Characterization of *CEBPA* mutations and promoter hypermethylation in pediatric acute myeloid leukemia

Iris H.I.M. Hollink,[1] Marry M. van den Heuvel-Eibrink,[1] Susan T.C.J.M. Arentsen-Peters,[1] Martin Zimmermann,[2] Justine K. Peeters,[3] Peter J.M. Valk,[3] Brian V. Balgobind,[1] Edwin Sonneveld,[4] Gertjan J.L. Kaspers,[4,5] Eveline S.J.M. de Bont,[6] Jan Trka,[7] Andre Baruchel,[8] Ursula Creutzig,[9] Rob Pieters,[1] Dirk Reinhardt,[2] and C. Michel Zwaan[1]

[1]Pediatric Oncology/Hematology, Erasmus MC-Sophia Children's Hospital, Rotterdam, the Netherlands; [2]AML-BFM Study Group, Pediatric Hematology/Oncology, Medical School Hannover, Germany; [3]Hematology, Erasmus University Medical Center, Rotterdam, the Netherlands; [4]Dutch Childhood Oncology Group (DCOG), The Hague, the Netherlands; [5]Pediatric Oncology/Hematology, VU University Medical Center, Amsterdam, the Netherlands; [6]Pediatric Oncology/Hematology, Beatrix Children's Hospital, University Medical Center Groningen, the Netherlands; [7]Pediatric Hematology/Oncology, 2nd Medical School, Charles University, Prague, Czech Republic; [8]Hematology, Hôpital Saint-Louis, Paris, France, and [9]AML-BFM Study Group, Pediatric Hematology/Oncology, University Hospital, Münster, Germany

Citation: Hollink IHIM, van den Heuvel-Eibrink MM, Arentsen-Peters STCJM, Zimmermann M, Peeters JK, Valk PJM, Balgobind BV, Sonneveld E, Kaspers GJL, de Bont ESJM, Trka J, Baruchel A, Creutzig U, Pieters R, Reinhardt D, and Zwaan CM. Characterization of CEBPA *mutations and promoter hypermethylation in pediatric acute myeloid leukemia. Haematologica 2011;96(3):384-392. doi:10.3324/haematol.2010.031336*

Online Supplementary Table S1. Primer sequences and PCR conditions used for *CEBPA* mutational screening.

Name	Sequence	PCR conditions
Primer 1	5'-CGCCATGCCGGGAGAACTCT-3'	10' 95ºC, 35 cycles of 1' 95ºC, 1' 60ºC and 1' 72ºC, 10' 72ºC
Primer 10	5'-CTTGGCTTCATCCTCCTCGC-3'	
Primer 4	5'-CGGCCGCTGGTGATCAAG-3'	10' 95ºC, touchdown of 20 cycles of 1' 95ºC, 1' 70-60ºC and 1' 72ºC plus 20 cycles of 1' 95ºC,
Primer 8	5'-CCAGGGCGGTCCCACAGC-3'	1' 60ºC and 1' 72ºC, 10' 72ºC

Online Supplementary Table S2. Detailed characteristics of *CEBPA*-mutant cases.

ID	Mutation(s)	Protein change(s)	# Mut.	Mutation type	Age (years)	Sex	WBC (x10⁹/L)	FAB	Karyotype	Molecular aberration	Survival analysis	Therapy protocol	Follow-up
#3401	c.276_306del31bp + c.937_939dupAAG	p.K92fsX150 + p.K313dup	double	N-term frame shift + in-frame ins bZIP	5.8	m	86	2	CN	*WT1*	yes	DCOG97	CCR for 7.5 yrs
#3412	c.69dupC + c.897_926dup30bp	p.23fsX107 + p.V308_E309ins10aa	double	N-term frame shift + in-frame ins bZIP	13.4	f	6	2	unknown	-	yes	DCOG/BFM87	CCR for 7.9 yrs
#3439	c.345_346insC + c.901_918dup18bp	p.G116fsX170 + p.D301_R306dup	double	N-term frame shift + in-frame ins bZIP	12	m	78.8	1	other	-	yes	DCOG/BFM87	Relapse at 2.7 yrs, CCR for 4.6 yrs
#3472	c.64_70del7bp + c.937_939dupAAG	p.P22fsX158 + p.K313dup	double	N-term frame shift + in-frame ins bZIP	15.6	m	388.3	2	CN	*FLT3*/ITD	yes	BFM98	Relapse at 0.8 yrs, died at 2.7 yrs
#4396	c.332_339del8bp + c.912_913insTTG	p.A111fsX167 + p.K304_Q305insL	double	N-term frame shift + in-frame ins bZIP	14.8	m	54	1	other	-	yes	BFM04	CCR for 2.5 yrs
#4445	c.319_322delGACT + C.913_945dup33bp	p.D107fsX159 + p.Q305_L315dup	double	N-term frame shift + in-frame ins bZIP	8.4	f	8.2	2	CN	-	yes	BFM98	CCR for 7.9 yrs
#4714	c.247delC + c.937_939dupAAG	p.Q83fsX160 + p.K313dup	double	N-term frame shift + in-frame ins bZIP	12.6	m	32.8	1	CN	-	yes	DCOG97	CCR for 3.8 yrs
#4746	c.69dupC + c.937_939dupAAG	p.23fsX107 + p.K313dup	double	N-term frame shift + in-frame ins bZIP	6.9	f	181.5	1	other	*FLT3*/ITD	yes	DCOG97	Died in CCR at 1.2 yrs
#5013	c.206_210dup5bp + c.937_939dupAAG	p.A71fsX162 + p.K313dup	double	N-term frame shift + in-frame ins bZIP	13.2	f	NA	2	CN	-	no	-	
#5047	c.351_352delGC + c.912_920dup9bp	p.G117fsX169 + p.K304_R306dup	double	N-term frame shift + in-frame ins bZIP	8.8	f	NA	1	other	*FLT3*/ITD	no	-	
#5061	c.198dupC + c.926_928dupAGA	p.Y67fsX107 + p.E309_T310insK	double	N-term frame shift + in-frame ins bZIP	4	m	NA	1	CN	*WT1*+N-*RAS*	no	-	
#5063	c.344_348del5bp + c.937_939dupAAG	p.P115fsX168 + p.K313dup	double	N-term frame shift + in-frame ins bZIP	14.6	f	NA	2	CN	*WT1*	no	-	
#5169	c.330_337del8bp + c.896_928dup33bp	p.G110fsX167 + p.S299_E309dup	double	N-term frame shift + in-frame ins bZIP	18.5	m	43.1	2	CN	-	yes	BFM04	Relapse at 2.3 yrs, CCR for 1.3 yrs
#4936	c.68_78del11bp + c.594dupC	p.23fsX103 + p.A199fsX321	double	N-term frame shift + frame shift before bZIP	9.6	m	66	2	other	N-*RAS*	yes	BFM98	CCR for 9.0 yrs
#1885	c.937_939dupAAG	p.K313dup	single	In-frame ins bZIP	2.3	m	15	4	CN	*WT1*+N-*RAS*	yes	DCOG97	CCR for 5.0 yrs
#4734	c.945_946insCAG	p.L315_E316insQ	single	In-frame ins bZIP	4.9	m	20	0	other	-	yes	DCOG97	Relapse at 1.3 yrs, died at 1.9 yrs
#4747	c.917_934del18bp	p.R306_Q311del	single	In-frame del bZIP	3.8	f	7.8	1	other	-	yes	DCOG97	CCR for 1.0 yrs
#5041	c.937_939dupAAG	p.K313dup	single	In-frame ins bZIP	13.6	f	NA	1	unknown	-	no	-	
#3465	c.382_383dupCC	p.P128fsX161	single	Frame shift TAD2	10	f	534.6	2	CN	*FLT3*/ITD	yes	BFM98	NR, died at 1.3 yrs
#4096	c.707_713dup7bp	p.A238fsX323	single	Frame shift before bZIP	12.9	f	214	4	other	*WT1*	yes	DCOG97	Relapse at 0.8 yrs, died at 1.6 yrs

CN: cytogenetically normal; CCR: continuous complete remission; NR: no response.

Online Supplementary Table S3. Characteristics of the study cohort included in the survival analysis (n=185).

	All	*CEBPA* single mutation	*CEBPA* double mutation	*CEBPA* wild-type	*P* value
Number	185	5	10	170	
Age, median (range) (yrs)	9.6 (0.0-18.8)	4.9 (2.3-12.9)	12.3 (5.8-18.5)	9.6 (0.0-18.8)	0.18
<3 yr, n (%)	38 (21%)	1 (20%)	-	37 (22%)	
≥3 yr, n (%)	147 (79%)	4 (80%)	10 (100%)	133 (78%)	0.25
Gender (% female)	43%	60%	30%	39%	0.53
WBC (x10⁹/L), median (range)	42 (1-535)	20 (8-535)	60 (6-388)	42 (1-483)	0.75
FAB classification, n(%)					0.10
M0	9 (5%)	1 (20%)	-	8 (5%)	
M1	21 (12%)	1 (20%)	4 (40%)	16 (10%)	
M2	48 (26%)	1 (20%)	6 (60%)	41 (24%)	
M3	3 (2%)	-	-	3 (2%)	
M4	50 (27%)	2 (40%)	-	48 (29%)	
M5	43 (24%)	-	-	43 (26%)	
M6	3 (2%)	-	-	3 (2%)	
M7	5 (3%)	-	-	5 (3%)	
other	1 (1%)	-	-	1 (1%)	
unknown	2 (1%)	-	-	2 (1%)	
Karyotype, n(%)					0.13
t(8;21)	26 (14%)	-	-	26 (15%)	
inv(16)	24 (13%)	-	-	24 (14%)	
t(15;17)	-	-	-	-	
11q23	36 (20%)	-	-	36 (21%)	
normal	46 (25%)	2 (40%)	5 (50%)	39 (23%)	
other	38 (21%)	3 (60%)	4 (40%)	31 (18%)	
unknown	12 (7%)	-	1 (10%)	11 (7%)	
FLT3/ITD, n(%) (n=185)	34 (18%)	1 (20%)	2 (20%)	31 (18%)	0.99
N- or *K-RAS*, n(%) (n=184)	41 (22%)	1 (20%)	1 (10%)	39 (23%)	0.62
c-KIT , n(%) (n=184)	16 (9%)	-	-	16 (9%)	0.46
MLL-PTD, n(%) (n=180)	3 (2%)	-	-	3 (2%)	0.87
NPM1, n(%) (n=182)	17 (9%)	-	-	17 (10%)	0.43
WT1, n(%) (n=185)	19 (10%)	2 (40%)	1 (10%)	16 (9%)	0.09

WBC: white blood cell count at diagnosis; FAB: French-American British.

Online Supplementary Table S4. Top 50 most significantly differentially expressed genes of the *CEBPA* cluster *versus* all other clusters. Probe sets are ordered from most significant (n.1) to least significant (n.50).

N.	Probe set	Gene Symbol	Gene Title
1	1555630_a_at	RAB34	RAB34, member RAS oncogene family
2	209191_at	TUBB6	tubulin, beta 6
3	214049_x_at	CD7	CD7 antigen (p41)
4	214551_s_at	CD7	CD7 antigen (p41)
5	219541_at	LIME1	Lck interacting transmembrane adaptor 1
6	225662_at	ZAK	sterile alpha motif and leucine zipper containing kinase AZK
7	225665_at	ZAK	sterile alpha motif and leucine zipper containing kinase AZK
8	206660_at	IGLL1	immunoglobulin lambda-like polypeptide 1
9	223724_s_at	DKFZP434A0131 /// LOC442582	DKFZp434A0131 protein /// STAG3-like
10	200765_x_at	CTNNA1	catenin (cadherin-associated protein), alpha 1, 102kDa
11	209185_s_at	IRS2	insulin receptor substrate 2
12	210844_x_at	CTNNA1	catenin (cadherin-associated protein), alpha 1, 102kDa
13	203542_s_at	KLF9	Kruppel-like factor 9
14	209184_s_at	IRS2	insulin receptor substrate 2
15	210448_s_at	P2RX5	purinergic receptor P2X, ligand-gated ion channel, 5
16	231982_at	LOC284422	similar to HSPC323
17	222786_at	CHST12	carbohydrate (chondroitin 4) sulfotransferase 12
18	202252_at	RAB13	RAB13, member RAS oncogene family
19	205383_s_at	ZBTB20	zinc finger and BTB domain containing 20
20	217143_s_at	TRA@ /// TRD@	T cell receptor alpha locus /// T cell receptor delta locus
21	214835_s_at	SUCLG2	succinate-CoA ligase, GDP-forming, beta subunit
22	235308_at	ZBTB20	zinc finger and BTB domain containing 20
23	227423_at	LRRC28	leucine rich repeat containing 28
24	229949_at	TRIM50A	DKFZp434A0131 protein
25	203543_s_at	KLF9	Kruppel-like factor 9
26	211846_s_at	PVRL1	poliovirus receptor-related 1 (herpesvirus entry mediator C; nectin)
27	213830_at	TRD@	T cell receptor delta locus
28	216191_s_at	TRA@ /// TRD@	T cell receptor alpha locus /// T cell receptor delta locus
29	218872_at	TSC	hypothetical protein FLJ20607
30	224710_at	RAB34	RAB34, member RAS oncogene family
31	231929_at	IKZF2	IKAROS family zinc finger 2 (Helios)
32	231935_at	ARPP-21	cyclic AMP-regulated phosphoprotein, 21 kD
33	1563468_at	ZAK	Hypothetical protein LOC339751
34	1556599_s_at	ARPP-21	cyclic AMP-regulated phosphoprotein, 21 kD
35	200764_s_at	CTNNA1	catenin (cadherin-associated protein), alpha 1, 102kDa
36	202242_at	TSPAN7	tetraspanin 7
37	205382_s_at	CFD	D component of complement (adipsin)
38	213156_at	---	MRNA; cDNA DKFZp586B211 (from clone DKFZp586B211)
39	213779_at	EMID1	EMI domain containing 1
40	240084_at	CBX2	chromobox homolog 2 (Pc class homolog, Drosophila)
41	202241_at	TRIB1	tribbles homolog 1 (Drosophila)
42	227121_at	---	MRNA; cDNA DKFZp586K1922 (from clone DKFZp586K1922)
43	208683_at	CAPN2	calpain 2, (m/II) large subunit
44	216268_s_at	JAG1	jagged 1 (Alagille syndrome)
45	218927_s_at	CHST12	carbohydrate (chondroitin 4) sulfotransferase 12
46	214298_x_at	SEPT6	septin 6
47	238151_at	AFG3L2	AFG3 ATPase family gene 3-like 2 (yeast)
48	212459_x_at	SUCLG2	succinate-CoA ligase, GDP-forming, beta subunit
49	201874_at	MPZL1	myelin protein zero-like 1
50	203233_at	IL4R	interleukin 4 receptor

Online Supplementary Table S5. Immunophenotype and gene expression of T-cell associated genes in the *CEBPA*-silenced cases.

Immunophenotype	#3451	#3496	#4728	#4736	#5033
T-lymphoid antigens					
CD7	+	NA	+	+	+
SmCD3	-	NA	-	-	-
CyCD3	-	NA	-	+	-
CD2	-	NA	-	+	-
TdT	-	NA	NA	NA	-
Myeloid antigens					
CD13	+	NA	-	+	+
CD33	+	NA	+	+	+
CD14	-	NA	-	-	-
CD65	-	NA	-	-	NA
CD11b	-	NA	+	+	NA
cyMPO	-	NA	-	-	-
Stem-cell antigens					
CD34	+	NA	+	+	+
CD117	+	NA	+	+ partly	+
CD133	NA	NA	+	+	NA
Other antigens					
CD45	+	NA	+	+	+
CD56	-	NA	-	-	-
Gene expression					
Gene (probe set ID)					
CD3D (213539_at)	1479.5	1248.5	1019.4	2129.4	1129.6
CD3G (206804_at)	96.,4	50.5	54.6	108.6	86.8
CD3Z (210031_at)	106.5	116.1	182.1	169.2	86.5
CD7 (214551_s_at)	227.9	364.2	386.8	618.9	516.1
LCK (204891_s_at)	228.7	110.4	175.4	716.1	167.7
NOTCH1 (218902_at)	188.1	145.3	276.5	279.2	139.6
TRD@ (213830_at)	158.9	114.1	1328.3	1055.9	92.9

NA indicates not available

Online Supplementary Table S6. Confusion matrix and prediction values for the 237 cases based on the previously described 21-probe set gene signature to predict *CEBPA* double-mutant cases (A) and the 9-probe set gene signature to predict *CEBPA*-silenced cases (B).

A

		Predicted			
		Wild-type	Double	Silenced	Single
Actual	Wild-type	209	0	6	1
	Double	0	10	1	1
	Silenced	1	0	4	0
	Single	1	1	0	2

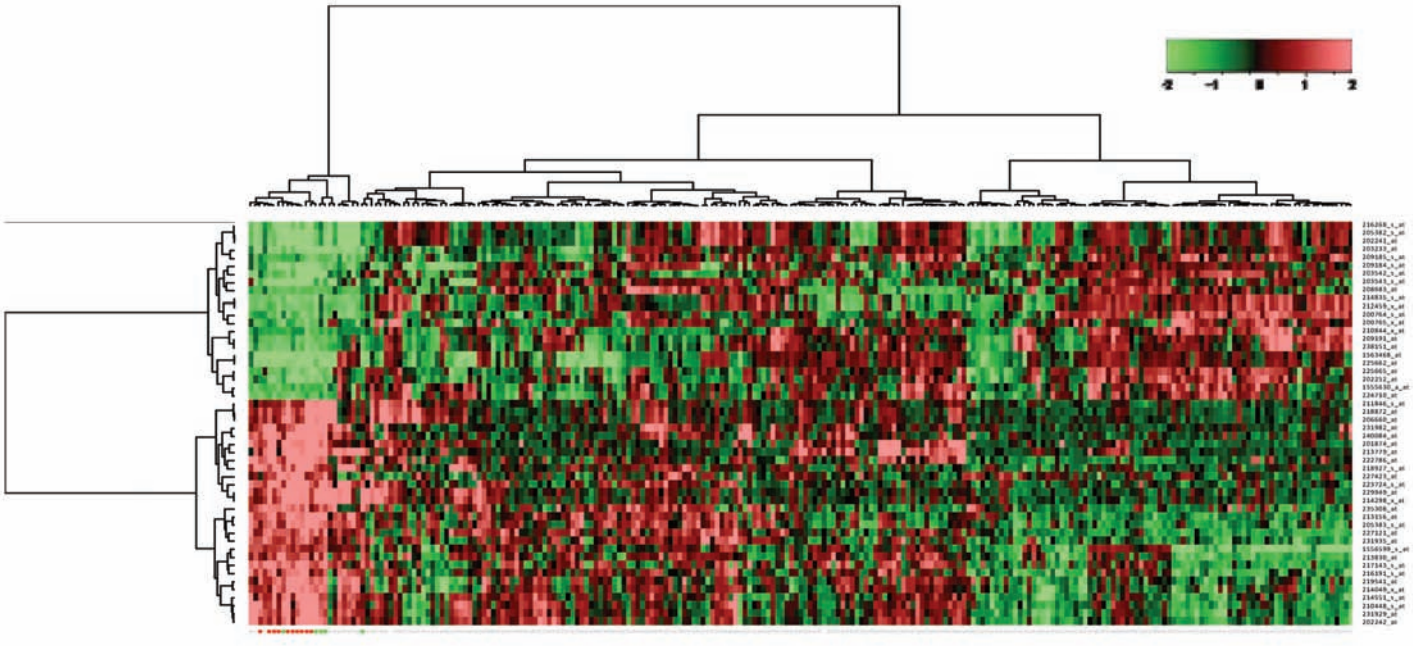
Sample	Sensitivity	Specificity	PPV	NPV
Wild-type	0.96	0.96	0.99	0.76
Double	0.83	0.99	0.90	0.99
Silenced	0.80	0.96	0.36	0.99
Single	0.50	0.99	0.50	0.99

B

		Predicted			
		Wild-type	Double	Silenced	Single
Actual	Wild-type	199	14	1	2
	Double	2	10	0	0
	Silenced	0	2	3	0
	Single	1	1	0	2

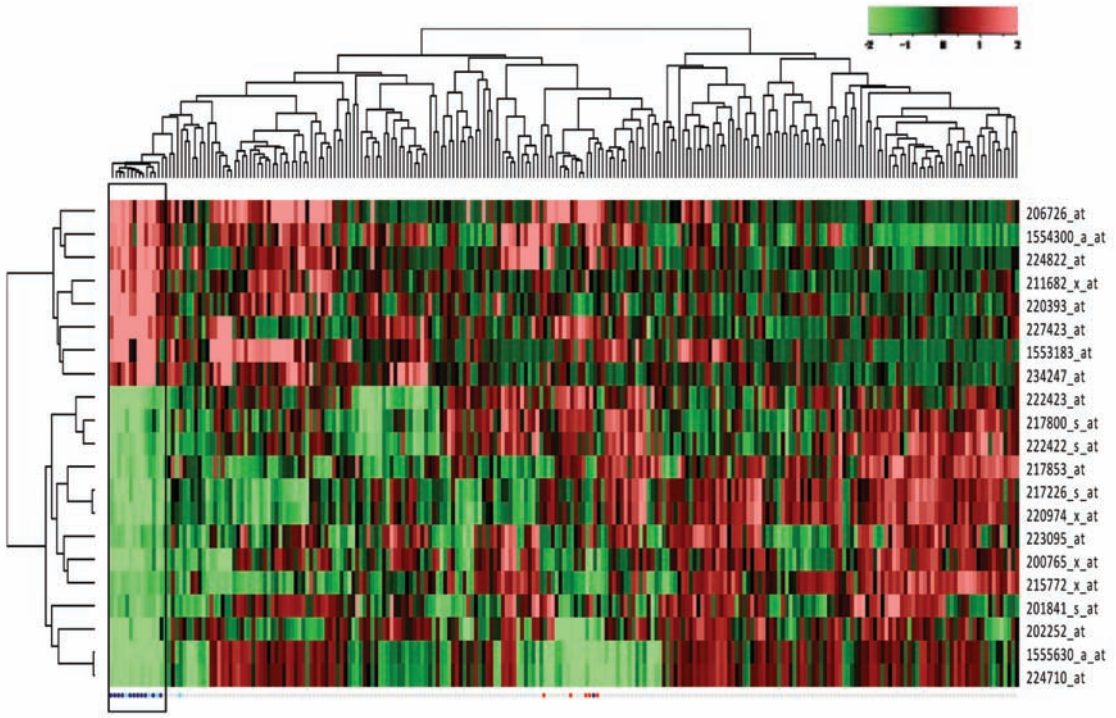
Sample	Sensitivity	Specificity	PPV	NPV
Wild-type	0.92	0.85	0.98	0.51
Double	0.83	0.92	0.37	0.99
Silenced	0.60	0.99	0.75	0.99
Single	0.50	0.99	0.50	0.99

PPV: positive predictive value; NPV: negative predictive value.



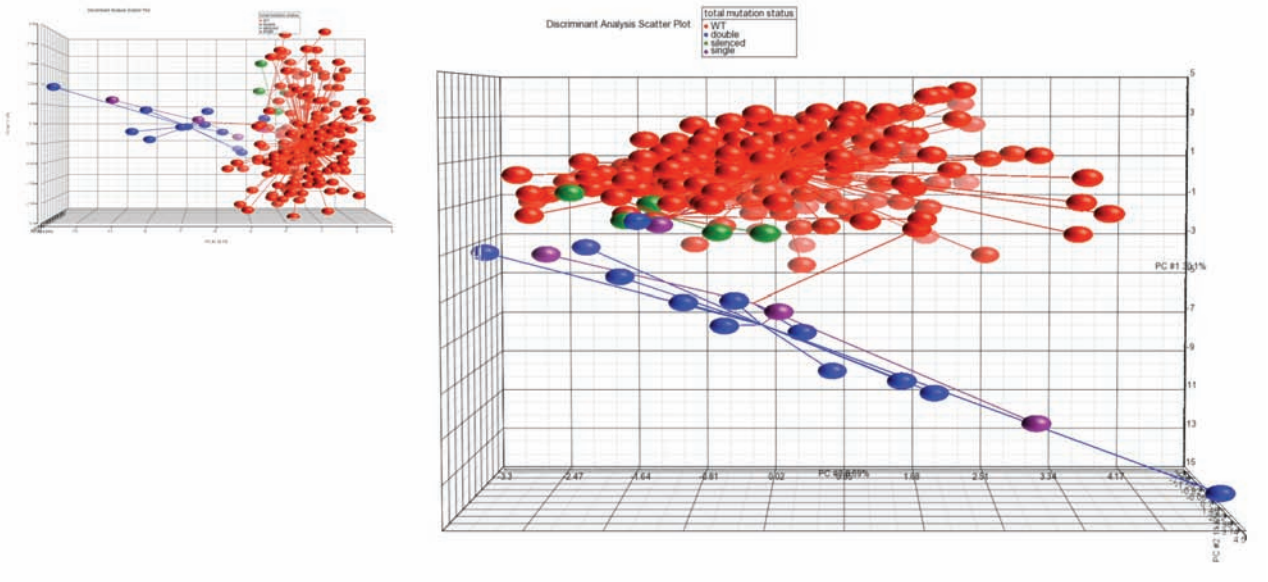
Online Supplementary Figure S1. Heat map of the top 50 most significantly differentially expressed genes between the main *CEBPA* cluster *versus* all other clusters. The heat map presents the top 50 probe sets on the vertical axis, and the 237 cases on the horizontal axis. Intensity values of the probe sets (log_2) are median centered, and probe sets and cases are hierarchically clustered. Cells represent relative log_2 expression values, and have been color-coded on a scale ranging from bright green (-2) to bright red (+2), with black indicating no change relative to the median. Below the heat map cases within the *CEBPA* cluster are marked with a red circle (*CEBPA*-mutant cases) and with a green circle (*CEBPA*-silenced cases), and cases outside the *CEBPA* cluster are marked with a gray circle. Patients within the *CEBPA* cluster show a distinct gene expression profile.

A

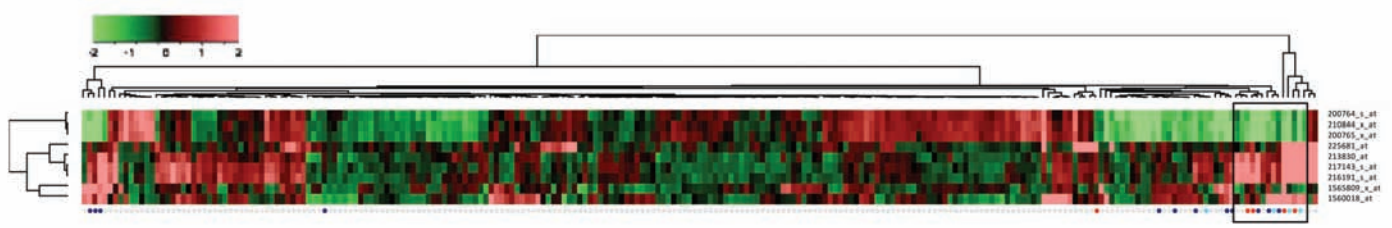


Online Supplementary Figure S2. Prediction and clustering of *CEBPA* double-mutant and silenced cases in our pediatric AML cohort based on predictive gene signatures derived from adult AML. (A) To identify our *CEBPA* double-mutant cases, a 21-probe set-containing signature previously described for adult AML was used for clustering.[27] In the heat map these probe sets are depicted on the vertical axis, and the 237 cases on the horizontal axis. Intensity values of the probe sets (log_2) are median centered, and probe sets and cases are hierarchically clustered. Cells represent relative log_2 expression values, and have been color coded on a scale ranging from bright green (-2) to bright red (+2), with black indicating no change relative to the median. Below the heat map *CEBPA* double-mutants are marked with a dark blue-colored circle, *CEBPA* single-mutants with a light blue-colored circle, *CEBPA*-silenced cases with a red-colored circle and *CEBPA* wild-type cases with a gray-colored circle. All but one double-mutant clustered together, as shown in the lined box; however three *CEBPA* single-mutants, with a mutation in the bZIP region, were also present in this cluster.

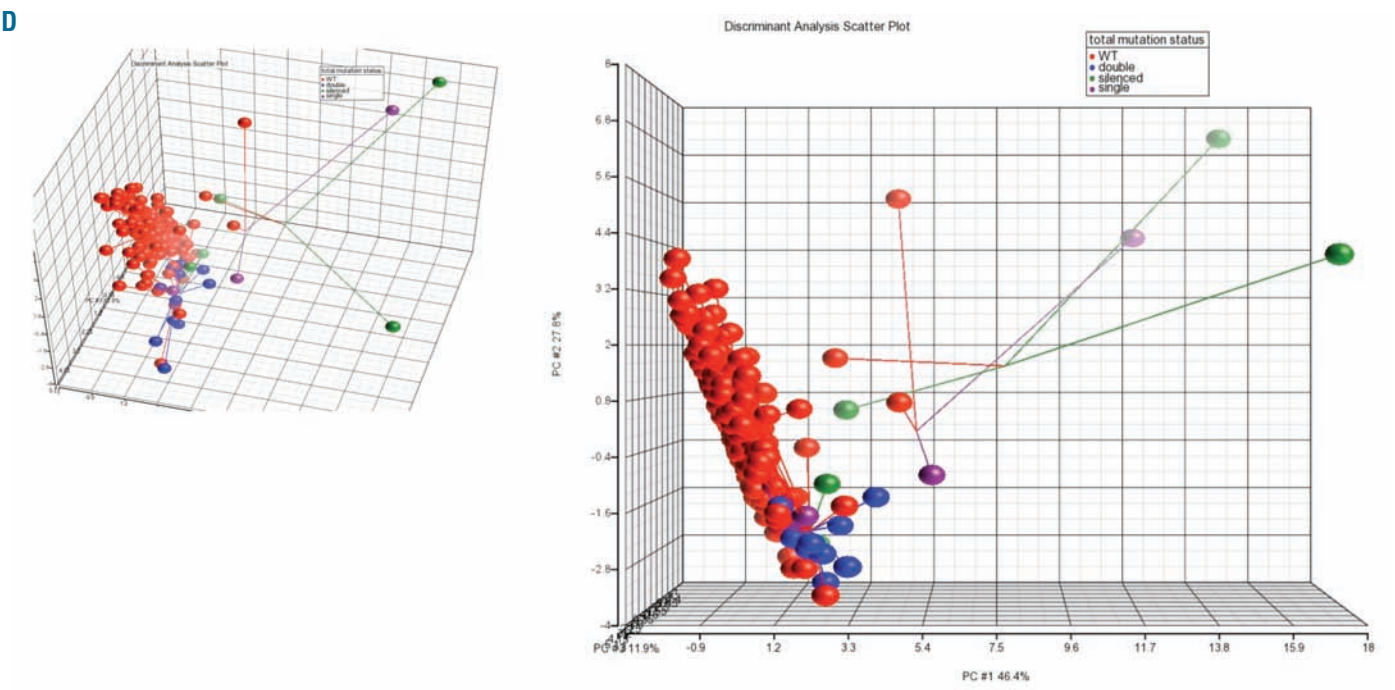
B



C



D



Online Supplementary Figure S2. (B) Principal component analysis (PCA) of the 237 cases was carried out based on the 21-probe set-containing signature predicting *CEBPA* double-mutants. Each circle represents an AML case. Cases are color-coded based on *CEBPA* status: *CEBPA* double-mutant (blue), *CEBPA* single-mutant (purple), *CEBPA* wild-type (red) and *CEBPA*-silenced AML (green). The first three principal components (PCA1, PCA2 and PCA3) are depicted. The figure illustrates that *CEBPA* double-mutants can be separated from *CEBPA* wild-type cases over the first principal component together with three single mutants (PCA1), while the *CEBPA*-silenced cases are scattered within the wild-type cohort. (C) To identify our *CEBPA*-silenced cases, we used the 9-probe set-containing gene signature previously described for adult AML.[15] The presented heat map is similar to *Online Supplementary Figure S2A*. Four of our five *CEBPA*-silenced cases clustered together, but six *CEBPA*-mutant cases and four wild-types clustered with these cases. (D) PCA of the 237 cases was carried out based on the 9-probe sets-containing signature predicting *CEBPA*-silenced cases. This figure is similar to *Online Supplementary Figure S2C*. The figure illustrates that only three out of the five *CEBPA*-silenced cases could be separated from the *CEBPA* wild-type cases. Moreover, two single-mutant cases and three wild-type cases aggregated together with these three silenced cases, while the other two *CEBPA*-silenced cases were scattered within the wild-type cohort.