Figure 1B, red color).

In addition, 2421 (Tw1) and 2955 (Tw2) unique somatic SNVs/indels were identified, out of which 915 (Tw1) and 1040 (Tw2) mapping to protein coding genes (supplemental Table 5). An in-frame insertion in *ETV6* (NM_001987.4: c.309_310insCGG CCTAGC, NP_001978.1:p.(R103_Y104insRPS)) was highlighted in Tw1. Further, a somatic SNV in *NSD2 (WHSC1)* p.E1099K (COSMIC ID: COSM379334)[58] (supplemental Table 5) was found in leukemias of both twins. This gene was reviewed manually due to its known recurrence in t(12;21)/*ETV6-RUNX1*+ BCP-ALL.[59] Applying chip-based dPCR, the *NSD2* variant was readily detected at diagnosis in Tw1 (29%) and Tw2 (15%) but not present at remission. At birth, it was undetectable in Tw1 but quantified to 0.1% in Tw2 (data not shown).

LOH analysis was also performed on WGS data for both shared and unique variants and it did not reveal any novel deletions, duplications, nor copy number–neutral LOH events.

Immunoglobulin rearrangements. IgCaller identified 1 IgH with shared D-J segment (IGHJ4 - IGHD3-22) but different V segments (IGHV3-69 and IGHV1-73) in the leukemias. All other IgH and immunoglobulin light chain rearrangements identified differed between the twins (supplemental Table 6).

Discussion

The *ETV6-RUNX1* fusion has long been considered an initiating event and disease driver in BCP-ALL.[6] The basis for this view has been its recurrence, its identification in cord blood of healthy neonates, the detection of identical *ETV6-RUNX1* fusion sequences in mz twins with concordant BCP-ALL, and the absence of any other damaging shared events detectable by WGS in concordant cases.[8-22,34,60] In twins with concordant BCP-ALL, a preleukemic clone is believed to arise in 1 twin and spread to the sibling by vascular transfusion in the shared placenta.[10-22] The detection of 58 shared SNVs/indels and, especially, a complex rearrangement t(11;12;21)(q23;p13;q22) with identical breakpoints in leukemias of both twins provided solid molecular evidence of a common in utero origin in this case also (Figure 3). Moreover, detection of a single identical IgH D-J rearrangement and otherwise only divergent IgH V(D)J and immunoglobulin light chain rearrangements in the twins' leukemias (supplemental Table 6) suggested the shared preleukemic clone was initiated in between pro- and pre-B cell stage.

Although detected at diagnosis by FISH analysis, the origin of the *ETV6-RUNX1* fusion in the shared complex rearrangement was revealed in our study by structural variant analysis of WGS data. This adds yet another example of complex rearrangements underlying the classical *ETV6-RUNX1* fusion[61-65] and further emphasizes the advantage of WGS-based applications to accurately detect and in detail characterize relevant genomic aberrations at leukemia diagnosis.[61,63,65]

Chip-based dPCR detected the deletion at 2.6% in the remission bone marrow samples of both twins, suggesting the clone involving the *UBA2* deletion remained in their respective bone marrows after ALL treatment and more than 2 years in complete remission (Figure 3). The variant allele frequency of the *UBA2* deletion was increased from 3% in Tw1 and 6% in Tw2% to 40% in Tw1 and 44% in Tw2 at diagnosis. The expansion of the clone during leukemia strongly supports that the leukemic clone arose in 1 of the cells of this clone and expanded over time. The complex rearrangement, harboring *ETV6-RUNX1*, was undetectable at birth using dPCR. We conclude that the level of this complex rearrangement at birth is below 0.1%, which is the limit of detection for the employed method. As the level of the *UBA2* deletion is at least 30 times more abundant than that of the complex rearrangement at birth, we infer that the *UBA2* variant precedes the complex chromosomal rearrangement. Thus, we suggest the shared complex rearrangement emerged in a del*UBA2*+ cell in 1 twin and then established in the bone marrows of both twins' following a second event of vascular transfusion. Considering the *ETV6-RUNX1* fusion's well-established role as a leukemia initiating event, discovering that the *UBA2* deletion preceded the shared complex rearrangement compelled us to reevaluate the assumption of *ETV6-RUNX1* as the initiating event in our cases.

UBA2 is predicted to be haploinsufficient and thus likely causative of a dominant disease when mutated. It is essential to and highly conserved across many organisms (https://varsome.com/gene/uba2) (accessed January 2021).[66] The *UBA2* frameshift deletion reported here introduced an early stop codon which most likely led to protein truncation and nonsense-mediated decay of any

transcribed product, resulting in heterozygous loss of *UBA2*. Constitutional loss-of-function variants and deletions affecting *UBA2* have recently been associated with congenital malformations.[67-74] No associated cancer predisposition has been reported; however, cases are few, and reduced penetrance of cancer phenotype is likely in dominant disease. Altogether, we assessed our reported deletion as likely pathogenic.

Interestingly, *UBA2* was recently uncovered as a recurrent target gene for somatic aberrations in BCP-ALL.[56,75] In line with our findings, *UBA2* variants have also been reported enriched in *ETV6-RUNX1*[+] cases.[76] However, the temporal relation between *UBA2* variants and the *ETV6-RUNX1* fusion has not been determined previously. To our knowledge, this is the first report of a *UBA2* variant preceding the *ETV6-RUNX1* fusion or any other known recurrent aberration associated to BCP-ALL.

Moreover, as reviewed by Han et al, the UBA2 protein is an essential component in posttranslational protein modification by small ubiquitin-like modifier (SUMO) attachment (SUMOylation).[77] As part of the SUMO E1-activating enzyme heterodimer, UBA2 activates the highly conserved SUMO proteins to exercise their effects on target proteins. SUMOylation regulates crucial cellular processes such as gene expression, cell signaling, DNA damage repair, cell cycle progression, apoptosis, etc, and is known to be exploited by viruses as a result interfering with diverse cellular mechanisms.[78] Consequently, it affects the function and activity of most intracellular pathways and thus nearly all physiological and pathological processes.[77] Importantly, a disrupted balance of SUMOylation has been implicated in cancer, including leukemia.[77,79] Given the somatic heterozygous loss of *UBA2* predicted in our cases, downregulation of SUMO E1-activating enzyme and thus of SUMOylation appeared the most likely functional outcome. Conversely, SUMOylation has been found upregulated in malignant cells of many cancer types,[77] speaking against a cancer-predisposing effect of our detected variant.

Nevertheless, SUMOylation is part of an intricate regulatory network, also cross-talking with other regulatory systems such as acetylation, phosphorylation, and ubiquitination.[80] Also, the contribution to carcinogenesis by dysregulated SUMOylation takes place in a context-dependent manner.[80] For example, functional effects of dysregulated SUMOylation in a cell before and after malignant transformation may differ. In that respect, cancer predisposing and/or promoting effects of dysregulated SUMOylation may encompass different causal relationships. Therefore, we argue that a possible association between downregulated SUMOylation and leukemia development should not yet be dismissed.

To us, the biological example posed by the twins provides the most compelling support for a possible leukemia-predisposing effect of *UBA2* deletion. The likelihood of the *ETV6-RUNX1* fusion arising in a del*UBA2* clone by pure chance appears inarguably small, even more so if the predicted downregulated SUMOylation would exert a protective effect against leukemogenesis. Moreover, other frameshift variants in *UBA2*, also likely causing downregulation of SUMOylation, can be found among somatic variants in B-cell ALL.[55]

SUMOylation has been implicated in both DNA damage response and DNA-repair pathways.[81] For example, SUMOylation is essential for the function of XRCC4, Ku70, and Ku80,[82] all key regulatory factors of nonhomologous end joining (NHEJ), the major repair mechanism for double-stranded breaks in DNA.[83,84] NHEJ is key to V(D)J recombination, the tight regulation of which is essential for normal T- and B-cell development and prevention of oncogenic genetic events.[85,86] V(D)J recombination is mainly driven by RAG endonucleases, the off-target activity of which has been found the main driver of secondary somatic events in *ETV6-RUNX1*[+] BCP-ALL.[87] However, the *ETV6-RUNX1* fusions themselves bear no signs of off-target RAG activity.[60,87] Rather, illegitimate recombination of multiple simultaneous double-stranded breaks by NHEJ is considered the mechanism behind *ETV6-RUNX1* fusion.[88] The above provides 1 theoretical mechanistic explanation to the effects of downregulated SUMOylation on the propensity of forming initiating and secondary genetic events leading to leukemogenesis.

In line with the established notion that additional somatic events are required for progression to overt BCP-ALL of *ETV6-RUNX1*[+] preleukemic clones,[32,33] WGS identified 10 deletions and 2421 SNVs/indels unique to Tw1's leukemia and 10 deletions, 2 translocations, 1 complex rearrangement, and 2955 SNVs/indels unique to Tw2's leukemia. This was well within range of the reported average of 3.5 (range 0-14) copy number variations in pediatric ALL, supporting gross genomic instability is not a common trait. Latency of 3.5 and 4 years, respectively, in the twins corresponded well to overall latency among *ETV6-RUNX1*[+] BCP-ALL.[25,89-91]

Six of the unique deletions in either leukemia were analogous, encompassing the same locus but with different breakpoints, targeting *ETV6, ATF7IP,* and *RAG1/RAG2*. These genes are known recurrent targets of secondary events in *ETV6-RUNX1*[+] BCP-ALL[92,93] and have been shown to predominantly result from aberrant RAG-mediated recombination.[59,87] Also, *RAG1/RAG2* and *ATF7IP* deletions have been reported to frequently cooccur in *ETV6-RUNX1*[+] BCP-ALL.[56,87] Nonetheless, the independent loss of both *ETV6, ATF7IP,* and *RAG1/RAG2* in both leukemias provides evidence of a convergent clonal evolution in the leukemias, which may only be explained by a strong selective pressure, most likely exerted by the genetic and consequential biological circumstances in the shared preleukemic clone.

Another known recurrent secondary event found in both twins was the SNV in *NSD2 (WHSC1)* p.E1099K (COSMIC ID: COSM-379334).[58] Being the most common *NSD2* variant in *ETV6-RUNX1*[+] BCP-ALL,[58] these variants were most likely independently acquired. This *NSD2* variant also frequently cooccurs with variants in *UBA2*.[56]

NF1 is an autosomal-dominant neurocutaneous disorder characterized by highly variable clinical features including multiple café-au-lait spots, neurofibromas, and ocular and skeletal abnormalities.[53] *NF1* being a tumor suppressor gene,[94] the syndrome also entails cancer predisposition to a range of tumors and malignancies, including leukemias.[95,96] In the studied twins, detection of a novel constitutional missense variant in *NF1*, predicted highly damaging but evaluated as a variant of unknown significance due to lack of phenotypic correlation, was intriguing. NF1 was dismissed in our cases due to complete lack of clinical symptoms obligate for this diagnosis.[53] *NF1* LOH in malignant cells of NF1 patients who developed cancer is common, in accordance with Knudson's 2-hit hypothesis, but not obligatory.[56] Also, the gene is considered to be haploinsufficient,[97] reflecting the cancer-predisposing effects of heterozygous loss of *NF1*. Cases with constitutional *NF1* variants and cancer but lacking other clinical signs of NF1 have also been reported.[98] Altogether,

we suggest that the absence of obligate NF1 symptoms and lack of LOH in leukemic cells does not exclude the possibility of a cancer-predisposing effect of our reported variant. Interestingly, a previous study has shown significant pairwise cooccurrence of somatic *NF1* and *UBA2* in childhood BCP-ALL.[56] Could this be an indication that emergence of *UBA2* deletion in the presence of the constitutional *NF1* variant was not a coincidence?

In summary, we report a detailed genetic characterization with partial temporal delineation of some central genetic aberrations in the concordant BCP-ALL of a mz twin pair. A common in utero origin of their preleukemic clone was supported by a shared complex rearrangement with identical breakpoint sequences in both twins and 58 shared SNV's/indels, including a frameshift deletion in *UBA2*. The complex rearrangement generated the well-known recurrent *ETV6-RUNX1* fusion gene, whereas the *UBA2* deletion was predicted to cause heterozygous loss of function. Typical to *ETV6-RUNX1*[+] BCP-ALL, additional copy number variants were required for progression to overt leukemia. Interestingly, these second hits targeted a number of recurrent second-hit genes. Surprisingly, *UBA2* deletion was retained in remission of both twins, providing indirect proof of its emergence preceding the shared complex rearrangement. *UBA2* variants are highly recurrent in *ETV6-RUNX1*[+] BCP-ALL. To our knowledge, this is the first report of a *UBA2* variant preceding the *ETV6-RUNX1* fusion or any other known recurrent aberration associated to BCP-ALL.

Acknowledgments

Authorship

Contribution: F.T., V.Z., B.B., and A.N. designed the study and interpreted the data; F.T. and J.E. contributed to data acquisition, and F.T. prepared the figures; F.T., B.B., and A.N. wrote the manuscript; M.H., G.B., A.H-S., and A.N. collected patient samples; F.T., B.B., M.H., G.B., V.Z., and A.N. contributed to data interpretation; and all authors revised the manuscript and approved the final version.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

ORCID profiles: B.B., 0000-0001-6907-4954; J.E., 0000-0003-3716-4917; G.B., 0000-0003-3185-2962; A.H-S., 0000-0003-2767-5828; M.H., 0000-0001-7637-1949; V.Z., 0000-0001-9360-9859; F.T., 0000-0002-2907-0235; A.N., 0000-0003-3285-4281.

Correspondence: Fulya Taylan, Department of Molecular Medicine and Surgery, Karolinska Institutet, BioClinicum J10:20, 171 64, Stockholm, Sweden; e-mail: fulya.taylan@ki.se.

References

1. Mulligan CG. Molecular genetics of B-precursor acute lymphoblastic leukemia. *J Clin Invest.* 2012;122(10):3407-3415.

2. Lim JY, Bhatia S, Robison LL, Yang JJ. Genomics of racial and ethnic disparities in childhood acute lymphoblastic leukemia. *Cancer.* 2014;120(7): 955-962.

3. Hunger SP, Mulligan CG. Acute lymphoblastic leukemia in children. *N Engl J Med.* 2015;373(16):1541-1552.

4. Chiaretti S, Zini G, Bassan R. Diagnosis and subclassification of acute lymphoblastic leukemia. *Mediterr J Hematol Infect Dis.* 2014;6(1): e2014073.

5. Bhojwani D, Yang JJ, Pui CH. Biology of childhood acute lymphoblastic leukemia. *Pediatr Clin North Am.* 2015;62(1):47-60.

6. Moorman AV. The clinical relevance of chromosomal and genomic abnormalities in B-cell precursor acute lymphoblastic leukaemia. *Blood Rev.* 2012;26(3):123-135.

7. Bloom M, Maciaszek JL, Clark ME, Pui CH, Nichols KE. Recent advances in genetic predisposition to pediatric acute lymphoblastic leukemia. *Expert Rev Hematol.* 2020;13(1):55-70.

8. Ford AM, Ridge SA, Cabrera ME, et al. In utero rearrangements in the trithorax-related oncogene in infant leukaemias. *Nature.* 1993;363(6427): 358-360.

9. Gill Super HJ, Rothberg PG, Kobayashi H, Freeman AI, Diaz MO, Rowley JD. Clonal, nonconstitutional rearrangements of the MLL gene in infant twins with acute lymphoblastic leukemia: in utero chromosome rearrangement of 11q23. *Blood.* 1994;83(3):641-644.

10. Ford AM, Bennett CA, Price CM, Bruin MC, Van Wering ER, Greaves M. Fetal origins of the TEL-AML1 fusion gene in identical twins with leukemia. *Proc Natl Acad Sci USA.* 1998;95(8):4584-4588.

11. Wiemels JL, Ford AM, Van Wering ER, Postma A, Greaves M. Protracted and variable latency of acute lymphoblastic leukemia after TEL-AML1 gene fusion in utero. *Blood.* 1999;94(3):1057-1062.

12. Alpar D, Wren D, Ermini L, et al. Clonal origins of ETV6-RUNX1[+] acute lymphoblastic leukemia: studies in monozygotic twins. *Leukemia.* 2015; 29(4):839-846.

13. Wiemels JL, Cazzaniga G, Daniotti M, et al. Prenatal origin of acute lymphoblastic leukaemia in children. *Lancet.* 1999;354(9189):1499-1503.

14. Yagi T, Hibi S, Tabata Y, et al. Detection of clonotypic IGH and TCR rearrangements in the neonatal blood spots of infants and children with B-cell precursor acute lymphoblastic leukemia. *Blood.* 2000;96(1):264-268.

15. Maia AT, Ford AM, Jalali GR, et al. Molecular tracking of leukemogenesis in a triplet pregnancy. *Blood.* 2001;98(2):478-482.

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

17. Maia AT, Koechling J, Corbett R, Metzler M, Wiemels JL, Greaves M. Protracted postnatal natural histories in childhood leukemia. *Genes Chromosomes Cancer.* 2004;39(4):335-340.

18. Olsen M, Hjalgrim H, Melbye M, Madsen HO, Schmiegelow K. RT-PCR screening for ETV6-RUNX1-positive clones in cord blood from newborns in the Danish National Birth Cohort. *J Pediatr Hematol Oncol.* 2012;34(4):301-303.

19. Ford AM, Greaves M. ETV6-RUNX1[+] acute lymphoblastic leukaemia in identical twins. *Adv Exp Med Biol.* 2017;962:217-228.

20. Schäfer D, Olsen M, Lahnemann D, et al. Five percent of healthy newborns have an ETV6-RUNX1 fusion as revealed by DNA-based GIPFEL screening. *Blood.* 2018;131(7):821-826.

21. Hein D, Dreisig K, Metzler M, et al. The preleukemic TCF3-PBX1 gene fusion can be generated in utero and is present in $\approx$0.6% of healthy newborns. *Blood.* 2019;134(16):1355-1358.

22. Hein D, Borkhardt A, Fischer U. Insights into the prenatal origin of childhood acute lymphoblastic leukemia. *Cancer Metastasis Rev.* 2020;39(1): 161-171.

23. Greaves MF, Maia AT, Wiemels JL, Ford AM. Leukemia in twins: lessons in natural history. *Blood.* 2003;102(7):2321-2333.

24. Bateman CM, Colman SM, Chaplin T, et al. Acquisition of genome-wide copy number alterations in monozygotic twins with acute lymphoblastic leukemia. *Blood.* 2010;115(17):3553-3558.

25. Mullighan CG, Goorha S, Radtke I, et al. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature.* 2007;446(7137): 758-764.

26. Dobbins SE, Sherborne AL, Ma YP, et al. The silent mutational landscape of infant MLL-AF4 pro-B acute lymphoblastic leukemia. *Genes Chromosomes Cancer.* 2013;52(10):954-960.

27. Andersson AK, Ma J, Wang J, et al; St. Jude Children's Research Hospital–Washington University Pediatric Cancer Genome Project. The landscape of somatic mutations in infant MLL-rearranged acute lymphoblastic leukemias. *Nat Genet.* 2015;47(4):330-337.

28. Golub TR, Barker GF, Stegmaier K, Gilliland DG. Involvement of the TEL gene in hematologic malignancy by diverse molecular genetic mechanisms. *Curr Top Microbiol Immunol.* 1996;211:279-288.

29. McLean TW, Ringold S, Neuberg D, et al. TEL/AML-1 dimerizes and is associated with a favorable outcome in childhood acute lymphoblastic leukemia. *Blood.* 1996;88(11):4252-4258.

30. Harbott J, Viehmann S, Borkhardt A, Henze G, Lampert F. Incidence of TEL/AML1 fusion gene analyzed consecutively in children with acute lymphoblastic leukemia in relapse. *Blood.* 1997;90(12):4933-4937.

31. Borkhardt A, Cazzaniga G, Viehmann S, et al; Associazione Italiana Ematologia Oncologia Pediatrica and the Berlin-Frankfurt-Münster Study Group. Incidence and clinical relevance of TEL/AML1 fusion genes in children with acute lymphoblastic leukemia enrolled in the German and Italian multicenter therapy trials. *Blood.* 1997;90(2):571-577.

32. Sun C, Chang L, Zhu X. Pathogenesis of ETV6/RUNX1-positive childhood acute lymphoblastic leukemia and mechanisms underlying its relapse. *Oncotarget.* 2017;8(21):35445-35459.

33. Greaves M. A causal mechanism for childhood acute lymphoblastic leukaemia [published correction appears in *Nat Rev Cancer.* 2018;18(8):526]. *Nat Rev Cancer.* 2018;18(8):471-484.

34. Ma Y, Dobbins SE, Sherborne AL, et al. Developmental timing of mutations revealed by whole-genome sequencing of twins with acute lymphoblastic leukemia. *Proc Natl Acad Sci USA.* 2013;110(18):7429-7433.

35. Bateman CM, Alpar D, Ford AM, et al. Evolutionary trajectories of hyperdiploid ALL in monozygotic twins. *Leukemia.* 2015;29(1):58-65.

36. Cibulskis K, Lawrence MS, Carter SL, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol.* 2013;31(3):213-219.

37. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20(9):1297-1303.

38. McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics.* 2010;26(16):2069-2070.

39. Paila U, Chapman BA, Kirchner R, Quinlan AR. GEMINI: integrative exploration of genetic variation and genome annotations. *PLOS Comput Biol.* 2013;9(7):e1003153.

40. Robinson JT, Thorvaldsdóttir H, Winckler W, et al. Integrative genomics viewer. *Nat Biotechnol.* 2011;29(1):24-26.

41. Martin AR, Williams E, Foulger RE, et al. PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. *Nat Genet.* 2019;51(11):1560-1565.

42. Abyzov A, Urban AE, Snyder M, Gerstein M. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res.* 2011;21(6):974-984.

43. Eisfeldt J, Vezzi F, Olason P, Nilsson D, Lindstrand A. *TIDDIT*, an efficient and comprehensive structural variant caller for massive parallel sequencing data. *F1000 Res.* 2017;6:664.

44. Nadeu F, Mas-de-Les-Valls R, Navarro A, et al. IgCaller for reconstructing immunoglobulin gene rearrangements and oncogenic translocations from whole-genome sequencing in lymphoid neoplasms. *Nat Commun.* 2020;11(1):3390.

45. Tran AN, Taylan F, Zachariadis V, et al. High-resolution detection of chromosomal rearrangements in leukemias through mate pair whole genome sequencing. *PLoS One.* 2018;13(3):e0193928.

46. Järviaho T, Bang B, Zachariadis V, et al. Predisposition to childhood acute lymphoblastic leukemia caused by a constitutional translocation disrupting *ETV6. Blood Adv.* 2019;3(18):2722-2731.

47. Krzywinski M, Schein J, Birol I, et al. Circos: an information aesthetic for comparative genomics. *Genome Res.* 2009;19(9):1639-1645.

48. Taylan F, Bang B, Öfverholm II, et al. Somatic structural alterations in childhood leukemia can be backtracked in neonatal dried blood spots by use of whole-genome sequencing and digital PCR. *Clin Chem.* 2019;65(2):345-347.

49. Karczewski KJ, Francioli LC, Tiao G, et al; Genome Aggregation Database Consortium. The mutational constraint spectrum quantified from variation in 141,456 humans [published corrections appear in *Nature.* 2021;590(7846):E53 and 2021;597(7874):E3-E4]. *Nature.* 2020;581(7809): 434-443.

50. Tate JG, Bamford S, Jubb HC, et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res.* 2019;47(D1):D941-D947.

51. Fokkema IF, Taschner PE, Schaafsma GC, Celli J, Laros JF, den Dunnen JT. LOVD v.2.0: the next generation in gene variant databases. *Hum Mutat.* 2011;32(5):557-563.

52. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet.* 2014;46(3):310-315.

53. Boyd KP, Korf BR, Theos A. Neurofibromatosis type 1. *J Am Acad Dermatol.* 2009;61(1):1-14, quiz 15-16.

54. Mitelman F, Johansson B, Mertens F (eds). Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer. https://mitelmandatabase.isb-cgc.org. Accessed 15 December 2020.

55. McLeod C, Gout AM, Zhou X, et al. St. Jude Cloud: A Pediatric Cancer Genomic Data-Sharing Ecosystem. *Cancer Discov.* 2021;11(5): 1082-1099.

56. Ma X, Liu Y, Liu Y, et al. Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. *Nature.* 2018; 555(7696):371-376.

57. Theda C, Hwang SH, Czajko A, Loke YJ, Leong P, Craig JM. Quantitation of the cellular content of saliva and buccal swab samples. *Sci Rep.* 2018;8(1):6944.

58. Jaffe JD, Wang Y, Chan HM, et al. Global chromatin profiling reveals NSD2 mutations in pediatric acute lymphoblastic leukemia. *Nat Genet.* 2013; 45(11):1386-1391.

59. Swaminathan S, Klemm L, Park E, et al. Mechanisms of clonal evolution in childhood acute lymphoblastic leukemia. *Nat Immunol.* 2015;16(7): 766-774.

60. Greaves MF, Wiemels J. Origins of chromosome translocations in childhood leukaemia. *Nat Rev Cancer.* 2003;3(9):639-649.

61. Reddy KS, Yang X, Mak L, Wang S, Johnston M. A child with ALL and ETV6/AML1 fusion on a chromosome 12 due to an insertion of AML1 and loss of ETV6 from the homolog involved in a t(12;15)(p13;q15). *Genes Chromosomes Cancer.* 2000;29(2):106-109.

62. Jalali GR, Martineau M, Ford AM, Greaves M, Stevens RF, Harrison CJ. A unique variant of ETV6/AML1 fusion in a child with acute lymphoblastic leukemia. *Leukemia.* 2003;17(5):993-995.

63. Jin Y, Wang X, Hu S, Tang J, Li B, Chai Y. Determination of ETV6-RUNX1 genomic breakpoint by next-generation sequencing. *Cancer Med.* 2016; 5(2):337-351.

64. Mathew S, Shurtleff SA, Raimondi SC. Novel cryptic, complex rearrangements involving ETV6-CBFA2 (TEL-AML1) genes identified by fluorescence in situ hybridization in pediatric patients with acute lymphoblastic leukemia. *Genes Chromosomes Cancer.* 2001;32(2):188-193.

65. Nordlund J, Marincevic-Zuniga Y, Cavelier L, et al. Refined detection and phasing of structural aberrations in pediatric acute lymphoblastic leukemia by linked-read whole-genome sequencing. *Sci Rep.* 2020;10(1):2512.

66. Kopanos C, Tsiolkas V, Kouris A, et al. VarSome: the human genomic variant search engine. *Bioinformatics.* 2019;35(11):1978-1980.

67. Wang Y, Dupuis L, Jobling R, Kannu P. Aplasia cutis congenita associated with a heterozygous loss-of-function UBA2 variant. *Br J Dermatol.* 2020;182(3):792-794.

68. Venegas-Vega C, Nieto-Martínez K, Martínez-Herrera A, et al. 19q13.11 microdeletion concomitant with ins(2;19)(p25.3;q13.1q13.4)dn in a boy: potential role of UBA2 in the associated phenotype. *Mol Cytogenet.* 2014;7(1):61.

69. Abe KT, Rizzo IMPO, Coelho ALV, Sakai N Jr, Carvalho DR, Speck-Martins CE. 19q13.11 microdeletion: clinical features overlapping ectrodactyly ectodermal dysplasia-clefting syndrome phenotype. *Clin Case Rep.* 2018;6(7):1300-1307.

70. Yamoto K, Saitsu H, Nishimura G, et al. Comprehensive clinical and molecular studies in split-hand/foot malformation: identification of two plausible candidate genes (LRP6 and UBA2). *Eur J Hum Genet.* 2019;27(12):1845-1857.

71. Melo JB, Estevinho A, Saraiva J, Ramos L, Carreira IM. Cutis Aplasia as a clinical hallmark for the syndrome associated with 19q13.11 deletion: the possible role for UBA2 gene. *Mol Cytogenet.* 2015;8(1):21.

72. Marble M, Guillen Sacoto MJ, Chikarmane R, Gargiulo D, Juusola J. Missense variant in UBA2 associated with aplasia cutis congenita, duane anomaly, hip dysplasia and other anomalies: a possible new disorder involving the SUMOylation pathway. *Am J Med Genet A.* 2017;173(3): 758-761.

73. Chowdhury S, Bandholz AM, Parkash S, et al. Phenotypic and molecular characterization of 19q12q13.1 deletions: a report of five patients. *Am J Med Genet A.* 2014;164A(1):62-69.

74. Schnur RE, Yousaf S, Liu J, et al. UBA2 variants underlie a recognizable syndrome with variable aplasia cutis congenita and ectrodactyly. *Genet Med.* 2021;23(9):1624-1635.

75. Loh ML, Zhang J, Pei D, et al. Whole exome sequencing of pediatric acute lymphoblastic leukemia patients identify mutations in 11 pathways: a report from the Children's Oncology Group. *Blood.* 2016;128(22):455.

76. Roberts KG, Brady SW, Gu Z, et al. The genomic landscape of childhood acute lymphoblastic leukemia. *Blood.* 2019;134(Supplement_1):649.

77. Han ZJ, Feng YH, Gu BH, Li YM, Chen H. The post-translational modification, SUMOylation, and cancer (review) [Review]. *Int J Oncol.* 2018; 52(4):1081-1094.

78. Mattoscio D, Segré CV, Chiocca S. Viral manipulation of cellular protein conjugation pathways: the SUMO lesson. *World J Virol.* 2013;2(2):79-90.

79. Zhao B, Zhang Z, Chen X, et al. The important roles of protein SUMOylation in the occurrence and development of leukemia and clinical implications. *J Cell Physiol.* 2021;236(5):3466-3480.

80. Lee JS, Choi HJ, Baek SH. Sumoylation and its contribution to cancer. *Adv Exp Med Biol.* 2017;963:283-298.

81. Dou H, Huang C, Van Nguyen T, Lu LS, Yeh ET. SUMOylation and de-SUMOylation in response to DNA damage. *FEBS Lett.* 2011;585(18): 2891-2896.

82. Yurchenko V, Xue Z, Sadofsky MJ. SUMO modification of human XRCC4 regulates its localization and function in DNA double-strand break repair. *Mol Cell Biol.* 2006;26(5):1786-1794.

83. Lees-Miller SP, Meek K. Repair of DNA double strand breaks by non-homologous end joining. *Biochimie.* 2003;85(11):1161-1173.

84. Lieber MR, Ma Y, Pannicke U, Schwarz K. Mechanism and regulation of human non-homologous DNA end-joining. *Nat Rev Mol Cell Biol.* 2003; 4(9):712-720.

85. Lieber MR, Ma Y, Pannicke U, Schwarz K. The mechanism of vertebrate nonhomologous DNA end joining and its role in V(D)J recombination. *DNA Repair (Amst).* 2004;3(8-9):817-826.

86. Rooney S, Chaudhuri J, Alt FW. The role of the non-homologous end-joining pathway in lymphocyte development. *Immunol Rev.* 2004;200(1): 115-131.

87. Papaemmanuil E, Rapado I, Li Y, et al. RAG-mediated recombination is the predominant driver of oncogenic rearrangement in ETV6-RUNX1 acute lymphoblastic leukemia. *Nat Genet.* 2014;46(2):116-125.

88. Wiemels JL, Alexander FE, Cazzaniga G, Biondi A, Mayer SP, Greaves M. Microclustering of TEL-AML1 translocation breakpoints in childhood acute lymphoblastic leukemia. *Genes Chromosomes Cancer.* 2000;29(3):219-228.

89. Mullighan CG, Downing JR. Global genomic characterization of acute lymphoblastic leukemia. *Semin Hematol.* 2009;46(1):3-15.

90. Lilljebjörn H, Soneson C, Andersson A, et al. The correlation pattern of acquired copy number changes in 164 ETV6/RUNX1-positive childhood acute lymphoblastic leukemias. *Hum Mol Genet.* 2010;19(16):3150-3158.

91. Borst L, Wesolowska A, Joshi T, et al. Genome-wide analysis of cytogenetic aberrations in ETV6/RUNX1-positive childhood acute lymphoblastic leukaemia. *Br J Haematol.* 2012;157(4):476-482.

92. Raynaud S, Cave H, Baens M, et al. The 12;21 translocation involving TEL and deletion of the other TEL allele: two frequently associated alterations found in childhood acute lymphoblastic leukemia. *Blood.* 1996;87(7):2891-2899.

93. Cavé H, Cacheux V, Raynaud S, et al. ETV6 is the target of chromosome 12p deletions in t(12;21) childhood acute lymphocytic leukemia. *Leukemia.* 1997;11(9):1459-1464.

94. Cichowski K, Jacks T. NF1 tumor suppressor gene function: narrowing the GAP. *Cell.* 2001;104(4):593-604.

95. Shannon KM, O'Connell P, Martin GA, et al. Loss of the normal NF1 allele from the bone marrow of children with type 1 neurofibromatosis and malignant myeloid disorders. *N Engl J Med.* 1994;330(9):597-601.

96. Garicochea B, Giorgi R, Odone VF, Dorlhiac-Llacer PE, Bendit I. Mutational analysis of N-RAS and GAP-related domain of the neurofibromatosis type 1 gene in chronic myelogenous leukemia. *Leuk Res.* 1998;22(11):1003-1007.

97. Inoue K, Fry EA. Haploinsufficient tumor suppressor genes. *Adv Med Biol.* 2017;118:83-122.

98. Gieldon L, Masjkur JR, Richter S, et al. Next-generation panel sequencing identifies *NF1* germline mutations in three patients with pheochromocytoma but no clinical diagnosis of neurofibromatosis type 1. *Eur J Endocrinol.* 2018;178(2):K1-K9.