

## Supplemental Information for:

### Gene Expression Profiling of Pediatric Acute Myelogenous Leukemia

Mary E. Ross, Rami Mahfouz, Mihaela Onciu, Hsi-Che Liu, Xiaodong Zhou, Guangchun Song, Sheila A. Shurtleff, Stanley Pounds, Cheng Cheng, Jing Ma, Raul C. Ribeiro, Jeffrey E. Rubnitz, Kevin Girtman, W. Kent Williams, Susana C. Raimondi, Der-Cherng Liang, Lee-Yung Shih, Ching-Hon Pui & James R. Downing

#### Table of Contents

##### Section I. Patient Datasets

- Table S1. Diagnostic AML characteristics
- Table S2. Cytogenetics Summary
- Table S3. Adult diagnostic AML characteristics
- Table S4. Additional T-ALL characteristics

##### Section II. Methods

- Table S5. Summary of filtered probe sets
- Table S6. *MLL*-PTD primers
- Additional Statistical Methods

##### Section III. Genetic Subtype Discriminating Genes

- Figure S1. Unsupervised Hierarchical clustering
- Figure S2. Hierarchical clustering with class discriminating genes
- Table S7. Top 100 probe sets selected by SAM for t(8;21)[*AML1-ETO*]
- Table S8. Top 100 probe sets selected by SAM for t(15;17) [*PML-RAR $\alpha$* ]
- Table S9. Top 63 probe sets selected by SAM for inv(16) [*CBF $\beta$ -MYH11*]
- Table S10. Top 100 probe sets selected by SAM for *MLL* chimeric fusion genes
- Table S11. Top 100 probe sets selected by SAM for FAB-M7
- Table S12. Top 100 probe sets selected by SAM for CBF leukemias (whole dataset)

##### Section IV. *MLL* in combined ALL and AML dataset

- Table S13. Top 100 probe sets selected by SAM for *MLL* chimeric fusions irrespective of blast lineage (whole dataset)
- Table S14. Class discriminating genes for cases with an *MLL* chimeric fusion gene that show uniform high expression, irrespective of blast lineage

##### Section V. Outcome Analysis

- Expression profiles as a predictor of time to relapse
- Evaluation of relapse genes identified by Yagi et. al. on this dataset
- Table S15. Probe sets most strongly associated with time to progression or relapse in the training set

##### Section VI. Additional References

**Table S1. Pediatric AML sample characteristics**

Case identifier	subtype <sup>a</sup>	MLL status <sup>b</sup>	FAB <sup>c</sup>	Sample <sup>d</sup>	% Blasts <sup>e</sup>	Protocol <sup>f</sup>
inv16-#01	<i>CBFβ-MYH11</i>		M4E	BM	94	AML83
inv16-#02	<i>CBFβ-MYH11</i>		M4	BM	98	AML87
inv16-#03	<i>CBFβ-MYH11</i>		M4E	BM	89	AML87
inv16-#04	<i>CBFβ-MYH11</i>		M4	BM	80	AML91
inv16-#05	<i>CBFβ-MYH11</i>		M4E	BM	85	AML91
inv16-#06	<i>CBFβ-MYH11</i>		M4	BM	77	AML91
inv16-#07	<i>CBFβ-MYH11</i>		M4E	BM	85	AML97
inv16-#08	<i>CBFβ-MYH11</i>		M4E	BM	88	AML97
inv16-#09	<i>CBFβ-MYH11</i>		M4E	BM	89	AML87
inv16-#10	<i>CBFβ-MYH11</i>		M4E	BM	79	AML97
inv16-#11	<i>CBFβ-MYH11</i>		M1	BM	88	AML97
inv16-#12	<i>CBFβ-MYH11</i>		M4E	BM	86	AML97
inv16-#13	<i>CBFβ-MYH11</i>		M1E	BM	96	AML83
inv16-#14	<i>CBFβ-MYH11</i>		M4E	BM	66	non protocol
MLL-#01	<i>MLL</i>	t(11;19)	M4	BM	91	AML87
MLL-#02	<i>MLL</i>	t(9;11)	M5a	BM	73	AML87
MLL-#03	<i>MLL</i>	t(6;11)	M5b	BM	96	AML91
MLL-#04	<i>MLL</i>	t(3;11)	M5a	BM	94	AML91
MLL-#05	<i>MLL</i>	t(1;11)	M5a	PB	88	AML91
MLL-#06	<i>MLL</i>	t(10;11)	M5a	BM	77	AML91
MLL-#07	<i>MLL</i>	t(9;11)	M5a	BM	96	AML91
MLL-#08	<i>MLL</i>	t(9;11)	M5a	BM	88	AML91
MLL-#09	<i>MLL</i>	t(7;9;11)	M5a	BM	92	AML91
MLL-#10	<i>MLL</i>	t(9;11)	M5a	BM	80	AML91
MLL-#11	<i>MLL</i>	t(11;19)	M5b	PB	90	AML91
MLL-#12	<i>MLL</i>	t(9;11)	M5a	BM	87	AML91
MLL-#13	<i>MLL</i>	t(1;11)	M5a	BM	98	AML91
MLL-#14	<i>MLL</i>	t(11;19)	M1	BM	95	AML97
MLL-#15	<i>MLL</i>	t(1;11)	M5a	BM	90	AML97
MLL-#16	<i>MLL</i>	t(10;11)	M5a	apheresis	97	AML97
MLL-#17	<i>MLL</i>	t(9;11)	M5a	BM	97	AML97
MLL-#18	<i>MLL</i>	t(9;11)	M1	BM	96	AML97
MLL-#19	<i>MLL</i>	t(10;11)	M5a	BM	70	AML97
MLL-#20	<i>MLL</i>	t(11;16)	M5b	BM	88	AML91
MLL-#21	<i>MLL</i>	t(9;11)	M5a	BM	96	AML87
MLL-#22	<i>MLL</i>	t(4;11)	M5a	PB	97	AML97
MLL-#23	<i>MLL</i>	t(10;11)	M5a	PB	71	AML97
t(15:17)-#01	<i>PML-RARα</i>		M3	BM	95	AML87
t(15:17)-#02	<i>PML-RARα</i>		M3	PB	88	AML83
t(15:17)-#03	<i>PML-RARα</i>		M3	BM	89	PG9220
t(15:17)-#04	<i>PML-RARα</i>		M3	BM	83	non protocol
t(15:17)-#05	<i>PML-RARα</i>		M3	BM	95	non protocol
t(15:17)-#06	<i>PML-RARα</i>		M3	BM	88	PG9710

t(15:17)-#07	<i>PML-RAR<math>\alpha</math></i>		M3	PB	87	El-Salvador
t(15:17)-#08	<i>PML-RAR<math>\alpha</math></i>		M3	BM	91	El-Salvador
t(15:17)-#09	<i>PML-RAR<math>\alpha</math></i>		M3	BM	97	El-Salvador
t(15:17)-#10	<i>PML-RAR<math>\alpha</math></i>		M3	BM	87	non protocol
t(15:17)-#11	<i>PML-RAR<math>\alpha</math></i>		M3	BM	94	PG9710
t(15:17)-#12	<i>PML-RAR<math>\alpha</math></i>		M3	BM	90	PG9710
t(15:17)-#13	<i>PML-RAR<math>\alpha</math></i>		M3	BM	93	non protocol
t(15:17)-#14	<i>PML-RAR<math>\alpha</math></i>		M3	BM	92	non protocol
t(15:17)-#15	<i>PML-RAR<math>\alpha</math></i>		M3	BM	93	AML87
t(8:21)-#01	<i>AML1-ETO</i>		M2	BM	74	AML87
t(8:21)-#02	<i>AML1-ETO</i>		M4E	BM	91	AML87
t(8:21)-#03	<i>AML1-ETO</i>		M2	BM	81	AML87
t(8:21)-#04	<i>AML1-ETO</i>		M2	BM	72	AML87
t(8:21)-#05	<i>AML1-ETO</i>		M2	BM	88	AML87
t(8:21)-#06	<i>AML1-ETO</i>		M2	BM	76	AML91
t(8:21)-#07	<i>AML1-ETO</i>		M2	BM	83	AML97
t(8:21)-#08	<i>AML1-ETO</i>		M2	BM	92	AML97
t(8:21)-#09	<i>AML1-ETO</i>		M2	BM	93	AML97
t(8:21)-#10	<i>AML1-ETO</i>		M2	BM	92	AML97
t(8:21)-#11	<i>AML1-ETO</i>		M2	BM	90	AML97
t(8:21)-#12	<i>AML1-ETO</i>		M2	BM	90	AML97
t(8:21)-#13	<i>AML1-ETO</i>		M2	BM	79	El-Salvador
t(8:21)-#14	<i>AML1-ETO</i>		M2	BM	95	El-Salvador
t(8:21)-#15	<i>AML1-ETO</i>		M2	BM	90	El-Salvador
t(8:21)-#16	<i>AML1-ETO</i>		M1	BM	92	El-Salvador
t(8:21)-#17	<i>AML1-ETO</i>		M2E	BM	80	El-Salvador
t(8:21)-#18	<i>AML1-ETO</i>		M1	BM	91	AML97
t(8:21)-#19	<i>AML1-ETO</i>		M2	BM	89	AML97
t(8:21)-#20	<i>AML1-ETO</i>		M2E	PB	99	AML87
t(8:21)-#21	<i>AML1-ETO</i>		M2	BM	76	non protocol
Other-#01	Other		M1	BM	93	AML87
Other-#02	Other		M2	BM	98	AML87
Other-#03	Other	<i>MLL-PTD</i>	M1	BM	93	AML87
Other-#04	Other		M4E	BM	80	AML80
Other-#05	Other		M1	BM	93	AML87
Other-#06	Other		M1	BM	88	AML87
Other-#07	Other	<i>MLL-PTD</i>	M2	PB	84	AML87
Other-#08	Other	<i>MLL-PTD</i>	M2	BM	92	AML87
Other-#09	Other	<i>MLL-PTD</i>	M2	BM	81	AML87
Other-#10	Other		M2	BM	93	AML84
Other-#11	Other		M1	BM	95	AML87
Other-#12	Other		M2	BM	85	AML87
Other-#13	Other	<i>MLL-PTD</i>	M2	BM	83	AML91
Other-#14	Other		M1	BM	93	AML91
Other-#15	Other	<i>MLL-PTD</i>	M2	BM	75	AML91
Other-#16	Other	<i>MLL-PTD</i>	M1	BM	85	AML91

Other-#17	Other		M1	BM	89	AML91
Other-#18	Other		M4	BM	93	AML97
Other-#19	Other		M2	BM	94	AML91
Other-#20	Other	<i>MLL</i> -PTD	M1	BM	96	AML91
Other-#21	Other		M1	BM	84	AML91
Other-#22	Other	<i>MLL</i> -PTD	M2	BM	78	AML91
Other-#23	Other		M1	BM	95	AML97
Other-#24	Other		M4E	BM	85	AML97
Other-#25	Other		M1	BM	87	AML97
Other-#26	Other	<i>MLL</i> -PTD	M4	BM	93	AML97
Other-#27	Other		M1	BM	92	AML97
Other-#28	Other		M1	BM	98	AML97
Other-#29	Other		M1	BM	74	AML97
Other-#30	Other		M2	BM	86	AML97
Other-#31	Other		M1	BM	80	AML97
Other-#32	Other		M2	BM	78	AML97
Other-#33	Other	<i>MLL</i> -PTD	M5a	BM	92	AML87
Other-#34	Other		M5a	BM	93	AML87
Other-#35	Other		M4	BM	82	AML91
Other-#36	Other		M5a	BM	72	AML97
Other-#37	Other		M1	BM	95	AML97
Other-#38	Other		M1	BM	92	AML97
Other-#39	Other		M5a	BM	90	AML97
Other-#40	Other		M5a	BM	76	AML97
Other-#41	Other		M1	BM	89	AML97
Other-#42	Other	<i>MLL</i> -PTD	M2	BM	85	AML97
Other-#43	Other	<i>MLL</i> -PTD	M4	BM	80	AML97
Other-#44	Other		M4	BM	81	AML91
Other-#45	Other		M2	BM	77	AML91
Other-#46	Other		M1	BM	92	AML83
Other-#47	Other		M2	PB	95	AML83
Other-M7-#01	FAB-M7		M7	BM	82	AML87
Other-M7-#02	FAB-M7		M7	BM	78	AML87
Other-M7-#03	FAB-M7		M7	BM	78	non protocol
Other-M7-#04	FAB-M7		M7	BM	50	AML97
Other-M7-#05	FAB-M7		M7	BM	80	AML97
Other-M7-#06	FAB-M7		M7	BM	68	AML97
Other-M7-#07	FAB-M7		M7	BM	90	AML91
Other-M7-#08	FAB-M7		M7	PB	73	AML91
Other-M7-#09	FAB-M7		M7	BM	97	non protocol
Other-M7-#10	FAB-M7		M0*	BM	80	AML83

<sup>a</sup> AML subtype cases were assigned to for this analysis. Translocation was determined by at least two methods: cytogenetics, FISH, and/or RT-PCR.

<sup>b</sup> *MLL* status: *MLL*-PTD is partial tandem duplication, or translocation partner is indicated for reciprocal translocations.

<sup>c</sup> French-American British type as reviewed by M.Onciu. No examples of FAB-M3 variant were included.

<sup>d</sup> type of specimen evaluated: BM = bone marrow; PB = peripheral blood.

<sup>e</sup> Percent blast in diagnostic specimen post-ficoll purification.

<sup>f</sup> treatment protocols: AML83, AML87, AML91, AML97 are St. Jude Children's Research Hospital institutional protocols. PG9710 Pediatric Oncology Cooperative Group study. El Salvador refers to patients treated in El Salvador through SJCRH's International Outreach Program.

\* morphologically and cytochemically this case was suggestive of FAB-M7. CD41 was negative but no other megakaryocytic markers were evaluated. This case was treated as FAB-M7 for purposes of these analyses.

**Table S2. Pediatric AML cytogenetics summary**

Case identifier	subtype	<i>MLL</i> status	Cytogenetic Summary
inv16-#01	<i>CBFβ-MYH11</i>		inv(16)
inv16-#02	<i>CBFβ-MYH11</i>		inv(16)
inv16-#03	<i>CBFβ-MYH11</i>		inv(16) + misc non-recurrent
inv16-#04	<i>CBFβ-MYH11</i>		inv(16) + misc non-recurrent
inv16-#05	<i>CBFβ-MYH11</i>		inv(16) + misc non-recurrent
inv16-#06	<i>CBFβ-MYH11</i>		inv(16) + misc non-recurrent
inv16-#07	<i>CBFβ-MYH11</i>		inv(16) + misc non-recurrent
inv16-#08	<i>CBFβ-MYH11</i>		inv(16)
inv16-#09	<i>CBFβ-MYH11</i>		inv(16) + misc non-recurrent
inv16-#10	<i>CBFβ-MYH11</i>		inv(16) + misc non-recurrent
inv16-#11	<i>CBFβ-MYH11</i>		inv(16) + misc non-recurrent
inv16-#12	<i>CBFβ-MYH11</i>		inv(16)
inv16-#13	<i>CBFβ-MYH11</i>		inv(16)
inv16-#14	<i>CBFβ-MYH11</i>		inv(16)
MLL-#01	<i>MLL</i>	t(11;19)	t(11;19)(q23;p13.1)
MLL-#02	<i>MLL</i>	t(9;11)	t(9;11)(p22;q23)
MLL-#03	<i>MLL</i>	t(6;11)	t(6;11)(q27;q23)
MLL-#04	<i>MLL</i>	t(3;11)	t(3;11;10)(q28;q23;p11.2)
MLL-#05	<i>MLL</i>	t(1;11)	t(1;11)(q21;q23)
MLL-#06	<i>MLL</i>	t(10;11)	t(10;11)(p11.2;q23), + misc non-recurrent
MLL-#07	<i>MLL</i>	t(9;11)	t(9;11)(p21;q23) + misc non-recurrent
MLL-#08	<i>MLL</i>	t(9;11)	t(9;11)(p22;q23) + misc non-recurrent
MLL-#09	<i>MLL</i>	t(7;9;11)	t(7;9;11)(p15;p22;q23)
MLL-#10	<i>MLL</i>	t(9;11)	t(9;11)(p22;q23)
MLL-#11	<i>MLL</i>	t(11;19)	t(11;19)(q23;p13.1)
MLL-#12	<i>MLL</i>	t(9;11)	t(9;11)(p22;q23) + misc non-recurrent
MLL-#13	<i>MLL</i>	t(1;11)	t(1;11)(q21;q23)
MLL-#14	<i>MLL</i>	t(11;19)	t(11;19)(q23;p13.1)
MLL-#15	<i>MLL</i>	t(1;11)	t(1;11)(q21;q23)
MLL-#16	<i>MLL</i>	t(10;11)	t(10;11)(p12;q23)
MLL-#17	<i>MLL</i>	t(9;11)	t(9;11)(p22;q23)
MLL-#18	<i>MLL</i>	t(9;11)	t(9;11)(p22;q23)
MLL-#19	<i>MLL</i>	t(10;11)	t(10;11)(p13;q23) + misc non-recurrent
MLL-#20	<i>MLL</i>	t(11;16)	t(11;16)(q23;q24)
MLL-#21	<i>MLL</i>	t(9;11)	t(9;11)(p22;q23)
MLL-#22	<i>MLL</i>	t(4;11)	t(4;11)(q21;q23)
MLL-#23	<i>MLL</i>	t(10;11)	t(10;11)(p13;q23) + misc non-recurrent
t(15;17)-#01	<i>PML-RARα</i>		t(15;17)

t(15;17)-#02	<i>PML-RAR<math>\alpha</math></i>		t(15;17), +8
t(15;17)-#03	<i>PML-RAR<math>\alpha</math></i>		t(15;17)
t(15;17)-#04	<i>PML-RAR<math>\alpha</math></i>		t(15;17)
t(15;17)-#05	<i>PML-RAR<math>\alpha</math></i>		t(15;17)
t(15;17)-#06	<i>PML-RAR<math>\alpha</math></i>		t(15;17)
t(15;17)-#07	<i>PML-RAR<math>\alpha</math></i>		ND
t(15;17)-#08	<i>PML-RAR<math>\alpha</math></i>		ND
t(15;17)-#09	<i>PML-RAR<math>\alpha</math></i>		ND
t(15;17)-#10	<i>PML-RAR<math>\alpha</math></i>		t(15;17)
t(15;17)-#11	<i>PML-RAR<math>\alpha</math></i>		t(15;17)
t(15;17)-#12	<i>PML-RAR<math>\alpha</math></i>		t(15;17)
t(15;17)-#13	<i>PML-RAR<math>\alpha</math></i>		t(15;17)
t(15;17)-#14	<i>PML-RAR<math>\alpha</math></i>		t(15;17)
t(15;17)-#15	<i>PML-RAR<math>\alpha</math></i>		t(15;17)
t(8;21)-#01	<i>AML1-ETO</i>		t(8;21)
t(8;21)-#02	<i>AML1-ETO</i>		t(8;21)
t(8;21)-#03	<i>AML1-ETO</i>		t(8;21) complex
t(8;21)-#04	<i>AML1-ETO</i>		t(8;21)
t(8;21)-#05	<i>AML1-ETO</i>		t(8;21)
t(8;21)-#06	<i>AML1-ETO</i>		t(8;21)
t(8;21)-#07	<i>AML1-ETO</i>		t(8;21)
t(8;21)-#08	<i>AML1-ETO</i>		t(8;21)
t(8;21)-#09	<i>AML1-ETO</i>		t(8;21) + misc non-recurrent
t(8;21)-#10	<i>AML1-ETO</i>		t(8;21) + misc non-recurrent
t(8;21)-#11	<i>AML1-ETO</i>		t(8;21)
t(8;21)-#12	<i>AML1-ETO</i>		t(8;21)
t(8;21)-#13	<i>AML1-ETO</i>		ND
t(8;21)-#14	<i>AML1-ETO</i>		ND
t(8;21)-#15	<i>AML1-ETO</i>		ND
t(8;21)-#16	<i>AML1-ETO</i>		ND
t(8;21)-#17	<i>AML1-ETO</i>		ND
t(8;21)-#18	<i>AML1-ETO</i>		t(8;21)
t(8;21)-#19	<i>AML1-ETO</i>		t(8;21) complex
t(8;21)-#20	<i>AML1-ETO</i>		t(8;21) complex
t(8;21)-#21	<i>AML1-ETO</i>		t(8;21)
Other-#01	Other		normal
Other-#02	Other		normal
Other-#03	Other	<i>MLL-PTD</i>	13q-
Other-#04	Other		normal
Other-#05	Other		misc non-recurrent
Other-#06	Other		normal
Other-#07	Other	<i>MLL-PTD</i>	+11
Other-#08	Other	<i>MLL-PTD</i>	+8
Other-#09	Other	<i>MLL-PTD</i>	normal
Other-#10	Other		misc non-recurrent
Other-#11	Other		5q- misc non-recurrent

Other-#12	Other		+4
Other-#13	Other	<i>MLL</i> -PTD	normal
Other-#14	Other		normal
Other-#15	Other	<i>MLL</i> -PTD	normal
Other-#16	Other	<i>MLL</i> -PTD	misc non-recurrent
Other-#17	Other		normal
Other-#18	Other		t(16;21)(p11.2;q22)
Other-#19	Other		+4
Other-#20	Other	<i>MLL</i> -PTD	t(3;5)(q25.1;q34)
Other-#21	Other		t(6;9)(p23;q34)
Other-#22	Other	<i>MLL</i> -PTD	normal
Other-#23	Other		t(16;21)(?p11.2;q22)
Other-#24	Other		-7
Other-#25	Other		normal
Other-#26	Other	<i>MLL</i> -PTD	+8, misc
Other-#27	Other		misc non-recurrent
Other-#28	Other		7q-
Other-#29	Other		+8
Other-#30	Other		t(6;9)(p23;q34)
Other-#31	Other		misc non-recurrent
Other-#32	Other		normal
Other-#33	Other	<i>MLL</i> -PTD	+8
Other-#34	Other		-7
Other-#35	Other		misc non-recurrent
Other-#36	Other		misc non-recurrent
Other-#37	Other		normal
Other-#38	Other		misc non-recurrent
Other-#39	Other		normal
Other-#40	Other		+8
Other-#41	Other		9q-
Other-#42	Other	<i>MLL</i> -PTD	misc non-recurrent
Other-#43	Other	<i>MLL</i> -PTD	normal
Other-#44	Other		5q-, misc non-recurrent
Other-#45	Other		normal
Other-#46	Other		misc non-recurrent
Other-#47	Other		misc non-recurrent
Other-M7-#01	FAB-M7		misc non-recurrent
Other-M7-#02	FAB-M7		misc non-recurrent (Down syndrome)
Other-M7-#03	FAB-M7		+8, misc non-recurrent
Other-M7-#04	FAB-M7		+21, misc non-recurrent
Other-M7-#05	FAB-M7		7q-, +21, misc non-recurrent
Other-M7-#06	FAB-M7		misc non-recurrent
Other-M7-#07	FAB-M7		misc non-recurrent (Down syndrome)
Other-M7-#08	FAB-M7		+8, misc non-recurrent
Other-M7-#09	FAB-M7		misc non-recurrent
Other-M7-#10	FAB-M7		misc non-recurrent

ND= not done.

**Table S3. Adult AML sample characteristics**

Case identifier <sup>a</sup>	Subtype <sup>b</sup>	Additional <sup>c</sup>	FAB <sup>d</sup>
Adult-#01	<i>CBFβ-MYH11</i>		M2
Adult-#02	<i>MLL</i>	t(6;11)	M1
Adult-#03	<i>MLL</i>	t(10;11)	M1
Adult-#04	<i>MLL</i>	t(9;11)	M5a
Adult-#05	<i>MLL</i>	t(9;11)	M5a
Adult-#06	<i>MLL</i>	<i>MLL-LARG</i>	M5a
Adult-#07	<i>MLL</i>	<i>MLL-ELL</i>	M4
Adult-#08	<i>PML-RARα</i>		M3
Adult-#09	<i>PML-RARα</i>		M3
Adult-#10	<i>PML-RARα</i>		M3
Adult-#11	<i>PML-RARα</i>		M3v
Adult-#12	<i>AML1-ETO</i>		M2
Adult-#13	<i>AML1-ETO</i>		M2
Adult-#14	<i>AML1-ETO</i>		M2
Adult-#15	<i>AML1-ETO</i>		M2
Adult-#16	<i>AML1-ETO</i>		M2
Adult-#17	<i>AML1-ETO</i>		M2
Adult-#18	Other	<i>MLL-PTD</i>	M2
Adult-#19	Other		M2
Adult-#20	<i>CBFβ-MYH11</i>	<i>BCR-ABL</i>	M4

<sup>a</sup>all samples were bone marrow aspirates obtained at the time of diagnosis of *de novo* AML.

<sup>b</sup>translocation as determined by cytogenetics, FISH, and/or RT-PCR.

<sup>c</sup>additional cytogenetic and molecular information.

<sup>d</sup>French-American British type as assigned at diagnosis in Taiwan.

**Table S4. Additional T-ALL samples with *MLL* gene rearrangements**

Case identifier <sup>a</sup>	subtype <sup>b</sup>	translocation <sup>c</sup>	Sample <sup>d</sup>	% blasts <sup>e</sup>
T-ALL#11	T-ALL, <i>MLL</i>	t(11;15)(q23;q15)	BM	73
T-ALL#12	T-ALL, <i>MLL</i>	t(11;19) <i>MLL-ENL</i>	PB	92
T-ALL#13	T-ALL, <i>MLL</i>	t(11;19) <i>MLL-ENL</i>	PB	97
T-ALL#14	T-ALL, <i>MLL</i>	t(11;19) <i>MLL-ENL</i>	PB	98
T-ALL-C20*	T-ALL, <i>MLL</i>	t(11;19) <i>MLL-ENL</i>	BM	98

<sup>a</sup>all samples were bone marrow aspirates obtained at the time of diagnosis of *de novo* AML

<sup>b</sup>translocation as determined by at least two tests: cytogenetics, FISH, and/or RT-PCR

<sup>c</sup>additional cytogenetic and molecular information

<sup>d</sup>sample obtained where BM = bone marrow; PB = peripheral blood

\*this sample is part of the ALL dataset previously published in Blood<sup>1</sup>. It is listed here to present information regarding *MLL-ENL* translocation.

**Table S5. Summary of genes removed by filter from AML training set**

	# probe sets
Absent in all samples	5606
Maximum signal < 100	0
Maximum signal – Minimum signal < 100	1

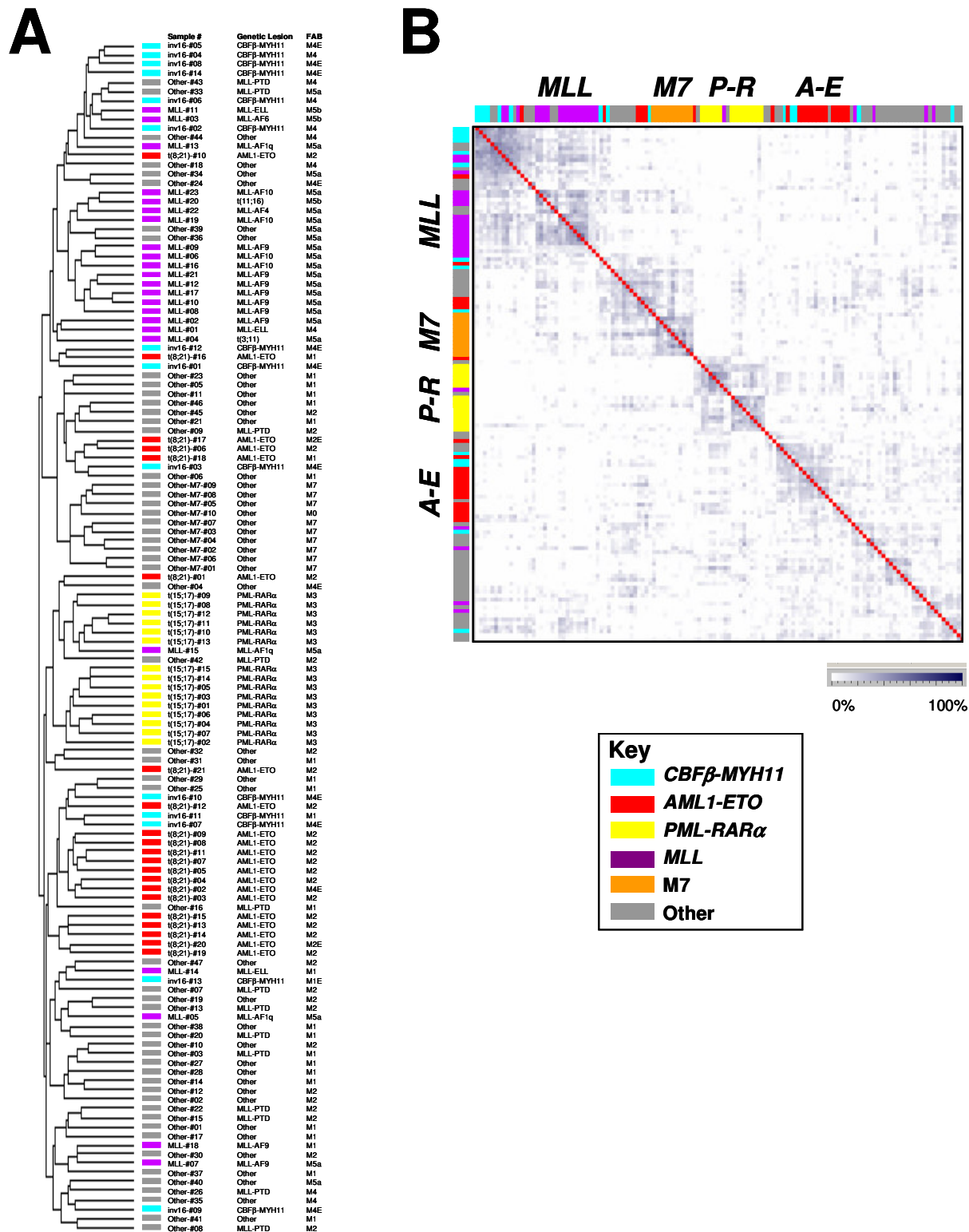
**Table S6. TacMan probes for detection of *MLL*-PTD**

Primer name	Primer direction	sequence
MLLe	forward	5'-CCTGAATCCAAACAGGCCACCACT-3'
MLL3.1C	reverse	5'-AGGAGAGAGTTTACCTGCTC-3'

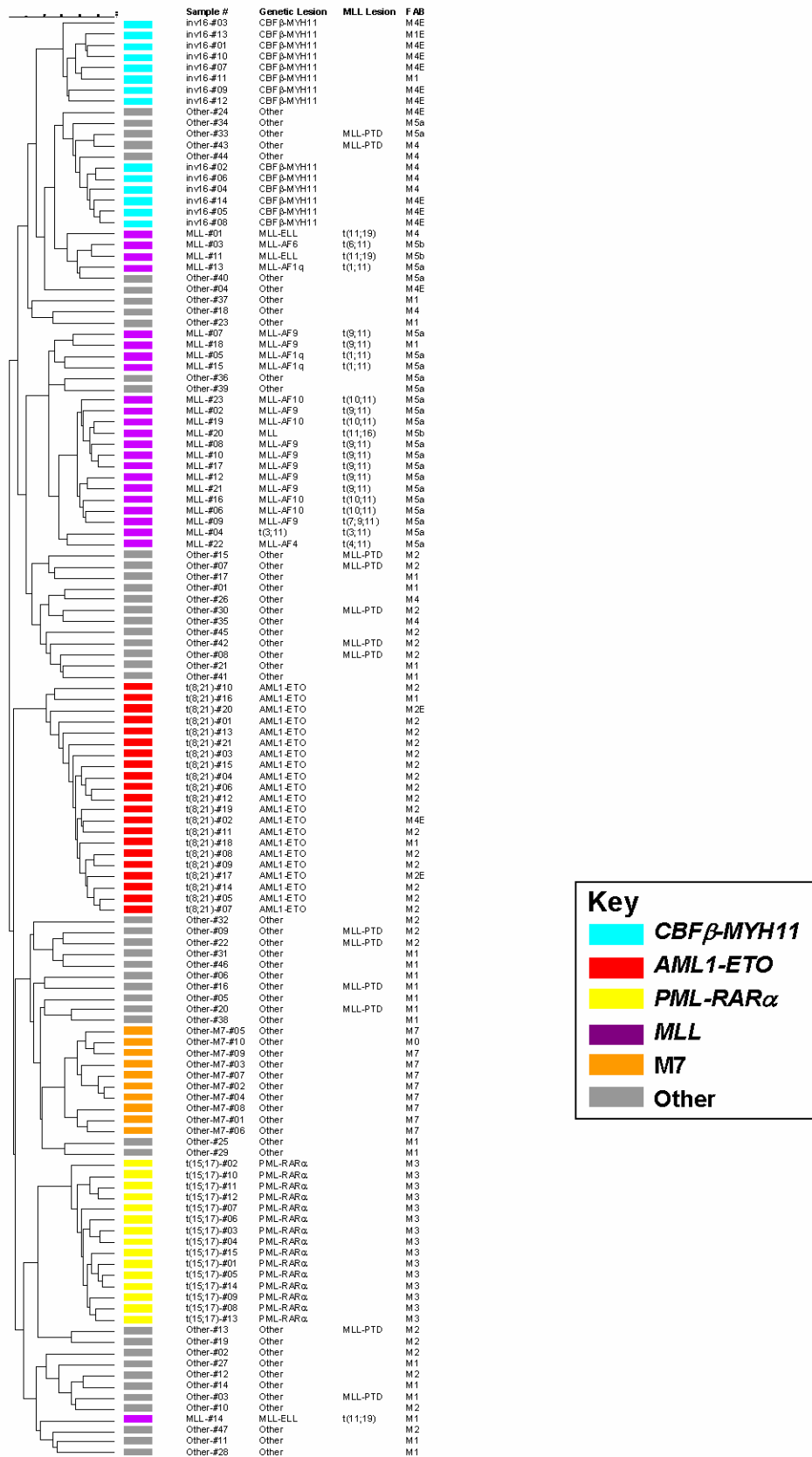
Amplification was carried out using cycle parameters of 98°C for 10 min; 94°C for 1 min; anneal at 60°C for 2 min; extend at 72°C for 2 min for 40 cycles.

### **Additional Statistical Methods**

The ANN supervised learning algorithms have been previously described.<sup>2</sup> To determine the performance of each model using ANN, a confidence threshold was built for each diagnostic subtype utilizing a modification of the method described by Khan et al.<sup>3</sup> Models were built with two possibilities: subgroup and non-subgroup. 3 ANN models were built by 3-fold cross validation utilizing only samples in the training set. The training set samples were then shuffled and 3 additional ANN models were built. 100 repetitions of the model building process were performed. An empirical probability distribution for the ANN output node value was summarized using only nodal values greater than 0.5. to determine the 95% confidence threshold. For each individual sample in the training set, the 100 validation subtype nodal values were averaged, the samples was assigned to the subgroup only when its average subtype nodal value was greater than the 95% confidence threshold. Similarly, nodal values for test set samples are averaged and assigned to a subgroup only when the nodal value exceed the 95% confidence level defined on the training set.



**Figure S1. Characteristics of AML cases clustered in Figure 1, and associated similarity plot. (A)** The dendrogram generated in Figure 1 is reproduced in this figure with the sample number, genetic lesion and FAB morphology indicated next to each case. **(B)** To further assess the relationship both within and between leukemia subtypes, the expression data for all of the cases were analyzed by pair wise comparisons. The data is displayed using a two-dimensional plot in which similarities are plotted using a scale that is based on Pearson correlation coefficients calculated for pair wise comparisons using the expression data for the 16,608 genes that passed the variation filter. The degree of similarity between cases is displayed using the blue color-scale at bottom of figure. The leukemia subtypes are organized identically to their distribution in the dendrogram shown in panel A and are color coded as noted in the Key.



**Figure S2. Characteristics of AML cases clustered in Figure 2 using the top 50 class discriminating genes.** The dendrogram generated in Figure 2 is reproduced in this figure, with the sample number, genetic lesion, karyotype, and FAB morphology indicated next to each case.

Close examination of the dendrogram shown in figure S2 above suggests the existence of several distinct subgroups within the cases with *MLL* chimeric fusion genes. Clustered within one of the major branches of the dendrogram are 80% of the *MLL* cases along with two non-*MLL* cases (Figure 2 and Figure 2S). The cases within this cluster are characterized by FAB-M5a morphology (90%) and include all t(9;11) and t(10;11) cases along with several other rarer *MLL* translocations. The two non-*MLL* cases each have M5a morphology. The remaining 4 *MLL* cases cluster together on a separate branch along with one non-*MLL* case. Within this cluster are both t(11;19) cases along with a t(6;11) and t(1;11). The single non-*MLL* case within this cluster had M5a morphology. Interestingly, the 4 non-*MLL* cases that separate the two major *MLL* clusters include two cases with a t(16;21)(p11.2;q22). The molecular targets of this translocation encode *FUS-ERG*<sup>4,5</sup>. Collectively, these data suggest that the variation in gene expression signatures among cases with *MLL* chimeric fusion genes is likely to result not only from differences in the nature of the *MLL* translocation, but also from differences in the extent of differentiation and the range of secondary mutations required for induction of a full leukemic phenotype. Although two cases lacking *MLL* chimeric fusion genes tightly clustered with the *MLL* cases, their expression signatures differed enough from true *MLL* chimeric fusion gene cases so that they were appropriately classified using supervised learning algorithms.

Similar to *MLL* cases, two subgroups of *CBFβ-MYH11* expressing leukemias were suggested by this analysis, along with a group of five “other” cases that lacked evidence of *CBFβ-MYH11* but nevertheless clustered with these cases (Figure 2 and Figure 2S). Within this extended cluster the majority (90%) of the cases had FAB-M4/M4E/M5 morphology, including the five non-*CBFβ-MYH11* cases. One group of the *CBFβ-MYH11* cases (6/14), expressed a set of genes that were also overexpressed in the five “other” cases, and in a subset of cases with *MLL* chimeric fusion genes (Figure 2), whereas the other subgroup of *CBFβ-MYH11* cases (8/14 cases), expressed these genes at a lower level. Moreover, the later cases also expressed a small set of genes that were overexpressed in *AML1-ETO* leukemias (Figure 2). A careful analysis of the morphology and cytogenetics of the *CBFβ-MYH11* expressing cases failed to reveal any significant differences between these two potential subgroups. Because of the marked variability in genomic breakpoint location in both the *CBFβ* and *MYH11* genes, we did not try to correlate these features with the observed pattern of clustering in this relatively small cohort of cases. Thus, the underlying reason for the observed heterogeneity in *CBFβ-MYH11* expressing cases remains unknown.

### Section III. Genetic Subtype Discriminating Genes

The following tables contain a listing of the top 100 probe sets for each diagnostic subgroup ranked by their SAM scores (Tables S7-S14). Each table contains the Affymetrix U133 series probe set number, a gene description, gene symbol, chromosomal location, primary GenBank reference, and UniGene reference as provided on the Affymetrix website (December 15, 2003 update). SAM scores were calculated utilizing only the samples in the training set (Tables S7-S11) or the whole dataset (Tables S12 and S13). Fold change was calculated using the entire data set, and comparing the mean signal value in the subgroup to the mean signal value in the non-subgroup.

**Table S7. Top 100 probe sets selected by SAM for t(8;21)[*AML1-ETO*]**

	U133 probe set	Gene description	Symbol	Location	GenBank reference	UniGene reference	SAM d-score	t(8;21) Above/ Below mean	Fold change
1	205529_s_at	core-binding factor, runt domain, alpha subunit 2; translocated to, 1; cyclin D-related	CBFA2T1	Chr:8q22	NM_004349	Hs.90858	30.0	Above	25.0
2	205528_s_at	core-binding factor, runt domain, alpha subunit 2; translocated to, 1; cyclin D-related	CBFA2T1	Chr:8q22	X79990	Hs.90858	25.2	Above	34.3
3	206940_s_at	POU domain, class 4, transcription factor 1	POU4F1	Chr:13q21.1-q22	NM_006237	Hs.458303	17.0	Above	32.2
4	211341_at	POU domain, class 4, transcription factor 1	POU4F1	Chr:13q21.1-q22	L20433	Hs.458303	16.0	Above	75.7
5	216831_s_at	core-binding factor, runt domain, alpha subunit 2; translocated to, 1; cyclin D-related	CBFA2T1	Chr:8q22	AF018283	Hs.90858	13.2	Above	36.4
6	212496_s_at	KIAA0876 protein	KIAA0876	Chr:19p13.3	BE256900	Hs.301011	13.1	Above	3.1
7	206622_at	thyrotropin-releasing hormone	TRH	Chr:3q13.3-q21	NM_007117	Hs.182231	12.3	Above	13.7
8	212097_at	caveolin 1, caveolae protein, 22kDa	CAV1	Chr:7q31.1	AU147399	Hs.74034	11.6	Above	16.6
9	204811_s_at	calcium channel, voltage-dependent, alpha 2/delta subunit 2	CACNA2D2	Chr:3p21.3	NM_006030	Hs.389415	11.3	Above	6.9
10	205173_x_at	CD58 antigen, (lymphocyte function-associated antigen 3)	CD58	Chr:1p13	NM_001779	Hs.75626	11.3	Above	3.4
11	211744_s_at	CD58 antigen, (lymphocyte function-associated antigen 3)	CD58	Chr:1p13	BC005930	Hs.75626	10.9	Above	3.6
12	211685_s_at	neurocalcin delta	NCALD	Chr:8q22-q23	AF251061	Hs.90063	10.4	Above	2.9
13	216942_s_at	CD58 antigen, (lymphocyte function-associated antigen 3)	CD58	Chr:1p13	D28586	Hs.75626	10.2	Above	3.8
14	206398_s_at	CD19 antigen	CD19	Chr:16p11.2	NM_001770	Hs.96023	10.0	Above	4.8
15	213194_at	roundabout, axon guidance receptor, homolog 1 (Drosophila)	ROBO1	Chr:3p14.2	NM_002941	Hs.301198	9.9	Above	15.0
16	206187_at	prostaglandin I2 (prostacyclin) receptor (IP)	PTGIR	Chr:19q13.3	NM_000960	Hs.458324	9.7	Above	4.3
17	210010_s_at	solute carrier family 25 (mitochondrial carrier; citrate transporter), member 1	SLC25A1	Chr:22q11.21	U25147	Hs.111024	9.5	Above	2.4
18	218858_at	hypothetical protein FLJ12428	FLJ12428	Chr:8q24.12	NM_022783	Hs.87729	9.2	Above	2.8
19	212495_at	KIAA0876 protein	KIAA0876	Chr:19p13.3	BE256900	Hs.301011	9.2	Above	2.3
20	208116_s_at	mannosidase, alpha, class 1A, member 1	MAN1A1	Chr:6q22	NM_005907	Hs.255149	9.1	Above	4.1

21	221760_at	mannosidase, alpha, class 1A, member 1	MAN1A1	Chr:6q22	BG287153	Hs.255149	9.0	Above	4.6
22	204073_s_at	chromosome 11 open reading frame 9	C11orf9	Chr:11q12-q13.1	NM_013279	Hs.184640	8.9	Above	3.9
23	219686_at	gene for serine/threonine protein kinase	HSA250839	Chr:4p16.2	NM_018401	Hs.58241	8.9	Above	5.4
24	206855_s_at	hyaluronoglucosaminidase 2	HYAL2	Chr:3p21.3	NM_003773	Hs.76873	8.9	Above	1.9
25	218613_at	hypothetical protein DKFZp761K1423	DKFZp761K1423	Chr:8p22	NM_018422	Hs.236438	8.8	Above	4.1
26	210744_s_at	interleukin 5 receptor, alpha	IL5RA	Chr:3p26-p24	M75914	Hs.68876	8.6	Above	9.9
27	216832_at	core-binding factor, runt domain, alpha subunit 2; translocated to, 1; cyclin D-related	CBFA2T1	Chr:8q22	AF018283	Hs.90858	8.5	Above	2.5
28	203355_s_at	ADP-ribosylation factor guanine nucleotide factor 6	EFA6R	Chr:8pter-p23.3	NM_015310	Hs.408177	8.3	Above	5.5
29	204698_at	interferon stimulated gene 20kDa	ISG20	Chr:15q26	NM_002201	Hs.105434	8.2	Above	6.8
30	218352_at	regulator of chromosome condensation (RCC1) and BTB (POZ) domain containing protein 1	RCBTB1	Chr:13q14	NM_018191	Hs.58452	8.2	Above	2.4
31	204990_s_at	integrin, beta 4	ITGB4	Chr:17q11-qter	NM_000213	Hs.85266	8.2	Above	4.2
32	211517_s_at	interleukin 5 receptor, alpha	IL5RA	Chr:3p26-p24	M96651	Hs.68876	8.2	Above	8.5
33	201281_at	adhesion regulating molecule 1	ADRM1	Chr:20q13.33	NM_007002	Hs.90107	8.1	Above	2.1
34	33304_at	interferon stimulated gene 20kDa	ISG20	Chr:15q26	U88964	Hs.105434	8.1	Above	2.7
35	204086_at	preferentially expressed antigen in melanoma	PRAME	Chr:22q11.22	NM_006115	Hs.30743	8.0	Above	4.5
36	214853_s_at	pygopus 2	PYGO2	Chr:1q22	AI091079	Hs.172084	8.0	Above	2.1
37	203065_s_at	caveolin 1, caveolae protein, 22kDa	CAV1	Chr:7q31.1	NM_001753	Hs.74034	7.8	Above	8.2
38	222067_x_at	histone 1, H2bd	HIST1H2BD	Chr:6p21.3	AL353759	Hs.180779	7.7	Above	3.5
39	218825_at	EGF-like-domain, multiple 7	EGFL7	Chr:9q34.3	NM_016215	Hs.91481	7.6	Above	3.1
40	217795_s_at	hypothetical protein MGC3222	MGC3222	Chr:3p25.1	W74580	Hs.130330	7.6	Above	1.5
41	206090_s_at	disrupted in schizophrenia 1	DISC1	Chr:1q42.1	NM_018662	Hs.26985	7.6	Above	1.8
42	209543_s_at	CD34 antigen	CD34	Chr:1q32	M81104	Hs.374990	7.6	Above	3.5
43	203859_s_at	paralemmin	PALM	Chr:19p13.3	NM_002579	Hs.78482	7.3	Above	2.2
44	219478_at	WAP four-disulfide core domain 1	WFDC1	Chr:16q24.3	NM_021197	Hs.36688	7.2	Above	3.3
45	205513_at	transcobalamin I (vitamin B12 binding protein, R binder family)	TCN1	Chr:11q11-q12	NM_001062	Hs.2012	7.2	Above	5.0
46	219911_s_at	solute carrier family 21 (organic anion transporter), member 12	SLC21A12	Chr:20q13.33	NM_016354	Hs.235782	7.1	Above	2.8
47	212706_at	DNA directed RNA polymerase II polypeptide J-related gene	POLR2J2	Chr:7q11.22	AB011110	Hs.406505	7.0	Above	2.1
48	222061_at	CD58 antigen, (lymphocyte function-associated antigen 3)	CD58	Chr:1p13	AA700015	Hs.75626	6.9	Above	3.5
49	219201_s_at	twisted gastrulation homolog 1 (Drosophila)	TWSG1	Chr:18p11.3	NM_020648	Hs.247302	6.7	Above	5.8

50	204494_s_at	DKFZP434H132 protein	DKFZP434H132	Chr:15q23	AW516789	Hs.17936	-6.6	Below	2.6
51	212370_x_at	hypothetical protein FLJ10824	FLJ10824	Chr:10q11.23	AL080183	Hs.375174	6.6	Above	1.7
52	216356_x_at	BAI1-associated protein 3	BAIAP3	Chr:16p13.3	AB018277	Hs.458427	6.6	Above	3.0
53	204495_s_at	DKFZP434H132 protein	DKFZP434H132	Chr:15q23	NM_015492	Hs.17936	-6.5	Below	2.8
54	206147_x_at	sex comb on midleg-like 2 (Drosophila)	SCML2	Chr:Xp22	NM_006089	Hs.171558	6.5	Above	1.7
55	206726_at	prostaglandin D2 synthase, hematopoietic	PGDS	Chr:4q22.3	NM_014485	Hs.128433	6.5	Above	4.0
56	212573_at	KIAA0830 protein	KIAA0830	Chr:11q21	AF131747	Hs.167115	6.4	Above	1.8
57	209911_x_at	histone 1, H2bd	HIST1H2BD	Chr:6p21.3	BC002842	Hs.180779	6.4	Above	2.3
58	219255_x_at	interleukin 17 receptor B	IL17RB	Chr:3p21.1	NM_018725	Hs.5470	6.4	Above	2.9
59	212492_s_at	KIAA0876 protein	KIAA0876	Chr:19p13.3	AW237172	Hs.301011	6.3	Above	2.5
60	218332_at	brain expressed, X-linked 1	BEX1	Chr:Xq21-q23	NM_018476	Hs.334370	6.3	Above	4.5
61	219132_at	pellino homolog 2 (Drosophila)	PELI2	Chr:14q21	NM_021255	Hs.44038	6.3	Above	2.1
62	201839_s_at	tumor-associated calcium signal transducer 1	TACSTD1	Chr:2p21	NM_002354	Hs.692	6.3	Above	4.2
63	201721_s_at	Lysosomal-associated multispinning membrane protein-5	LAPTM5	Chr:1p34	NM_006762	Hs.436200	-6.2	Below	2.2
64	218899_s_at	brain and acute leukemia, cytoplasmic	BAALC	Chr:8q22.3	NM_024812	Hs.169395	6.2	Above	3.6
65	219890_at	C-type (calcium dependent, carbohydrate-recognition domain) lectin, superfamily member 5	CLECSF5	Chr:7q33	NM_013252	Hs.126355	6.1	Above	2.9
66	209137_s_at	ubiquitin specific protease 10	USP10	Chr:16q24.1	BC000263	Hs.78829	6.1	Above	1.5
67	218742_at	protein related to Narf	HPRN	Chr:16p13.3	NM_022493	Hs.22158	6.1	Above	2.4
68	218205_s_at	MAP kinase-interacting serine/threonine kinase 2	MKNK2	Chr:19p13.3	NM_017572	Hs.512094	6.0	Above	1.7
69	204319_s_at	regulator of G-protein signalling 10	RGS10	Chr:10q25	NM_002925	Hs.82280	6.0	Above	2.3
70	211068_x_at	KIAA0592 protein	KIAA0592	Chr:10q11.1	BC006456	Hs.439367	6.0	Above	1.6
71	203379_at	ribosomal protein S6 kinase, 90kDa, polypeptide 1	RPS6KA1	Chr:3	NM_002953	Hs.149957	-6.0	Below	2.0
72	218557_at	Nit protein 2	NIT2	Chr:3q12.3	NM_020202	Hs.439152	5.9	Above	1.7
73	208712_at	cyclin D1 (PRAD1: parathyroid adenomatosis 1)	CCND1	Chr:11q13	M73554	Hs.371468	5.9	Above	4.2
74	214349_at	Homo sapiens transcribed sequences	---	---	AV764378	Hs.464403	5.9	Above	3.9
75	204929_s_at	vesicle-associated membrane protein 5 (myobrevin)	VAMP5	Chr:2p11.2	NM_006634	Hs.74669	5.9	Above	2.3
76	208546_x_at	histone 1, H2bh	HIST1H2BH	Chr:6p21.3	NM_003524	Hs.247815	5.9	Above	2.6
77	220307_at	natural killer cell receptor 2B4	CD244	Chr:1q23.1	NM_016382	Hs.157872	5.9	Above	2.2
78	202081_at	immediate early protein	ETR101	Chr:19p13.13	NM_004907	Hs.737	5.8	Above	2.0
79	202117_at	Rho GTPase activating protein 1	ARHGAP1	Chr:11p12-q12	BG468434	Hs.138860	-5.8	Below	1.6
80	214946_x_at	hypothetical protein FLJ10824	FLJ10824	Chr:10q11.23	AV728658	Hs.375174	5.8	Above	1.9

81	202006_at	protein tyrosine phosphatase, non-receptor type 12	PTPN12	Chr:7q11.23	NM_002835	Hs.62	-5.8	Below	2.2
82	208490_x_at	histone 1, H2bf	HIST1H2BF	Chr:6p21.3	NM_003522	Hs.182137	5.8	Above	2.8
83	208885_at	lymphocyte cytosolic protein 1 (L-plastin)	LCP1	Chr:13q14.3	J02923	Hs.381099	-5.7	Below	2.5
84	215411_s_at	chromosome 6 open reading frame 4	C6orf4	Chr:6q21	AL008730	Hs.437508	5.7	Above	2.3
85	221581_s_at	Williams-Beuren syndrome chromosome region 5	WBSCR5	Chr:7q11.23	AF257135	Hs.56607	-5.7	Below	10.6
86	208091_s_at	hypothetical protein DKFZp564K0822	DKFZP564K0822	Chr:7p11.2	NM_030796	Hs.4750	-5.7	Below	3.2
87	219752_at	RAS protein activator like 1 (GAP1 like)	RASAL1	Chr:12q23-q24	NM_004658	Hs.198312	5.7	Above	2.1
88	209539_at	Rac/Cdc42 guanine nucleotide exchange factor (GEF) 6	ARHGEF6	Chr:Xq26	D25304	Hs.372800	5.7	Above	1.6
89	221257_x_at	hypothetical protein SP329	SP329	Chr:5q33.1	NM_030793	Hs.163825	5.7	Above	1.8
90	208523_x_at	histone 1, H2bi	HIST1H2BI	Chr:6p21.3	NM_003525	Hs.182140	5.7	Above	3.2
91	201720_s_at	Lysosomal-associated multispanning membrane protein-5	LAPTM5	Chr:1p34	AI589086	Hs.436200	-5.7	Below	2.8
92	205423_at	adaptor-related protein complex 1, beta 1 subunit	AP1B1	Chr:22q12	NM_001127	Hs.331602	5.6	Above	1.9
93	212929_s_at	KIAA0592 protein	KIAA0592	Chr:10q11.1	W68158	Hs.439367	5.6	Above	1.7
94	220606_s_at	x 006 protein	MDS006	Chr:17p13.1	NM_020233	Hs.47668	5.6	Above	2.0
95	212707_s_at	polymerase (RNA) II (DNA directed) polypeptide J, 13.3kDa	POLR2J	Chr:7q11.2	AI738591	Hs.220255	5.6	Above	2.0
96	205101_at	MHC class II transactivator	MHC2TA	Chr:16p13	NM_000246	Hs.126714	5.6	Above	2.5
97	218486_at	TGFB inducible early growth response 2	TIEG2	Chr:2p25	AA149594	Hs.12229	5.6	Above	2.1
98	211323_s_at	inositol 1,4,5-triphosphate receptor, type 1	ITPR1	Chr:3p26-p25	L38019	Hs.149900	5.6	Above	2.0
99	210377_at	SA hypertension-associated homolog (rat)	SAH	Chr:16p13.11	D16350	Hs.512678	5.6	Above	2.2
100	336_at	---	---	---	D38081	---	5.6	Above	2.0

**Table S8. Top 100 probe sets selected by SAM for t(15;17)[PML-RAR $\alpha$ ]**

U133 probe set	Gene description	Symbol	Location	GenBank reference	UniGene reference	SAM d-score	t(15;17) Above /Below mean	Fold change
1 210755_at	hepatocyte growth factor (hepapoietin A; scatter factor)	HGF	Chr:7q21.1	U46010	Hs.396530	20.1	Above	11.1
2 204150_at	stabilin 1	STAB1	Chr:3p21.31	NM_015136	Hs.301989	19.7	Above	16.7
3 38487_at	stabilin 1	STAB1	Chr:3p21.31	D87433	Hs.301989	18.8	Above	13.8
4 209960_at	hepatocyte growth factor (hepapoietin A; scatter factor)	HGF	Chr:7q21.1	X16323	Hs.396530	16.7	Above	9.6

5	221804_s_at	uncharacterized hypothalamus protein HT011	HT011	Chr:Xq26.1	BE565675	Hs.434241	16.0	Above	3.9
6	211748_x_at	prostaglandin D2 synthase 21kDa (brain)	PTGDS	Chr:9q34.2- q34.3	BC005939	Hs.446429	15.4	Above	20.7
7	211663_x_at	prostaglandin D2 synthase 21kDa (brain)	PTGDS	Chr:9q34.2- q34.3	M61900	Hs.446429	15.1	Above	13.4
8	210997_at	hepatocyte growth factor (hepapoietin A; scatter factor)	HGF	Chr:7q21.1	M77227	Hs.396530	15.0	Above	17.8
9	218618_s_at	FAD104	FAD104	Chr:3q26.31	NM_022763	Hs.299883	14.8	Above	4.5
10	212187_x_at	prostaglandin D2 synthase 21kDa (brain)	PTGDS	Chr:9q34.2- q34.3	NM_000954	Hs.446429	14.6	Above	20.6
11	219225_at	piggyBac transposable element derived 5	PGBD5	Chr:1q42.2	NM_024554	Hs.12247	14.3	Above	9.8
12	209961_s_at	hepatocyte growth factor (hepapoietin A; scatter factor)	HGF	Chr:7q21.1	M60718	Hs.396530	14.2	Above	4.0
13	37408_at	mannose receptor, C type 2	MRC2	Chr:17q24.1	AB014609	Hs.7835	14.1	Above	4.5
14	212285_s_at	agrin	AGRN	Chr:1p36.33	AW008051	Hs.273330	13.9	Above	4.3
15	210998_s_at	hepatocyte growth factor (hepapoietin A; scatter factor)	HGF	Chr:7q21.1	M77227	Hs.396530	13.9	Above	12.8
16	39248_at	aquaporin 3	AQP3	Chr:9p13	N74607	Hs.234642	13.2	Above	6.0
17	201069_at	matrix metalloproteinase 2 (gelatinase A, 72kDa gelatinase, 72kDa type IV collagenase)	MMP2	Chr:16q13- q21	NM_004530	Hs.367877	12.9	Above	8.3
18	217419_x_at	agrin	AGRN	Chr:1p36.33	AK021586	Hs.273330	12.8	Above	3.4
19	210140_at	cystatin F (leukocystatin)	CST7	Chr:20p11.21	AF031824	Hs.143212	11.8	Above	6.7
20	214228_x_at	tumor necrosis factor receptor superfamily, member 4	TNFRSF4	Chr:1p36	AJ277151	Hs.129780	11.6	Above	10.6
21	201063_at	reticulocalbin 1, EF-hand calcium binding domain	RCN1	Chr:11p13	NM_002901	Hs.167791	11.2	Above	2.8
22	203074_at	annexin A8	ANXA8	Chr:10q11.2	NM_001630	Hs.87268	11.2	Above	6.3
23	212953_x_at	calreticulin	CALR	Chr:19p13.3- p13.2	BE251303	Hs.353170	11.1	Above	3.1
24	216320_x_at	macrophage stimulating 1 (hepatocyte growth factor-like)	MST1	Chr:3p21	U37055	Hs.512587	11.0	Above	5.6
25	205614_x_at	macrophage stimulating 1 (hepatocyte growth factor-like)	MST1	Chr:3p21	NM_020998	Hs.512587	10.9	Above	7.2
26	214315_x_at	calreticulin	CALR	Chr:19p13.3- p13.2	AI348935	Hs.353170	10.8	Above	2.6
27	205110_s_at	fibroblast growth factor 13	FGF13	Chr:Xq26.3	NM_004114	Hs.6540	10.6	Above	10.5
28	221971_x_at	Homo sapiens similar to ARF GTPase-activating protein (LOC283018), mRNA	---	---	BE672818	Hs.448612	10.3	Above	3.1
29	212509_s_at	Homo sapiens transcribed sequence with strong similarity to protein sp:P05023 (H.sapiens)	---	---	BF968134	Hs.356623	10.2	Above	3.7

		A1A1_HUMAN Sodium/potassium-transporting ATPase alpha-1 chain precursor (Sodium pump 1) (Na+/K+ ATPase 1)							
30	213492_at	collagen, type II, alpha 1 (primary osteoarthritis, spondyloepiphyseal dysplasia, congenital)	COL2A1	Chr:12q13.11	X06268	Hs.408182	10.1	Above	4.3
31	202260_s_at	syntaxin binding protein 1	STXBP1	Chr:9q34.1	NM_003165	Hs.325862	10.1	Above	4.1
32	205899_at	cyclin A1	CCNA1	Chr:13q12.3-q13	NM_003914	Hs.417050	10.0	Above	7.2
33	213380_x_at	macrophage stimulating, pseudogene 9	MSTP9	Chr:1p36.13	AA911235	Hs.475654	9.9	Above	2.9
34	202718_at	insulin-like growth factor binding protein 2, 36kDa	IGFBP2	Chr:2q33-q34	NM_000597	Hs.433326	9.8	Above	7.0
35	45687_at	hypothetical protein MGC3121	MGC3121	Chr:16p11.2	AA161130	Hs.293629	9.8	Above	2.1
36	214450_at	cathepsin W (lymphopain)	CTSW	Chr:11q13.1	NM_001335	Hs.416848	9.8	Above	4.8
37	218922_s_at	hypothetical protein FLJ12089	FLJ12089	Chr:19p13.3	NM_024552	Hs.300439	9.6	Above	5.3
38	219527_at	hypothetical protein FLJ20605	FLJ20605	Chr:1q42.11	NM_017898	Hs.4932	9.6	Above	4.0
39	221850_x_at	Homo sapiens cDNA FLJ38788 fis, clone LIVER2002638.	---	---	AI826075	Hs.446716	9.4	Above	2.6
40	218714_at	hypothetical protein MGC3121	MGC3121	Chr:16p11.2	NM_024031	Hs.293629	9.4	Above	2.3
41	216614_at	---	---	---	AL049988	---	9.3	Above	4.7
42	201591_s_at	nischarin	NISCH	Chr:3p21.1	NM_007184	Hs.435290	9.3	Above	2.1
43	204425_at	Rho GTPase activating protein 4	ARHGAP4	Chr:Xq28	NM_001666	Hs.3109	-9.3	Below	17.6
44	212599_at	autism susceptibility candidate 2	AUTS2	Chr:7q11.23	AK025298	Hs.296720	9.3	Above	3.0
45	203916_at	N-deacetylase/N-sulfotransferase (heparan glucosaminyl) 2	NDST2	Chr:10q22	NM_003635	Hs.225129	9.3	Above	2.3
46	205624_at	carboxypeptidase A3 (mast cell)	CPA3	Chr:3q21-q25	NM_001870	Hs.646	9.2	Above	6.0
47	207895_at	N-acetylated alpha-linked acidic dipeptidase-like	NAALADASEL	Chr:11q12	NM_005468	Hs.13967	9.2	Above	8.6
48	203151_at	microtubule-associated protein 1A	MAP1A	Chr:15q13-qter	AW296788	Hs.194301	9.1	Above	2.9
49	203949_at	myeloperoxidase	MPO	Chr:17q23.1	NM_000250	Hs.458272	9.1	Above	3.7
50	201952_at	activated leukocyte cell adhesion molecule	ALCAM	Chr:3q13.1	AA156721	Hs.10247	9.1	Above	3.0
51	209574_s_at	chromosome 18 open reading frame 1	C18orf1	Chr:18p11.2	AI349506	Hs.285091	9.0	Above	3.2
52	212012_at	Melanoma associated gene	D2S448	Chr:2pter-p25.1	BF342851	Hs.118893	9.0	Above	5.7
53	207106_s_at	leukocyte tyrosine kinase	LTK	Chr:15q15.1-q21.1	NM_002344	Hs.434481	9.0	Above	4.1
54	202655_at	arginine-rich, mutated in early stage tumors	ARMET	Chr:3p21.1	NM_006010	Hs.436446	9.0	Above	2.2
55	219491_at	hypothetical protein MGC3103	MGC3103	Chr:11q13.1	NM_024036	Hs.148438	9.0	Above	3.8
56	212204_at	DKFZP564G2022 protein	DKFZP564G2022	Chr:15q14	AL049944	Hs.200692	9.0	Above	1.9

57	212013_at	Melanoma associated gene	D2S448	Chr:2pter-p25.1	D86983	Hs.118893	8.9	Above	7.5
58	216268_s_at	jagged 1 (Alagille syndrome)	JAG1	Chr:20p12.1-p11.23	U77914	Hs.409202	8.8	Above	5.0
59	213355_at	alpha2,3-sialyltransferase	ST3GALVI	Chr:3q12.2	AI989567	Hs.440913	8.8	Above	3.2
60	203221_at	transducin-like enhancer of split 1 (E(sp1) homolog, Drosophila)	TLE1	Chr:9q21.32	AI758763	Hs.406491	8.7	Above	3.5
61	217653_x_at	Homo sapiens transcribed sequence with weak similarity to protein ref:NP_060312.1 (H.sapiens) hypothetical protein FLJ20489 [Homo sapiens]	---	---	AW150065	Hs.499531	8.7	Above	2.9
62	214203_s_at	proline dehydrogenase (oxidase) 1	PRODH	Chr:22q11.21	AA074145	Hs.343874	8.6	Above	5.1
63	221799_at	KIAA1402 protein	CSG1cA-T	Chr:7q36.1	AB037823	Hs.86392	8.6	Above	2.3
64	212810_s_at	solute carrier family 1 (glutamate/neutral amino acid transporter), member 4	SLC1A4	Chr:2p15-p13	W72527	Hs.323878	8.6	Above	2.5
65	200986_at	serine (or cysteine) proteinase inhibitor, clade G (C1 inhibitor), member 1, (angioedema, hereditary)	SERPING1	Chr:11q12-q13.1	NM_000062	Hs.384598	8.5	Above	4.5
66	201596_x_at	keratin 18	KRT18	Chr:12q13	NM_000224	Hs.406013	8.5	Above	4.0
67	221636_s_at	hypothetical protein FLJ20605	FLJ20605	Chr:1q42.11	AL136931	Hs.4932	8.5	Above	3.2
68	205047_s_at	asparagine synthetase	ASNS	Chr:7q21.3	NM_001673	Hs.446546	8.4	Above	3.6
69	200654_at	procollagen-proline, 2-oxoglutarate 4-dioxygenase (proline 4-hydroxylase), beta polypeptide (protein disulfide isomerase; thyroid hormone binding protein p55)	P4HB	Chr:17q25	J02783	Hs.410578	8.3	Above	2.4
70	203364_s_at	KIAA0652 gene product	KIAA0652	Chr:11p11.2	NM_014741	Hs.410092	8.3	Above	1.7
71	219167_at	Ras family member Ris	RIS	Chr:15q11.2-q22.33	NM_016563	Hs.27018	8.3	Above	4.5
72	214582_at	phosphodiesterase 3B, cGMP-inhibited	PDE3B	Chr:11p15.1	NM_000753	Hs.337616	8.3	Above	2.2
73	204446_s_at	arachidonate 5-lipoxygenase	ALOX5	Chr:10q11.2	NM_000698	Hs.89499	8.2	Above	4.1
74	200935_at	calreticulin	CALR	Chr:19p13.3-p13.2	NM_004343	Hs.353170	8.2	Above	3.6
75	210794_s_at	maternally expressed 3	MEG3	Chr:14q32	AF119863	Hs.418271	8.2	Above	26.0
76	203948_s_at	myeloperoxidase	MPO	Chr:17q23.1	J02694	Hs.458272	8.1	Above	3.3
77	213942_at	EGF-like-domain, multiple 3	EGFL3	Chr:1p36.3	AL134303	Hs.56186	8.1	Above	2.7
78	200945_s_at	yeast Sec31p homolog	KIAA0905	Chr:4q21.3	NM_014933	Hs.436549	8.1	Above	1.5
79	204099_at	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily d, member 3	SMARCD3	Chr:7q35-q36	NM_003078	Hs.444445	8.0	Above	3.2

80	206335_at	galactosamine (N-acetyl)-6-sulfate sulfatase (Morquio syndrome, mucopolysaccharidosis type IVA)	GALNS	Chr:16q24.3	NM_000512	Hs.159479	8.0	Above	2.1
81	209686_at	S100 calcium binding protein, beta (neural)	S100B	Chr:21q22.3	BC001766	Hs.422181	8.0	Above	10.0
82	217961_at	hypothetical protein FLJ20551	FLJ20551	Chr:3p21.33	NM_017875	Hs.7994	7.9	Above	1.8
83	202650_s_at	KIAA0195 gene product	KIAA0195	Chr:17q25.3	NM_014738	Hs.301132	7.9	Above	2.1
84	206105_at	fragile X mental retardation 2	FMR2	Chr:Xq28	NM_002025	Hs.54472	7.9	Above	2.4
85	205382_s_at	D component of complement (adipsin)	DF	Chr:19p13.3	NM_001928	Hs.155597	7.9	Above	3.5
86	202956_at	brefeldin A-inhibited guanine nucleotide-exchange protein 1	BIG1	Chr:8q13	NM_006421	Hs.94631	7.9	Above	1.9
87	205518_s_at	Homo sapiens mRNA for CMP-N-acetylneuraminic acid hydroxylase, complete cds.	---	---	NM_003570	Hs.24697	7.8	Above	2.6
88	219868_s_at	ankyrin repeat and FYVE domain containing 1	ANKFY1	Chr:17p13.3	NM_016376	Hs.512604	7.8	Above	1.9
89	209099_x_at	jagged 1 (Alagille syndrome)	JAG1	Chr:20p12.1-p11.23	U73936	Hs.409202	7.8	Above	4.3
90	212566_at	microtubule-associated protein 4	MAP4	Chr:3p21	AL523310	Hs.31095	7.8	Above	1.6
91	211936_at	heat shock 70kDa protein 5 (glucose-regulated protein, 78kDa)	HSPA5	Chr:9q33-q34.1	AF216292	Hs.310769	7.7	Above	2.6
92	202660_at	Homo sapiens transcribed sequences	---	---	AA834576	Hs.406751	7.6	Above	2.9
93	213348_at	cyclin-dependent kinase inhibitor 1C (p57, Kip2)	CDKN1C	Chr:11p15.5	N33167	Hs.106070	7.6	Above	2.7
94	221042_s_at	calmin (calponin-like, transmembrane)	CLMN	Chr:14q32.2	NM_024734	Hs.301478	7.6	Above	3.8
95	213385_at	chimerin (chimaerin) 2	CHN2	Chr:7p15.3	AK026415	Hs.407520	7.6	Above	3.0
96	202662_s_at	inositol 1,4,5-triphosphate receptor, type 2	ITPR2	Chr:12p11	NM_002223	Hs.512235	7.6	Above	2.5
97	200825_s_at	hypoxia up-regulated 1	HYOU1	Chr:11q23.1-q23.3	NM_006389	Hs.277704	7.6	Above	2.0
98	205505_at	glucosaminyl (N-acetyl) transferase 1, core 2 (beta-1,6-N-acetylglucosaminyltransferase)	GCNT1	Chr:9q13	NM_001490	Hs.159642	7.6	Above	2.3
99	210999_s_at	growth factor receptor-bound protein 10	GRB10	Chr:7p12-p11.2	U66065	Hs.512118	7.6	Above	2.9
100	201581_at	hypothetical protein DJ971N18.2	DJ971N18.2	Chr:20p12	BF572868	Hs.169358	7.5	Above	2.0

**Table S9. Top 60 Probe sets selected by SAM for *inv(16)[CBF $\beta$ -MYH11]***

U133 probe set	Gene description	Symbol	Location	GenBank reference	UniGene reference	SAM d-score	Inv(16) Above/Below mean	Fold change
----------------	------------------	--------	----------	-------------------	-------------------	-------------	--------------------------	-------------

1	201497_x_at	myosin, heavy polypeptide 11, smooth muscle	MYH11	Chr:16p13.13 -p13.12	NM_022844	Hs.78344	23.3	Above	12.9
2	205330_at	meningioma (disrupted in balanced translocation) 1	MN1	Chr:22q11	NM_002430	Hs.268515	9.7	Above	4.6
3	200665_s_at	secreted protein, acidic, cysteine-rich (osteonectin)	SPARC	Chr:5q31.3- q32	NM_003118	Hs.111779	7.4	Above	4.0
4	212298_at	neuropilin 1	NRP1	Chr:10p12	BE620457	Hs.173548	7.3	Above	5.5
5	202370_s_at	core-binding factor, beta subunit	CBFB	Chr:16q22.1	NM_001755	Hs.179881	-6.9	Below	2.5
6	219243_at	immunity associated protein 4	HIMAP4	Chr:7q36.1	NM_018326	Hs.30822	6.8	Above	5.2
7	204661_at	CDW52 antigen (CAMPATH-1 antigen)	CDW52	Chr:1p36	NM_001803	Hs.276770	6.4	Above	3.4
8	34210_at	CDW52 antigen (CAMPATH-1 antigen)	CDW52	Chr:1p36	N90866	Hs.276770	6.4	Above	3.5
9	204912_at	interleukin 10 receptor, alpha	IL10RA	Chr:11q23	NM_001558	Hs.327	6.3	Above	2.8
10	206978_at	chemokine (C-C motif) receptor 2	CCR2	Chr:3p21	NM_000647	Hs.511794	6.3	Above	3.9
11	206135_at	suppression of tumorigenicity 18 (breast carcinoma) (zinc finger protein)	ST18	Chr:8q11.22	NM_014682	Hs.151449	6.2	Above	6.6
12	209395_at	chitinase 3-like 1 (cartilage glycoprotein-39)	CHI3L1	Chr:1q32.1	M80927	Hs.382202	6.2	Above	11.2
13	219308_s_at	adenylate kinase 5	AK5	Chr:1p31	NM_012093	Hs.18268	6.1	Above	3.6
14	204588_s_at	solute carrier family 7 (cationic amino acid transporter, y+ system), member 7	SLC7A7	Chr:14q11.2	NM_003982	Hs.194693	5.9	Above	3.8
15	207794_at	chemokine (C-C motif) receptor 2	CCR2	Chr:3p21	NM_000648	Hs.511794	5.7	Above	3.8
16	218454_at	hypothetical protein FLJ22662	FLJ22662	Chr:12p13.2	NM_024829	Hs.178470	5.7	Above	5.7
17	203535_at	S100 calcium binding protein A9 (calgranulin B)	S100A9	Chr:1q21	NM_002965	Hs.112405	5.7	Above	3.9
18	218237_s_at	solute carrier family 38, member 1	SLC38A1	Chr:12q13.11	NM_030674	Hs.132246	5.6	Above	2.6
19	202902_s_at	cathepsin S	CTSS	Chr:1q21	NM_004079	Hs.181301	5.6	Above	3.2
20	205174_s_at	glutaminyl-peptide cyclotransferase (glutaminyl cyclase)	QPCT	Chr:2p22.3	NM_012413	Hs.79033	5.6	Above	4.9
21	203329_at	protein tyrosine phosphatase, receptor type, M	PTPRM	Chr:18p11.2	NM_002845	Hs.154151	5.5	Above	3.0
22	202901_x_at	cathepsin S	CTSS	Chr:1q21	BC002642	Hs.181301	5.5	Above	2.7
23	212552_at	hippocalcin-like 1	HPCAL1	Chr:2p25.1	BE617588	Hs.3618	5.4	Above	1.8
24	201496_x_at	myosin, heavy polypeptide 11, smooth muscle	MYH11	Chr:16p13.13 -p13.12	S67238	Hs.78344	5.4	Above	4.9
25	204036_at	endothelial differentiation, lysophosphatidic acid G-protein-coupled receptor, 2	EDG2	Chr:9q32	AW269335	Hs.75794	5.3	Above	2.6
26	202944_at	N-acetylgalactosaminidase, alpha-	NAGA	Chr:22q13- qter	NM_000262	Hs.75372	5.3	Above	2.4
27	222144_at	kinesin family member 17	KIF17	Chr:1p36.12	AA909345	Hs.130411	5.3	Above	3.8
28	209396_s_at	chitinase 3-like 1 (cartilage glycoprotein-39)	CHI3L1	Chr:1q32.1	M80927	Hs.382202	5.1	Above	5.3
29	212192_at	hypothetical protein BC013764	LOC115207	Chr:13q22.1	AI718937	Hs.109438	5.0	Above	3.9

30	205987_at	CD1C antigen, c polypeptide	CD1C	Chr:1q22-q23	NM_001765	Hs.1311	5.0	Above	3.8
31	202917_s_at	S100 calcium binding protein A8 (calgranulin A)	S100A8	Chr:1q21	NM_002964	Hs.416073	5.0	Above	2.8
32	208450_at	lectin, galactoside-binding, soluble, 2 (galectin 2)	LGALS2	Chr:22q12-q13	NM_006498	Hs.113987	5.0	Above	5.3
33	208890_s_at	plexin B2	PLXNB2	Chr:22q13.33	BC004542	Hs.3989	4.9	Above	2.1
34	203923_s_at	cytochrome b-245, beta polypeptide (chronic granulomatous disease)	CYBB	Chr:Xp21.1	NM_000397	Hs.88974	4.9	Above	2.7
35	212535_at	MADS box transcription enhancer factor 2, polypeptide A (myocyte enhancer factor 2A)	MEF2A	Chr:15q26	AA142929	Hs.415033	4.8	Above	1.6
36	215784_at	CD1E antigen, e polypeptide	CD1E	Chr:1q22-q23	AA309511	Hs.249217	4.8	Above	4.6
37	205076_s_at	cisplatin resistance associated	CRA	Chr:1q12-q21	NM_006697	Hs.425144	4.7	Above	3.1
38	202510_s_at	tumor necrosis factor, alpha-induced protein 2	TNFAIP2	Chr:14q32	NM_006291	Hs.101382	4.7	Above	2.4
39	204037_at	endothelial differentiation, lysophosphatidic acid G-protein-coupled receptor, 2	EDG2	Chr:9q32	BF055366	Hs.75794	4.7	Above	2.6
40	212188_at	hypothetical protein BC013764	LOC115207	Chr:13q22.1	AA551075	Hs.109438	4.7	Above	4.6
41	206682_at	C-type (calcium dependent, carbohydrate-recognition domain) lectin, superfamily member 13 (macrophage-derived)	CLECSF13	Chr:17p13.2	NM_006344	Hs.54403	4.7	Above	3.9
42	216984_x_at	immunoglobulin lambda joining 3	IGLJ3	Chr:22q11.1-q11.2	D84143	Hs.449592	4.6	Above	3.3
43	214684_at	MADS box transcription enhancer factor 2, polypeptide A (myocyte enhancer factor 2A)	MEF2A	Chr:15q26	X63381	Hs.415033	4.6	Above	1.7
44	221698_s_at	C-type (calcium dependent, carbohydrate-recognition domain) lectin, superfamily member 12	CLECSF12	Chr:12p13.2-p12.3	AF313468	Hs.161786	4.6	Above	4.2
45	208891_at	dual specificity phosphatase 6	DUSP6	Chr:12q22-q23	BC003143	Hs.298654	4.6	Above	2.3
46	206200_s_at	annexin A11	ANXA11	Chr:10q23	NM_001157	Hs.75510	4.6	Above	1.5
47	201887_at	interleukin 13 receptor, alpha 1	IL13RA1	Chr:Xq24	NM_001560	Hs.285115	4.6	Above	2.1
48	201968_s_at	phosphoglucomutase 1	PGM1	Chr:1p31	NM_002633	Hs.1869	-4.5	Below	2.4
49	201999_s_at	t-complex-associated-testis-expressed 1-like 1	TCTEL1	Chr:6q25.2-q25.3	NM_006519	Hs.266940	4.5	Above	1.5
50	206380_s_at	properdin P factor, complement	PFC	Chr:Xp11.3-p11.23	NM_002621	Hs.53155	4.5	Above	3.0
51	212993_at	hypothetical protein MGC23427	MGC23427	Chr:9q34.3	AA114166	Hs.244847	4.5	Above	2.7
52	211675_s_at	I-mfa domain-containing protein	HIC	Chr:7q31.2	AF054589	Hs.132739	4.5	Above	2.0
53	201506_at	transforming growth factor, beta-induced, 68kDa	TGFBI	Chr:5q31	NM_000358	Hs.421496	4.5	Above	4.9
54	211794_at	FYN binding protein (FYB-120/130)	FYB	Chr:5p13.1	AF198052	Hs.276506	4.5	Above	3.1

55	221724_s_at	C-type (calcium dependent, carbohydrate-recognition domain) lectin, superfamily member 6	CLECSF6	Chr:12p13	AF200738	Hs.115515	4.5	Above	3.9
56	206788_s_at	core-binding factor, beta subunit	CBFB	Chr:16q22.1	AF294326	Hs.179881	-4.5	Below	2.2
57	203973_s_at	KIAA0146 protein	KIAA0146	Chr:8q11.21	NM_005195	Hs.381058	4.5	Above	2.0
58	214864_s_at	glyoxylate reductase/hydroxypyruvate reductase	GRHPR	Chr:9q12	AK024386	Hs.155742	4.4	Above	1.5
59	212830_at	EGF-like-domain, multiple 5	EGFL5	Chr:9q32-q33.3	W68084	Hs.236216	4.4	Above	2.2
60	211645_x_at	Homo sapiens isolate donor Z clone Z55K immunoglobulin kappa light chain variable region mRNA, partial cds	---	---	M85256	Hs.512133	4.4	Above	2.7
61	211798_x_at	immunoglobulin lambda joining 3	IGLJ3	Chr:22q11.1-q11.2	AB001733	Hs.102950	4.4	Above	2.92
62	202720_at	testis derived transcript (3 LIM domains)	TES	Chr:7q31.2	NM_015641	Hs.129129	4.4	Above	1.73
63	210314_x_at	tumor necrosis factor (ligand) superfamily, member 13	TNFSF13	Chr:17p13.1	AF114013	Hs.54673	4.4	Above	2.22

**Table S10. Top 100 probe sets selected by SAM for *MLL* chimeric fusion gene**

U133 probe set	Gene description	Symbol	Location	GenBank reference	UniGene reference	SAM d-score	<i>MLL</i> Above/ Below mean	Fold change
1 205471_s_at	dachshund homolog (Drosophila)	DACH	Chr:13q22	AW772082	Hs.63931	14.8	Above	11.9
2 205472_s_at	dachshund homolog (Drosophila)	DACH	Chr:13q22	NM_004392	Hs.63931	13.9	Above	13.7
3 218376_s_at	NEDD9 interacting protein with calponin homology and LIM domains	NICAL	Chr:6q21	NM_022765	Hs.33476	11.6	Above	3.0
4 204082_at	pre-B-cell leukemia transcription factor 3	PBX3	Chr:9q33-q34	NM_006195	Hs.294101	11.3	Above	4.5
5 202053_s_at	aldehyde dehydrogenase 3 family, member A2	ALDH3A2	Chr:17p11.2	L47162	Hs.440662	10.6	Above	2.4
6 212174_at	adenylate kinase 2	AK2	Chr:1p34	W02312	Hs.294008	10.4	Above	2.5
7 212173_at	adenylate kinase 2	AK2	Chr:1p34	AU154985	Hs.294008	9.7	Above	2.7
8 209616_s_at	carboxylesterase 1 (monocyte/macrophage serine esterase 1)	CES1	Chr:16q13-q22.1	S73751	Hs.278997	9.2	Above	22.3
9 201858_s_at	proteoglycan 1, secretory granule	PRG1	Chr:10q22.1	J03223	Hs.1908	9.2	Above	2.7
10 208967_s_at	adenylate kinase 2	AK2	Chr:1p34	U39945	Hs.294008	9.1	Above	2.3
11 201850_at	capping protein (actin filament), gelsolin-like	CAPG	Chr:2cen-q24	NM_001747	Hs.82422	9.0	Above	3.7
12 212175_s_at	adenylate kinase 2	AK2	Chr:1p34	AL513611	Hs.294008	9.0	Above	2.2
13 204655_at	chemokine (C-C motif) ligand 5	CCL5	Chr:17q11.2-q12	NM_002985	Hs.489044	8.4	Above	5.4

14	205996_s_at	adenylate kinase 2	AK2	Chr:1p34	NM_013411	Hs.294008	8.3	Above	2.3
15	203372_s_at	suppressor of cytokine signaling 2	SOCS2	Chr:12q	AB004903	Hs.405946	8.3	Above	4.3
16	206111_at	ribonuclease, RNase A family, 2 (liver, eosinophil-derived neurotoxin)	RNASE2	Chr:14q24-q31	NM_002934	Hs.728	8.3	Above	3.8
17	206851_at	ribonuclease, RNase A family, 3 (eosinophil cationic protein)	RNASE3	Chr:14q24-q31	NM_002935	Hs.73839	8.2	Above	8.0
18	1405_i_at	chemokine (C-C motif) ligand 5	CCL5	Chr:17q11.2-q12	M21121	Hs.489044	8.2	Above	6.9
19	216950_s_at	Fc fragment of IgG, high affinity Ia, receptor for (CD64)	FCGR1A	Chr:1q21.2-q21.3	X14355	Hs.77424	8.2	Above	3.4
20	218217_at	likely homolog of rat and mouse retinoid-inducible serine carboxypeptidase	RISC	Chr:17q23.2	NM_021626	Hs.431107	8.2	Above	3.8
21	209199_s_at	MADS box transcription enhancer factor 2, polypeptide C (myocyte enhancer factor 2C)	MEF2C	Chr:5q14	N22468	Hs.368950	8.2	Above	2.7
22	203373_at	suppressor of cytokine signaling 2	SOCS2	Chr:12q	NM_003877	Hs.405946	8.1	Above	4.1
23	213915_at	natural killer cell group 7 sequence	NKG7	Chr:19q13.41	NM_005601	Hs.10306	8.1	Above	2.9
24	214511_x_at	Fc fragment of IgG, high affinity Ia, receptor for (CD64)	FCGR1A	Chr:1q21.2-q21.3	L03419	Hs.77424	8.0	Above	3.6
25	208270_s_at	arginyl aminopeptidase (aminopeptidase B)	RNPEP	Chr:1q32	NM_020216	Hs.283667	8.0	Above	2.1
26	200782_at	annexin A5	ANXA5	Chr:4q26-q28	NM_001154	Hs.145741	7.9	Above	3.6
27	204674_at	lymphoid-restricted membrane protein	LRMP	Chr:12p12.1	NM_006152	Hs.124922	7.9	Above	2.4
28	219360_s_at	transient receptor potential cation channel, subfamily M, member 4	TRPM4	Chr:19q13.33	NM_017636	Hs.31608	7.7	Above	9.0
29	201003_x_at	ubiquitin-conjugating enzyme E2 variant 1	UBE2V1	Chr:20q13.2	NM_003349	Hs.381025	7.7	Above	2.0
30	201944_at	hexosaminidase B (beta polypeptide)	HEXB	Chr:5q13	NM_000521	Hs.69293	7.7	Above	2.5
31	202054_s_at	aldehyde dehydrogenase 3 family, member A2	ALDH3A2	Chr:17p11.2	NM_000382	Hs.440662	7.7	Above	2.8
32	201050_at	phospholipase D3	PLD3	Chr:19q13.2	NM_012268	Hs.257008	7.6	Above	8.0
33	209696_at	fructose-1,6-bisphosphatase 1	FBP1	Chr:9q22.3	D26054	Hs.360509	7.5	Above	3.4
34	202139_at	aldo-keto reductase family 7, member A2 (aflatoxin aldehyde reductase)	AKR7A2	Chr:1p35.1-p36.23	NM_003689	Hs.512807	7.4	Above	2.3
35	204633_s_at	ribosomal protein S6 kinase, 90kDa, polypeptide 5	RPS6KA5	Chr:14q31-q32.1	AF074393	Hs.109058	7.4	Above	2.3
36	216331_at	integrin, alpha 7	ITGA7	Chr:12q13	AK022548	Hs.74369	7.4	Above	9.1
37	217861_s_at	prolactin regulatory element binding	PREB	Chr:2p23.3	NM_013388	Hs.279784	7.3	Above	1.6
38	206050_s_at	ribonuclease/angiogenin inhibitor	RNH	Chr:11p15.5	NM_002939	Hs.130958	7.2	Above	2.2
39	209663_s_at	integrin, alpha 7	ITGA7	Chr:12q13	AF072132	Hs.74369	7.2	Above	5.6
40	35974_at	lymphoid-restricted membrane protein	LRMP	Chr:12p12.1	U10485	Hs.124922	7.2	Above	2.8
41	204497_at	adenylate cyclase 9	ADCY9	Chr:16p13.3	AB011092	Hs.20196	7.2	Above	6.5
42	209439_s_at	phosphorylase kinase, alpha 2 (liver)	PHKA2	Chr:Xp22.2-p22.1	D38616	Hs.54941	7.1	Above	1.9

43	205613_at	B/K protein	LOC51760	Chr:16p13.11	NM_016524	Hs.258326	7.1	Above	4.5
44	218109_s_at	hypothetical protein FLJ14153	FLJ14153	Chr:3q25.32	NM_022736	Hs.7503	7.1	Above	1.6
45	210548_at	chemokine (C-C motif) ligand 23	CCL23	Chr:17q21.1	U58913	Hs.169191	7.1	Above	10.2
46	205418_at	feline sarcoma oncogene	FES	Chr:15q26.1	NM_002005	Hs.7636	7.0	Above	2.0
47	210549_s_at	chemokine (C-C motif) ligand 23	CCL23	Chr:17q21.1	U58913	Hs.169191	7.0	Above	20.1
48	212080_at	myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila)	MLL	Chr:11q23	AV714029	Hs.258855	-7.0	Below	2.3
49	203052_at	complement component 2	C2	Chr:6p21.3	NM_000063	Hs.2253	7.0	Above	4.0
50	206100_at	carboxypeptidase M	CPM	Chr:12q15	NM_001874	Hs.511820	7.0	Above	3.9
51	213737_x_at	Homo sapiens transcribed sequences	---	---	AI620911	Hs.50787	7.0	Above	2.3
52	211067_s_at	growth arrest-specific 7	GAS7	Chr:17p	BC006454	Hs.226133	6.9	Above	2.2
53	207968_s_at	MADS box transcription enhancer factor 2, polypeptide C (myocyte enhancer factor 2C)	MEF2C	Chr:5q14	NM_002397	Hs.368950	6.9	Above	2.2
54	200742_s_at	ceroid-lipofuscinosis, neuronal 2, late infantile (Jansky-Bielschowsky disease)	CLN2	Chr:11p15	BG231932	Hs.429658	6.9	Above	1.9
55	217520_x_at	similar to HP95	na	Chr:15q12	BG396614	Hs.374397	6.8	Above	3.8
56	203781_at	mitochondrial ribosomal protein L33	MRPL33	Chr:2p21	NM_004891	Hs.14454	6.8	Above	2.7
57	204561_x_at	apolipoprotein C-II	APOC2	Chr:19q13.2	NM_000483	Hs.75615	6.8	Above	10.7
58	218176_at	melanoma antigen, family F, 1	MAGEF1	Chr:3q13	NM_022149	Hs.306123	6.8	Above	2.1
59	218404_at	sorting nexin 10	SNX10	Chr:7p15.2	NM_013322	Hs.418132	6.8	Above	2.8
60	203733_at	dexamethasone-induced transcript	DEXI	Chr:16p13.2	NM_014015	Hs.157925	6.7	Above	2.9
61	213361_at	tudor repeat associator with PCTAIRE 2	PCTAIRE2BP	Chr:9q22.33	AW129593	Hs.416543	6.7	Above	2.1
62	201859_at	proteoglycan 1, secretory granule	PRG1	Chr:10q22.1	NM_002727	Hs.1908	6.7	Above	1.8
63	213622_at	collagen, type IX, alpha 2	COL9A2	Chr:1p33-p32	AI733465	Hs.418012	6.7	Above	2.3
64	221692_s_at	mitochondrial ribosomal protein L34	MRPL34	Chr:19p13.1	AB049652	Hs.238808	6.7	Above	1.6
65	207704_s_at	growth arrest-specific 7	GAS7	Chr:17p	NM_003644	Hs.226133	6.6	Above	2.5
66	201119_s_at	cytochrome c oxidase subunit VIII	COX8	Chr:11q12-q13	NM_004074	Hs.433901	6.6	Above	1.6
67	221036_s_at	anterior pharynx defective 1B-like	PSFL	Chr:15q22.1	NM_031301	Hs.42954	6.6	Above	1.5
68	214259_s_at	aldo-keto reductase family 7, member A2 (aflatoxin aldehyde reductase)	AKR7A2	Chr:1p35.1-p36.23	AI144075	Hs.512807	6.6	Above	2.0
69	220057_at	G antigen, family D, 2	GAGED2	Chr:Xp11.22-p11.21	NM_020411	Hs.112208	6.5	Above	17.3
70	203909_at	solute carrier family 9 (sodium/hydrogen exchanger), isoform 6	SLC9A6	Chr:Xq26.3	NM_006359	Hs.62185	6.5	Above	1.4
71	210658_s_at	golgi associated, gamma adaptin ear containing, ARF binding protein 2	GGA2	Chr:16p12	BC000284	Hs.133340	6.5	Above	2.1
72	202087_s_at	cathepsin L	CTSL	Chr:9q21-q22	NM_001912	Hs.418123	6.5	Above	2.9
73	203562_at	fasciculation and elongation protein zeta 1 (zygin I)	FEZ1	Chr:11q24.2	NM_005103	Hs.79226	6.5	Above	4.6

74	201105_at	lectin, galactoside-binding, soluble, 1 (galectin 1)	LGALS1	Chr:22q13.1	NM_002305	Hs.407909	6.4	Above	2.5
75	214196_s_at	ceroid-lipofuscinosis, neuronal 2, late infantile (Jansky-Bielschowsky disease)	CLN2	Chr:11p15	AA602532	Hs.429658	6.4	Above	2.1
76	210184_at	integrin, alpha X (antigen CD11C (p150), alpha polypeptide)	ITGAX	Chr:16p11.2	M81695	Hs.385521	6.4	Above	2.9
77	220010_at	potassium voltage-gated channel, Isk-related family, member 1-like	KCNE1L	Chr:Xq22.3	NM_012282	Hs.146372	6.4	Above	6.5
78	201152_s_at	muscleblind-like (Drosophila)	MBNL1	Chr:3q25	N31913	Hs.28578	6.3	Above	1.7
79	202592_at	GCN5 general control of amino-acid synthesis 5-like 1 (yeast)	GCN5L1	Chr:12q13-q14	NM_001487	Hs.94672	6.3	Above	1.5
80	221666_s_at	apoptosis-associated speck-like protein containing a CARD	ASC	Chr:16p12-p11.2	BC004470	Hs.197875	6.3	Above	1.9
81	205936_s_at	hexokinase 3 (white cell)	HK3	Chr:5q35.2	NM_002115	Hs.411695	6.3	Above	5.7
82	202605_at	glucuronidase, beta	GUSB	Chr:7q21.11	NM_000181	Hs.255230	6.3	Above	1.7
83	205859_at	lymphocyte antigen 86	LY86	Chr:6p24.3	NM_004271	Hs.184018	6.3	Above	2.5
84	213908_at	Homo sapiens hypothetical protein LOC339005, mRNA (cDNA clone IMAGE:4837016), partial cds	---	---	AI824078	Hs.212670	6.3	Above	4.2
85	213844_at	homeo box A5	HOXA5	Chr:7p15-p14	NM_019102	Hs.37034	6.2	Above	5.8
86	202180_s_at	major vault protein	MVP	Chr:16p13.1-p11.2	NM_017458	Hs.80680	6.2	Above	2.0
87	209093_s_at	glucosidase, beta; acid (includes glucosylceramidase)	GBA	Chr:1q21	K02920	Hs.511984	6.2	Above	1.5
88	200886_s_at	phosphoglycerate mutase 1 (brain)	PGAM1	Chr:10q25.3	NM_002629	Hs.447492	6.2	Above	1.8
89	220793_at	sarcoma antigen	SAGE	Chr:Xq26	NM_018666	Hs.195292	6.2	Above	7.7
90	214651_s_at	homeo box A9	HOXA9	Chr:7p15-p14	U41813	Hs.127428	6.2	Above	4.3
91	214730_s_at	golgi apparatus protein 1	GLG1	Chr:16q22-q23	AK025457	Hs.78979	6.2	Above	1.6
92	221726_at	ribosomal protein L22	RPL22	Chr:1p36.3-p36.2	BE250348	Hs.326249	-6.2	Below	1.7
93	211004_s_at	aldehyde dehydrogenase 3 family, member B1	ALDH3B1	Chr:11q13	BC002553	Hs.274235	6.2	Above	1.8
94	217827_s_at	acid cluster protein 33	ACP33	Chr:15q21-q22	NM_016630	Hs.242458	6.2	Above	1.6
95	204498_s_at	adenylate cyclase 9	ADCY9	Chr:16p13.3	NM_001116	Hs.20196	6.1	Above	2.6
96	202897_at	protein tyrosine phosphatase, non-receptor type substrate 1	PTPNS1	Chr:20p13	AB023430	Hs.156114	6.1	Above	2.1
97	207614_s_at	cullin 1	CUL1	Chr:7q36.1	NM_003592	Hs.348153	-6.1	Below	2.0
98	218262_at	hypothetical protein FLJ22318	FLJ22318	Chr:5q35.3	NM_022762	Hs.22753	6.1	Above	2.1
99	209367_at	syntaxin binding protein 2	STXBP2	Chr:19p13.3-	AB002559	Hs.379204	6.1	Above	2.1

100	204174_at	arachidonate 5-lipoxygenase-activating protein	ALOX5AP	p13.2 Chr:13q12	NM_001629	Hs.100194	6.0	Above	3.1
-----	-----------	--	---------	--------------------	-----------	-----------	-----	-------	-----

**Table S11. Top 100 Probe sets selected by SAM for FAB-M7**

U133 probe set	Gene description	Symbol	Location	GenBank reference	UniGene reference	SAM d-score	M7 Above/ Below mean	Fold change
1 206493_at	integrin, alpha 2b (platelet glycoprotein IIb of IIb/IIIa complex, antigen CD41B)	ITGA2B	Chr:17q21.32	NM_000419	Hs.411312	24.4	Above	26.1
2 206494_s_at	integrin, alpha 2b (platelet glycoprotein IIb of IIb/IIIa complex, antigen CD41B)	ITGA2B	Chr:17q21.32	NM_000419	Hs.411312	23.9	Above	35.8
3 216956_s_at	integrin, alpha 2b (platelet glycoprotein IIb of IIb/IIIa complex, antigen CD41B)	ITGA2B	Chr:17q21.32	AF098114	Hs.411312	19.5	Above	53.0
4 201147_s_at	tissue inhibitor of metalloproteinase 3 (Sorsby fundus dystrophy, pseudoinflammatory)	TIMP3	Chr:22q12.1-q13.2	BF347089	Hs.245188	16.6	Above	23.8
5 210762_s_at	deleted in liver cancer 1	DLC1	Chr:8p22	AF026219	Hs.8700	16.1	Above	8.4
6 221896_s_at	likely ortholog of mouse hypoxia induced gene 1	HIG1	Chr:3p21.33	BE739519	Hs.7917	16.0	Above	5.2
7 217845_x_at	likely ortholog of mouse hypoxia induced gene 1	HIG1	Chr:3p21.33	NM_014056	Hs.7917	16.0	Above	3.9
8 206655_s_at	glycoprotein Ib (platelet), beta polypeptide	GP1BB	Chr:22q11.21-q11.23	NM_000407	Hs.512000	15.5	Above	21.7
9 206283_s_at	T-cell acute lymphocytic leukemia 1	TAL1	Chr:1p32	NM_003189	Hs.73828	14.3	Above	7.8
10 209839_at	KIAA0820 protein	KIAA0820	Chr:1q24.1	AL136712	Hs.327347	13.8	Above	6.8
11 201150_s_at	tissue inhibitor of metalloproteinase 3 (Sorsby fundus dystrophy, pseudoinflammatory)	TIMP3	Chr:22q12.1-q13.2	NM_000362	Hs.245188	13.3	Above	13.2
12 210446_at	GATA binding protein 1 (globin transcription factor 1)	GATA1	Chr:Xp11.23	M30601	Hs.765	13.3	Above	7.2
13 205352_at	serine (or cysteine) proteinase inhibitor, clade I (neuroserpin), member 1	SERPINI1	Chr:3q26.2	NM_005025	Hs.78589	12.8	Above	13.6
14 208971_at	uroporphyrinogen decarboxylase	UROD	Chr:1p34	M14016	Hs.78601	12.5	Above	3.2
15 206145_at	Rhesus blood group-associated glycoprotein	RHAG	Chr:6p21.1-p11	NM_000324	Hs.368178	12.4	Above	20.3
16 208767_s_at	lysosomal associated protein transmembrane 4 beta	LAPTM4B	Chr:8q22.1	AW149681	Hs.296398	12.3	Above	8.0
17 208970_s_at	uroporphyrinogen decarboxylase	UROD	Chr:1p34	M14016	Hs.78601	12.3	Above	3.7
18 214039_s_at	lysosomal associated protein transmembrane 4 beta	LAPTM4B	Chr:8q22.1	T15777	Hs.296398	12.0	Above	9.7

19	212372_at	myosin, heavy polypeptide 10, non-muscle	MYH10	Chr:17p13	AK026977	Hs.280311	12.0	Above	8.1
20	213107_at	Traf2 and NCK interacting kinase	KIAA0551	Chr:3q26.31	R59093	Hs.252550	11.5	Above	5.0
21	216054_x_at	myosin, light polypeptide 4, alkali; atrial, embryonic	MYL4	Chr:17q21-qter	X58851	Hs.356717	11.5	Above	6.2
22	211254_x_at	Rhesus blood group-associated glycoprotein	RHAG	Chr:6p21.1-p11	AF031549	Hs.368178	11.4	Above	8.1
23	212412_at	LIM protein (similar to rat protein kinase C-binding enigma)	LIM	Chr:4q22	AV715767	Hs.154103	11.4	Above	6.8
24	221021_s_at	catenin, beta like 1	CTNBL1	Chr:20q11.23-q12	NM_030877	Hs.178576	11.2	Above	3.3
25	210395_x_at	myosin, light polypeptide 4, alkali; atrial, embryonic	MYL4	Chr:17q21-qter	AF116676	Hs.356717	11.2	Above	6.0
26	205390_s_at	ankyrin 1, erythrocytic	ANK1	Chr:8p11.1	NM_000037	Hs.443711	11.2	Above	15.0
27	212694_s_at	propionyl Coenzyme A carboxylase, beta polypeptide	PCCB	Chr:3q21-q22	NM_000532	Hs.63788	11.1	Above	2.8
28	208029_s_at	lysosomal associated protein transmembrane 4 beta	LAPTM4B	Chr:8q22.1	NM_018407	Hs.296398	11.0	Above	12.1
29	218711_s_at	serum deprivation response (phosphatidylserine binding protein)	SDPR	Chr:2q32-q33	NM_004657	Hs.26530	10.9	Above	6.7
30	210088_x_at	myosin, light polypeptide 4, alkali; atrial, embryonic	MYL4	Chr:17q21-qter	M36172	Hs.356717	10.7	Above	5.7
31	219546_at	BMP2 inducible kinase	BMP2K	Chr:4q21.23	NM_017593	Hs.20137	10.6	Above	4.6
32	206146_s_at	Rhesus blood group-associated glycoprotein	RHAG	Chr:6p21.1-p11	AF178841	Hs.368178	10.6	Above	28.7
33	209585_s_at	multiple inositol polyphosphate histidine phosphatase, 1	MINPP1	Chr:10q23	AF084943	Hs.95907	10.5	Above	4.8
34	207808_s_at	protein S (alpha)	PROS1	Chr:3p11-q11.2	NM_000313	Hs.64016	10.4	Above	9.1
35	211828_s_at	Traf2 and NCK interacting kinase	KIAA0551	Chr:3q26.31	AF172268	Hs.252550	10.4	Above	7.1
36	204720_s_at	DnaJ (Hsp40) homolog, subfamily C, member 6	DNAJC6	Chr:1pter-q31.3	AV729634	Hs.44896	10.3	Above	10.0
37	204416_x_at	apolipoprotein C-I	APOC1	Chr:19q13.2	NM_001645	Hs.268571	10.3	Above	31.3
38	210215_at	transferrin receptor 2	TFR2	Chr:7q22	AF067864	Hs.63758	10.3	Above	17.2
39	205128_x_at	prostaglandin-endoperoxide synthase 1 (prostaglandin G/H synthase and cyclooxygenase)	PTGS1	Chr:9q32-q33.3	NM_000962	Hs.88474	10.2	Above	3.7
40	203382_s_at	apolipoprotein E	APOE	Chr:19q13.2	NM_000041	Hs.110675	10.2	Above	16.2
41	209881_s_at	linker for activation of T cells	LAT	Chr:16p12.1	AF036905	Hs.498997	10.2	Above	10.1
42	210504_at	Kruppel-like factor 1 (erythroid)	KLF1	Chr:19p13.13-p13.12	U65404	Hs.37860	10.0	Above	5.4

43	201830_s_at	neuroepithelial cell transforming gene 1	NET1	Chr:10p15	NM_005863	Hs.25155	10.0	Above	5.1
44	207194_s_at	intercellular adhesion molecule 4, Landsteiner-Wiener blood group	ICAM4	Chr:19p13.2-	NM_001544	Hs.512159	9.9	Above	4.5
45	202753_at	proteasome regulatory particle subunit p44S10	p44S10	Chr:3p21.1	NM_014814	Hs.350939	9.8	Above	1.8
46	205127_at	prostaglandin-endoperoxide synthase 1 (prostaglandin G/H synthase and cyclooxygenase)	PTGS1	Chr:9q32-q33.3	NM_000962	Hs.88474	9.8	Above	6.6
47	203243_s_at	LIM protein (similar to rat protein kinase C-binding enigma)	LIM	Chr:4q22	NM_006457	Hs.154103	9.7	Above	5.6
48	210986_s_at	tropomyosin 1 (alpha)	TPM1	Chr:15q22.1	Z24727	Hs.133892	9.6	Above	6.0
49	210987_x_at	---	---	---	M19267	---	9.6	Above	3.5
50	200642_at	superoxide dismutase 1, soluble (amyotrophic lateral sclerosis 1 (adult))	SOD1	Chr:21q22.1	NM_000454	Hs.443914	9.6	Above	2.6
51	209586_s_at	TcD37 homolog	HTCD37	Chr:1q21	AF123539	Hs.78524	9.6	Above	2.5
52	215813_s_at	prostaglandin-endoperoxide synthase 1 (prostaglandin G/H synthase and cyclooxygenase)	PTGS1	Chr:9q32-q33.3	S36219	Hs.88474	9.5	Above	3.6
53	217963_s_at	nerve growth factor receptor (TNFRSF16) associated protein 1	NGFRAP1	Chr:Xq22.2	NM_014380	Hs.448588	9.4	Above	7.4
54	218111_s_at	cytidine monophosphate N-acetylneuraminic acid synthetase	CMAS	Chr:12p12.2	NM_018686	Hs.311346	9.3	Above	2.7
55	210036_s_at	potassium voltage-gated channel, subfamily H (eag-related), member 2	KCNH2	Chr:7q35-q36	AB044806	Hs.188021	9.3	Above	6.5
56	204331_s_at	mitochondrial ribosomal protein S12	MRPS12	Chr:19q13.1-q13.2	NM_021107	Hs.411125	9.3	Above	3.1
57	207389_at	glycoprotein Ib (platelet), alpha polypeptide	GP1BA	Chr:17pter-p12	NM_000173	Hs.1472	9.3	Above	6.6
58	205262_at	potassium voltage-gated channel, subfamily H (eag-related), member 2	KCNH2	Chr:7q35-q36	NM_000238	Hs.188021	9.3	Above	21.6
59	202218_s_at	fatty acid desaturase 2	FADS2	Chr:11q12-q13.1	NM_004265	Hs.388164	9.3	Above	3.7
60	203196_at	ATP-binding cassette, sub-family C (CFTR/MRP), member 4	ABCC4	Chr:13q32	AI948503	Hs.307915	9.2	Above	5.4
61	206077_at	Kell blood group	KEL	Chr:7q33	NM_000420	Hs.420322	9.1	Above	6.4
62	213553_x_at	apolipoprotein C-I	APOC1	Chr:19q13.2	W79394	Hs.268571	9.1	Above	5.9
63	211005_at	linker for activation of T cells	LAT	Chr:16p12.1	AF036906	Hs.498997	9.1	Above	6.7
64	204321_at	neogenin homolog 1 (chicken)	NEO1	Chr:15q22.3-q23	NM_002499	Hs.388613	9.0	Above	7.8
65	59644_at	BMP2 inducible kinase	BMP2K	Chr:4q21.23	AI735391	Hs.20137	8.9	Above	3.4
66	210907_s_at	programmed cell death 10	PDCD10	Chr:3q26.2	BC002506	Hs.424619	8.9	Above	1.9

67	203242_s_at	LIM protein (similar to rat protein kinase C-binding enigma)	LIM	Chr:4q22	BG054550	Hs.154103	8.9	Above	5.7
68	212224_at	aldehyde dehydrogenase 1 family, member A1	ALDH1A1	Chr:9q21.13	NM_000689	Hs.76392	8.8	Above	26.8
69	206306_at	ryanodine receptor 3	RYR3	Chr:15q14-q15	NM_001036	Hs.427717	8.8	Above	11.6
70	216804_s_at	LIM protein (similar to rat protein kinase C-binding enigma)	LIM	Chr:4q22	AK027217	Hs.154103	8.8	Above	10.2
71	209218_at	squalene epoxidase	SQLE	Chr:8q24.1	AF098865	Hs.71465	8.8	Above	2.7
72	201148_s_at	tissue inhibitor of metalloproteinase 3 (Sorsby fundus dystrophy, pseudo-inflammatory)	TIMP3	Chr:22q12.1-q13.2	AW338933	Hs.245188	8.8	Above	5.6
73	219215_s_at	solute carrier family 39 (zinc transporter), member 4	SLC39A4	Chr:8q24.3	NM_017767	Hs.411274	8.8	Above	3.4
74	201149_s_at	tissue inhibitor of metalloproteinase 3 (Sorsby fundus dystrophy, pseudo-inflammatory)	TIMP3	Chr:22q12.1-q13.2	U67195	Hs.245188	8.7	Above	3.0
75	202279_at	chromosome 14 open reading frame 2	C14orf2	Chr:14q32.33	NM_004894	Hs.109052	8.7	Above	2.0
76	202949_s_at	four and a half LIM domains 2	FHL2	Chr:2q12-q14	NM_001450	Hs.8302	8.6	Above	13.8
77	204904_at	gap junction protein, alpha 4, 37kDa (connexin 37)	GJA4	Chr:1p35.1	NM_002060	Hs.296310	8.6	Above	15.7
78	209710_at	GATA binding protein 2	GATA2	Chr:3q21.3	AL563460	Hs.367725	8.5	Above	4.7
79	212397_at	radixin	RDX	Chr:11q23	AL137751	Hs.263671	8.5	Above	2.8
80	205389_s_at	ankyrin 1, erythrocytic	ANK1	Chr:8p11.1	AI659683	Hs.443711	8.5	Above	4.8
81	210621_s_at	RAS p21 protein activator (GTPase activating protein) 1	RASA1	Chr:5q13.3	M23612	Hs.758	8.4	Above	2.1
82	220994_s_at	syntaxin binding protein 6 (amisyn)	STXBP6	Chr:14q11.2	NM_014178	Hs.99291	8.4	Above	29.9
83	202620_s_at	procollagen-lysine, 2-oxoglutarate 5-dioxygenase (lysine hydroxylase) 2	PLOD2	Chr:3q23-q24	NM_000935	Hs.41270	8.4	Above	10.1
84	203258_at	DR1-associated protein 1 (negative cofactor 2 alpha)	DRAP1	Chr:11q13.3	NM_006442	Hs.356742	8.4	Above	2.2
85	206702_at	TEK tyrosine kinase, endothelial (venous malformations, multiple cutaneous and mucosal)	TEK	Chr:9p21	NM_000459	Hs.89640	8.3	Above	10.6
86	206793_at	phenylethanolamine N-methyltransferase	PNMT	Chr:17q21-q22	NM_002686	Hs.1892	8.3	Above	9.2
87	209049_s_at	protein kinase C binding protein 1	PRKCBP1	Chr:20q13.13	BC001004	Hs.191990	8.3	Above	2.5
88	210532_s_at	chromosome 14 open reading frame 2	C14orf2	Chr:14q32.33	AF116639	Hs.109052	8.3	Above	1.9
89	203486_s_at	HSPC056 protein	HSPC056	Chr:3q22.3	BF195973	Hs.102708	8.2	Above	3.3
90	202983_at	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 3	SMARCA3	Chr:3q25.1-q26.1	AI760760	Hs.3068	8.2	Above	2.8
91	221488_s_at	chromosome 6 open reading frame 82	C6orf82	Chr:6pter-	AF230924	Hs.107187	8.1	Above	1.7

92	213088_s_at	DnaJ (Hsp40) homolog, subfamily C, member 9	DNAJC9	p21.31 Chr:10q22.3	BE551340	Hs.400676	8.1	Above	2.2
93	216925_s_at	T-cell acute lymphocytic leukemia 1	TAL1	Chr:1p32	X51990	Hs.73828	8.1	Above	4.5
94	208353_x_at	ankyrin 1, erythrocytic	ANK1	Chr:8p11.1	NM_020480	Hs.443711	8.0	Above	4.9
95	208352_x_at	ankyrin 1, erythrocytic	ANK1	Chr:8p11.1	NM_020479	Hs.443711	7.9	Above	4.9
96	221704_s_at	hypothetical protein FLJ12750	FLJ12750	Chr:12q24.31	BC005882	Hs.77870	7.9	Above	3.0
97	208653_s_at	CD164 antigen, sialomucin	CD164	Chr:6q21	AF263279	Hs.43910	7.9	Above	3.1
98	203613_s_at	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 6, 17kDa	NDUFB6	Chr:9p13.3	NM_002493	Hs.109646	7.9	Above	2.0
99	219737_s_at	Homo sapiens transcribed sequence with strong similarity to protein ref:NP_065136.1 (H.sapiens) protocadherin 9 precursor; cadherin superfamily protein VR4-11 [Homo sapiens]	---	---	AI524125	Hs.458282	7.9	Above	6.8
100	40687_at	gap junction protein, alpha 4, 37kDa (connexin 37)	GJA4	Chr:1p35.1	M96789	Hs.296310	7.8	Above	17.8

**Table S12. Top 100 probe sets selected by SAM for CBF leukemias**

U133 probe set	Gene description	Symbol	Location	GenBank reference	UniGene reference	SAM d-score	CBF Above/ Below mean	Fold change
1 206187_at	prostaglandin I2 (prostacyclin) receptor (IP)	PTGIR	Chr:19q13.3	NM_000960	Hs.458324	11.2	Above	4.3
2 205529_s_at	core-binding factor, runt domain, alpha subunit 2; translocated to, 1; cyclin D-related	CBFA2T1	Chr:8q22	NM_004349	Hs.90858	10.3	Above	13.7
3 205528_s_at	core-binding factor, runt domain, alpha subunit 2; translocated to, 1; cyclin D-related	CBFA2T1	Chr:8q22	X79990	Hs.90858	9.9	Above	19.6
4 206622_at	thyrotropin-releasing hormone	TRH	Chr:3q13.3-q21	NM_007117	Hs.182231	9.7	Above	14.4
5 204495_s_at	DKFZP434H132 protein	DKFZP434H132	Chr:15q23	NM_015492	Hs.17936	-9.6	Below	2.3
6 204494_s_at	DKFZP434H132 protein	DKFZP434H132	Chr:15q23	AW516789	Hs.17936	-9.2	Below	2.2
7 206090_s_at	disrupted in schizophrenia 1	DISC1	Chr:1q42.1	NM_018662	Hs.26985	8.7	Above	1.7
8 206940_s_at	POU domain, class 4, transcription factor 1	POU4F1	Chr:13q21.1-q22	NM_006237	Hs.458303	8.7	Above	16.6
9 34210_at	CDW52 antigen (CAMPATH-1 antigen)	CDW52	Chr:1p36	N90866	Hs.276770	8.7	Above	3.5
10 211341_at	POU domain, class 4, transcription factor 1	POU4F1	Chr:13q21.1-q22	L20433	Hs.458303	8.5	Above	31.1

11	218825_at	EGF-like-domain, multiple 7	EGFL7	Chr:9q34.3	NM_016215	Hs.91481	8.4	Above	3.0
12	204661_at	CDW52 antigen (CAMPATH-1 antigen)	CDW52	Chr:1p36	NM_001803	Hs.276770	8.3	Above	3.4
13	208891_at	dual specificity phosphatase 6	DUSP6	Chr:12q22-q23	BC003143	Hs.298654	8.2	Above	2.5
14	208893_s_at	dual specificity phosphatase 6	DUSP6	Chr:12q22-q23	BC005047	Hs.298654	8.2	Above	2.8
15	218858_at	hypothetical protein FLJ12428	FLJ12428	Chr:8q24.12	NM_022783	Hs.87729	8.1	Above	2.4
16	204811_s_at	calcium channel, voltage-dependent, alpha 2/delta subunit 2	CACNA2D2	Chr:3p21.3	NM_006030	Hs.389415	7.9	Above	5.3
17	204020_at	purine-rich element binding protein A	PURA	Chr:5q31	BF739943	Hs.29117	-7.8	Below	2.5
18	218899_s_at	brain and acute leukemia, cytoplasmic	BAALC	Chr:8q22.3	NM_024812	Hs.169395	7.6	Above	3.8
19	211685_s_at	neurocalcin delta	NCALD	Chr:8q22-q23	AF251061	Hs.90063	7.6	Above	2.3
20	216831_s_at	core-binding factor, runt domain, alpha subunit 2; translocated to, 1; cyclin D-related	CBFA2T1	Chr:8q22	AF018283	Hs.90858	7.6	Above	21.4
21	208892_s_at	dual specificity phosphatase 6	DUSP6	Chr:12q22-q23	BC003143	Hs.298654	7.5	Above	2.2
22	212496_s_at	KIAA0876 protein	KIAA0876	Chr:19p13.3	BE256900	Hs.301011	7.5	Above	2.2
23	220560_at	chromosome 11 open reading frame 21	C11orf21	Chr:11p15.5	NM_014144	Hs.272100	-7.4	Below	3.0
24	221760_at	mannosidase, alpha, class 1A, member 1	MAN1A1	Chr:6q22	BG287153	Hs.255149	7.4	Above	3.7
25	209543_s_at	CD34 antigen	CD34	Chr:1q32	M81104	Hs.374990	7.4	Above	3.2
26	212097_at	caveolin 1, caveolae protein, 22kDa	CAV1	Chr:7q31.1	AU147399	Hs.74034	7.2	Above	9.1
27	206855_s_at	hyaluronoglucosaminidase 2	HYAL2	Chr:3p21.3	NM_003773	Hs.76873	7.1	Above	1.6
28	210010_s_at	solute carrier family 25 (mitochondrial carrier; citrate transporter), member 1	SLC25A1	Chr:22q11.21	U25147	Hs.111024	7.1	Above	1.9
29	219686_at	gene for serine/threonine protein kinase	HSA250839	Chr:4p16.2	NM_018401	Hs.58241	7.0	Above	4.2
30	205173_x_at	CD58 antigen, (lymphocyte function-associated antigen 3)	CD58	Chr:1p13	NM_001779	Hs.75626	7.0	Above	2.3
31	208978_at	cysteine-rich protein 2	CRIP2	Chr:14q32.3	U36190	Hs.70327	7.0	Above	4.2
32	219478_at	WAP four-disulfide core domain 1	WFDC1	Chr:16q24.3	NM_021197	Hs.36688	6.8	Above	2.8
33	216356_x_at	BAI1-associated protein 3	BAIAP3	Chr:16p13.3	AB018277	Hs.458427	6.8	Above	2.8
34	212136_at	ATPase, Ca <sup>++</sup> transporting, plasma membrane 4	ATP2B4	Chr:1q25-q32	AW517686	Hs.343522	-6.8	Below	1.7
35	208116_s_at	mannosidase, alpha, class 1A, member 1	MAN1A1	Chr:6q22	NM_005907	Hs.255149	6.8	Above	3.0
36	336_at	---	---	---	D38081	---	6.7	Above	1.9
37	203865_s_at	adenosine deaminase, RNA-specific, B1 (RED1 homolog rat)	ADARB1	Chr:21q22.3	NM_015833	Hs.148822	-6.7	Below	5.3
38	204021_s_at	purine-rich element binding protein A	PURA	Chr:5q31	NM_005859	Hs.29117	-6.7	Below	2.9

39	219271_at	UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase T10	GalNAc-T10	Chr:2p23.2	NM_024572	Hs.15830	6.7	Above	2.7
40	200989_at	hypoxia-inducible factor 1, alpha subunit (basic helix-loop-helix transcription factor)	HIF1A	Chr:14q21-q24	NM_001530	Hs.412416	6.6	Above	1.7
41	201968_s_at	phosphoglucomutase 1	PGM1	Chr:1p31	NM_002633	Hs.1869	-6.6	Below	1.8
42	204285_s_at	phorbol-12-myristate-13-acetate-induced protein 1	PMAIP1	Chr:18q21.32	AI857639	Hs.96	6.6	Above	2.3
43	211744_s_at	CD58 antigen, (lymphocyte function-associated antigen 3)	CD58	Chr:1p13	BC005930	Hs.75626	6.6	Above	2.4
44	221581_s_at	Williams-Beuren syndrome chromosome region 5	WBSCR5	Chr:7q11.23	AF257135	Hs.56607	-6.5	Below	3.5
45	201137_s_at	major histocompatibility complex, class II, DP beta 1	HLA-DPB1	Chr:6p21.3	NM_002121	Hs.368409	6.5	Above	2.3
46	206631_at	prostaglandin E receptor 2 (subtype EP2), 53kDa	PTGER2	Chr:14q22	NM_000956	Hs.2090	6.4	Above	1.9
47	204794_at	dual specificity phosphatase 2	DUSP2	Chr:2q11	NM_004418	Hs.1183	6.4	Above	2.8
48	221865_at	DKFZp547P234 protein	DKFZp547P234	Chr:9q33.1	BF969986	Hs.170226	6.4	Above	1.7
49	204086_at	preferentially expressed antigen in melanoma	PRAME	Chr:22q11.22	NM_006115	Hs.30743	6.3	Above	3.3
50	212135_s_at	ATPase, Ca <sup>++</sup> transporting, plasma membrane 4	ATP2B4	Chr:1q25-q32	AW517686	Hs.343522	-6.3	Below	1.8
51	213194_at	hypothetical protein FLJ10539	FLJ10539	Chr:3p14.2	BF059159	Hs.301198	6.3	Above	7.8
52	218352_at	regulator of chromosome condensation (RCC1) and BTB (POZ) domain containing protein 1	RCBTB1	Chr:13q14	NM_018191	Hs.58452	6.3	Above	2.1
53	200665_s_at	secreted protein, acidic, cysteine-rich (osteonectin)	SPARC	Chr:5q31.3-q32	NM_003118	Hs.111779	6.3	Above	2.8
54	206726_at	prostaglandin D2 synthase, hematopoietic	PGDS	Chr:4q22.3	NM_014485	Hs.128433	6.3	Above	3.5
55	207375_s_at	interleukin 15 receptor, alpha	IL15RA	Chr:10p15-p14	NM_002189	Hs.12503	6.2	Above	2.1
56	204990_s_at	integrin, beta 4	ITGB4	Chr:17q11-qter	NM_000213	Hs.85266	6.2	Above	3.0
57	212099_at	ras homolog gene family, member B	ARHB	Chr:2pter-p12	AI263909	Hs.406064	6.2	Above	2.6
58	212895_s_at	active BCR-related gene	ABR	Chr:17p13.3	AL527773	Hs.434004	-6.2	Below	2.1
59	211991_s_at	major histocompatibility complex, class II, DP alpha 1	HLA-DPA1	Chr:6p21.3	M27487	Hs.914	6.2	Above	2.2
60	210744_s_at	interleukin 5 receptor, alpha	IL5RA	Chr:3p26-p24	M75914	Hs.68876	6.2	Above	6.8
61	205101_at	MHC class II transactivator	MHC2TA	Chr:16p13	NM_000246	Hs.126714	6.1	Above	2.4
62	203773_x_at	biliverdin reductase A	BLVRA	Chr:7p14-cen	NM_000712	Hs.435726	6.1	Above	1.8
63	216942_s_at	CD58 antigen, (lymphocyte function-associated antigen 3)	CD58	Chr:1p13	D28586	Hs.75626	6.1	Above	2.4
64	201497_x_at	myosin, heavy polypeptide 11, smooth muscle	MYH11	Chr:16p13.13	NM_022844	Hs.78344	6.1	Above	4.9

65	218742_at	protein related to Narf	HPRN	-p13.12 Chr:16p13.3	NM_022493	Hs.22158	6.1	Above	2.1
66	211031_s_at	cytoplasmic linker 2	CYLN2	Chr:7q11.23	BC006259	Hs.104717	-6.1	Below	2.6
67	201700_at	cyclin D3	CCND3	Chr:6p21	NM_001760	Hs.83173	-6.0	Below	1.9
68	203741_s_at	adenylate cyclase 7	ADCY7	Chr:16q12- q13	NM_001114	Hs.172199	-6.0	Below	2.3
69	203932_at	major histocompatibility complex, class II, DM beta	HLA-DMB	Chr:6p21.3	NM_002118	Hs.1162	6.0	Above	2.1
70	218205_s_at	MAP kinase-interacting serine/threonine kinase 2	MKNK2	Chr:19p13.3	NM_017572	Hs.512094	6.0	Above	1.5
71	217989_at	retinal short-chain dehydrogenase/reductase 2	RETSR2	Chr:4q22.1	NM_016245	Hs.282984	6.0	Above	1.4
72	211729_x_at	biliverdin reductase A	BLVRA	Chr:7p14-cen	BC005902	Hs.435726	6.0	Above	1.7
73	202664_at	Wiskott-Aldrich syndrome protein interacting protein	WASPIP	Chr:2q31.2	AW058622	Hs.401414	6.0	Above	1.4
74	207134_x_at	tryptase beta 2	TPSB2	Chr:16p13.3	NM_024164	Hs.405479	6.0	Above	3.8
75	217795_s_at	hypothetical protein MGC3222	MGC3222	Chr:3p25.1	W74580	Hs.130330	6.0	Above	1.3
76	205760_s_at	8-oxoguanine DNA glycosylase	OGG1	Chr:3p26.2	NM_016821	Hs.380271	-5.9	Below	2.3
77	219890_at	C-type (calcium dependent, carbohydrate-recognition domain) lectin, superfamily member 5	CLECSF5	Chr:7q33	NM_013252	Hs.126355	5.9	Above	2.5
78	212495_at	KIAA0876 protein	KIAA0876	Chr:19p13.3	BE256900	Hs.301011	5.9	Above	1.8
79	208178_x_at	triple functional domain (PTPRF interacting)	TRIO	Chr:5p15.1- p14	NM_007118	Hs.367689	5.9	Above	2.3
80	210613_s_at	synaptogyrin 1	SYNGR1	Chr:22q13.1	BC000731	Hs.414343	5.9	Above	1.7
81	33304_at	interferon stimulated gene 20kDa	ISG20	Chr:15q26	U88964	Hs.105434	5.9	Above	2.0
82	218613_at	hypothetical protein DKFZp761K1423	DKFZp761K1423	Chr:8p22	NM_018422	Hs.236438	5.9	Above	2.7
83	221012_s_at	tripartite motif-containing 8	TRIM8	Chr:10q24.3	NM_030912	Hs.54580	-5.8	Below	1.8
84	203355_s_at	ADP-ribosylation factor guanine nucleotide factor 6	EFA6R	Chr:8pter- p23.3	NM_015310	Hs.408177	5.8	Above	3.5
85	205683_x_at	tryptase beta 2	TPSB2	Chr:16p13.3	NM_003294	Hs.405479	5.8	Above	3.5
86	221218_s_at	thiamin pyrophosphokinase 1	TPK1	Chr:7q34-q35	NM_022445	Hs.127548	5.8	Above	1.6
87	205330_at	meningioma (disrupted in balanced translocation) 1	MN1	Chr:22q11	NM_002430	Hs.268515	5.8	Above	2.8
88	202255_s_at	signal-induced proliferation-associated 1-like 1	KIAA0440	Chr:14q24.1	NM_015556	Hs.172180	5.8	Above	2.6
89	220668_s_at	DNA (cytosine-5-)-methyltransferase 3 beta	DNMT3B	Chr:20q11.2	NM_006892	Hs.251673	-5.8	Below	2.5

90	204929_s_at	vesicle-associated membrane protein 5 (myobrevin)	VAMP5	Chr:2p11.2	NM_006634	Hs.74669	5.7	Above	2.0
91	221497_x_at	egl nine homolog 1 (C. elegans)	EGLN1	Chr:1q42.1	BC005369	Hs.130946	5.7	Above	1.8
92	218749_s_at	hypothetical protein FLJ22233	FLJ22233	Chr:12q24.21	NM_024959	Hs.286194	-5.7	Below	1.8
93	216832_at	core-binding factor, runt domain, alpha subunit 2; translocated to, 1; cyclin D-related	CBFA2T1	Chr:8q22	AF018283	Hs.90858	5.7	Above	1.8
94	202370_s_at	core-binding factor, beta subunit	CBFB	Chr:16q22.1	NM_001755	Hs.179881	-5.7	Below	1.5
95	212706_at	DNA directed RNA polymerase II polypeptide J-related gene	POLR2J2	Chr:7q11.22	AB011110	Hs.406505	5.7	Above	1.7
96	207081_s_at	phosphatidylinositol 4-kinase, catalytic, alpha polypeptide	PIK4CA	Chr:22q11.21	NM_002650	Hs.334874	-5.7	Below	1.4
97	217080_s_at	homer homolog 2 (Drosophila)	HOMER2	Chr:15q24.3	Y19026	Hs.93564	5.7	Above	1.9
98	203733_at	dexamethasone-induced transcript	DEXI	Chr:16p13.2	NM_014015	Hs.157925	-5.7	Below	4.3
99	201473_at	jun B proto-oncogene	JUNB	Chr:19p13.2	NM_002229	Hs.400124	5.7	Above	2.0
100	217023_x_at	---	---	---	AF099143	---	5.7	Above	3.7

#### Section IV. *MLL* in combined ALL and AML datasets

**Table S13. Top 100 probe sets selected by SAM for *MLL* chimeric fusion gene irrespective of blast lineage**

U133 probe set	Gene description	Symbol	Location	GenBank reference	UniGene reference	SAM d-score	MLL Above/ Below	Fold change
1 204069_at	Meis1, myeloid ecotropic viral integration site 1 homolog (mouse)	MEIS1	Chr:2p14-p13	NM_002398	Hs.170177	14.6	Above	6.0
2 201153_s_at	muscleblind-like (Drosophila)	MBNL1	Chr:3q25	NM_021038	Hs.28578	12.7	Above	1.9
3 201152_s_at	muscleblind-like (Drosophila)	MBNL1	Chr:3q25	N31913	Hs.28578	12.6	Above	1.9
4 214651_s_at	homeo box A9	HOXA9	Chr:7p15-p14	U41813	Hs.127428	11.5	Above	6.6
5 219463_at	chromosome 20 open reading frame 103	C20orf103	Chr:20p12	NM_012261	Hs.22920	11.3	Above	11.3
6 209905_at	homeo box A9	HOXA9	Chr:7p15-p14	AI246769	Hs.127428	10.2	Above	6.8
7 201850_at	capping protein (actin filament), gelsolin-like	CAPG	Chr:2cen-q24	NM_001747	Hs.82422	9.9	Above	3.1
8 204082_at	pre-B-cell leukemia transcription factor 3	PBX3	Chr:9q33-q34	NM_006195	Hs.294101	9.9	Above	3.0
9 201151_s_at	muscleblind-like (Drosophila)	MBNL1	Chr:3q25	BF512200	Hs.28578	9.6	Above	2.2
10 213147_at	homeo box A10	HOXA10	Chr:7p15-p14	AI375919	Hs.110637	9.4	Above	4.9
11 206847_s_at	homeo box A7	HOXA7	Chr:7p15-p14	AF026397	Hs.446318	9.2	Above	3.6
12 213150_at	homeo box A10	HOXA10	Chr:7p15-p14	BF792917	Hs.110637	8.9	Above	5.7

13	218217_at	likely homolog of rat and mouse retinoid-inducible serine carboxypeptidase	RISC	Chr:17q23.2	NM_021626	Hs.431107	8.8	Above	3.2
14	204497_at	adenylate cyclase 9	ADCY9	Chr:16p13.3	AB011092	Hs.20196	8.7	Above	5.8
15	218376_s_at	NEDD9 interacting protein with calponin homology and LIM domains	NICAL	Chr:6q21	NM_022765	Hs.33476	8.7	Above	2.1
16	213844_at	homeo box A5	HOXA5	Chr:7p15-p14	NM_019102	Hs.37034	8.5	Above	5.5
17	211733_x_at	sterol carrier protein 2	SCP2	Chr:1p32	BC005911	Hs.75760	8.4	Above	1.4
18	201105_at	lectin, galactoside-binding, soluble, 1 (galectin 1)	LGALS1	Chr:22q13.1	NM_002305	Hs.407909	8.2	Above	3.3
19	204824_at	endonuclease G	ENDOG	Chr:9q34.1	NM_004435	Hs.420106	8.2	Above	2.5
20	212174_at	adenylate kinase 2	AK2	Chr:1p34	W02312	Hs.294008	8.1	Above	2.1
21	219099_at	chromosome 12 open reading frame 5	C12orf5	Chr:12p13.3	NM_020375	Hs.24792	8.0	Above	1.9
22	218312_s_at	hypothetical protein FLJ12895	FLJ12895	Chr:19q13.43	NM_023926	Hs.235390	-8.0	Below	2.3
23	205472_s_at	dachshund homolog (Drosophila)	DACH	Chr:13q22	NM_004392	Hs.63931	8.0	Above	6.1
24	214297_at	chondroitin sulfate proteoglycan 4 (melanoma-associated)	CSPG4	Chr:15q23	BE857703	Hs.436301	7.9	Above	5.2
25	203733_at	dexamethasone-induced transcript	DEXI	Chr:16p13.2	NM_014015	Hs.157925	7.9	Above	2.6
26	202603_at	a disintegrin and metalloproteinase domain 10	ADAM10	Chr:15q2	N51370	Hs.172028	7.8	Above	1.7
27	205471_s_at	dachshund homolog (Drosophila)	DACH	Chr:13q22	AW772082	Hs.63931	7.8	Above	5.4
28	209167_at	glycoprotein M6B	GPM6B	Chr:Xp22.2	AI419030	Hs.5422	7.6	Above	4.7
29	205821_at	DNA segment on chromosome 12 (unique) 2489 expressed sequence	D12S2489E	Chr:12p13.2-p12.3	NM_007360	Hs.387787	7.6	Above	3.0
30	203781_at	mitochondrial ribosomal protein L33	MRPL33	Chr:2p21	NM_004891	Hs.14454	7.6	Above	2.4
31	218966_at	myosin VC	MYO5C	Chr:15q21	NM_018728	Hs.111782	-7.5	Below	2.8
32	201163_s_at	insulin-like growth factor binding protein 7	IGFBP7	Chr:4q12	NM_001553	Hs.435795	7.5	Above	2.5
33	212175_s_at	adenylate kinase 2	AK2	Chr:1p34	AL513611	Hs.294008	7.5	Above	1.8
34	211066_x_at	protocadherin gamma subfamily C, 3	PCDHGC3	Chr:5q31	BC006439	Hs.283794	7.4	Above	2.6
35	204839_at	RNase MRP/RNase P protein-like	POP5	Chr:12q24.31	NM_015918	Hs.279913	7.4	Above	1.3
36	219577_s_at	ATP-binding cassette, sub-family A (ABC1), member 7	ABCA7	Chr:19p13.3	NM_019112	Hs.134514	7.4	Above	1.7
37	212173_at	adenylate kinase 2	AK2	Chr:1p34	AU154985	Hs.294008	7.4	Above	2.1
38	204937_s_at	zinc finger protein 274	ZNF274	Chr:19qter	NM_016325	Hs.83761	-7.3	Below	1.7
39	219033_at	hypothetical protein FLJ21308	FLJ21308	Chr:5q11.2	NM_024615	Hs.310185	7.3	Above	2.4
40	212306_at	cytoplasmic linker associated protein 2	CLASP2	Chr:3p22.3	AI741784	Hs.108614	7.3	Above	1.4
41	206289_at	homeo box A4	HOXA4	Chr:7p15-p14	NM_002141	Hs.147747	7.2	Above	2.1
42	201944_at	hexosaminidase B (beta polypeptide)	HEXB	Chr:5q13	NM_000521	Hs.69293	7.1	Above	1.8
43	214259_s_at	aldo-keto reductase family 7, member A2 (aflatoxin aldehyde reductase)	AKR7A2	Chr:1p35.1-p36.23	AI144075	Hs.512807	7.1	Above	1.7

44	215836_s_at	protocadherin gamma subfamily C, 3	PCDHGC3	Chr:5q31	AK026188	Hs.283794	7.1	Above	2.3
45	52164_at	chromosome 11 open reading frame 24	C11orf24	Chr:11q13	AA065185	Hs.303025	7.0	Above	1.6
46	217520_x_at	similar to HP95	na	Chr:15q12	BG396614	Hs.374397	7.0	Above	2.9
47	202976_s_at	Rho-related BTB domain containing 3	RHOBTB3	Chr:5q15	NM_014899	Hs.31653	7.0	Above	2.5
48	211926_s_at	myosin, heavy polypeptide 9, non-muscle	MYH9	Chr:22q13.1	AI827941	Hs.146550	7.0	Above	1.5
49	215204_at	Homo sapiens cDNA FLJ14090 fis, clone MAMMA1000264.	---	---	AU147295	Hs.288575	7.0	Above	1.9
50	205996_s_at	adenylate kinase 2	AK2	Chr:1p34	NM_013411	Hs.294008	7.0	Above	2.0
51	221763_at	thyroid hormone receptor interactor 8	TRIP8	Chr:10q22.1	AI694023	Hs.442675	7.0	Above	1.8
52	200897_s_at	palladin	KIAA0992	Chr:4q32.3	NM_016081	Hs.194431	6.9	Above	4.2
53	213915_at	natural killer cell group 7 sequence	NKG7	Chr:19q13.41	NM_005601	Hs.10306	6.9	Above	2.2
54	210549_s_at	chemokine (C-C motif) ligand 23	CCL23	Chr:17q21.1	U58913	Hs.169191	6.8	Above	11.8
55	209079_x_at	protocadherin gamma subfamily C, 3	PCDHGC3	Chr:5q31	AF152318	Hs.283794	6.8	Above	2.3
56	209616_s_at	carboxylesterase 1 (monocyte/macrophage serine esterase 1)	CES1	Chr:16q13-q22.1	S73751	Hs.278997	6.8	Above	9.0
57	209822_s_at	very low density lipoprotein receptor	VLDLR	Chr:9p24	L22431	Hs.370422	6.8	Above	2.6
58	214790_at	SUMO-1-specific protease	SUSP1	Chr:6q13-q14.3	AK001406	Hs.435628	6.7	Above	3.0
59	201050_at	phospholipase D3	PLD3	Chr:19q13.2	NM_012268	Hs.257008	6.7	Above	5.6
60	221458_at	5-hydroxytryptamine (serotonin) receptor 1F	HTR1F	Chr:3p12	NM_000866	Hs.248136	6.7	Above	1.8
61	203836_s_at	mitogen-activated protein kinase kinase kinase 5	MAP3K5	Chr:6q22.33	D84476	Hs.151988	6.7	Above	1.7
62	210660_at	leukocyte immunoglobulin-like receptor, subfamily B (with TM and ITIM domains), member 1	LILRB1	Chr:19q13.4	AF025529	Hs.149924	6.7	Above	2.8
63	200953_s_at	cyclin D2	CCND2	Chr:12p13	NM_001759	Hs.376071	-6.7	Below	3.7
64	213908_at	Homo sapiens hypothetical protein LOC339005, mRNA (cDNA clone IMAGE:4837016), partial cds	---	---	AI824078	Hs.212670	6.6	Above	3.3
65	218824_at	hypothetical protein FLJ10781	FLJ10781	Chr:19q13.33	NM_018215	Hs.8395	6.6	Above	2.7
66	204304_s_at	prominin 1	PROM1	Chr:4p15.33	NM_006017	Hs.370052	6.6	Above	3.6
67	208679_s_at	actin related protein 2/3 complex, subunit 2, 34kDa	ARPC2	Chr:2q36.1	AF279893	Hs.83583	6.6	Above	1.3
68	200046_at	defender against cell death 1	DAD1	Chr:14q11-q12	NM_001344	Hs.82890	6.5	Above	1.6
69	212588_at	protein tyrosine phosphatase, receptor type, C	PTPRC	Chr:1q31-q32	Y00062	Hs.444324	6.5	Above	1.6
70	206851_at	ribonuclease, RNase A family, 3 (eosinophil cationic protein)	RNASE3	Chr:14q24-q31	NM_002935	Hs.73839	6.5	Above	6.1

71	203603_s_at	zinc finger homeobox 1b	ZFHX1B	Chr:2q22	NM_014795	Hs.34871	6.5	Above	1.6
72	202139_at	aldo-keto reductase family 7, member A2 (aflatoxin aldehyde reductase)	AKR7A2	Chr:1p35.1-p36.23	NM_003689	Hs.512807	6.5	Above	1.8
73	202957_at	hematopoietic cell-specific Lyn substrate 1	HCLS1	Chr:3q13	NM_005335	Hs.14601	6.5	Above	1.4
74	204736_s_at	chondroitin sulfate proteoglycan 4 (melanoma-associated)	CSPG4	Chr:15q23	NM_001897	Hs.436301	6.5	Above	3.2
75	200989_at	hypoxia-inducible factor 1, alpha subunit (basic helix-loop-helix transcription factor)	HIF1A	Chr:14q21-q24	NM_001530	Hs.412416	-6.5	Below	1.9
76	202054_s_at	aldehyde dehydrogenase 3 family, member A2	ALDH3A2	Chr:17p11.2	NM_000382	Hs.440662	6.5	Above	2.1
77	201818_at	hypothetical protein FLJ12443	FLJ12443	Chr:5p15.33	NM_024830	Hs.179882	6.5	Above	1.5
78	56256_at	transgelin	TAGLN	Chr:11q23.2	AA150165	Hs.410977	6.5	Above	1.7
79	200760_s_at	cytoskeleton related vitamin A responsive protein	JWA	Chr:3p14	N92494	Hs.92384	6.4	Above	1.4
80	218147_s_at	glycosyltransferase AD-017	AD-017	Chr:3p21.31	NM_018446	Hs.297304	6.4	Above	1.5
81	210548_at	chemokine (C-C motif) ligand 23	CCL23	Chr:17q21.1	U58913	Hs.169191	6.4	Above	6.3
82	212587_s_at	protein tyrosine phosphatase, receptor type, C	PTPRC	Chr:1q31-q32	AI809341	Hs.444324	6.4	Above	1.8
83	206606_at	lipase, hepatic	LIPC	Chr:15q21-q23	NM_000236	Hs.9994	6.4	Above	3.6
84	220643_s_at	Fas apoptotic inhibitory molecule	FAIM	Chr:3q22.3	NM_018147	Hs.173438	6.4	Above	1.9
85	208967_s_at	adenylate kinase 2	AK2	Chr:1p34	U39945	Hs.294008	6.3	Above	1.6
86	203562_at	fasciculation and elongation protein zeta 1 (zygin I)	FEZ1	Chr:11q24.2	NM_005103	Hs.79226	6.3	Above	2.7
87	200782_at	annexin A5	ANXA5	Chr:4q26-q28	NM_001154	Hs.145741	6.3	Above	2.3
88	57540_at	ribokinase	RBSK	Chr:2p23.3	AI823980	Hs.11916	6.2	Above	2.1
89	209170_s_at	glycoprotein M6B	GPM6B	Chr:Xp22.2	AF016004	Hs.5422	6.2	Above	3.3
90	1405_i_at	chemokine (C-C motif) ligand 5	CCL5	Chr:17q11.2-q12	M21121	Hs.489044	6.2	Above	4.7
91	205717_x_at	protocadherin gamma subfamily C, 3	PCDHGC3	Chr:5q31	NM_002588	Hs.283794	6.2	Above	1.8
92	201705_at	proteasome (prosome, macropain) 26S subunit, non-ATPase, 7 (Mov34 homolog)	PSMD7	Chr:16q23-q24	NM_002811	Hs.440604	6.2	Above	1.2
93	221676_s_at	coronin, actin binding protein, 1C	CORO1C	Chr:12q24.1	BC002342	Hs.17377	6.2	Above	1.6
94	202265_at	B lymphoma Mo-MLV insertion region (mouse)	BMI1	Chr:10p11.23	NM_005180	Hs.380403	6.1	Above	2.1
95	201278_at	Homo sapiens cDNA FLJ35517 fis, clone SPLEN2000698.	---	---	N21202	Hs.519270	-6.1	Below	1.8
96	203216_s_at	myosin VI	MYO6	Chr:6q13	NM_004999	Hs.118483	6.1	Above	5.2
97	201280_s_at	disabled homolog 2, mitogen-responsive phosphoprotein (Drosophila)	DAB2	Chr:5p13	NM_001343	Hs.81988	-6.1	Below	3.3

98	207419_s_at	ras-related C3 botulinum toxin substrate 2 (rho family, small GTP binding protein Rac2)	RAC2	Chr:22q13.1	NM_002872	Hs.301175	6.1	Above	1.6
99	216331_at	integrin, alpha 7	ITGA7	Chr:12q13	AK022548	Hs.74369	6.1	Above	7.1
100	209168_at	glycoprotein M6B	GPM6B	Chr:Xp22.2	AW148844	Hs.5422	6.1	Above	2.1

In the analysis below, genes were selected that showed high expression across the majority of *MLL* chimeric fusion gene containing cases, irrespective of their lineage. This was accomplished by selecting those genes from Table S13, for which  $\geq 60\%$  of *MLL* containing cases in each lineage expressed the gene at a level greater than the mean expression in non-*MLL* cases. Twenty-one genes were identified that met these criteria are listed below.

**Table S14. Class discriminating genes for cases with an *MLL* chimeric fusion gene that show uniform high expression, irrespective of blast lineage**

	U133 probe set	Gene description	Symbol	Location	GenBank reference	UniGene reference	Rank <sup>a</sup>	Fold Change vs CD34 <sup>b</sup>	$p \leq 0.001^c$
1	204069_at	Meis1, myeloid ecotropic viral integration site 1 homolog (mouse)	MEIS1	Chr:2p14-p13	NM_002398	Hs.170177	1	1.0	
2	201153_s_at	muscleblind-like (Drosophila)	MBNL1	Chr:3q25	NM_021038	Hs.28578	2	1.4	
3	214651_s_at	homeo box A9	HOXA9	Chr:7p15-p14	U41813	Hs.127428	4	3.5	*
4	201850_at	capping protein (actin filament), gelsolin-like	CAPG	Chr:2cen-q24	NM_001747	Hs.82422	7	6.5	*
5	213147_at	homeo box A10	HOXA10	Chr:7p15-p14	AI375919	Hs.110637	10	2.4	*
6	206847_s_at	homeo box A7	HOXA7	Chr:7p15-p14	AF026397	Hs.446318	11	1.4	
7	218376_s_at	NEDD9 interacting protein with calponin homology and LIM domains	NICAL	Chr:6q21	NM_022765	Hs.33476	15	3.4	*
8	213844_at	homeo box A5	HOXA5	Chr:7p15-p14	NM_019102	Hs.37034	16	2.5	
9	211733_x_at	sterol carrier protein 2	SCP2	Chr:1p32	BC005911	Hs.75760	17	1.0	
10	204824_at	endonuclease G	ENDOG	Chr:9q34.1	NM_004435	Hs.420106	19	1.6	
11	204839_at	RNase MRP/RNase P protein-like	POP5	Chr:12q24.31	NM_015918	Hs.279913	35	0.9	
12	219577_s_at	ATP-binding cassette, sub-family A (ABC1), member 7	ABCA7	Chr:19p13.3	NM_019112	Hs.134514	36	2.4	*
13	206289_at	homeo box A4	HOXA4	Chr:7p15-p14	NM_002141	Hs.147747	41	1.8	
14	211926_s_at	myosin, heavy polypeptide 9, non-muscle	MYH9	Chr:22q13.1	AI827941	Hs.146550	48	2.6	*
15	200897_s_at	palladin	KIAA0992	Chr:4q32.3	NM_016081	Hs.194431	52	2.1	
16	209822_s_at	very low density lipoprotein receptor	VLDLR	Chr:9p24	L22431	Hs.370422	57	6.1	*
17	208679_s_at	actin related protein 2/3 complex, subunit 2, 34kDa	ARPC2	Chr:2q36.1	AF279893	Hs.83583	67	1.6	*
18	212588_at	protein tyrosine phosphatase, receptor type, C	PTPRC	Chr:1q31-q32	Y00062	Hs.444324	69	1.5	*
19	201818_at	hypothetical protein FLJ12443	FLJ12443	Chr:5p15.33	NM_024830	Hs.179882	77	1.2	

20	201705_at	proteasome (prosome, macropain) 26S subunit, non-ATPase, 7 (Mov34 homolog)	PSMD7	Chr:16q23-q24	NM_002811	Hs.440604	92	1.0	
21	207419_s_at	ras-related C3 botulinum toxin substrate 2 (rho family, small GTP binding protein Rac2)	RAC2	Chr:22q13.1	NM_002872	Hs.301175	98	2.1	*

<sup>a</sup>Indicated rankings correspond to those shown in Table S13.

<sup>b</sup>Fold change was calculated as the ratio of the mean signal in cases containing *MLL* chimeric fusion gene to mean signal of sorted normal bone marrow CD34+ cells (n=4).

<sup>c</sup>an \* marks genes with a fold change that was statistically significant at a level of  $p \leq 0.001$  by a t-test.

## Section V. Outcome Analysis

### Expression profile as a predictor of time to relapse

Analysis of the association of gene expression with clinical outcome pediatric patients treated on the AML 87, AML 91, and AML 97 protocols conducted at St. Jude Children's Research Hospital and included in the microarray study. After exclusion of *PML-RAR $\alpha$*  cases, 98 cases had sufficient follow-up for evaluation. Time to relapse or progression was defined as zero for patients never achieving complete remission. For other patients, time to relapse or progression was defined as the time elapsed from study enrollment date to: relapse, death, or most recent follow-up. Patients still living and disease free at last follow-up were considered censored in this analysis. Additionally, patients who died while in first complete remission were censored at the date of death.

16,134 probes sets remained after application of the variation filter. A protocol-stratified randomization divided the 98 patients into a training cohort (n=68) and a validation cohort (n=30). For each probe set and within each protocol, a generalized Mantel statistic (GMS) measured the strength of the association of expression with time to progression or relapse in the training cohort.<sup>6</sup> Our implementation of the GMS is concisely described by analogy to the log-rank test.<sup>7</sup> The log-rank test computes a series of contingency-table chi-square test statistics comparing the distribution of group memberships within the set of individuals known to have failed with that of those individuals known to have not failed prior to each unique observed failure time. Our implementation of the GMS replaces the series of chi-square tests with a series of rank-sum tests comparing the median expression of those having failed to that of those known not to have failed.<sup>8</sup> We assessed the significance of the observed GMS by simulation of the null hypothesis in a series of 10,000 independent replications. The simulation was conducted by computing the GMS statistic for data created by coupling randomly generated "expression" values with the observed failure times and censoring indicators. The p-value for an observed GMS is the proportion of simulated GMS statistics with greater or equal absolute value.

For each probe set, one p-value represented the significance of the association of expression with outcome under each protocol. For each probe set, the three protocol-specific p-values were combined into an across-study summary p-value by comparing the negative sum of the log of the three p-values with the gamma distribution that describes the distribution of three similarly transformed independent uniform (0,1) random variables.<sup>9</sup>

The spacings LOESS histogram was used to estimate the conditional false discovery rate (cFDR) corresponding to each of the summary p-values.<sup>10,11</sup> Table S15 lists the 50 most significant probe sets and their corresponding summary p-values and cFDR estimates. The cFDR estimates imply that approximately half of the probe sets represent false discoveries arising solely due to chance mechanisms. However, these cFDR estimates also clearly indicate that several probe sets' expressions may be truly associated with time to relapse or progression. Therefore, a leave-one-out jackknife was used to identify probe sets whose significance (in the traditional sense) was robust against the exclusion of one patient from the analysis.<sup>12</sup> The jackknife identified three probe sets

having p-values less than or equal  $\alpha = 0.001$  in all 68 leave-one-out GMS assessments, indicated by an asterisk in Table S15.

A multivariable, protocol-stratified, Cox proportional hazards regression model simultaneously examined the association of the three jackknife-selected probe sets with time to progression or relapse within the training cohort.<sup>7</sup> The multivariable Cox analysis found that increased expression of the probe sets 60471\_at and 203063\_at were significantly associated with decreased time to relapse or progression ( $p < 0.0001$  and  $p = 0.0409$  respectively). A prognostic score function based on these two probe sets' expressions was developed by using them as outcome predictors in a second Cox model fit to the training cohort data. An increased score was found to be significantly associated ( $p = 0.0200$ ) with decreased time to relapse or progression in the validation cohort. More specifically, a unit increase in the score is associated with a 1.54 fold increase in the hazard of relapse or progression in the validation cohort (95% CI = 1.05 - 2.27).

The association of the score with time to relapse or progression in the adult cohort was also examined. The power of this analysis was severely limited by the small sample size. A total of 6 patients were excluded: three t(15;17) patients, two patients who refused therapy, and one patient with an extremely rare and complex karyotype (containing both *BCR-ABL* and *CBF $\beta$ -MYH11*) Consequently, only 14 adult patients were available for analysis. Nevertheless, Cox regression analysis suggested that time to relapse or progression in adults also tended to decrease as score increases ( $p = 0.0837$ ).

#### Evaluation of relapse genes identified by Yagi et. al. on this dataset

Recently Yagi et. al utilized the students t-test to isolate a gene expression profile which apparently correlated with prognosis (no remission or relapse in less than 1 year vs. complete response for more than 3 years).<sup>13</sup> When an additional 36 cases were added, hierarchical clustering separated the cases into good prognosis and poor prognosis groups. Utilizing HG-U95 to HG-U133 Best Match table available on the Affymetrix website, HG-U133 probe sets were identified for the 35 genes listed in Table 2 of Yagi et.al.. When applied to the 98 cases of our dataset with adequate time to follow-up, resulting clusters did not correlate with prognosis and resulting Kaplan-Meier curves were not different.

**Table S15. Probe sets most strongly associated with time to progression or relapse in the training set**

U133 probe set	Gene description	Symbol	Location	GenBank reference	UniGene reference	P value <sup>a</sup>	cFDR <sup>b</sup>
1 60471_at <sup>§*</sup>	Homo sapiens transcribed sequence with strong similarity to protein ref:NP_079108.1 hypothetical protein FLJ22439	---	---	AA625133	Hs.512736	0.0001	0.4864
2 202283_at <sup>§</sup>	serine (or cysteine) proteinase inhibitor, clade F (alpha-2 antiplasmin, pigment epithelium derived factor), member 1	SERPINF1	Chr:17p13.1	NM_002615	Hs.173594	0.0001	0.4872
3 200682_s_at	ubiquitin-conjugating enzyme E2L 3	UBE2L3	Chr:22q11.21	BG531983	Hs.108104	0.0004	0.4885
4 217208_s_at	discs, large homolog 1 (Drosophila)	DLG1	Chr:3q29	AL121981	Hs.389893	0.0004	0.4886
5 203063_at <sup>§*</sup>	protein phosphatase 1F (PP2C domain containing)	PPM1F	Chr:22q11.22	NM_014634	Hs.278441	0.0004	0.4886
6 40420_at	serine/threonine kinase 10	STK10	Chr:5q35.1	AB015718	Hs.16134	0.0004	0.4888
7 210720_s_at	amyloid beta (A4) precursor protein-binding, family A, member 2 binding protein	APBA2BP	Chr:20q11.22	AB039947	Hs.324104	0.0004	0.4889
8 209011_at	triple functional domain (PTPRF interacting)	TRIO	Chr:5p15.1-p14	BF223718	Hs.367689	0.0005	0.4891
9 218117_at	ring-box 1	RBX1	Chr:22q13.2	NM_014248	Hs.373992	0.0005	0.4891
10 60528_at	phospholipase A2, group IVB (cytosolic)	PLA2G4B	Chr:15q11.2-q21.3	N71116	Hs.198161	0.0005	0.4896
11 38269_at	protein kinase D2	PRKD2	Chr:19q13.2	AL050147	Hs.205431	0.0005	0.4896
12 219457_s_at	Ras and Rab interactor 3	RIN3	Chr:14q32.13	NM_024832	Hs.413374	0.0008	0.4911
13 209707_at	phosphatidylinositol glycan, class K	PIGK	Chr:1p31.1	AF022913	Hs.293653	0.0008	0.4914
14 218184_at	tubby like protein 4	TULP4	Chr:6q25-q26	NM_020245	Hs.102237	0.0009	0.4917
15 214682_at	polycystic kidney disease 1 (autosomal dominant)	PKD1	Chr:16p13.3	AK023376	Hs.75813	0.0009	0.4917
16 212656_at	hepatocellularcarcinoma-associated antigen HCA557a	DKFZP586D0919	Chr:12q13.2	AF110399	Hs.435643	0.0009	0.4919
17 212129_at	non-imprinted in Prader-Willi/Angelman syndrome 2	NIPA2	Chr:15q11.2	AI589507	Hs.512761	0.0010	0.4921
18 202650_s_at	KIAA0195 gene product	KIAA0195	Chr:17q25.3	NM_014738	Hs.301132	0.0010	0.4923
19 209553_at	KIAA0804 protein	KIAA0804	Chr:3q28	BC001001	Hs.361493	0.0010	0.4924
20 203241_at	UV radiation resistance associated gene	UVRAG	Chr:11q13.5	NM_003369	Hs.13137	0.0010	0.4925
21 217329_x_at	---	---	---	AF042164	---	0.0011	0.4927
22 201381_x_at	Siah-interacting protein	SIP	Chr:1q24-q25	AF057356	Hs.27258	0.0012	0.4930
23 209039_x_at	EH-domain containing 1	EHD1	Chr:11q13	AF001434	Hs.155119	0.0012	0.4933

24	37384_at	protein phosphatase 1F (PP2C domain containing)	PPM1F	Chr:22q11.22	D86995	Hs.278441	0.0013	0.4935
25	202308_at	sterol regulatory element binding transcription factor 1	SREBF1	Chr:17p11.2	NM_004176	Hs.426528	0.0013	0.4935
26	212096_s_at	mitochondrial tumor suppressor gene 1	MTSG1	Chr:8p22	AL096842	Hs.7946	0.0013	0.4935
27	221945_at	KIAA1940 protein	KIAA1940	Chr:2p13.2	AA906578	Hs.23158	0.0013	0.4936
28	200773_x_at	prothymosin, alpha (gene sequence 28)	PTMA	Chr:2q35-q36	NM_002823	Hs.459927	0.0013	0.4937
29	219933_at	glutaredoxin 2	GLRX2	Chr:1q31.2-q31.3	NM_016066	Hs.458283	0.0014	0.4940
30	219416_at	scavenger receptor class A, member 3	SCARA3	Chr:8p21	NM_016240	Hs.128856	0.0014	0.4941
31	218336_at	prefoldin 2	PFDN2	Chr:1q23.1	NM_012394	Hs.298229	0.0014	0.4942
32	211113_s_at	ATP-binding cassette, sub-family G (WHITE), member 1	ABCG1	Chr:21q22.3	U34919	Hs.369055	0.0014	0.4943
33	219033_at	hypothetical protein FLJ21308	FLJ21308	Chr:5q11.2	NM_024615	Hs.310185	0.0015	0.4943
34	216033_s_at	FYN oncogene related to SRC, FGR, YES	FYN	Chr:6q21	S74774	Hs.390567	0.0015	0.4944
35	202530_at	mitogen-activated protein kinase 14	MAPK14	Chr:6p21.3-p21.2	NM_001315	Hs.79107	0.0017	0.4954
36	44146_at	glucocorticoid modulatory element binding protein 2	GMEB2	Chr:20q13.33	AA045183	Hs.28906	0.0017	0.4955
37	214061_at	unknown MGC21654 product	MGC21654	Chr:8q24.13	AI017564	Hs.95631	0.0017	0.4955
38	211096_at	---	---	---	BC003111	---	0.0018	0.4959
39	203662_s_at	tropomodulin 1	TMOD1	Chr:9q22.3	NM_003275	Hs.374849	0.0019	0.4960
40	45653_at	polymerase delta-interacting protein 1	PDIP1	Chr:16p11.2	AW026481	Hs.106817	0.0019	0.4961
41	203047_at	serine/threonine kinase 10	STK10	Chr:5q35.1	NM_005990	Hs.16134	0.0019	0.4962
42	217586_x_at	Homo sapiens transcribed sequence with weak similarity to protein ref:NP_062553.1 (H.sapiens) hypothetical protein FLJ11267 [Homo sapiens]	---	---	N35922	Hs.269852	0.0019	0.4962
43	218890_x_at	mitochondrial ribosomal protein L35	MRPL35	Chr:2p11.2	NM_016622	Hs.433439	0.0020	0.4963
44	218226_s_at	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 4, 15kDa	NDUFB4	Chr:3q13.33	NM_004547	Hs.227750	0.0020	0.4964
45	218896_s_at	ELG protein	HSA277841	Chr:17p13.3	NM_018553	Hs.120963	0.0021	0.4967
46	213880_at	G protein-coupled receptor 49	GPR49	Chr:12q22-q23	AL524520	Hs.166705	0.0022	0.4970
47	219539_at	gem (nuclear organelle) associated protein 6	GEMIN6	Chr:2p22.3	NM_024775	Hs.143818	0.0022	0.4972
48	201840_at	neural precursor cell expressed,	NEDD8	Chr:14q11.2	NM_006156	Hs.75512	0.0023	0.4975

developmentally down-regulated 8							
49	218461_at	protein x 0004	MGC14560	Chr:12q24.13	NM_016301	Hs.284164	0.0024 0.4976
50	212477_at	centaurin, beta 2	CENTB2	Chr:3q29	D26069	Hs.410715	0.0024 0.4976

<sup>a</sup> summary p represents a combined p value for the significance of association of expression with outcome in each treatment protocol as compared with the gamma distribution of three similarly transformed independent variables.

<sup>b</sup> conditional false discovery rate (cFDR) estimated using the LOESS histogram.

<sup>§</sup> Probe sets with p-values less than or equal  $\alpha = 0.001$  in all 68 leave-one-out GMS assessments (jack-knife procedure).

\* Significant in a multivariable Cox regression analysis including all probes selected by the jackknife procedure.

## Section VI. Additional References

- (1) Ross ME, Zhou X, Song G et al. Classification of pediatric acute lymphoblastic leukemia by gene expression profiling. *Blood*. 2003;102:2951-2959.
- (2) Yeoh EJ, Ross ME, Shurtleff SA et al. Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell*. 2002;1:133-143.
- (3) Khan J, Wei JS, Ringner M et al. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat Med*. 2001;7:673-679.
- (4) Ichikawa H, Shimizu K, Hayashi Y, Ohki M. An RNA-binding protein gene, TLS/FUS, is fused to ERG in human myeloid leukemia with t(16;21) chromosomal translocation. *Cancer Res*. 1994;54:2865-2868.
- (5) Morgan R, Riske CB, Meloni A et al. t(16;21)(p11.2;q22): a recurrent primary rearrangement in ANLL. *Cancer Genet Cytogenet*. 1991;53:83-90.
- (6) Mantel N. Chi-Square Tests with 1 Degree of Freedom - Extensions of Mantel-Haenszel Procedure. *J Am Stat Assoc*. 1963;58:690-&.
- (7) Klein JP, Moeschberger M.L. *Survival Analysis*. New York, New York: Springer; 2003.
- (8) Mason RL, Gunst RF, Hess JL. *Statistical Design and Analysis of Experiments*. New York, NY: John Wiley and Sons; 1989.
- (9) Casella G, Berger R. *Statistical Inference*. Pacific Grove, CA: Wadsworth & Brooks/Cole; 1990.
- (10) Pounds S, Cheng C. Improving False Discovery Rate Estimation. *Bioinformatics*. 2004;in press.
- (11) Tsai C-A, Hsueh H-M, Chen JJ. Estimation of false discovery rates in multiple testing: application to gene microarray data. *Biometrics*. 2003;59:1071-1081.
- (12) Klein JP, Moeschberger M.L. A leisurely look at the bootstrap, the jackknife, and cross-validation. *J Am Stat Assoc*. 2003;96:1151-1160.
- (13) Yagi T, Morimoto A, Eguchi M et al. Identification of a gene expression signature associated with pediatric AML prognosis. *Blood*. 2003;102:1849-1856.