Evaluation of gene expression signatures predictive of cytogenetic and molecular subtypes of pediatric acute myeloid leukemia

Brian V. Balgobind,[1] Marry M. Van den Heuvel-Eibrink,[1] Renee X. De Menezes,[1,2] Dirk Reinhardt,[3] Iris H.I.M. Hollink,[1] Susan T.J.C.M. Arentsen-Peters,[1] Elisabeth R. van Wering,[4] Gertjan J.L. Kaspers,[5] Jacqueline Cloos,[5] Evelien S.J.M. de Bont,[6] Jean-Michel Cayuela,[7] Andre Baruchel,[7] Claus Meyer,[8] Rolf Marschalek,[8] Jan Trka,[9] Jan Stary,[9] H. Berna Beverloo,[10] Rob Pieters,[1] C. Michel Zwaan,[1] and Monique L. den Boer[1]

[1]Pediatric Oncology/Hematology, Erasmus MC, Sophia Children's Hospital, Rotterdam, The Netherlands; [2]Center for Human and Clinical Genetics, LUMC, Leiden, The Netherlands; [3]AML-BFM Study Group, Hannover, Germany; [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, Groningen, The Netherlands; [7]Hematology, St. Louis Hospital, Paris, France; [8]Institute of Pharmaceutical Biology, ZAFES, Diagnostic Center of Acute Leukemias (DCAL), Frankfurt, Germany; [9]Pediatric Hematology/Oncology, 2nd Medical School, Charles University, Prague, Czech Republic, and [10]Clinical Genetics, Erasmus MC, Rotterdam, The Netherlands

Citation: Balgobind BV, Van den Heuvel-Eibrink MM, De Menezes RX, Reinhardt D, Hollink IHIM, Arentsen-Peters STJCM, van Wering ER, Kaspers GJL, Cloos J, de Bont ESJM, Cayuela J-M, Baruchel A, Meyer C, Marschalek R, Trka J, Stary J, Beverloo HB, Pieters R, Zwaan CM, and den Boer ML. Evaluation of gene expression signatures predictive of cytogenetic and molecular subtypes of pediatric acute myeloid leukemia. Haematologica 2011;96(2):221-230.
doi:10.3324/haematol.2010.029660

SEE PDF OF DATA SUPPLEMENTARY

Supplementary Table S1. Distribution of cytogenetic subtypes in the current study compared to the average frequency observed in other pediatric AML studies.

	Current Study (%)	Pediatric AML studies (%)[1]
Cytogenetic subtypes		
MLL-rearrangements	*20*	*18*
t(8;21)(q22;q22)	*11*	*13*
inv(16)(p13q22)	*11*	*6*
t(15;17)(q21;q22)	*9*	*4*
t(7;12)(q36;p13)	*3*	*ND*
CN-AML	*15*	*24*

ND = not determined.

[1]Based on the average frequency of the following studies:

1. Creutzig U, Zimmermann M, Ritter J, et al. Treatment strategies and long-term results in paediatric patients treated in four consecutive AML-BFM trials. Leukemia. 2005;19:2030-42.

2. Dluzniewska A, Balwierz W, Armata J, et al. Twenty years of Polish experience with three consecutive protocols for treatment of childhood acute myelogenous leukemia. Leukemia. 2005;19:2117-24.

3. Entz-Werle N, Suciu S, van der Werff ten Bosch J, et al. Results of 58872 and 58921 trials in acute myeloblastic leukemia and relative value of chemotherapy vs allogeneic bone marrow transplantation in first complete remission: the EORTC Children Leukemia Group report. Leukemia. 2005;19:2072-81.

4. Gibson BE, Wheatley K, Hann IM, et al. Treatment strategy and long-term results in paediatric patients treated in consecutive UK AML trials. Leukemia. 2005;19:2130-8.

5. Kardos G, Zwaan CM, Kaspers GJ, et al. Treatment strategy and results in children treated on three Dutch Childhood Oncology Group acute myeloid leukemia trials. Leukemia. 2005;19:2063-71.

6. Lie SO, Abrahamsson J, Clausen N, et al. Long-term results in children with AML: NOPHO-AML Study Group--report of three consecutive trials. Leukemia. 2005;19:2090-100.

7. Perel Y, Auvrignon A, Leblanc T, et al. Treatment of childhood acute myeloblastic leukemia: dose intensification improves outcome and maintenance therapy is of no benefit--multicenter studies of the French LAME (Leucemie Aigue Myeloblastique Enfant) Cooperative Group. Leukemia. 2005;19:2082-9.

8. Pession A, Rondelli R, Basso G, et al. Treatment and long-term results in children with acute myeloid leukaemia treated according to the AIEOP AML protocols. Leukemia. 2005;19:2043-53.

9. Ravindranath Y, Chang M, Steuber CP, et al. Pediatric Oncology Group (POG) studies of acute myeloid leukemia (AML): a review of four consecutive childhood AML trials conducted between 1981 and 2000. Leukemia. 2005;19:2101-16.

10. Ribeiro RC, Razzouk BI, Pounds S, Hijiya N, Pui CH, Rubnitz JE. Successive clinical trials for childhood acute myeloid leukemia at St Jude Children's Research Hospital, from 1980 to 2000. Leukemia. 2005;19:2125-9.

11. Smith FO, Alonzo TA, Gerbing RB, Woods WG, Arceci RJ. Long-term results of children with acute myeloid leukemia: a report of three consecutive Phase III trials by the Children's Cancer Group: CCG 251, CCG 213 and CCG 2891. Leukemia. 2005;19:2054-62.

Supplementary Table S2. Number of probe sets at different cut-off *P* values per cytogenetic subtype (total cohort).

	MLL	**t(8;21)**	**inv(16)**	**t(15;17)**	**t(7;12)**	**CN-AML**	**Remaining cytogenetics**	**AML-unknown**
FDR adjusted p-values								
p<1.0E-08	1171	247	138	604	31	2	0	0
p<1.0E-06	2031	431	260	1027	59	4	1	0
p<1.0E-04	3886	853	610	2125	118	6	2	0
p<0.001	5829	1431	994	3477	182	12	6	1
p<0.005	7943	2210	1482	5478	325	25	47	2
p<0.01	9164	2752	1843	6586	453	47	93	5
p<0.05	13866	4804	3506	11491	1349	239	525	42

Supplementary Table S3. Overview of the 75 probe sets used to classify pediatric AML in the present study.

Probe Set	Gene	Gene Symbol	Chromosome
244536_at	tumor protein p53 binding protein, 2 isoform 2	TP53BP2	chr1q42.12
228740_at	---	---	chr14q23.1
227949_at	phosphatase and actin regulator 3	PHACTR3	chr20q13.32-q13.33
1552665_at	hypothetical LOC84989	LOC84989	chr10q21.3
1567101_at	---	---	chr13q22.1
1557261_at	WAS protein homolog associated with actin, golgi membranes and microtubules-like 1 /// WAS protein homolog associated with actin, golgi membranes and microtubules-like 2 (pseudogene)	WHAMML1 /// WHAMML2	chr15q11.2 /// chr15q13.1
213908_at	WAS protein homolog associated with actin, golgi membranes and microtubules-like 1 /// WAS protein homolog associated with actin, golgi membranes and microtubules-like 2 (pseudogene)	WHAMML1 /// WHAMML2	chr15q11.2 /// chr15q13.1
1559265_at	chromosome 10 open reading frame 140	C10orf140	chr10p12.31
1559266_s_at	chromosome 10 open reading frame 140	C10orf140	chr10p12.31
1555923_a_at	chromosome 10 open reading frame 114	C10orf114	chr10p12.31
239503_at	---	---	chr10p12.31
209616_s_at	carboxylesterase 1 (monocyte/macrophage serine esterase 1)	CES1	chr16q13-q22.1
221858_at	TBC1 domain family, member 12	TBC1D12	chr10q23.33
235273_at	dyslexia susceptibility 1 candidate 1	DYX1C1	chr15q21.3
202746_at	integral membrane protein 2A	ITM2A	chrXq13.3-Xq21.2
202747_s_at	integral membrane protein 2A	ITM2A	chrXq13.3-Xq21.2
206059_at	zinc finger protein 91	ZNF91	chr19p13.1-p12
219765_at	zinc finger protein 329	ZNF329	chr19q13.43
201496_x_at	myosin, heavy chain 11, smooth muscle	MYH11	chr16p13.11
201497_x_at	myosin, heavy chain 11, smooth muscle	MYH11	chr16p13.11
207961_x_at	myosin, heavy chain 11, smooth muscle	MYH11	chr16p13.11
232716_at	lysophosphatidic acid receptor 1	LPAR1	chr9q31.3
241773_at	lysophosphatidic acid receptor 1	LPAR1	chr9q31.3
209386_at	transmembrane 4 L six family member 1	TM4SF1	chr3q21-q25
212850_s_at	low density lipoprotein receptor-related protein 4	LRP4	chr11p11.2-p12
238091_at	---	---	chr17p13.3
1553994_at	5'-nucleotidase, ecto (CD73)	NT5E	chr6q14-q21
227486_at	5'-nucleotidase, ecto (CD73)	NT5E	chr6q14-q21
212358_at	CAP-GLY domain containing linker protein 3	CLIP3	chr19q13.12
205330_at	meningioma (disrupted in balanced translocation) 1	MN1	chr22q11|22q12.1
212667_at	secreted protein, acidic, cysteine-rich (osteonectin)	SPARC	chr5q31.3-q32
222862_s_at	adenylate kinase 5	AK5	chr1p31
239519_at	neuropilin 1 isoform a	NRP1	chr10p11.22
203074_at	annexin A8 /// annexin A8-like 1 /// annexin A8-like 2	ANXA8 /// ANXA8L1 /// ANXA8L2	chr10q11.22
230244_at	chromosome 2 open reading frame 82	C2orf82	chr2q37.1
204150_at	stabilin 1	STAB1	chr3p21.1
38487_at	stabilin 1	STAB1	chr3p21.1
223828_s_at	lectin, galactoside-binding, soluble, 12	LGALS12	chr11q13
206634_at	SIX homeobox 3	SIX3	chr2p16-p21
210997_at	hepatocyte growth factor (hepapoietin A; scatter factor)	HGF	chr7q21.1
210998_s_at	hepatocyte growth factor (hepapoietin A; scatter	HGF	chr7q21.1

	factor)		
205110_s_at	fibroblast growth factor 13	FGF13	chrXq26.3
229459_at	family with sequence similarity 19 (chemokine (C-C motif)-like), member A5	FAM19A5	chr22q13.32
229655_at	family with sequence similarity 19 (chemokine (C-C motif)-like), member A5	FAM19A5	chr22q13.32
214228_x_at	tumor necrosis factor receptor superfamily, member 4	TNFRSF4	chr1p36
227870_at	immunoglobulin superfamily, DCC subclass, member 4	IGDCC4	chr15q22.31
205614_x_at	macrophage stimulating 1 (hepatocyte growth factor-like)	MST1	chr3p21
219225_at	piggyBac transposable element derived 5	PGBD5	chr1q42.13
225275_at	EGF-like repeats and discoidin I-like domains 3	EDIL3	chr5q14
229349_at	lin-28 homolog B (C. elegans)	LIN28B	chr6q16.3-q21
203304_at	BMP and activin membrane-bound inhibitor homolog (Xenopus laevis)	BAMBI	chr10p12.3-p11.2
206363_at	v-maf musculoaponeurotic fibrosarcoma oncogene homolog (avian)	MAF	chr16q22-q23
227370_at	family with sequence similarity 171, member B	FAM171B	chr2q32.1
209173_at	anterior gradient homolog 2 (Xenopus laevis)	AGR2	chr7p21.3
243339_at	---	---	chr2q14.3
207802_at	cysteine-rich secretory protein 3	CRISP3	chr6p12.3
214614_at	motor neuron and pancreas homeobox 1	MNX1	chr7q36
232136_s_at	cortactin binding protein 2	CTTNBP2	chr7q31
1564435_a_at	keratin 72	KRT72	chr12q13.13
203936_s_at	matrix metallopeptidase 9 (gelatinase B, 92kDa gelatinase, 92kDa type IV collagenase)	MMP9	chr20q11.2-q13.1
210744_s_at	interleukin 5 receptor, alpha	IL5RA	chr3p26-p24
211517_s_at	interleukin 5 receptor, alpha	IL5RA	chr3p26-p24
206622_at	thyrotropin-releasing hormone	TRH	chr3q13.3-q21
1555943_at	phosphoglycerate mutase family member 5	PGAM5	chr12q24.33
216832_at	runt-related transcription factor 1; translocated to, 1 (cyclin D-related)	RUNX1T1	chr8q22
242845_at	---	---	chr8q21.3
216831_s_at	runt-related transcription factor 1; translocated to, 1 (cyclin D-related)	RUNX1T1	chr8q22
205529_s_at	runt-related transcription factor 1; translocated to, 1 (cyclin D-related)	RUNX1T1	chr8q22
205528_s_at	runt-related transcription factor 1; translocated to, 1 (cyclin D-related)	RUNX1T1	chr8q22
228827_at	---	---	chr8q21.3
225056_at	signal-induced proliferation-associated 1 like 2	SIPA1L2	chr1q42.2
233587_s_at	signal-induced proliferation-associated 1 like 2	SIPA1L2	chr1q42.2
204811_s_at	calcium channel, voltage-dependent, alpha 2/delta subunit 2	CACNA2D2	chr3p21.3
206940_s_at	POU class 4 homeobox 1	POU4F1	chr13q31.1
211341_at	POU class 4 homeobox 1	POU4F1	chr13q31.1

Legend: Probe sets are listed in the same order as in Figure 1B of the main paper.

Supplementary Table S4. Detailed cytogenetic data of misclassified cases in the discovery cohort.

ID	Subtype	Predicted	Karyotype	FISH	LDI-PCR
#1	AML-Other	*MLL*	51~53,XX,+2,+4,+6,+10,+add(11)(p?15), +13,-18,+21,+22,inc[cp10]	negative	*t(10;11)*
#2	AML-Other	*MLL*	46,XY	negative	*t(10;11)*
#3	AML-Other	*MLL*	46,XX	negative	*t(11;19)*

Supplementary Table S5. Predictive value of the classifier for relapsed and secondary cases of pediatric AML.

SVM Predicted Subtype	*Subtype according to cytogenetic screening*					
	MLL	**t(8;21)**	**inv(16)**	**t(15;17)**	**t(7;12)**	**AML-Other**
***MLL*-rearranged**	9	0	0	0	0	0
t(8;21)	0	5	0	0	0	0
inv(16)	0	0	0	0	0	0
t(15;17)	0	0	0	0	0	0
t(7;12)	0	0	0	0	0	0
AML-Other	0	0	0	0	0	27

Supplementary Table S6. Diagnostic test values for the classification of cytogenetic subtypes in the validation cohort of pediatric AML cases.

	Independent validation cohort (n=80)		
	Pediatric AML		***Adult AML***
	Balgobind *et al*	Ross *et al* [ref16]	Valk *et al* [ref17]
	75 probe sets	150 probe sets	2856 probe sets
	%	%	%
***MLL*-rearranged**			
% sensitivity	94	88	81
% specificity	100	98	98
% PPV	100	93	93
% NPV	98	96	95
% accuracy	99	96	95
t(8;21)			
% sensitivity	100	100	100
% specificity	100	100	100
% PPV	100	100	100
% NPV	100	100	100
% accuracy	100	100	100
inv(16)			
% sensitivity	100	100	100
% specificity	100	100	100
% PPV	100	100	100
% NPV	100	100	100
% accuracy	100	100	100
t(15;17)			
% sensitivity	100	100	100
% specificity	100	100	100
% PPV	100	100	100
% NPV	100	100	100
% accuracy	100	100	100
t(7;12)			
% sensitivity	100	0	50
% specificity	100	100	100
% PPV	100	ND	100
% NPV	100	98	99
% accuracy	100	98	99
AML-Other			
% sensitivity	100	97	97
% specificity	98	91	91
% PPV	97	90	90
% NPV	100	98	98
% accuracy	99	94	94
Overall			
% sensitivity	98	91	91
% specificity	100	97	97
% PPV	100	98	98
% NPV	97	90	90
% accuracy	99	94	94

ND= not determined, since no samples were classified as t(7;12)

Supplementary Table S7. Molecular aberrations found in 237 children with newly diagnosed AML according to cytogenetic subtype.

Molecular aberration	Subtype according to cytogenetic screening					
	MLL (N=47)	*t(8;21) (N=28)*	*inv(16) (N=27)*	*t(15;17) (N=19)*	*t(7;12) (N=7)*	*AML-other (N=92)*
FLT3-ITD	1	3	1	12	0	31
RAS pathway	11	4	6	0	1	24
KIT	1	8	8	0	1	0
MLL-PTD	0	0	0	0	0	5
NPM1	0	0	0	0	0	17
CEBPA	0	0	0	0	0	16

Supplementary Table S8. Number of probe sets at different cut-off *P* values per molecular subtype (total cohort).

	NPM1	***CEBPA***	***MLL*-PTD**	***FLT3*-ITD**	***KIT***	***N/KRAS***	***PTPN11***
FDR adjusted p-values							
p<1.0E-08	7	7	0	54	8	0	1
p<1.0E-06	12	13	0	127	11	0	1
p<1.0E-04	27	66	1	387	46	0	6
p<0.001	46	131	4	933	86	0	10
p<0.005	75	222	7	1732	155	0	12
p<0.01	106	320	9	2308	220	0	19
p<0.05	266	852	12	4896	510	0	57

Supplementary Table S9. Diagnostic test values for the prediction of mutations in *NPM1*, *CEBPA* and *MLL*-PTD by a gene expression signature consisting of 45 probe sets.

A)

	Discovery cohort [a] 3-fold cross-validation (100 iterations)				
	% sensitivity	**% specificity**	**% PPV**	**% NPV**	**% accuracy**
NPM1	16 (0-33)	100 (100-100)	100 (100-100)	95 (94-96)	94 (94-96)
***MLL*-PTD**	0 (0-0)	100 (100-100)	ND	98 (98-98)	98 (98-98)
CEBPA	67 (67-100)	100 (100-100)	100 (100-100)	98 (98-100)	96 (96-98)
Remaining cases	100 (100-100)	43 (26-43)	92 (90-92)	100 (100-100)	92 (90-92)
All groups	43 (26-43)	100 (100-100)	100 (100-100)	92 (90-92)	92 (90-92)

ND= Not determined

[a] Values represent the median and 25^{th}-75^{th} percentiles (in parentheses) obtained by 3-fold cross-validation using the discovery cohort of 157 cases (100 iterations).

B)

	Validation cohort independent validation group, N=80				
	% sensitivity	**% specificity**	**% PPV**	**% NPV**	**% accuracy**
NPM1	13	99	50	91	90
***MLL*-PTD**	0	100	ND	96	96
CEBPA	33	100	100	95	95
Remaining cases	98	18	82	75	81
All groups	18	98	75	82	81

Supplementary Table S10. Diagnostic test values for the prediction of cytogenetic and molecular subtypes (*NPM1*, *CEBPA* and *MLL*-PTD) by a gene expression signature consisting of 120 probe sets.

A)

	Discovery cohort [a] 3-fold cross-validation (100 iterations)				
	% sensitivity	**% specificity**	**% PPV**	**% NPV**	**% accuracy**
***MLL*-rearranged**	90 (80-90)	97 (95-97)	88 (82-90)	97 (95-98)	94 (92-96)
t(8;21)	100 (100-100)	100 (100-100)	100 (100-100)	100 (100-100)	100 (100-100)
inv(16)	100 (80-100)	100 (100-100)	100 (100-100)	100 (98-100)	100 (98-100)
t(15;17)	100 (100-100)	100 (100-100)	100 (100-100)	100 (100-100)	100 (100-100)
t(7;12)	100 (100-100)	100 (100-100)	100 (100-100)	100 (100-100)	100 (100-100)
NPM1	0 (0-33)	100 (100-100)	100 (50-100)	96 (94-96)	94 (94-96)
***MLL*-PTD**	0 (0-0)	100 (100-100)	ND	98 (98-98)	98 (98-98)
CEBPA	100 (66-100)	100 (100-100)	100 (100-100)	100 (98-100)	100 (98-100)
Remaining cases[#]	91 (88-94)	82 (79-85)	70 (67-74)	95 (93-97)	85 (82-88)
All groups	82 (79-85)	91 (88-94)	95 (93-97)	70 (67-74)	85 (82-88)

ND= Not determined

[a] Values represent the median and 25^{th}-75^{th} percentiles (in parentheses) obtained by 3-fold cross-validation using the discovery cohort of 157 cases (100 iterations).

[#] Including three cases predicted as *MLL*-rearranged AML and confirmed to be so by LDI-PCR.

B)

	Validation cohort independent validation group, N=80				
	% sensitivity	**% specificity**	**% PPV**	**% NPV**	**% accuracy**
***MLL*-rearranged**	94	98	94	98	98
t(8;21)	100	100	100	100	100
inv(16)	100	100	100	100	100
t(15;17)	100	100	100	100	100
t(7;12)	100	100	100	100	100
NPM1	13	99	50	91	90
***MLL*-PTD**	0	100	ND	96	96
CEBPA	33	100	100	95	95
Remaining cases	90	75	55	96	78
All groups	75	90	96	55	78

ND= Not determined

Supplementary Table S11. Diagnostic test values for the prediction of type I molecular subtypes by a gene expression signature consisting of 30 probe sets.

A)

	Discovery cohort [a] 3-fold cross-validation (100 iterations)				
	% sensitivity	**% specificity**	**% PPV**	**% NPV**	**% accuracy**
***FLT3*-ITD**	40 (30-53)	95 (93-98)	67 (60-80)	87 (85-90)	85 (83-87)
KIT	25 (0-25)	100 (98-100)	100 (50-100)	94 (92-94)	92 (92-94)
Remaining cases	95 (92-97)	35 (29-43)	80 (78-82)	70 (58-80)	77 (75-81)
All groups	35 (29-43)	95 (92-97)	70 (58-80)	80 (78-82)	77 (75-81)

[a] Values represent the median and 25^{th}-75^{th} percentiles (in parentheses) obtained by 3-fold cross-validation using the discovery cohort of 157 cases (100 iterations).

B)

	Validation cohort independent validation group, N=80				
	% sensitivity	**% specificity**	**% PPV**	**% NPV**	**% accuracy**
***FLT3*-ITD**	72	100	100	93	94
KIT	33	99	66	95	94
Remaining cases	89	63	86	94	88
All groups	63	89	94	86	88

Supplementary Table S12. Diagnostic test values for the prediction of type I molecular subtypes by a gene expression signature consisting of 45 probe sets.

A)

	Discovery cohort [a] 3-fold cross-validation (100 iterations)				
	% sensitivity	**% specificity**	**% PPV**	**% NPV**	**% accuracy**
***FLT3*-ITD**	40 (30-50)	95 (93-98)	67 (55-83)	87 (85-89)	85 (83-87)
RAS-pathway[b]	0 (0-0)	95 (95-98)	0 (0-0)	84 (84-85)	82 (81-83)
KIT	25 (0-25)	100 (98-100)	100 (50-100)	94 (92-94)	92 (92-94)
Remaining cases	87 (83-93)	27 (18-32)	62 (59-64)	60 (50-78)	62 (58-65)
All groups	27 (18-32)	87 (83-93)	60 (50-78)	62 (59-64)	62 (58-65)

[a] Values represent the median and 25th-75th percentiles (in parentheses) obtained by 3-fold cross-validation using the discovery cohort of 157 cases (100 iterations).
[b] Includes cases with mutations in *NRAS*, *KRAS* and *PTPN11*.

B)

	Validation cohort independent validation group, N=80				
	% sensitivity	**% specificity**	**% PPV**	**% NPV**	**% accuracy**
***FLT3*-ITD**	61	98	92	90	90
RAS-pathway[a]	5	94	20	76	73
KIT	17	99	50	94	93
Remaining cases	89	35	54	79	60
All groups	30	89	76	52	58

[a] Includes cases with mutations in *NRAS*, *KRAS* and *PTPN11*.

Supplementary Table S13. The 126 probe sets used for the hierarchical clustering of *FLT3*-ITD positive cases in t(15;17) and CN-AML ranked according to hierarchical order in Figure 2 (*HOXB*-genes highlighted in yellow).

Probe Set	Gene	Gene Symbol	Chromosome
219201_s_at	twisted gastrulation homolog 1 (Drosophila)	TWSG1	chr18p11.3
217848_s_at	pyrophosphatase (inorganic) 1	PPA1	chr10q11.1-q24
204429_s_at	solute carrier family 2 (facilitated glucose/fructose transporter), member 5	SLC2A5	chr1p36.2
203710_at	inositol 1,4,5-triphosphate receptor, type 1	ITPR1	chr3p26-p25
216944_s_at	inositol 1,4,5-triphosphate receptor, type 1	ITPR1	chr3p26-p25
225337_at	abhydrolase domain containing 2	ABHD2	chr15q26.1
201432_at	catalase	CAT	chr11p13
211922_s_at	catalase	CAT	chr11p13
201468_s_at	NAD(P)H dehydrogenase, quinone 1	NQO1	chr16q22.1
220658_s_at	aryl hydrocarbon receptor nuclear translocator-like 2	ARNTL2	chr12p12.2-p11.2
228011_at	family with sequence similarity 92, member A1	FAM92A1	chr8q22.1
224204_x_at	aryl hydrocarbon receptor nuclear translocator-like 2	ARNTL2	chr12p12.2-p11.2
228624_at	transmembrane protein 144	TMEM144	chr4q32.1
202890_at	microtubule-associated protein 7	MAP7	chr6q23.3
210145_at	phospholipase A2, group IVA (cytosolic, calcium-dependent)	PLA2G4A	chr1q25
204030_s_at	schwannomin interacting protein 1	SCHIP1	chr3q25.32-q25.33
229309_at	adrenergic, beta-1-, receptor	ADRB1	chr10q24-q26
218445_at	H2A histone family, member Y2	H2AFY2	chr10q22
236738_at	Similar to LOC166075	LOC401097	chr3q25.33
204082_at	pre-B-cell leukemia homeobox 3	PBX3	chr9q33-q34
232088_x_at	hypothetical LOC100271722	hCG_2039027	chr22q13.31
228365_at	copine VIII	CPNE8	chr12q12
204779_s_at	homeobox B7	HOXB7	chr17q21.3
216973_s_at	homeobox B7	HOXB7	chr17q21.3
231767_at	homeobox B4	HOXB4	chr17q21-q22
205453_at	homeobox B2	HOXB2	chr17q21-q22
205600_x_at	homeobox B5	HOXB5	chr17q21.3
230743_at	hypothetical LOC404266	LOC404266	chr17q21.32
239791_at	Hypothetical LOC404266	LOC404266	chr17q21.32
228904_at	homeobox B3	HOXB3	chr17q21.3
236892_s_at	---	---	chr17q21.33
1553808_a_at	NK2 transcription factor related, locus 3 (Drosophila)	NKX2-3	chr10q24.2
205366_s_at	homeobox B6	HOXB6	chr17q21.3
205601_s_at	homeobox B5	HOXB5	chr17q21.3
232979_at	---	---	chr17q21.33
242426_at	neuregulin 4	NRG4	chr15q24.2
232424_at	PR domain containing 16	PRDM16	chr1p36.23-p33
242269_at	hypothetical LOC440556	FLJ42875	chr1p36.32
226500_at	zinc finger and BTB domain containing 47	ZBTB47	chr3p22.1
237108_x_at	hypothetical LOC440556	FLJ42875	chr1p36.32
210327_s_at	alanine-glyoxylate aminotransferase	AGXT	chr2q36-q37
204501_at	nephroblastoma overexpressed gene	NOV	chr8q24.1
214321_at	nephroblastoma overexpressed gene	NOV	chr8q24.1

218164_at	spermatogenesis associated 20	SPATA20	chr17q21.33
225097_at	homeodomain interacting protein kinase 2	HIPK2	chr7q32-q34
201618_x_at	glycosylphosphatidylinositol anchor attachment protein 1 homolog (yeast)	GPAA1	chr8q24.3
210338_s_at	heat shock 70kDa protein 8	HSPA8	chr11q24.1
58696_at	exosome component 4	EXOSC4	chr8q24.3
219919_s_at	slingshot homolog 3 (Drosophila)	SSH3	chr11q13.2
227400_at	nuclear factor I/X (CCAAT-binding transcription factor)	NFIX	chr19p13.3
224968_at	coiled-coil domain containing 104	CCDC104	chr2p16.1
238583_at	methionine sulfoxide reductase B3	MSRB3	chr12q14.3
1554127_s_at	methionine sulfoxide reductase B3	MSRB3	chr12q14.3
225790_at	methionine sulfoxide reductase B3	MSRB3	chr12q14.3
230520_at	androgen-induced 1	AIG1	chr6q24.2
205382_s_at	complement factor D (adipsin)	CFD	chr19p13.3
204548_at	steroidogenic acute regulatory protein	STAR	chr8p11.2
222462_s_at	beta-site APP-cleaving enzyme 1	BACE1	chr11q23.2-q23.3
205204_at	neuromedin B	NMB	chr15q22-qter
1553723_at	G protein-coupled receptor 97	GPR97	chr16q21
222043_at	clusterin	CLU	chr8p21-p12
208791_at	clusterin	CLU	chr8p21-p12
208792_s_at	clusterin	CLU	chr8p21-p12
214615_at	purinergic receptor P2Y, G-protein coupled, 10	P2RY10	chrXq21.1
236280_at	---	---	chrXq21.1
224964_s_at	guanine nucleotide binding protein (G protein), gamma 2	GNG2	chr14q21
202759_s_at	A kinase (PRKA) anchor protein 2 /// PALM2-AKAP2 readthrough transcript	AKAP2 /// PALM2-AKAP2	chr9q31-q33
226694_at	A kinase (PRKA) anchor protein 2 /// PALM2-AKAP2 readthrough transcript	AKAP2 /// PALM2-AKAP2	chr9q31-q33
219955_at	LINE-1 type transposase domain containing 1	L1TD1	chr1p31.3
238292_at	---	---	chr6q27
242078_at	---	---	chr2q11.2
206682_at	C-type lectin domain family 10, member A	CLEC10A	chr17p13.1
227006_at	protein phosphatase 1, regulatory (inhibitor) subunit 14A	PPP1R14A	chr19q13.1
207675_x_at	artemin	ARTN	chr1p33-p32
208216_at	distal-less homeobox 4	DLX4	chr17q21.33
235434_at	---	---	chr22q13.2
241975_at	Hypothetical gene supported by BX647608	LOC399959	chr11q24.1
201069_at	matrix metallopeptidase 2 (gelatinase A, 72kDa gelatinase, 72kDa type IV collagenase)	MMP2	chr16q13-q21
224839_s_at	glutamic pyruvate transaminase (alanine aminotransferase) 2	GPT2	chr16q12.1
200986_at	serpin peptidase inhibitor, clade G (C1 inhibitor), member 1	SERPING1	chr11q12-q13.1
201564_s_at	fascin homolog 1, actin-bundling protein (Strongylocentrotus purpuratus)	FSCN1	chr7p22
202718_at	insulin-like growth factor binding protein 2, 36kDa	IGFBP2	chr2q33-q34
207895_at	N-acetylated alpha-linked acidic dipeptidase-like 1	NAALADL1	chr11q12
222693_at	fibronectin type III domain containing 3B	FNDC3B	chr3q26.31
225032_at	fibronectin type III domain containing 3B	FNDC3B	chr3q26.31
230486_at	Poly(rC) binding protein 3	PCBP3	chr21q22.3
242393_x_at	ArfGAP	AGAP10 /// AGAP4 ///	chr10q11.22

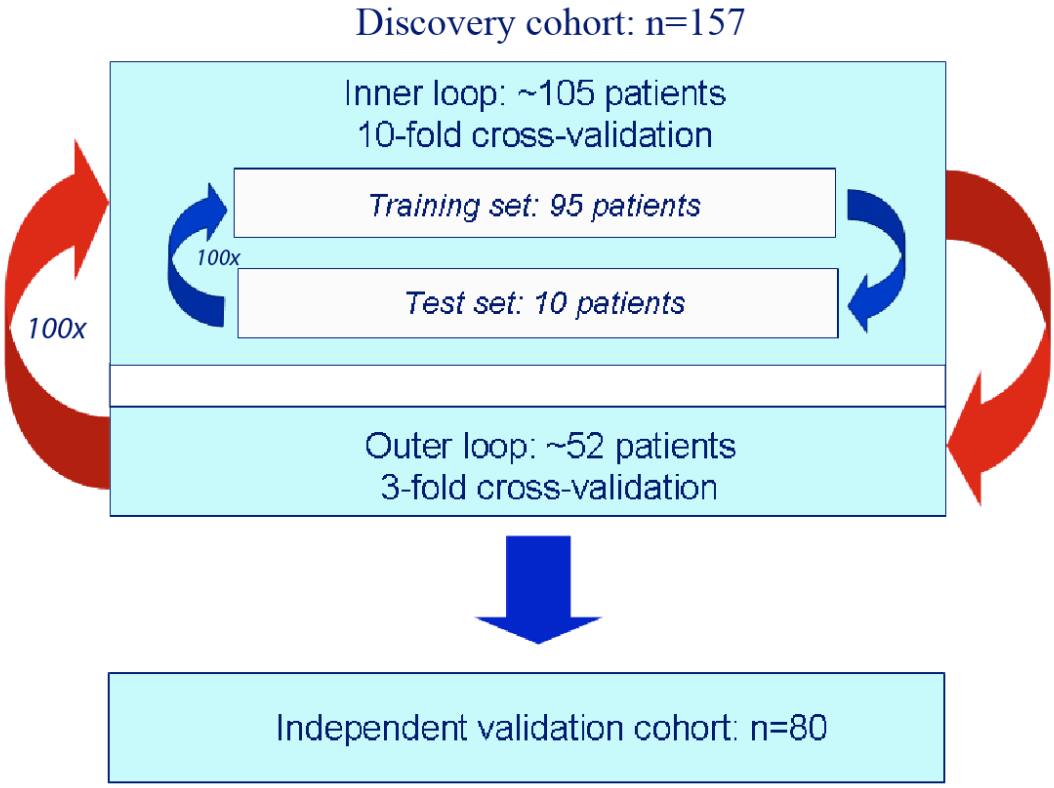
		AGAP9 /// BMS1P5	
243937_x_at	ArfGAP	AGAP10 /// AGAP4 /// AGAP9 /// BMS1P5	chr10q11.22
238553_at	ArfGAP	AGAP10 /// AGAP4 /// AGAP9 /// BMS1P5	chr10q11.22
239151_at	hypothetical LOC399753	RP11-144G6.7	chr10q11.22
205110_s_at	fibroblast growth factor 13	FGF13	chrXq26.3
234269_at	---	---	chrXq27.1
227410_at	family with sequence similarity 43, member A	FAM43A	chr3q29
213125_at	olfactomedin-like 2B	OLFML2B	chr1q23.3
207031_at	NK3 homeobox 2	NKX3-2	chr4p16.1
205614_x_at	macrophage stimulating 1 (hepatocyte growth factor-like)	MST1	chr3p21
216320_x_at	macrophage stimulating 1 (hepatocyte growth factor-like)	MST1	chr3p21
205944_s_at	clathrin, heavy chain-like 1	CLTCL1	chr22q11.2|22q11.21
221636_s_at	MOCO sulphurase C-terminal domain containing 2	MOSC2	chr1q41
204537_s_at	gamma-aminobutyric acid (GABA) A receptor, epsilon	GABRE	chrXq28
228285_at	tudor domain containing 9	TDRD9	chr14q32.33
236787_at	---	---	chr2p11.2
203074_at	annexin A8 /// annexin A8-like 1 /// annexin A8-like 2	ANXA8 /// ANXA8L1 /// ANXA8L2	chr10q11.22
214228_x_at	tumor necrosis factor receptor superfamily, member 4	TNFRSF4	chr1p36
227185_at	hypothetical LOC643988	LOC643988	chr1p36.33
225203_at	protein phosphatase 1, regulatory (inhibitor) subunit 16A	PPP1R16A	chr8q24.3
233072_at	netrin G2	NTNG2	chr9q34
214203_s_at	proline dehydrogenase (oxidase) 1	PRODH	chr22q11.21
228550_at	reticulon 4 receptor	RTN4R	chr22q11.21
206634_at	SIX homeobox 3	SIX3	chr2p16-p21
209815_at	patched homolog 1 (Drosophila)	PTCH1	chr9q22.3
227145_at	lysyl oxidase-like 4	LOXL4	chr10q24
235468_at	hexaribonucleotide binding protein 3	hCG_1776007	chr17q25.3
223828_s_at	lectin, galactoside-binding, soluble, 12	LGALS12	chr11q13
224794_s_at	cerebral endothelial cell adhesion molecule	CERCAM	chr9q34.11
210755_at	hepatocyte growth factor (hepapoietin A; scatter factor)	HGF	chr7q21.1
210997_at	hepatocyte growth factor (hepapoietin A; scatter factor)	HGF	chr7q21.1
210998_s_at	hepatocyte growth factor (hepapoietin A; scatter factor)	HGF	chr7q21.1
230244_at	chromosome 2 open reading frame 82	C2orf82	chr2q37.1
229459_at	family with sequence similarity 19 (chemokine (C-C motif)-like), member A5	FAM19A5	chr22q13.32
229655_at	family with sequence similarity 19 (chemokine (C-C motif)-like), member A5	FAM19A5	chr22q13.32
204150_at	stabilin 1	STAB1	chr3p21.1
38487_at	stabilin 1	STAB1	chr3p21.1
204163_at	elastin microfibril interfacer 1	EMILIN1	chr2p23.3-p23.2
212285_s_at	agrin	AGRN	chr1p36.33
217419_x_at	agrin	AGRN	chr1p36.33

Supplementary Table S14. Diagnostic test values for the prediction of the cytogenetic subtypes and different subgroups of *FLT3*-ITD for the independent validation cohort by a gene expression signature consisting of 99 probe sets.

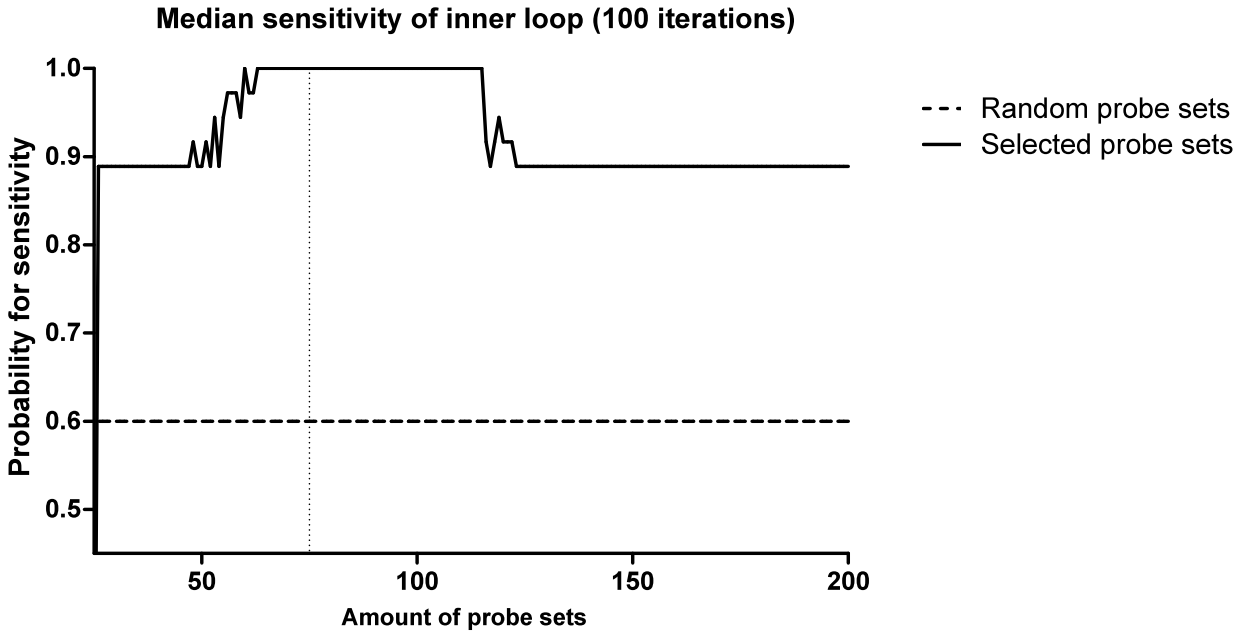
	Validation cohort independent validation group, N=80				
	% sensitivity	**% specificity**	**% PPV**	**% NPV**	**% accuracy**
***MLL*-rearranged**	94	98	94	98	98
t(8;21)(q22;q22)	100	100	100	100	100
inv(16)(p13q22)	100	100	100	100	100
t(15;17)(q21;q22)	0	99	0	99	98
t(15;17)/*FLT3*-ITD	75	99	75	99	98
t(7;12)(q36;p13)	100	100	100	100	100
CN-AML/*FLT3*-ITD	0	100	ND	90	90
AML-other [a]	76	100	100	82	89
All groups	80	97	98	74	86

[a] Without the CN-AML/*FLT3*-ITD cases. ND = not determined

Supplementary Figure S1. Identification of a gene-expression signature enabling classification of pediatric AML .
The discovery cohort was used to estimate the number of probe sets in the inner loop (2/3 of the patients) and the prediction accuracy in the outer loop (remaining 1/3 of the patients). The final classifier was constructed on the total discovery cohort (n=157) and this was tested on the validation cohort (n=80) to determine the true accuracy of the classifier.



Supplementary Figure S2. Estimated sensitivity within the inner loop. The median sensitivity was calculated using 100 iterations. Probe sets selected for the different subgroups yielded a higher sensitivity compared with random selected probe sets. A minimum of 75 probe sets was needed to have a median sensitivity of 100%



Supplementary Figure S3. Gene expressions of the *HOXB* cluster between *FLT3*-ITD-positive and *FLT3*-ITD-negative cases in CN-AML. Among patients with CN-AML, all those positive for *FLT3*-ITD show higher expression of *HOXB2* to *HOXB9*. Three patients without *FLT3*-ITD also showed this expression, including two patients with an *NPM1* mutation.