Frequency and prognostic impact of *ZEB2* H1038 and Q1072 mutations in childhood B-other acute lymphoblastic leukemia

Marketa Zaliova,[1,2,3] Eliska Potuckova,[1,2] Julius Lukes Jr,[1,2] Lucie Winkowska,[1,2] Julia Starkova,[1,2] Iveta Janotova,[3] Lucie Sramkova,[1,2,3] Jan Stary,[2,3] Jan Zuna,[1,2,3] Martin Stanulla,[4] Martin Zimmermann,[4] Beat Bornhauser,[5] Jean-Pierre Bourquin,[5] Cornelia Eckert,[6,7] Gunnar Cario[8] and Jan Trka[1,2,3]

[1]CLIP: Childhood Leukemia Investigation Prague, Prague, Czech Republic; [2]Department of Pediatric Hematology and Oncology, Second Faculty of Medicine, Charles University, Prague, Czech Republic; [3]University Hospital Motol, Prague, Czech Republic; [4]Department of Pediatric Hematology and Oncology, Hannover Medical School, Hannover, Germany; [5]Division of Pediatric Oncology, University Children's Hospital, Zürich, Switzerland; [6]Department of Pediatric Oncology/Hematology, Charité - Universitätsmedizin Berlin, Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany; [7]German Cancer Consortium and German Cancer Research Center, Heidelberg, Germany and [8]Department of Pediatrics, University Medical Center Schleswig-Holstein, Kiel, Germany

Correspondence: MARKETA ZALIOVA - marketa.zaliova@lfmotol.cuni.cz

JAN TRKA -jan.trka@lfmotol.cuni.cz

doi:10.3324/haematol.2020.249094

Frequency and prognostic impact of ZEB2 H1038 and Q1072 mutations in childhood B-other acute lymphoblastic leukemia

Marketa Zaliova, Eliska Potuckova, Julius Lukes Jr., Lucie Winkowska, Julia Starkova, Iveta Janotova, Lucie Sramkova, Jan Stary, Jan Zuna, Martin Stanulla, Martin Zimmermann, Beat Bornhauser, Jean-Pierre Bourquin, Cornelia Eckert, Gunnar Cario, Jan Trka

SUPPLEMENTARY DATA

PATIENTS, MATERIAL AND METHODS

Patients

The discovery cohort of this study included 231 and 36 children (age 1-18 years) with newly diagnosed and relapsed, respectively, B-cell precursor ALL negative for routinely screened aberrations (B-other ALL – ALL negative for hypodiploidy (≤ 44 chromosomes), high hyperdiploidy (51-67 chromosomes), *ETV6-RUNX1*, *BCR-ABL1*, *TCF3-PBX1* and *KMT2A*-involving gene fusions), who were diagnosed and treated in the Czech Republic between November 2002 and December 2017. The cohort of children with newly diagnosed B-other ALL represents 94% of all consecutively diagnosed children with this ALL subtype from the respective time period; in the remaining 6% of children no material was available for this study. These children were treated according to 3 consecutive protocols: ALLIC-BFM 2002 (ClinicalTrials.gov Identifier: NCT00764907), AIEOP-BFM ALL 2000 (ClinicalTrials.gov Identifier: NCT00430118) and AIEOP-BFM ALL 2009 (ClinicalTrials.gov Identifier: NCT01117441); the last two protocols utilized minimal residual disease monitoring for risk stratification. The median follow up time of the children who remained in continuous complete remission was 5.2 years (range 1-15 years). The cohort of children with relapsed B-other ALL represents 97% of all consecutive pediatric B-other ALL relapses from the respective time period, only a single child had no relapse material available for this study. Isolated extramedullary relapses were not excluded from the study. Thirty children are included in both cohorts (the cohort with newly diagnosed and the cohort of relapsed B-other ALL).

Validation cohorts comprised 626 and 102 children (age 1-18 years; single patient was a 20 years-old adult) with newly diagnosed and relapsed ALL, respectively, from Germany. Patients were selected based on the negativity of routinely screened aberrations (hypodiploidy, hyperdiploidy, *ETV6-RUNX1*, *BCR-ABL1*, *TCF3-PBX1* and *KMT2A-AFF1* gene fusions) and availability of material. The children with newly diagnosed ALL were treated on the AIEOP-BFM ALL 2000 (n=235) or AIEOP-BFM ALL 2009 (n=391) protocols. Only children with bone-marrow relapse or combined relapse with morphologically-detectable bone-marrow involvement were included in the validation relapse cohort.

The study was approved by ethical review boards of participating institutions and the informed consent was obtained from the patient's parents in accordance with the Declaration of Helsinki.

Biological samples processing

Mononuclear cells (MNCs) were separated from diagnostic and remission bone marrow aspirates and peripheral blood samples by Ficoll-Paque (Pharmacia, Germany) gradient centrifugation. Total DNA and RNA were isolated from MNCs, and RNA was transcribed into cDNA as a part of the routine sample processing procedure. DNA isolated from samples collected in remission or from separated T lymphocytes was used as germline for whole-exome sequencing.

Analysis of ZEB2 H1038 and Q1072 mutations

The presence of mutations was analyzed at DNA and/or RNA levels using whole-exome sequencing (WES), whole-transcriptome sequencing (RNAseq) and amplicon sequencing of DNA or cDNA (AmpliSeq-DNA, AmpliSeq-cDNA). Supplementary Table 2 shows methods used for the discovery cohorts. The choice of the method was influenced by the availability of already existing WES/RNAseq data and/or of material. Part of the leukemia cases were analyzed by more methods; in all but one case, results obtained by different methods were concordant. In one case, *ZEB2* Q1072R mutation with allele frequency 1.3% was detected by AmpliSeq-cDNA, while it was not detected by WES, RNAseq and AmpliSeq-DNA. Validation cohorts were analyzed by AmpliSeq-DNA.

Whole-exome and whole-transcriptome sequencing, subtype classification

Sequencing and data analysis (variant calling, fusion detection, gene expression analysis and hierarchical clustering analysis) were performed as described previously (Zaliova et al., Haematologica 2019; doi: 10.3324/haematol.2018.204974).

In addition to routinely screened genetic aberrations, *DUX4* rearrangement (r), *ZNF384*r, *MEF2D*r, *NUTM1*r, *TCF3-HLF*, *PAX5* P80R, *IKZF1* N159Y and *BCR-ABL1*-like and *ETV6-RUNX1*-like gene expression signatures were considered as subtype-defining.

Presence of gene expression signatures was analyzed by hierarchical clustering analyses (HCA). For HCA, we pooled 30 and 10 samples of *ZEB2*mut-positive patients from initial manifestations and relapses, respectively, with the cohort of 110 patients with B-other ALL presented in our previous study (Zaliova et al., Haematologica 2019; doi: 10.3324/haematol.2018.204974). To verify co-clustering of BCR-ABL1-like and ETV6-RUNX1-like subtypes with the respective genetic groups, we amended the cohort with 2 BCR-ABL1-positive and 9 ETV6-RUNX1-positive ALL cases.

Deep amplicon sequencing

To perform bi-directional sequencing, part of the *ZEB2* coding region (chr2:145147414-145147587; hg19) spanning over codons H1038 and Q1072 (NM _014795) was amplified from DNA/cDNA in 2 separate single-round PCR reactions using primer pairs listed in Supplementary Table 3. Resulting indexed libraries were sequenced on Ion Torrent PGM using 400bp chemistry according to the manufacturer's instructions (Life Technologies, USA). Fastq files were processed from raw data and reads were mapped to hg19 using Torrent Suite software (Life Technologies). Variant calling was performed using Variant Caller plugin in Torrent Suite software (parameter settings: somatic variant frequency, low stringency). Mapped reads were visually inspected in IGV.

Error rate at the positions of interest was repeatedly analyzed by sequencing of DNA from buffy coats collected from healthy individuals. It reached on average 0.09% and 0.06% (medians 0.05% and 0.07%) for codons H1038 (chr2:145147549-145147551) and Q1072 (chr2: 145147447-145147449), respectively. The median sequencing depths were 10263 and 10966 for codons H1038 and Q1072, respectively (ranges 1230-52681 and 1328-53561 reads). The mutation screening was performed using sensitivity of detection set to 1% mutated allele frequency (which corresponds to 2% cells with the mutation).

Statistical analysis

The Mann-Whitney U test was used to compare numerical parameters. The Fisher exact probability test was used to compare frequencies. The Kaplan-Meier method was used to estimate survival rates, differences were compared with the 2-sided log-rank test. Event-free survival (EFS) was defined as the time from diagnosis to the date of last follow-up in complete remission or to the first event. Events were resistance to therapy (non-response), relapse, secondary neoplasm or death from any cause. Failure to achieve remission due to early death or non-response was considered as events at time zero. Patients lost to follow-up were censored at the time of their withdrawal. Cumulative incidence functions for competing events were estimated according to Kalbfleisch and Prentice and were compared with Gray's test.

SUPPLEMENTARY TABLES

Table 1. Demographic, clinical and genetic characteristics of *ZEB2mut*-positive patients.

Case ID	Cohort	Gender	Age (years)	WBC (x10^9/l)	Treatment (DG)	Risk (DG)	ALL subtype	Outcome	Follow-up time for CCR / Time to event (years)	ZEB2mut in DG	ZEB2mut VAF in DG - DNA	ZEB2mut VAF in DG - cDNA	ZEB2mut in REL	ZEB2mut VAF in REL - DNA	ZEB2mut VAF in REL - cDNA	RNAseq of DG performed?	Additional genetic findings in DG	RNAseq of REL performed?	Additional genetic findings in REL
865	D-DG	M	17	4	ALL IC-BFM 2002	MR	B-other	REL-BM	3,8	H1038R	n.a.	94%	n.a.	n.a.	n.a.	Yes	_	No	n.ap.
1098	D-DG, D-REL	F	14	20	ALL IC-BFM 2002	MR	DUX4r	REL-BM	3,4	H1038R	14%	20%	H1038R	44%	18%	Yes	_	No	n.ap.
1883	D-DG, D-REL	M	12	7	ALL-BFM 2009	HR	B-other	REL-BM	2,1	H1038R	79%	25%	H1038R	67%	93%	Yes	_	Yes	NT5C2 R238W, NT5C2 R367Q
1114	D-DG, D-REL	M	17	16	ALL IC-BFM 2002	MR	DUX4r	REL-EM [3]	2,1	Q1072R	7%	7%	Q1072R	37%	42%	Yes	_	No	n.ap.
2058	D-DG, D-REL	M	3	5	ALL-BFM 2009	SR	B-other	REL-BM+CNS	3,2	Q1072R	18%	22%	Q1072R	30%	37%	Yes	P2RY8-CRLF2	Yes	SETD2 Y1666C
2134	D-DG, D-REL	F	4	10	ALL-BFM 2009	MR	DUX4r	REL-EM [4]	2,6	Q1072R	0%	1%	Q1072R	28%	42%	Yes	IGH-MYC	Yes	_
835	D-DG	F	7	6	ALL IC-BFM 2002	MR	B-other	CCR	9,2	H1038R	n.a.	32%	n.ap.	n.ap.	n.ap.	Yes	ZEB2-TEX41 (of) [5]	No	n.ap.
1323	D-DG	F	16	2	ALL-BFM 2000	MR	DUX4r	CCR	9,0	Q1072K	n.a.	94%	n.ap.	n.ap.	n.ap.	Yes	_	No	n.ap.
1154	D-DG	M	2	19	ALL-BFM 2000	SR	B-other	CCR	6,0	Q1072R	28%	39%	n.ap.	n.ap.	n.ap.	Yes	P2RY8-CRLF2	No	n.ap.
206	D-REL	M	6	8	ALL-BFM 95	SR	B-other	REL-BM	5,3	H1038R	9%	28%	H1038R	86%	na	Yes	ZEB2-GTDC1 (of) [5]	No	n.ap.
961	D-DG, D-REL	M	12	3	ALL IC-BFM 2002	HR	iAMP21	REL-BM	2,5	No	n.ap.	n.ap.	Q1072R	9%	17%	Yes	PAX5-BCAS4, KMT2D R5027*, SETD2 I2482fs	No	n.ap.
1768	D-DG, D-REL	M	1	3	ALL-BFM 2009	SR	B-other	REL-BM	1,3	No	n.ap.	n.ap.	H1038R	41%	43%	Yes	PAX5-PML, P2RY8-CRLF2, CBFA2T3-SLC7A5, JAK1 V658I	Yes	PAX5-PML, P2RY8-CRLF2, CBFA2T3-SLC7A5, NT5C2 K25E
FB62	V-DG, V-REL	M	7	137	ALL BFM 2000	MR	B-other	REL-BM	2,7	H1038R	39%	36%	H1038R	42%	40%	Yes	_	Yes	_
H 46	V-DG, V-REL	F	12	4	ALL BFM 2000	HR	B-other	REL-BM	3,8	H1038R	28%	20%	H1038R	37%	42%	Yes	NRAS Q61K	Yes	CSDE1-ST7L, KRAS A146V
B 150	V-DG	F	5	40	ALL BFM 2000	MR	B-other	REL-BM	1,8	Q1072R	16%	19%	No	n.ap.	n.ap.	Yes	_	Yes	_
MA5	V-DG	F	5	3	ALL-BFM 2009	SR	B-other	REL-CNS	2,0	Q1072R	9%	23%	n.a.	n.a.	n.a.	Yes	ATG4D-PDE4A	No	n.ap.
U 131	V-DG	M	16	9	ALL BFM 2000	MR	B-other	CCR	6,8	H1038R	40%	41%	n.ap.	n.ap.	n.ap.	Yes	FLT3 D839G, L576R, N676K	No	n.ap.
GI9	V-DG	F	9	15	ALL-BFM 2009	SR	B-other	CCR	6,1	H1038R	38%	44%	n.ap.	n.ap.	n.ap.	Yes	IGH-CEBPA	No	n.ap.
RG31	V-DG	M	6	20	ALL-BFM 2009	HR	ZNF384r [2]	CCR	6,0	H1038R	22%	22%	n.ap.	n.ap.	n.ap.	Yes	_	No	n.ap.
WB35	V-DG	F	15	41	ALL-BFM 2009	HR	B-other	CCR	4,3	H1038R	39%	40%	n.ap.	n.ap.	n.ap.	Yes	TP53 P278L, NRAS Y64D	No	n.ap.
WB23	V-DG	F	9	3	ALL-BFM 2009	HR	B-other	Died	0,7	H1038R	42%	44%	n.ap.	n.ap.	n.ap.	Yes	PTPN11 G60V	No	n.ap.
KI15	V-DG	F	9	6	ALL BFM 2000	MR	DUX4r	CCR	10,2	Q1072K	2%	1%	n.ap.	n.ap.	n.ap.	Yes	_	No	n.ap.
B 65	V-DG	F	4	4	ALL-BFM 2009	SR	B-other	CCR	1,7	Q1072R	36%	40%	n.ap.	n.ap.	n.ap.	Yes	P2RY8-CRLF2, JAK2 D873N, TP53 E258G, ZEB2 A1035G [6]	No	n.ap.
B 80	V-DG	F	3	8	ALL-BFM 2009	MR	B-other	CCR	3,3	Q1072R	6%	10%	n.ap.	n.ap.	n.ap.	Yes	P2RY8-CRLF2, CRLF2 F232C, CRLF2 V244M	No	n.ap.
E 46	V-DG	F	3	3	ALL-BFM 2009	SR	B-other	CCR	3,6	Q1072R	7%	30%	n.ap.	n.ap.	n.ap.	Yes	PAX5 E105*	No	n.ap.
HV70	V-DG	F	2	41	ALL-BFM 2009	MR	B-other	CCR	5,1	Q1072R	20%	39%	n.ap.	n.ap.	n.ap.	Yes	P2RY8-CRLF2	No	n.ap.
KI16	V-DG	F	4	35	ALL-BFM 2009	SR	B-other	CCR	6,2	Q1072R	29%	44%	n.ap.	n.ap.	n.ap.	Yes	P2RY8-CRLF2	No	n.ap.
S 29	V-DG	F	3	30	ALL-BFM 2009	MR	B-other	CCR	6,3	Q1072R	4%	7%	n.ap.	n.ap.	n.ap.	Yes	P2RY8-CRLF2, PAX5 V319_P320fs, PAX5 A111T	No	n.ap.
RG51	V-DG	F	9	8	ALL-BFM 2009	HR	TCF3/HLF	Died	0,0	Q1072R	8%	20%	n.ap.	n.ap.	n.ap.	Yes	NRAS G13D	No	n.ap.
B 207	V-REL	M	14	6	ALL BFM 2000	MR	B-other	REL-BM+EM	2,5	H1038R	25%	88%	H1038R	34%	94%	Yes	NRAS G12D	Yes	_
KI63	V-REL	F	20	22	ALL BFM 2000	MR	B-other [1]	REL-BM	0,9	H1038R	36%	n.a.	H1038R	22%	28%	No	n.ap.	Yes	IGH-CEBPA
HV62	V-REL	F	13	111	ALL BFM 2000	HR	DUX4r [1]	REL-BM	3,3	No	n.ap.	n.ap.	H1038R	38%	42%	No	n.ap.	Yes	KRAS G12V

D-discovery, V - validation, DG - initial ALL manifestation, REL - ALL relapse, F -female, M - male, WBC - white blood cell count, SR/MR/HR - standard/medium/high risk, CCR - continuous complete remission, r-rearrangement, BM - bone marrow, EM - extramedullary, CNS - central nervous system, n.a. - not analyzed, n.ap. - not applicable, RNAseq - whole transcriptome sequencing, of -out-of-frame;
1 - Subtype assigned using RNAseq data from relapse; 2 - EP300-ZNF384; 3 - isolated testicular relapse; 4 - isolated relapse in uterus; 5 - the rearranged allele carries H1038R mutation; 6 - affects second ZEB2 allele

Supplementary Table 2. Methods used for *ZEB2* mutation screening in discovery cohorts.

Method				Initial manifestation		Relapse	
WES	RNAseq	AmpliSeq-DNA	AmpliSeq-cDNA	Cases analyzed (n)	*Positive out of analyzed (n)*	Cases analyzed (n)	*Positive out of analyzed (n)*
✓	✓	✓	✓	**1**	*1**	**2**	
✓	✓	✓		**15**		**6**	*1*
✓	✓			**63**		**3**	*2*
✓		✓		**1**		**2**	
	✓	✓	✓	**6**	*5*		
	✓	✓		**5**			
	✓		✓	**4**	*3*		
	✓			**47**			
		✓	✓			**4**	*4*
		✓		**41**		**19**	*1*
			✓	**48**			

** Positive result only by AmpliSeq-cDNA, considered as positive*

Supplementary Table 3. Primers used to prepare libraries for amplicon sequencing.

primer pair 1	P1-Fow	CCTCTCTATGGGCAGTCGGT-GAT-***CCACATCAGTGTCAGATTTGTAAGAAAGC***
	A-Rev	CCATCTCATCCCTGCGTGTCTCCGACTCAG-index-GAT-***CCGCTTGCAGTAGGAATACCTGTG***

primer pair 2	A-Fow	CCATCTCATCCCTGCGTGTCTCCGACTCAGC-index-GAT-***CCGCTTGCAGTAGGAATACCTGTG***
	P1-Rev	CCTCTCTATGGGCAGTCGGT-GAT-***CCACATCAGTGTCAGATTTGTAAGAAAGC***

Dashes separate primer elements: P1 and A adapters, index, key, template specific primer (in bold italics)

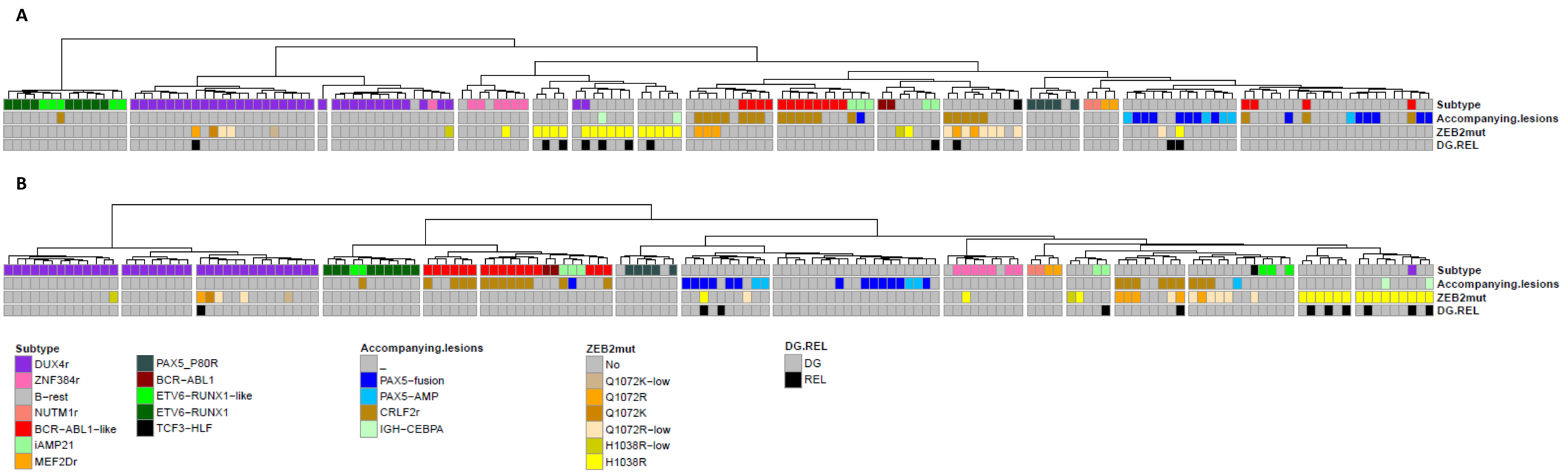
Supplementary Table 4. Ordered list of patients analyzed in unsupervised HCA presented on Figure 2.

Patient ID	DG/REL	ZEB2mut	Accompanying-lesions	Subtype	Left-to-right order in Figure 2	Cluster (out of 16 clusters from left to rigt)
2290*	DG	No	_	DUX4r	1	1
2349*	DG	No	_	DUX4r	2	1
2584*	DG	No	_	DUX4r	3	1
2684*	DG	No	_	DUX4r	4	1
2788*	DG	No	_	DUX4r	5	1
2522*	DG	No	_	DUX4r	6	1
2634*	DG	No	_	DUX4r	7	1
1980*	DG	No	_	DUX4r	8	1
2488*	DG	No	_	DUX4r	9	1
1699*	DG	No	_	DUX4r	10	1
1848*	DG	No	_	DUX4r	11	1
2037*	DG	No	_	DUX4r	12	2
1098	DG	H1038R-low	_	DUX4r	13	2
2441*	DG	No	_	DUX4r	14	2
2062*	DG	No	_	DUX4r	15	2
2650*	DG	No	_	DUX4r	16	2
2613*	DG	No	_	DUX4r	17	2
2673*	DG	No	_	DUX4r	18	2
2350*	DG	No	_	DUX4r	19	2
1638*	DG	No	_	DUX4r	20	2
2731*	DG	No	_	DUX4r	21	2
2134	REL	Q1072R	_	DUX4r	22	3
1323	DG	Q1072K	_	DUX4r	23	3
2065*	DG	No	_	DUX4r	24	3
1114	DG	Q1072R-low	_	DUX4r	25	3
2134*	DG	Q1072R-low	_	DUX4r	26	3
1726*	DG	No	_	DUX4r	27	3
1952*	DG	No	_	DUX4r	28	3
2694*	DG	No	_	DUX4r	29	3
KI15	DG	Q1072K-low	_	DUX4r	30	3
2617*	DG	No	_	DUX4r	31	3
2052*	DG	No	_	DUX4r	32	3
2028*	DG	No	_	DUX4r	33	3
2287*	DG	No	_	DUX4r	34	3
2629*	DG	No	_	DUX4r	35	3
2212*	DG	No	_	ETV6-RUNX1	36	4
1552*	DG	No	_	ETV6-RUNX1-like	37	4
2640*	DG	No	CRLF2r	ETV6-RUNX1-like	38	4
1674*	DG	No	_	ETV6-RUNX1	39	4
2466*	DG	No	_	ETV6-RUNX1-like	40	4
2046*	DG	No	_	ETV6-RUNX1-like	41	4
912*	DG	No	_	ETV6-RUNX1	42	4
1127*	DG	No	_	ETV6-RUNX1	43	4
1236*	DG	No	_	ETV6-RUNX1	44	4
1640*	DG	No	_	ETV6-RUNX1	45	4
2475*	DG	No	_	ETV6-RUNX1	46	4
2155*	DG	No	_	ETV6-RUNX1-like	47	4
1738*	DG	No	_	ETV6-RUNX1	48	4
2534*	DG	No	_	ETV6-RUNX1	49	4
B150	DG	Q1072R-low	_	B-rest	50	5
B150	REL	No	_	B-rest	51	5
1641*	DG	No	PAX5-AMP	B-rest	52	5
2544*	DG	No	PAX5-fusion	B-rest	53	5
1768	REL	H1038R	PAX5-fusion	B-rest	54	5
2548*	DG	No	PAX5-fusion	B-rest	55	5
2564*	DG	No	PAX5-fusion	B-rest	56	5
2359*	DG	No	PAX5-AMP	B-rest	57	5
2701*	DG	No	PAX5-fusion	B-rest	58	5
2760*	DG	No	PAX5-AMP	B-rest	59	5
2097*	DG	No	_	B-rest	60	5
1645*	DG	No	_	MEF2Dr	61	5
2651*	DG	No	_	MEF2Dr	62	5
1722*	DG	No	_	B-rest	63	5
2730*	DG	No	_	B-rest	64	5
2486*	DG	No	_	B-rest	65	5
2606*	DG	No	_	B-rest	66	5
2543*	DG	No	CRLF2r	BCR-ABL1-like	67	5
2103*	DG	No	_	BCR-ABL1-like	68	5
1682*	DG	No	PAX5-AMP	B-rest	69	5
2663*	DG	No	PAX5-fusion	B-rest	70	5
2319*	DG	No	PAX5-fusion	B-rest	71	5
2578*	DG	No	PAX5-fusion	B-rest	72	5
1929*	DG	No	_	B-rest	73	5
2621*	DG	No	PAX5-fusion	B-rest	74	5
1826*	DG	No	PAX5-fusion	B-rest	75	5
2460*	DG	No	_	B-rest	76	5
2524*	DG	No	_	B-rest	77	5
2727*	DG	No	_	B-rest	78	5
2596*	DG	No	CRLF2r	BCR-ABL1-like	79	5
2618*	DG	No	_	B-rest	80	5
1794*	DG	No	CRLF2r	BCR-ABL1-like	81	6
1741*	DG	No	_	BCR-ABL1-like	82	6
961	DG	No	PAX5-fusion	iAMP21	83	6
2724*	DG	No	CRLF2r	iAMP21	84	6
1551*	DG	No	CRLF2r	BCR-ABL1-like	85	6
2479*	DG	No	CRLF2r	BCR-ABL1-like	86	6
2078*	DG	No	CRLF2r	BCR-ABL1-like	87	6
2689*	DG	No	CRLF2r	BCR-ABL1-like	88	6
1838*	DG	No	CRLF2r	BCR-ABL1-like	89	6
2623*	DG	No	CRLF2r	BCR-ABL1-like	90	6
2058*	DG	Q1072R-low	CRLF2r	B-rest	91	7
2058	REL	Q1072R	CRLF2r	B-rest	92	7
1154	DG	Q1072R	CRLF2r	B-rest	93	7
1755*	DG	No	_	B-rest	94	7
FB62	DG	H1038R	_	B-rest	95	7
FB62	REL	H1038R	_	B-rest	96	7
U131	DG	H1038R	_	B-rest	97	7
2001*	DG	No	_	B-rest	98	7
2509*	DG	No	_	BCR-ABL1-like	99	7
1886*	DG	No	_	PAX5_P80R	100	8
2607*	DG	No	_	PAX5_P80R	101	8
2221*	DG	No	_	PAX5_P80R	102	8
2677*	DG	No	_	PAX5_P80R	103	8
2779*	DG	No	_	B-rest	104	8
2026*	DG	No	_	PAX5_P80R	105	8
1852*	DG	No	_	NUTM1r	106	9
2068*	DG	No	_	NUTM1r	107	9
1768*	DG	No	PAX5-fusion	B-rest	108	9
2415*	DG	No	PAX5-fusion	B-rest	109	9
1846*	DG	No	CRLF2r	B-rest	110	9
2377*	DG	No	_	B-rest	111	9
E46	DG	Q1072R-low	_	B-rest	112	10
MA5	DG	Q1072R-low	_	B-rest	113	10
G51	DG	Q1072R-low	_	TCF3-HLF	114	10
HV70	DG	Q1072R	CRLF2r	B-rest	115	10
KI16	DG	Q1072R	CRLF2r	B-rest	116	10
B80	DG	Q1072R-low	CRLF2r	B-rest	117	10
B65	DG	Q1072R	CRLF2r	B-rest	118	10
S29	DG	Q1072R-low	CRLF2r	B-rest	119	10
2515*	DG	No	_	B-rest	120	11
2188*	DG	No	PAX5-fusion	B-rest	121	11
2328*	DG	No	PAX5-AMP	B-rest	122	11
1867*	DG	No	CRLF2r	BCR-ABL1-like	123	11
2042*	DG	No	CRLF2r	BCR-ABL1-like	124	11
GI9	DG	H1038R	IGH-CEBPA	B-rest	125	12
WB35	DG	H1038R	_	B-rest	126	12
WB23	DG	H1038R	_	B-rest	127	12
1633*	DG	No	_	BCR-ABL1-like	128	12
2368*	DG	No	_	BCR-ABL1-like	129	12
2720*	DG	No	_	B-rest	130	13
2612*	DG	No	_	ZNF384r	131	13
1861*	DG	No	_	B-rest	132	13
2786*	DG	No	_	ZNF384r	133	13
1584*	DG	No	_	ZNF384r	134	13
1847*	DG	No	_	ZNF384r	135	13
2473*	DG	No	_	iAMP21	136	14
1733	REL	No	_	iAMP21	137	14
1678*	DG	No	_	iAMP21	138	14
1915*	DG	No	_	B-rest	139	14
2299*	DG	No	_	B-rest	140	14
206	DG	H1038R-low	_	B-rest	141	14
865	DG	H1038R	_	B-rest	142	14
1304*	DG	No	_	BCR-ABL1	143	14
2726*	DG	No	_	BCR-ABL1	144	14
G31	DG	H1038R	_	ZNF384r	145	14
2141*	DG	No	_	ZNF384r	146	14
2208*	DG	No	_	ZNF384r	147	14
HV62	REL	H1038R	_	DUX4r	148	15
H46	DG	H1038R	_	B-rest	149	15
KI63	REL	H1038R	IGH-CEBPA	B-rest	150	15
835	DG	H1038R	_	B-rest	151	15
H46	REL	H1038R	_	B-rest	152	15
B207	DG	H1038R	_	B-rest	153	16
B207	REL	H1038R	_	B-rest	154	16
1883*	DG	H1038R	_	B-rest	155	16
1883	REL	H1038R	_	B-rest	156	16

* included in Zaliova et al., Haematologica 2019 (doi: 10.3324/haematol.2018.204974)
ZEB2mut with VAF <20% annotated as "low"

Supplementary Figure 1. Hierarchical clustering based on expression of genes defining *ETV6-RUNX1*-like and *BCR-ABL1*-like gene expression signature.

Hierarchical clustering (ward.D method and Euclidean distance linkage) was performed using gene sets defining *ETV6-RUNX1*-like (A) and *BCR-ABL1*-like (B) subtypes described in Zaliova et al. (Haematologica 2019; doi: 10.3324/haematol.2018.204974). Figure shows resulting dendrograms.
DG – diagnosis, REL – relapse, r – rearrangement, low – variant allele frequency at the DNA level (or cDNA level if DNA not analyzed) < 20%



Supplementary Figure 2. Scheme of the ZEB2 protein and its aberrations.

Figure shows positions of codons H1038 and Q1072 in C-terminal Zinc fingers and point of junction to two identified fusion partners.
Both *TEX41* and *GTDC1* genes are direct neighbors of *ZEB2* (*GTDC1* in centromeric and *TEX41* in telomeric direction). The *GTDC1* gene has the same orientation as *ZEB2* and the fusion (transcript fusion point corresponds to the positions chr2:45147126 and chr2:144832796 at the genomic level) likely results from a deletion. The *TEX41* gene has the opposite orientation and, thus, the fusion (transcript fusion point corresponds to the positions chr2:145147280 and chr2:145552498 at the genomic level) likely results from a more complex rearrangement.
The *ZEB2-TEX41* transcript is predicted to be translated into an aberrant protein where the last 103 amino acids of ZEB2 are replaced by 10 amino acids translated by readthrough into the *TEX41* intron on the non-coding strand. Similarly, the last 51 ZEB2 amino acids are replaced by 4 amino acids resulting from readthrough into the *GTDC1* intron in an aberrant protein potentially resulting from the *ZEB2-GTDC1* fusion.