

**SUPPLEMENTARY DATA  
FOR**

**CLASSIFICATION, SUBTYPE DISCOVERY, AND PREDICTION OF OUTCOME  
IN PEDIATRIC ACUTE LYMPHOBLASTIC LEUKEMIA BY GENE  
EXPRESSION PROFILING**

Eng-Juh Yeoh, Mary E. Ross, Sheila A. Shurtleff, W. Kent Williams, Divyen Patel, Rami Mahfouz, Fred G. Behm, Susana C. Raimondi, Mary V. Relling, Anami Patel, Cheng Cheng, Dario Campana, Dawn Wilkins, Xiaodong Zhou, Jinyan Li, Huiqing Liu, Ching-Hon Pui, William E. Evans, Clayton Naeve, Limsoon Wong, & James R. Downing

**TABLE OF CONTENTS**

**Section I: Expanded Methods**

- RNA extraction procedure, labeling, hybridization, and data analysis**
- Reproducibility of microarray data**
- Comparison of microarray data between PB and BM leukemic samples**
- Real-time RT-PCR results**
- Comparison of real-time RT-PCR and Affymetrix data**
- Comparison of Affymetrix and immunophenotype results**

**Section II: Patient Dataset**

**Section III: Detailed Analysis**

- Hierarchical cluster analysis of diagnostic cases using genes that passed the variation filter**
- Methods for gene selection**
  - Chi-square**
  - Correlation-based Feature Selection (CFS)**
  - T-statistics**
  - Wilkins'**
  - SOM/DAV**
- Comparison of genes selected by the different metrics**
- Decision tree for the diagnosis of genetic subtypes**
- Description of supervised learning algorithms**
  - Support Vector Machine (SVM)**
  - Prediction by Collective Likelihood of Emerging Patterns (PCL)**
  - K-Nearest Neighbors (K-NN)**
  - Artificial Neural Networks (ANN)**
- Table of results using the different algorithms to predict genetic subgroups**
- Absence of correlation of expression data for genetic subtypes with stage of B-cell differentiation**
- Results for relapse prediction**
  - Permutations test results**
- Results for secondary AML prediction**
- FISH analysis**
- References**

## Section I: Expanded Methods

### RNA extraction, labeling, hybridization, and data analysis

Mononuclear cell suspensions from diagnostic BM aspirates or peripheral blood (PB) samples were prepared from each patient and an aliquot cryopreserved. RNA was extracted using the Trizol reagent (Gibco BRL Life Technologies, Gaithersburg, MD) following the manufacturer's recommended protocol. RNA integrity was assessed by electrophoresis on the Agilent 2100 Bioanalyzer (Agilent, Palo Alto, CA).

First and second strand cDNA were synthesized from 5-15  $\mu$ g of total RNA using the SuperScript Double-Stranded cDNA Synthesis Kit (Gibco Life Technologies) and the oligo-dT<sub>24</sub>-T7 (5'-GGC CAG TGA ATT GTA ATA CGA CTC ACT ATA GGG AGG CGG-3') primer according to the manufacturer's instructions. cRNA was synthesized and labeled with biotinylated UTP and CTP by in vitro transcription using the T7 promoter coupled double stranded cDNA as template and the T7 RNA Transcript Labeling Kit (ENZO Diagnostics Inc., Farmingdale NY). Briefly, double stranded cDNA synthesized from the previous steps was washed twice with 70% ethanol and resuspended in 22  $\mu$ l RNase-free H<sub>2</sub>O. The cDNA was incubated with 4  $\mu$ l of 10X each reaction buffer, 1 $\mu$ l of biotin labeled ribonucleotides, 2 $\mu$ l of DTT, 1 $\mu$ l of RNase inhibitor mix and 2  $\mu$ l 20X T7 RNA Polymerase for 5 hr at 37°C. The labeled cRNA was separated from unincorporated ribonucleotides by passing through a CHROMA SPIN-100 column (Clontech, Palo Alto, CA) and precipitated at -20°C for 1 hr to overnight.

The cRNA pellet was resuspended in 10  $\mu$ l Rnase-free H<sub>2</sub>O and 10.0  $\mu$ g was fragmented by heat and ion-mediated hydrolysis at 95°C for 35 minutes in 200 mM Tris-acetate, pH 8.1, 500 mM KOAc, 150 mM MgOAc. The fragmented cRNA was hybridized for 16 hr at 45°C to HG\_U95Av2 oligonucleotide arrays (Affymetrix, Santa Clara, CA) containing 12,600 probe sets from full length annotated genes together with additional probe sets designed to represent EST sequences. Arrays were washed at 25°C with 6 X SSPE (0.9M NaCl, 60 mM NaH<sub>2</sub>PO<sub>4</sub>, 6 mM EDTA + 0.01% Tween 20) followed by a stringent wash at 50°C with 100 mM MES, 0.1M NaCl<sub>2</sub>, 0.01% Tween 20. The arrays were then stained with phycoerythrin conjugated streptavidin (Molecular Probes, Eugene, OR). Arrays were scanned using a laser confocal scanner (Agilent, Palo Alto, CA) and the expression value for each gene was calculated using Affymetrix Microarray software (MAS 4.0). The signal intensity for each gene was calculated as the average intensity difference (AID), represented by  $[\Sigma(\text{PM} - \text{MM})/(\text{number of probe pairs})]$ , where PM and MM denote perfect-match and mismatch probes, respectively. Expression values were normalized across the sample set by scaling the average of the fluorescent intensities of all genes on an array to a constant target intensity of 2500, then any AID over 45,000 was capped to a value of 45,000. All AID's less than 100, including negative values and absent calls were converted to a value of 1. In addition, a variation filter was used that eliminated any probe set in which less than 1% of the samples had a present call, or if the Max AID - Min AID across the sample set was less than 100. The average intensity differences for each of the remaining genes were analyzed. For some metrics the data was log transformed prior to analysis. The minimum quality control values required for inclusion of a sample's hybridization data in the study were 10% or greater present calls, a GAPDH/Actin 3'/5' ratio <5, and use of a scaling factor that was within 3 standard deviations from the mean of the scaling values of all chips analyzed. The average percent present calls for our overall dataset was 29.7%, and for each of the genetic subgroups was *BCR-ABL* (31.1%), *E2A-PBX1* (28.9%), Hyper >50 (31%), *MLL* (29.8%), T-ALL (29.1%), *TEL-AML1* (28.5%), Novel (30.2%), others (31.1%). In addition, we required each sample to have >75% blasts. The average percentage blasts for the overall dataset used to define the genetic subtypes was 93%, and for each genetic subtype was *BCR-ABL* (92%), *E2A-PBX1* (96%), Hyper >50 (93%), *MLL* (93%), T-ALL (91%), *TEL-AML1* (92%), Novel (95%), and others (94%).

The data was analyzed using the MAS 4.0 software. Recently, a new version of the Affymetrix software became available (Affymetrix MAS 5.0). Affymetrix reports an overall 94% correlation between data analyzed using these two software programs. We have performed a limited analysis to directly compare the results of our data analyzed with these two different software programs. We took 74

randomly selected cases and reanalyzed the primary data using the MAS 5.0 software. Using the default parameters set by Affymetrix, the average percent present call was lower for the MAS 5.0 analyzed data as compare to the values obtained using MAS 4.0; however, with exceedingly rare exception, all of the 271 Chi-squared genes selected as discriminators of the seven genetic subtypes (see below) remained within the new dataset. Moreover, when we used these 271 in a 2-D hierarchical clustering analysis (as described below), the data looked identical to the results obtained using the MAS 4.0 dataset (data not shown).

### Reproducibility of microarray data

The reproducibility of the Affymetrix microarray system was assessed by comparing the gene expression profiles of the following samples: (i) RNA extracted from duplicate cryopreserved diagnostic leukemic samples from 23 patients; (ii) Single RNA samples from 13 patients analyzed on two separate arrays. Results from these analyses are presented in Table 2. As shown, the mean number of probe sets that displayed a  $\geq 2$ -fold difference in expression between separately extracted but paired RNA samples was 144, and for single RNA samples analyzed on two separate occasions was 133. Moreover, very few probe sets were found to have a  $\geq 3$ -fold difference in expression levels between replicate samples. The observed number of probe sets showing a difference in expression values represents  $< 2\%$  of the total number of probe sets on the microarray, and thus these data suggest that the Affymetrix microarray system has a very high degree of reproducibility.

**Table 2. Comparison data from replicate bone marrow samples**

#>2 fold	2.1-3.0	3.1-4.0	>4	Chip	Extraction
42	38	4	0	T-ALL-#3	Different
45	43	2	0	BCR-ABL-#7	Different
59	53	1	5	E2A-PBX-#7	Different
61	55	2	4	T-ALL C25	Different
68	61	6	1	Hyperdip47-50-C17	Different
69	65	3	1	MLL-C6	Different
84	81	1	2	Hyperdip47-50-C18	Different
103	99	4	0	Pseudodip-C10	Different
107	93	11	3	Pseudodip-C12	Different
109	106	3	0	Normal-R1	Different
116	106	7	3	TEL-AML1-C33	Different
117	109	4	4	TEL-AML1-2M#2	Different
126	117	8	1	T-ALL-C3 replicate #1	Different
133	122	7	4	Normal-#3	Different
153	146	5	2	Hypodip-#4 replicate #1	Different
159	117	26	16	Hyperdip47-50-C13	Different
163	121	15	27	TEL-AML1-#10	Different
166	154	12	0	T-ALL-C3 replicate #2	Different
183	168	12	3	Pseudodip-#6	Different
187	155	23	9	T-ALL-RR1	Different
239	200	19	20	Hypodip-#4 replicate #2	Different
256	204	29	23	Hypodip-#4 replicate #3	Different
568	295	137	136	Hyperdip>50-R2M1	Different
21	19	2	0	T-ALL-#7	Same
24	22	2	0	T-ALL-R2	Same
34	33	1	0	T-ALL-C3 replicate #3	Same
52	52	0	0	TEL-AML1-C29	Same
64	58	2	4	Hyperdip>50-C34	Same

64	56	5	3	Hyperdip>50-R2M2	Same
76	74	1	1	T-ALL-R2M1	Same
132	129	1	2	Pseudodip-#7	Same
158	147	9	2	Hyperdip>50-C1	Same
161	139	19	3	TEL-AML1-C20	Same
197	189	6	2	T-ALL-#8	Same
210	169	25	16	Normal-C30-N	Same
539	139	123	277	Pseudodip-C4	Same

	<u>Mean</u>	<u>Median</u>	<u>Stdev</u>
Different Extraction	144.0	117.0	109.0
Same Extraction	133.2	76.0	138.4

### Comparison of microarray data between PB and BM leukemia samples

Matched BM and PB samples that contained  $\geq 80\%$  leukemic blasts were obtained from 10 patients and the RNA extracted and assessed by microarray analysis. As shown in Table 3, a very high level of correlation was observed for the expression profiles of BM and PB, with only 189 probe sets having a >2-fold difference in expression. More importantly, no genes were found to be consistently over- or under-expressed in one sample type. These data suggest that there are minimal differences in the gene expression profiles of leukemic blasts obtained from BM or PB, and raise the possibility that diagnostic gene expression profiling might be possible on samples obtained from the PB.

**Table 3. Comparison of peripheral blood and bone marrow samples**

#>2 fold	2.1-3.0	3.1-4.0	>4	Peripheral Blood	Bone Marrow
98	90	6	2	E2A-PBX1-C7 PB	E2A-PBX1-C7
108	93	9	6	Pseudodip-NHR1 PB	Pseudodip-NHR1
113	105	4	4	BCR-ABL-#1 PB	BCR-ABL-#1
139	117	18	4	Hyperdip>50-C8 PB	Hyperdip>50-C8
141	135	6	0	E2A-PBX1-C3 PB	E2A-PBX1-C3
159	139	16	4	T-ALL-C9 PB	T-ALL-C9
244	223	14	7	Hyperdip>50-C6 PB	Hyperdip>50-C6
245	194	33	18	MLL-C3 PB	MLL-C3
272	246	20	6	MLL-RR3 PB	MLL-RR3
377	300	50	27	E2A-PBX1-C2 PB	E2A-PBX1-C2

	<u>Mean</u>	<u>Median</u>	<u>Stdev</u>
Differences between PB vs BM	189.6	150.0	91.1

### Real-time RT-PCR results

Real-time RT-PCR assays (Taqman; Perkin-Elmer/Applied Biosystems, Foster City, CA) were developed to independently determine the level of mRNA for five genes that were found by microarray analysis to be predictive of either T-lineage ALL (*CD3 $\delta$* , CD3D antigen delta polypeptide TiT3 complex; *MAL*, mal T-Cell differentiation protein; and *PRKCO*, protein kinase C theta) or *E2A-PBX1* expressing ALL (*MERTK*, *c-MER* proto-oncogene tyrosine kinase and KIAA802). RNA samples analyzed included four samples each of *E2A-PBX1* and T-ALL, and two samples each from the remaining subtypes (*BCR-ABL*, *MLL*, *TEL-AML1*, Hyperdiploid >50, Hyperdiploid 47-50, Hypodiploid, Pseudodiploid, and Normal). Whenever possible, the forward and reverse primers were designed in different exons so that DNA contamination would not be a concern. In the case of *MAL* where this was not clear, the RNA was treated

for 15 minutes at room temperature with 1.0 unit of DNase I (Invitrogen, Carlsbad, California) using the Invitrogen protocol to remove any contaminating DNA.

Thirty-three ng of RNA from each sample was reverse transcribed using random hexamers and Multiscribe Reverse Transcriptase (Perkin-Elmer/Applied Biosystems) in a total volume of 10  $\mu$ l. Real Time PCR was performed on a PE Applied Biosystems 7700 prism using oligonucleotide primers and probes sequences designed using Primer Express. The following primers and probes were used: (1) *MERTK*, 5'-GGC GTG CTA ACT GTT CCA GG-3'(forward primer), 5'-CCT TTG TCA TTG TGG GCC TC-3'(reverse primer), and 5'-CAA CTG AAG ACC GCC ATC TCC GTC AG-3'(probe); (2) *KIAA802*, 5'- CCA AGC TGA AGG AGT CGG AC-3'(forward), 5'-AGA TTT CCT TTG TGA TTT TCT TCT TCC-3'(reverse), and 5'-TGC TCG GCC AGT GAG AAT CTC TAC CTG-3'(probe); (3) *PRKCQ*, 5'-AAG CTG CCA CAA GTT CGA CC-3'(forward), 5'-ATG AAG GAA CTG CAG ACC AAG AA-3' (reverse), and 5'-CCA GAG CGA CGT TTT ATG CTG CTG AAA TC-3'(probe); (4) *MAL*, 5'-CCA GTG GCT TCT CGG TCT TC-3' (forward), 5'-CAG GAG CCA CTC ACA AAC TCA A-3'(reverse), and 5'-CCA CCT TGC CCG ACT TGC TCT TCA-3'(probe); and (5) *CD3 $\delta$* , 5'-TCA ACA GAG CTT GTG TGT CGG-3'(forward), 5'-CAT TGC CAC TCT GCT CCT TG-3(reverse), and 5'-TGT CCA GCA AAG CAG AAG ACT CCC AAA-3'(probe). All probes were labeled at the 5' end with FAM (6-carboxy-fluorescein) and at the 3' end with TAMRA (6-carboxy-tetramethyl-rhodamine).

The PCR reactions were performed in a total volume of 50  $\mu$ l containing 10  $\mu$ l of the RT product, 300 nm each of the forward and reverse primers, 100 nM of probe, 1X master mix (PE Applied Biosystems) and 1  $\mu$ l of Taq Gold. Following a 10 minute incubation at 95°C to activate the Taq Gold, samples were denatured at 95°C for 15 seconds, then annealed and extended at 60°C for 1 minute, for a total of 40 cycles. The RNA from each sample was also amplified using primers and probes to RNase P (PE Applied Biosystems) for use in normalization. Appropriate negative controls were included in each run. Standard curves were generated for T-cell markers and RNase P using MOLT4 RNA, a T-cell leukemia cell line, and for the *E2A-PBX1* markers and RNase P using a leukemia cell line, 697, that contains an *E2A-PBX1* fusion. The expression level of the predictive genes and RNase P were determined in each of the 24 ALL samples (Figures 7-11). A ratio was then calculated by taking the expression value for the specific gene and dividing it by the expression level of RNase P in the sample. These ratios were then compared to the values obtained from the Affymetrix data from the same RNA sample. The raw Affymetrix data were scaled as described and then normalized using the 3'GAPDH value for each sample, yielding a normalized ratio. The Taqman and Affymetrix ratios were then log transformed and compared. Since the markers selected for Taqman analysis were predictors for either *E2A-PBX1* or T-ALLs, each gene was expected to have four RNA samples with high and 20 samples with low expression. For each gene evaluated, an average expression value for both the Taqman and Affymetrix data was calculated for all samples in the up-regulated group, and similarly, for the samples in the down-regulated group. These values were then compared against each other as shown in Figure 12.

## Analysis of *CD3δ*

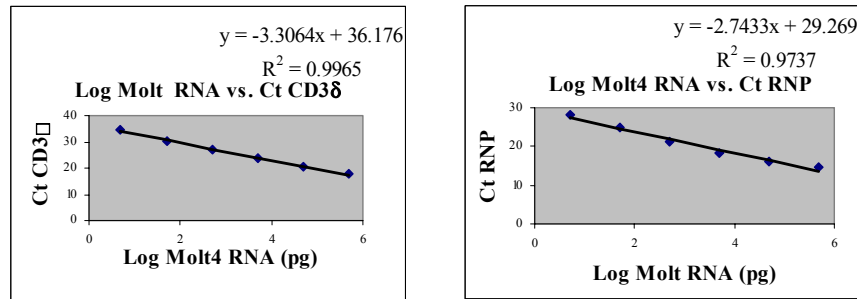


Figure 7. The standard curves of critical threshold values (Ct) versus log of the concentration of RNA in picograms (pg).

**Table 4. Results of Taqman assay for *CD3δ* and RNase P on twenty-four selected ALL diagnostic bone marrow samples representing the various leukemia subgroups.**

	Ct <i>CD3δ</i> <sup>@</sup>	Ct RNP <sup>+</sup>	Input Log <i>CD3δ</i> <sup>++</sup>	Input <i>CD3δ</i> <sup>%</sup>	Input Log RNP <sup>*</sup>	Input RNP <sup>**</sup>	Input <i>CD3δ</i> /RNP <sup>&amp;</sup>
E2A-PBX1-#1	28.26	17.05	2.394	247.825	4.454	28452.773	<b>0.0087</b>
E2A-PBX1-#3	30	16.69	1.868	73.772	4.585	38490.487	<b>0.0019</b>
E2A-PBX1-#7	28.94	17.36	2.188	154.342	4.341	21934.21	<b>0.0070</b>
E2A-PBX1-C10	27.53	16.27	2.615	412.030	4.738	54758.77	<b>0.0075</b>
T-ALL-R3	23.47	19.43	3.843	6963.863	3.587	3859.725	<b><u>1.8042</u></b>
T-ALL-C20	21.76	17.14	4.360	22910.208	4.421	26382.588	<b><u>0.8684</u></b>
T-ALL-R1	24.79	20.06	3.444	2777.311	3.357	2274.605	<b><u>1.2210</u></b>
T-ALL-#3	22.39	18.24	4.170	14773.707	4.020	10479.498	<b><u>1.4098</u></b>
BCR-ABL-R3	29.68	20.38	1.965	92.188	3.240	1738.834	<b>0.0530</b>
BCR-ABL-#7	25.8	17.51	3.138	1374.537	4.286	19339.415	<b>0.0711</b>
Hyperdip>50-C36	27.65	17.44	2.579	378.997	4.312	20509.735	<b>0.0185</b>
Hyperdip>50-#5	30.04	17.36	1.856	71.7455	4.341	21934.21	<b>0.0033</b>
Hyperdip47-50-C17	30.18	20.08	1.813	65.081	3.350	2236.74	<b>0.0291</b>
Hyperdip47-50-C8	30.49	17.95	1.7197	52.444	4.126	13367.573	<b>0.0039</b>
Hypodip-#1	28.75	18.09	2.246	176.176	4.075	11885.546	<b>0.0148</b>
Hypodip-C4	29.91	18.28	1.895	78.544	4.006	10133.5	<b>0.0078</b>
MLL-C2	28.53	17.26	2.312	205.345	4.378	23854.727	<b>0.0086</b>
MLL-2M#2	29.88	17.11	1.904	80.202	4.432	27055.348	<b>0.0030</b>
Normal-C5	27.34	17.58	2.672	470.319	4.261	18235.876	<b>0.0258</b>
Normal-C6	31.51	18.79	1.411	25.77523	3.820	6604.684	<b>0.0040</b>
Pseudodip-#4	28.49	18.73	2.325	211.146	3.842	6945.82	<b>0.0304</b>
Pseudodip-C16-N	28.21	18.18	2.409	256.606	4.042	11020.77	<b>0.0233</b>
TEL-AML1-#7	29.05	18.08	2.155	142.960	4.079	11985.727	<b>0.0119</b>
TEL-AML1-C39	30.89	19.6	1.599	39.693	3.525	3346.472	<b>0.0119</b>

<sup>@</sup>Critical threshold value for *CD3δ*

<sup>+</sup> Critical threshold value for RNase P

<sup>++</sup> Log of the pg *CD3δ* in patient sample calculated from standard curve

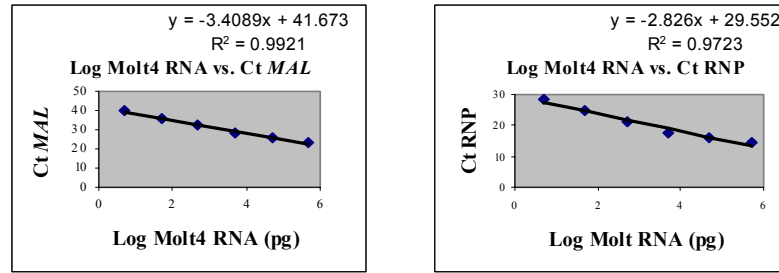
<sup>%</sup> Exponentiated value of log *CD3δ*

<sup>\*</sup> Log of the pg RNase P in patient sample calculated from standard curve

<sup>\*\*</sup> Exponentiated value of log RNase P

<sup>&</sup> Ratio of *CD3δ* to RNase P

## Analysis of *MAL*



**Figure 8.** The standard curves of critical threshold values (Ct) versus log of the concentration of RNA in picograms (pg).

**Table 5. Results of Taqman assay for *MAL* and RNase P on twenty-four selected ALL diagnostic bone marrow samples representing the various leukemia subgroups.**

	Ct <i>MAL</i> @	Ct RNP <sup>+</sup>	Input Log <i>MAL</i> <sup>++</sup>	Input <i>MAL</i> %	Input Log RNP <sup>+</sup>	Input RNP <sup>**</sup>	Input <i>MAL</i> /RNP <sup>&amp;</sup>
E2A-PBX1-#1	30.11	16.69	3.392	2466.058	4.551	35588.466	<b>0.0693</b>
E2A-PBX1-#3	32.63	16.28	2.653	449.533	4.696	49703.922	<b>0.0090</b>
E2A-PBX1-#7	32.01	16.46	2.835	683.343	4.633	42923.625	<b>0.0159</b>
E2A-PBX1-C10	32	18.69	2.838	687.974	3.844	6975.819	<b>0.0986</b>
T-ALL-R3	31.16	19.91	3.084	1213.350	3.412	2581.604	<b>0.4700</b>
T-ALL-C20	27.23	17.39	4.237	17252.443	4.304	20119.137	<b>0.8575</b>
T-ALL-R1	31.11	19.42	3.099	1255.028	3.585	3848.394	<b>0.3261</b>
T-ALL-#3	26.3	17.5	4.510	32334.48	4.265	18394.376	<b>1.7578</b>
BCR-ABL-R3	35.47	22.1	1.820	66.0160	2.637	433.454	<b>0.1523</b>
BCR-ABL-#7	29.43	16.29	3.591	3903.742	4.693	49300.587	<b>0.0792</b>
Hyperdip>50-C36	31.31	16.54	3.040	1096.437	4.604	40214.977	<b>0.0273</b>
Hyperdip>50-#5	32.3	16.73	2.750	561.781	4.537	34447.284	<b>0.0163</b>
Hyperdip47-50-C17	33.59	18.56	2.371	235.042	3.890	7755.264	<b>0.0303</b>
Hyperdip47-50-C8	31.87	17.27	2.876	751.117	4.346	22185.652	<b>0.0339</b>
Hypodip-#1	31.5	16.93	2.984	964.378	4.466	29267.362	<b>0.0330</b>
Hypodip-C4	32.47	16.19	2.700	500.839	4.728	53485.713	<b>0.0094</b>
MLL-C2	31.43	16.58	3.005	1011.071	4.590	38925.441	<b>0.0260</b>
MLL-2M#2	33.14	16.22	2.503	318.532	4.718	52194.18	<b>0.0061</b>
Normal-C5	30.09	16.68	3.398	2499.599	4.555	35879.621	<b>0.0697</b>
Normal-C6	35.59	17.94	1.784	60.876	4.109	12852.511	<b>0.0047</b>
Pseudodip-#4	32.11	16.99	2.805	638.710	4.445	27870.975	<b>0.0229</b>
Pseudodip-C16-N	35.14	17.51	1.916	82.500	4.2611	18245.11	<b>0.0045</b>
TEL-AML1-#7	31.78	17.22	2.902	798.195	4.3638	23108.143	<b>0.0345</b>
TEL-AML1-C39	33.1	17.21	2.515	327.256	4.3673	23297.194	<b>0.0140</b>

@Critical threshold value for *MAL*

+ Critical threshold value for RNase P

++ Log of the pg *MAL* in patient sample calculated from standard curve

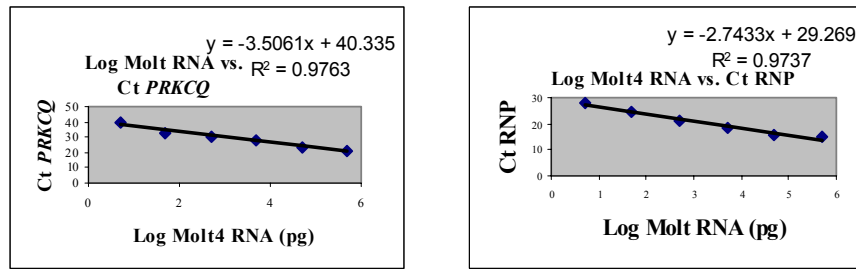
% Exponentiated value of log *MAL*

\* Log of the pg RNase P in patient sample calculated from standard curve

\*\* Exponentiated value of log RNase P

& Ratio of *MAL* to RNase P

## Analysis of *PRKCQ*



**Figure 9.** The standard curves of critical threshold values (Ct) versus log of the concentration of RNA in picograms (pg).

**Table 6. Results of Taqman assay for *PRKCQ* and RNase P on twenty-four selected ALL diagnostic bone marrow samples representing the various leukemia subgroups.**

	Ct <i>PRKCQ</i> @	Ct RNP <sup>+</sup>	Input Log <i>PRKCQ</i> <sup>++</sup>	Input <i>PRKCQ</i> %	Input Log RNP <sup>+</sup>	Input RNP <sup>**</sup>	Input <i>PRKCQ</i> /RNP <sup>&amp;</sup>
E2A-PBX1-#1	27.43	17.05	3.681	4794.330	4.454	28452.773	<b>0.1685</b>
E2A-PBX1-#3	28.18	16.69	3.467	2929.645	4.585	38490.487	<b>0.0761</b>
E2A-PBX1-#7	28.5	17.36	3.376	2374.354	4.341	21934.21	<b>0.1082</b>
E2A-PBX1-C10	28.73	16.27	3.310	2041.482	4.738	54758.77	<b>0.0373</b>
T-ALL-R3	26.54	19.43	3.935	8601.440	3.587	3859.725	<b>2.2285</b>
T-ALL-C20	26.3	17.14	4.003	10069.857	4.421	26382.588	<b>0.3817</b>
T-ALL-R1	27.21	20.06	3.743	5539.565	3.357	2274.605	<b>2.4354</b>
T-ALL-#3	25.98	18.24	4.094	12424.897	4.020	10479.498	<b>1.1856</b>
BCR-ABL-R3	32.34	20.38	2.280	190.683	3.240	1738.834	<b>0.1097</b>
BCR-ABL-#7	27.59	17.51	3.635	4316.116	4.286	19339.415	<b>0.2232</b>
Hyperdip>50-C36	28.17	17.44	3.470	2948.948	4.312	20509.735	<b>0.1438</b>
Hyperdip>50-#5	28.57	17.36	3.356	2267.672	4.341	21934.21	<b>0.1033</b>
Hyperdip47-50-C17	29.84	20.08	2.993	984.815	3.350	2236.740	<b>0.4403</b>
Hyperdip47-50-C8	28.67	17.95	3.327	2123.531	4.126	13367.573	<b>0.1589</b>
Hypodip-#1	30.12	18.09	2.913	819.396	4.075	11885.546	<b>0.0689</b>
Hypodip-C4	28.58	18.28	3.353	2252.828	4.006	10133.5	<b>0.2223</b>
MLL-C2	27.68	17.26	3.609	4068.399	4.378	23854.727	<b>0.1705</b>
MLL-2M#2	27.92	17.11	3.541	3475.133	4.432	27055.348	<b>0.1284</b>
Normal-C5	28.04	17.58	3.507	3211.777	4.261	18235.876	<b>0.1761</b>
Normal-C6	31.68	18.79	2.469	294.1405	3.820	6604.683	<b>0.0445</b>
Pseudodip-#4	30.01	18.73	2.945	880.780	3.842	6945.819	<b>0.1268</b>
Pseudodip-C16-N	30.02	18.18	2.942	875.015	4.042	11020.77	<b>0.0794</b>
TEL-AML1-#7	29.18	18.08	3.182	1519.14	4.079	11985.727	<b>0.1267</b>
TEL-AML1-C39	35.32	19.6	1.430	26.938	3.525	3346.472	<b>0.0080</b>

@Critical threshold value for *PRKCQ*

<sup>+</sup> Critical threshold value for RNase P

<sup>++</sup> Log of the pg *PRKCQ* in patient sample calculated from standard curve

% Exponentiated value of log *PRKCQ*

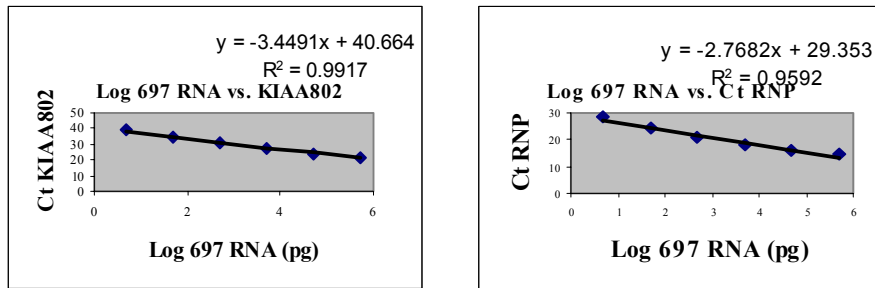
\* Log of the pg RNase P in patient sample calculated from standard curve

\*\* Exponentiated value of log RNase P

& Ratio of *PRKCQ* to RNase P



## Analysis of KIAA802



**Figure 10.** The standard curves of critical threshold values (Ct) versus log of the concentration of RNA in picograms (pg).

**Table 7. Results of Taqman assay for KIAA802 and RNase P on twenty-four selected ALL diagnostic bone marrow samples representing the various leukemia subgroups.**

	Ct KIAA802 <sup>@</sup>	Ct RNP <sup>+</sup>	Input Log KIAA802 <sup>++</sup>	Input KIAA802 <sup>%</sup>	Input Log RNP <sup>*</sup>	Input RNP <sup>**</sup>	Input KIAA802/RNP <sup>&amp;</sup>
E2A-PBX1-#1	24.4	16.999	4.715	51931.756	4.463	29028.716	<b><u>1.7890</u></b>
E2A-PBX1-#3	24.55	16.168	4.672	46983.288	4.763	57945.925	<b><u>0.8108</u></b>
E2A-PBX1-#7	26.05	16.55	4.237	17260.26	4.625	42172.281	<b><u>0.4093</u></b>
E2A-PBX1-C10	26.05	16.985	4.237	17260.26	4.468	29368.737	<b><u>0.5877</u></b>
T-ALL-R3	33.72	19.879	2.013	103.10479	3.422	2645.091	<b><u>0.0390</u></b>
T-ALL-C20	31.81	16.86	2.567	369.017	4.513	32586.744	<b><u>0.0113</u></b>
T-ALL-R1	37.045	19.078	1.049	11.201	3.712	5149.894	<b><u>0.0022</u></b>
T-ALL-#3	33.64	17.55	2.036	108.761	4.264	18356.147	<b><u>0.0059</u></b>
BCR-ABL-R3	34.815	20.79	1.696	49.637	3.093	1239.784	<b><u>0.0400</u></b>
BCR-ABL-#7	30.21	17.04	3.031	1073.830	4.448	28055.419	<b><u>0.0383</u></b>
Hyperdip>50-C36	30.68	17.04	2.895	784.636	4.448	28055.419	<b><u>0.0280</u></b>
Hyperdip>50-#5	31.135	16.52	2.763	579.095	4.636	43237.887	<b><u>0.0134</u></b>
Hyperdip47-50-C17	31.06	19.19	2.784	608.828	3.672	4691.792	<b><u>0.1298</u></b>
Hyperdip47-50-C8	31.56	17.2	2.640	436.043	4.390	24559.384	<b><u>0.0178</u></b>
Hypodip-#1	31.07	17.02	2.782	604.777	4.455	28526.052	<b><u>0.0212</u></b>
Hypodip-C4	33.397	17.157	2.107	127.916	4.406	25453.706	<b><u>0.0050</u></b>
MLL-C2	30.03	16.74	3.083	1210.941	4.556	36007.29	<b><u>0.0336</u></b>
MLL-2M#2	34.23	16.167	1.865	73.352	4.763	57994.145	<b><u>0.0013</u></b>
Normal-C5	36.1	16.41	1.323	21.050	4.676	47380.7	<b><u>0.0004</u></b>
Normal-C6	33.03	17.24	2.213	163.430	4.376	23755.69	<b><u>0.0069</u></b>
Pseudodip-#4	29.2	17.01	3.324	2107.491	4.459	28764.321	<b><u>0.0733</u></b>
Pseudodip-C16-N	28.51	16.8	3.524	3340.549	4.535	34254.348	<b><u>0.0975</u></b>
TEL-AML1-#7	29.83	16.986	3.141	1383.915	4.468	29344.318	<b><u>0.0472</u></b>
TEL-AML1-C39	30.165	18.187	3.044	1106.579	4.034	10806.078	<b><u>0.1024</u></b>

<sup>@</sup>Critical threshold value for KIAA802

<sup>+</sup>Critical threshold value for RNase P

<sup>++</sup> Log of the pg KIAA802 in patient sample calculated from standard curve

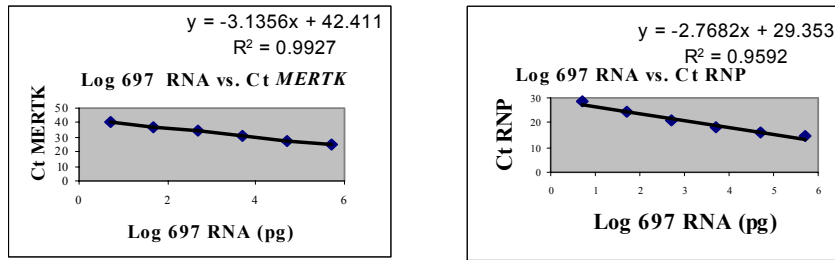
<sup>%</sup> Exponentiated value of log KIAA802

<sup>\*</sup> Log of the pg RNase P in patient sample calculated from standard curve

<sup>\*\*</sup> Exponentiated value of log RNase P

<sup>&</sup> Ratio of KIAA802 to RNase P

## Analysis of *MERTK*



**Figure 11.** The standard curves of critical threshold values (Ct) versus log of the concentration of RNA in picograms (pg).

**Table 8. Results of Taqman assay for *MERTK* and RNase P on twenty-four selected ALL diagnostic bone marrow samples representing the various leukemia subgroups.**

	Ct <i>MERTK</i> <sup>@</sup>	Ct RNP <sup>+</sup>	Input Log <i>MERTK</i> <sup>++</sup>	Input <i>MERTK</i> <sup>%</sup>	Input Log RNP <sup>*</sup>	Input RNP <sup>**</sup>	Input <i>MERTK</i> /RNP <sup>&amp;</sup>
E2A-PBX1-#1	23.202	16.999	6.126	1336904.1	4.463	29028.716	<b>46.0545</b>
E2A-PBX1-#3	23.816	16.168	5.930	851695.74	4.763	57945.925	<b>14.6981</b>
E2A-PBX1-#7	24.164	16.55	5.819	659630.81	4.625	42172.281	<b>15.6413</b>
E2A-PBX1-C10	23.198	16.985	6.127	1340836.9	4.468	29368.737	<b>45.6552</b>
T-ALL-R3	33.649	19.879	2.794	622.819	3.422	2645.091	<b>0.2355</b>
T-ALL-C20	30.594	16.86	3.769	5870.252	4.513	32586.744	<b>0.1801</b>
T-ALL-R1	28.06	19.078	4.577	37739.445	3.712	5149.894	<b>7.3282</b>
T-ALL-#3	31.38	17.55	3.518	3295.998	4.264	18356.147	<b>0.1800</b>
BCR-ABL-R3	34.5	20.79	2.523	333.397	3.093	1239.784	<b>0.2689</b>
BCR-ABL-#7	28.06	17.04	4.577	37739.445	4.448	28055.419	<b>1.3452</b>
Hyperdip>50-C36	28.46	17.04	4.449	28133.788	4.448	28055.419	<b>1.0028</b>
Hyperdip>50-#5	28.43	16.52	4.459	28760.456	4.636	43237.887	<b>0.6652</b>
Hyperdip47-50-C17	29.767	19.19	4.032	10774.623	3.671	4691.792	<b>2.2965</b>
Hyperdip47-50-C8	33.36	17.2	2.887	770.068	4.390	24559.384	<b>0.0314</b>
Hypodip-#1	30.67	17.02	3.744	5551.610	4.455	28526.052	<b>0.1946</b>
Hypodip-C4	28.88	17.157	4.315	20667.244	4.406	25453.706	<b>0.8120</b>
MLL-C2	30.415	16.74	3.826	6694.884	4.556	36007.29	<b>0.1859</b>
MLL-2M#2	33.44	16.167	2.861	726.132	4.763	57994.145	<b>0.0125</b>
Normal-C5	32.13	16.41	3.279	1900.196	4.676	47380.7	<b>0.0401</b>
Normal-C6	33.376	17.24	2.881	761.073	4.376	23755.69	<b>0.0320</b>
Pseudodip-#4	31.98	17.01	3.327	2121.466	4.459	28764.321	<b>0.0738</b>
Pseudodip-C16-N	28.497	16.8	4.437	27379.672	4.535	34254.348	<b>0.7993</b>
TEL-AML1-#7	34.15	16.986	2.635	431.105	4.468	29344.318	<b>0.0147</b>
TEL-AML1-C39	34.3	18.187	2.587	386.141	4.034	10806.078	<b>0.0357</b>

<sup>@</sup>Critical threshold value for *MERTK*

<sup>+</sup> Critical threshold value for RNase P

<sup>++</sup> Log of the pg *MERTK* in patient sample calculated from standard curve

<sup>%</sup> Exponentiated value of log *MERTK*

<sup>\*</sup> Log of the pg RNase P in patient sample calculated from standard curve

<sup>\*\*</sup> Exponentiated value of log RNase P

<sup>&</sup> Ratio of *MERTK* to RNase P

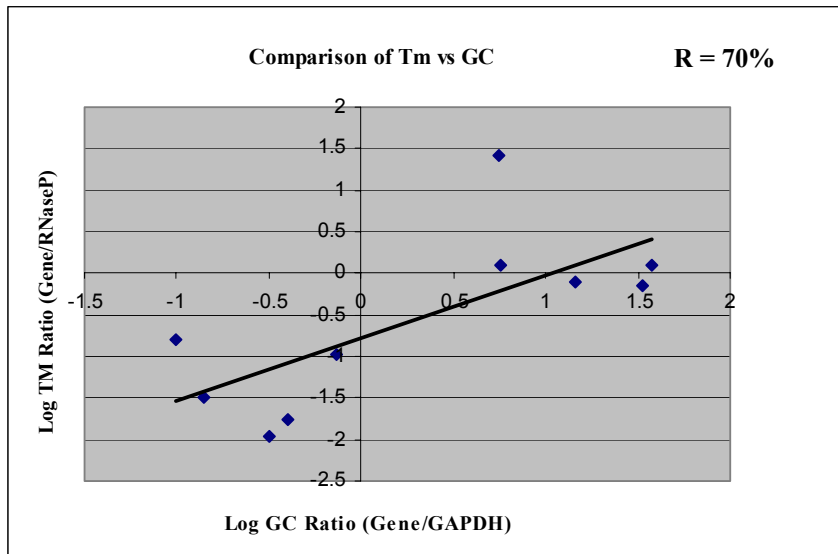
### Comparison of real-time RT-PCR and Affymetrix data

The normalized gene expression ratios for the TaqMan data (TM) (gene/RNase P) and for the Affymetrix microarray data (GC) (AID for a gene/AID for GAPDH) were log transformed and then the average expression values for each gene was calculated in the four samples in which its expression was expected to be up-regulated and separately in the 20 samples in which its expression was expected to be down-regulated. For example, for genes that were expected to be up-regulated in T-ALL (*CD3 $\delta$* , *MAL*, and *PRKCQ*), the log expression ratios in the T-ALL samples were averaged to give the up regulated (Up Reg) values and the log expression ratios of each gene in the non-T-ALL cases were averaged to give the down regulated value (Down Reg).

**Table 9. Comparison of real-time RT-PCR and Affymetrix data**

		Log ave GC	Log ave TM
<i>CD3<math>\delta</math></i>	Up Reg (4)	1.578	0.108
	Down Reg (20)	-0.502	-1.958
<i>MAL</i>	Up Reg (4)	1.524	-0.159
	Down Reg (20)	-0.855	-1.502
<i>PRKCQ</i>	Up Reg (4)	0.7521	0.098
	Down Reg (20)	-0.136	-0.972
KIAA802	Up Reg (4)	1.16	-0.114
	Down Reg (20)	-0.393	-1.772
<i>MERTK</i> (1)	Up Reg (4)	0.732	1.42
	Down Reg (20)	-0.934	-0.794
<i>MERTK</i> (2)	Up Reg (4)	0.762	1.421
	Down Reg (20)	-1.068	-0.792
<i>MERTK</i> (ave)	Up Reg (4)	0.747	1.4205
	Down Reg (20)	-1.001	-0.793

In the case of *MERTK*, there were two probe sets on the Affymetrix chip. The normalized log values for each probe sets is shown in Table 9, as well as the average. The average value was used in Figure 12.



**Figure 12.** The log ratio averages from the up regulated and down regulated genes from the Taqman data plotted against the up regulated and down regulated genes from the microarray data (CG).

In both the Taqman and the microchip array analysis, *MERTK* and *KIAA802*, were very highly expressed in the diagnostic samples containing *E2A-PBX1*, and expressed at low levels in all of the other samples. Likewise, *PRKCQ*, *CD3 $\delta$* , and *MAL*, showed high levels of expression in T cells by both methodologies as compared to non T-cells. The normalized ratios from the Taqman assay were plotted against the normalized ratios from the microchip array for both the up-regulated and down-regulated genes. The correlation between Taqman and the microchip array was 70%, indicating that the same pattern of gene expression was seen in both analyses. The *MERTK* was extremely high in two of the *E2A-PBX1* patient samples by Taqman analysis. It would be necessary to repeat the assay several times to validate the two high figures, but there was not sufficient sample to do so. Removal of the *MERTK* gene from the analysis resulted in a correlation of 91% between the Taqman and Microchip array (data not shown).

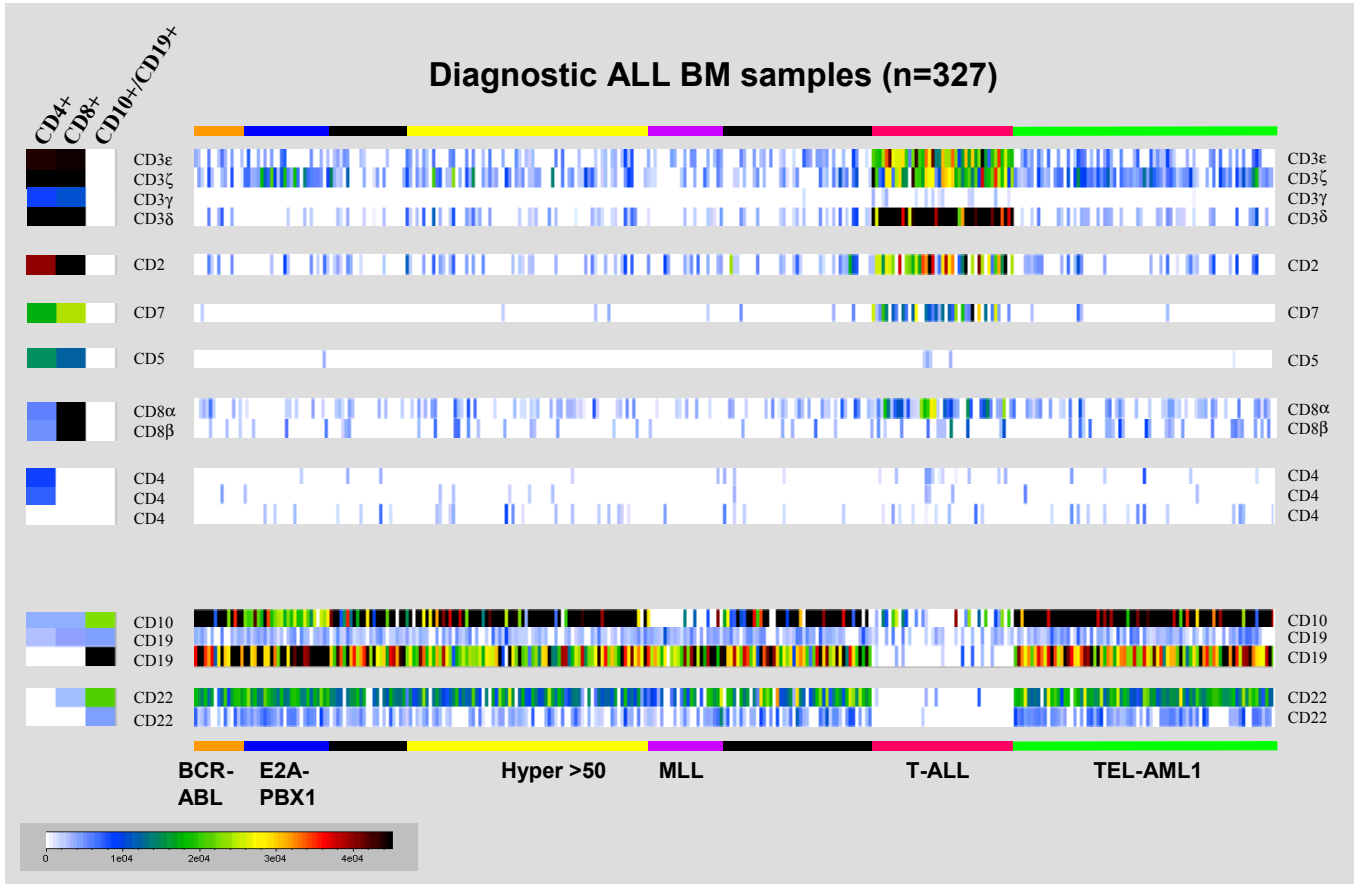
### Comparisons of Affymetrix and immunophenotype results

Leukemic blasts at the time of diagnosis were analyzed for expression of lineage restricted cell surface antigens using phycoerythrin- or fluorescein isothiocyanate-conjugated monoclonal antibodies against CD2, CD3 $\epsilon$ , CD4, CD5, CD7, CD8, CD10, CD19, and CD22 (Becton Dickinson Immunocytometry Systems, San Jose, CA, USA). Data were obtained using a Coulter Epic XL (Beckman Coulter, Miami, Fla, USA), a Coulter Elite, or a FACScalibur flow cytometer (Becton Dickinson, San Jose, CA, USA). The expression patterns for these antigens were then compared to gene expression patterns for the Affymetrix chip sites specified for *CD2* (1 probe set, 40738\_at), *CD3 $\delta$*  (1 probe set, 38319\_at), *CD3 $\epsilon$*  (1 probe set, 36277\_at), *CD3 $\zeta$*  (1 probe set, 37078\_at), *CD3 $\gamma$*  (1 probe set, 39226\_at), *CD4* (5 probe sets, 856\_at, 1146\_at, 35517\_at, 34003\_at, and 37942\_at), *CD5* (1 probe set, 32953\_at), *CD7* (1 probe set, 771\_s\_at), *CD8 $\alpha$*  (1 probe set, 40699\_at), *CD8 $\beta$*  (1 probe set, 39239\_at), *CD10* (1 probe set, 1389\_at), *CD19* (2 probe sets, 1096\_g\_at and 1116\_at), and *CD22* (2 probe sets, 38521\_at and 38522\_s\_at). As a control, the performance of the Affymetrix microarray probe sets were also assessed using RNA isolated from flow sorted single positive CD4+ and CD8+ thymocytes, and CD10+/CD19+ bone marrow cells. The results from the microarray analysis are illustrated in Figure 2 in the paper and Figure 13 below. As shown, high RNA expression was observed in T-ALL for the T-lineage restricted genes *CD2*, *CD3 $\delta$* ,  $\epsilon$ , and  $\zeta$ , *CD8 $\alpha$* , and *CD7*, and in B-lineage ALLs for the B-cell restricted genes *CD19*, and *CD22*. A similar high level of correlation was observed between RNA and protein expression for CD10. The observed low expression levels of T-cell restricted genes in B-cell cases, and B-cell restricted genes in T-ALLs, is consistent with the low level of normal contaminating lymphocytes present in the diagnostic marrow samples analyzed.

Interestingly, in T-lineage ALLs, although high expression was observed for *CD3 $\delta$* ,  $\epsilon$ , and  $\zeta$ , only trace levels of *CD3 $\gamma$*  RNA expression were detected. Verification of the absence of the expression of the encoded protein will require analysis with an antibody specific for this chain.

Although a high correlation was detected between RNA expression by microarray analysis and protein expression by flow cytometry, the absence of RNA expression by microarray analysis did not always correlate with the absence of protein. This resulted from a number of different causes. For example, the Affymetrix probe set for *CD5* was able to detect expression in normal thymocytes, but failed to detect expression in T-lineage ALLs, despite the presence of low levels of CD5 protein as assessed by flow cytometry. The most likely interpretation for this observation is that this feature performs poorly under standard hybridization conditions. Thus, high expression levels as seen in normal T-cells can be detected, whereas the lower expression observed in T-ALLs appears to fall below the detection threshold. Two of the five probe sets representing the *CD4* gene (1146\_at and 35517\_at) performed similarly, detecting mRNA in CD4+ thymocytes, but failing to identify the lower *CD4* expression levels seen in 23/43 of the T-ALL cases. By contrast, two *CD4* probe sets, 856\_at (shown) and 37942\_at, failed to detect expression in either CD4+ thymocyte or T-ALLs, whereas another probe set (34003\_at) detected a transcript that was expressed in the majority of both B- and T-ALLs. A search of the target sequence supplied by Affymetrix against NCBI and CELERA<sup>®</sup> databases provided some insights into the poor

hybridization characteristics of these probe sets. The 34003\_at probe set lacked homology to *CD4* but demonstrated homology with *triosephosphate isomerase 1*, and thus represented an unrelated gene. This



**Figure 13:** B and T cell lineage-associated gene expression in ALL Diagnostic BM . Results are displayed as quantitative values.

finding is consistent with recently updated information in the Affymetrix database. The *CD4* probe set 856\_at had 3 high probability matches, only one that was *CD4*, however, this sequence was located in an intron. Similarly, the *CD8β* target sequence was found to localize 3' of the last exon, and thus the absence of an mRNA able to hybridize to this feature must be cautiously interpreted. Lastly, the *CD19* probe set 1116\_at was found to have many high probability matches, only one that was homologous to *CD19*; however, it localized to a region 5' of exon 1. Taken together, these data suggest that absent or low expression of a gene by microarray analysis must be verified by an alternative method.

## Section II: Patient Dataset

A total of 389 Pediatric acute leukemia samples were analyzed in this study, from which high quality gene expression data were obtained on 360 (93%). The successfully analyzed samples included: 332 diagnostic bone marrows (BM), 3 diagnostic peripheral bloods (PB), and 25 relapse ALL samples from BM or PB. 264 (79%) of the diagnostic ALL BM samples and all relapse samples were from patients treated on St. Jude Children's Research Hospital Total Therapy Studies XIII A or XIII B and correspond to 64% of the patients treated on these protocols. The details of these protocols have been previously published.<sup>1,2</sup> XIII A ran from 12-20-91 to 8-23-94 and enrolled 165 patients, whereas XIII B ran from 8-24-94 to 7-27-98 and enrolled 247 patients. No patients were lost to follow-up during treatment. When the databases were frozen for analysis, 100% and 93% of event-free survivors in studies XIII A and XIII B, respectively, had been seen within 12 months. The median (minimum, maximum)

follow-up of the event-free survivors was 8.09 (6.59, 9.94) and 4.52 (2.37, 7.06) years for XIII A and XIII B, respectively. All other samples were obtained from patients treated on St. Jude Total Therapy Studies XI, XII, XIV, XV, or by best clinical management.

For the identification of gene expression profiles that predict specific genetic subtypes of ALL, 327 diagnostic BM samples were used. Inclusion in this dataset required availability of a cryopreserved diagnostic BM sample containing  $\geq 75\%$  blasts, and complete data from each of the following diagnostic studies: morphology, immunophenotype, cytogenetics, DNA ploidy, Southern blot for *MLL* gene rearrangements, and RT-PCR analysis for *MLL-AF4*, *MLL-AF9*, *E2A-PBX1*, *TEL-AML1*, and *BCR-ABL*. This final dataset includes diagnostic BM samples from XV (38), XIV (4), XIII A (100), XIII B (161), or from patients treated on one of our older protocols or by best clinical management (24).

The data sets used to identify expression profiles predictive of hematologic relapse and the development of therapy-induced AML are described below.

**Table 10. Patient database**

**Diagnostic samples used for subtype classification (n=327)**

<b><i>BCR-ABL</i> subgroup (n=15)</b>					
<b>Label<sup>@</sup></b>	<b>Protocol<sup>#</sup></b>	<b>Outcome<sup>%</sup></b>	<b>Label<sup>@</sup></b>	<b>Protocol<sup>#</sup></b>	<b>Outcome<sup>%</sup></b>
BCR-ABL-C1	T13B	CCR	BCR-ABL-#4	T11	NA
BCR-ABL-R1	T13A	Heme Relapse	BCR-ABL-#5	T12	NA
BCR-ABL-R2	T13A	Heme Relapse	BCR-ABL-#6	T12	NA
BCR-ABL-R3	T13B	Heme Relapse	BCR-ABL-#7	T12	NA
BCR-ABL-Hyperdip-R5	T13B	Heme Relapse	BCR-ABL-#8	T14	NA
BCR-ABL-#1	T13A	Censored	BCR-ABL-#9	T15	NA
BCR-ABL-#2	T13B	Censored	BCR-ABL-Hyperdip-#10	T12	NA
BCR-ABL-#3	T13B	Censored			
<b><i>E2A-PBX1</i> subgroup (n=27)</b>					
E2A-PBX1-C1	T13A	CCR	E2A-PBX1-#1	Others	NA
E2A-PBX1-C2	T13A	CCR	E2A-PBX1-#2	Others	NA
E2A-PBX1-C3	T13A	CCR	E2A-PBX1-#3	Others	NA
E2A-PBX1-C4	T13A	CCR	E2A-PBX1-#4	Others	NA
E2A-PBX1-C5	T13A	CCR	E2A-PBX1-#5	Others	NA
E2A-PBX1-C6	T13B	CCR	E2A-PBX1-#6	Others	NA
E2A-PBX1-C7	T13B	CCR	E2A-PBX1-#7	T11	NA
E2A-PBX1-C8	T13B	CCR	E2A-PBX1-#8	T11	NA
E2A-PBX1-C9	T13B	CCR	E2A-PBX1-#9	T12	NA
E2A-PBX1-C10	T13B	CCR	E2A-PBX1-#10	T12	NA
E2A-PBX1-C11	T13B	CCR	E2A-PBX1-#11	T14	NA
E2A-PBX1-C12	T13B	CCR	E2A-PBX1-#12	T15	NA
E2A-PBX1-R1	T13B	Heme Relapse	E2A-PBX1-#13	T15	NA
E2A-PBX1-2M#1	T13B	2nd AML			
<b>Hyperdip&gt;50 subgroup (n=64)</b>					
Hyperdip>50-C1	T13A	CCR	Hyperdip>50-C33	T13B	CCR
Hyperdip>50-C2	T13A	CCR	Hyperdip>50-C34	T13B	CCR
Hyperdip>50-C3	T13A	CCR	Hyperdip>50-C35	T13B	CCR
Hyperdip>50-C4	T13A	CCR	Hyperdip>50-C36	T13B	CCR
Hyperdip>50-C5	T13A	CCR	Hyperdip>50-C37	T13B	CCR
Hyperdip>50-C6	T13A	CCR	Hyperdip>50-C38	T13B	CCR
Hyperdip>50-C7	T13A	CCR	Hyperdip>50-C39	T13B	CCR
Hyperdip>50-C8	T13A	CCR	Hyperdip>50-C40	T13B	CCR
Hyperdip>50-C9	T13A	CCR	Hyperdip>50-C41	T13B	CCR
Hyperdip>50-C10	T13A	CCR	Hyperdip>50-C42	T13B	CCR

Hyperdip>50-C11	T13A	CCR	Hyperdip>50-C43	T13B	CCR
Hyperdip>50-C12	T13A	CCR	Hyperdip>50-R1	T13A	Heme Relapse
Hyperdip>50-C13	T13A	CCR	Hyperdip>50-R2	T13A	Heme Relapse
Hyperdip>50-C14	T13A	CCR	Hyperdip>50-R3	T13A	Heme Relapse
Hyperdip>50-C15	T13B	CCR	Hyperdip>50-R4	T13B	Heme Relapse
Hyperdip>50-C16	T13B	CCR	Hyperdip>50-R5	T13B	Heme Relapse
Hyperdip>50-C17	T13B	CCR	Hyperdip>50-2M#1	T13A	2nd AML
Hyperdip>50-C18	T13B	CCR	Hyperdip>50-2M#2	T13B	2nd AML
Hyperdip>50-C19	T13B	CCR	Hyperdip>50-#1	T13A	Censored
Hyperdip>50-C20	T13B	CCR	Hyperdip>50-#2	T13B	Censored
Hyperdip>50-C21	T13B	CCR	Hyperdip>50-#3	Others	NA
Hyperdip>50-C22	T13B	CCR	Hyperdip>50-#4	Others	NA
Hyperdip>50-C23	T13B	CCR	Hyperdip>50-#5	T12	NA
Hyperdip>50-C24	T13B	CCR	Hyperdip>50-#6	T15	NA
Hyperdip>50-C25	T13B	CCR	Hyperdip>50-#7	T15	NA
Hyperdip>50-C26	T13B	CCR	Hyperdip>50-#8	T15	NA
Hyperdip>50-C27-N	T13B	CCR	Hyperdip>50-#9	T15	NA
Hyperdip>50-C28	T13B	CCR	Hyperdip>50-#10	T15	NA
Hyperdip>50-C29	T13B	CCR	Hyperdip>50-#11	T15	NA
Hyperdip>50-C30	T13B	CCR	Hyperdip>50-#12	T15	NA
Hyperdip>50-C31	T13B	CCR	Hyperdip>50-#13	T15	NA
Hyperdip>50-C32	T13B	CCR	Hyperdip>50-#14	T15	NA

**Hyperdip47-50 subgroup (n=23)**

Hyperdip47-50-C1	T13A	CCR	Hyperdip47-50-C13	T13B	CCR
Hyperdip47-50-C2	T13A	CCR	Hyperdip47-50-C14-N	T13B	CCR
Hyperdip47-50-C3-N	T13A	CCR	Hyperdip47-50-C15	T13B	CCR
Hyperdip47-50-C4	T13A	CCR	Hyperdip47-50-C16	T13B	CCR
Hyperdip47-50-C5	T13A	CCR	Hyperdip47-50-C17	T13B	CCR
Hyperdip47-50-C6	T13B	CCR	Hyperdip47-50-C18	T13B	CCR
Hyperdip47-50-C7	T13B	CCR	Hyperdip47-50-C19	T13B	CCR
Hyperdip47-50-C8	T13B	CCR	Hyperdip47-50-2M#1	T13A	2nd AML
Hyperdip47-50-C9	T13B	CCR	Hyperdip47-50-#1	T15	NA
Hyperdip47-50-C10	T13B	CCR	Hyperdip47-50-#2	T15	NA
Hyperdip47-50-C11	T13B	CCR	Hyperdip47-50-#3	T15	NA
Hyperdip47-50-C12	T13B	CCR			

**Hypodip subgroup (n=9)**

Hypodip-C1	T13A	CCR	Hypodip-C6	T13B	CCR
Hypodip-C2	T13A	CCR	Hypodip-2M#1	T13A	2nd AML
Hypodip-C3	T13B	CCR	Hypodip-#1	T15	NA
Hypodip-C4	T13B	CCR	Hypodip-#2	T15	NA
Hypodip-C5	T13B	CCR			

**MLL subgroup (n=20)**

MLL-C1	T13A	CCR	MLL-2M#1	T13A	2nd AML
MLL-C2	T13B	CCR	MLL-2M#2	T13A	2nd AML
MLL-C3	T13B	CCR	MLL-#1	T13B	Censored
MLL-C4	T13B	CCR	MLL-#2	T13B	Censored
MLL-C5	T13B	CCR	MLL-#3	Others	NA
MLL-C6	T13B	CCR	MLL-#4	Others	NA
MLL-R1	T13A	Heme Relapse	MLL-#5	Others	NA
MLL-R2	T13A	Heme Relapse	MLL-#6	T12	NA
MLL-R3	T13B	Heme Relapse	MLL-#7	T14	NA
MLL-R4	T13B	Heme Relapse	MLL-#8	T14	NA

**Normal subgroup (n=18)**

Normal-C1-N	T13A	CCR	Normal-C10	T13B	CCR
Normal-C2-N	T13A	CCR	Normal-C11-N	T13B	CCR
Normal-C3-N	T13A	CCR	Normal-C12	T13B	CCR
Normal-C4-N	T13B	CCR	Normal-R1	T13A	Heme Relapse
Normal-C5	T13B	CCR	Normal-R2-N	T13B	Heme Relapse
Normal-C6	T13B	CCR	Normal-R3	T13B	Heme Relapse
Normal-C7-N	T13B	CCR	Normal-#1	T13A	Censored
Normal-C8	T13B	CCR	Normal-#2	T13B	Censored
Normal-C9	T13B	CCR	Normal-#3	T13B	Censored

**Pseudodip subgroup (n=29)**

Pseudodip-C1	T13A	CCR	Pseudodip-C16-N	T13B	CCR
Pseudodip-C2-N	T13A	CCR	Pseudodip-C17	T13B	CCR
Pseudodip-C3	T13A	CCR	Pseudodip-C18	T13B	CCR
Pseudodip-C4	T13A	CCR	Pseudodip-C19	T13B	CCR
Pseudodip-C5	T13A	CCR	Pseudodip-R1-N	T13A	Heme Relapse
Pseudodip-C6	T13A	CCR	Pseudodip-#1	T13B	Other Relapse
Pseudodip-C7	T13A	CCR	Pseudodip-#2	T13B	Censored
Pseudodip-C8	T13A	CCR	Pseudodip-#3	Others	NA
Pseudodip-C9	T13A	CCR	Pseudodip-#4	Others	NA
Pseudodip-C10	T13B	CCR	Pseudodip-#5	T15	NA
Pseudodip-C11	T13B	CCR	Pseudodip-#6	T15	NA
Pseudodip-C12	T13B	CCR	Pseudodip-#7	T15	NA
Pseudodip-C13	T13B	CCR	Pseudodip-#8-N	T15	NA
Pseudodip-C14	T13B	CCR	Pseudodip-#9	T15	NA
Pseudodip-C15	T13B	CCR			

**T-ALL subgroup (n=43)**

T-ALL-C1	T13A	CCR	T-ALL-C23	T13B	CCR
T-ALL-C2	T13A	CCR	T-ALL-C24	T13B	CCR
T-ALL-C3	T13A	CCR	T-ALL-C25	T13B	CCR
T-ALL-C4	T13A	CCR	T-ALL-C26	T13B	CCR
T-ALL-C5	T13A	CCR	T-ALL-R1	T13A	Heme Relapse
T-ALL-C6	T13A	CCR	T-ALL-R2	T13B	Heme Relapse
T-ALL-C7	T13A	CCR	T-ALL-R3	T13B	Heme Relapse
T-ALL-C8	T13A	CCR	T-ALL-R4	T13B	Heme Relapse
T-ALL-C9	T13B	CCR	T-ALL-R5	T13B	Heme Relapse
T-ALL-C10	T13B	CCR	T-ALL-R6	T13B	Heme Relapse
T-ALL-C11	T13B	CCR	T-ALL-2M#1	T13B	2nd AML
T-ALL-C12	T13B	CCR	T-ALL-#1	T13B	Other Relapse
T-ALL-C13	T13B	CCR	T-ALL-#2	T13B	Other Relapse
T-ALL-C14	T13B	CCR	T-ALL-#3	T13B	Censored
T-ALL-C15	T13B	CCR	T-ALL-#4	T13B	Censored
T-ALL-C16	T13B	CCR	T-ALL-#5	T15	NA
T-ALL-C17	T13B	CCR	T-ALL-#6	T15	NA
T-ALL-C18	T13B	CCR	T-ALL-#7	T15	NA
T-ALL-C19	T13B	CCR	T-ALL-#8	T15	NA
T-ALL-C20	T13B	CCR	T-ALL-#9	T15	NA
T-ALL-C21	T13B	CCR	T-ALL-#10	T15	NA
T-ALL-C22	T13B	CCR			

**TEL-AML1 subgroup (n=79)**

TEL-AML1-C1	T13A	CCR	TEL-AML1-C41	T13B	CCR
-------------	------	-----	--------------	------	-----



TEL-AML1-C2	T13A	CCR	TEL-AML1-C42	T13B	CCR
TEL-AML1-C3	T13A	CCR	TEL-AML1-C43	T13B	CCR
TEL-AML1-C4	T13A	CCR	TEL-AML1-C44	T13B	CCR
TEL-AML1-C5	T13A	CCR	TEL-AML1-C45	T13B	CCR
TEL-AML1-C6	T13A	CCR	TEL-AML1-C46	T13B	CCR
TEL-AML1-C7	T13A	CCR	TEL-AML1-C47	T13B	CCR
TEL-AML1-C8	T13A	CCR	TEL-AML1-C48	T13B	CCR
TEL-AML1-C9	T13A	CCR	TEL-AML1-C49	T13B	CCR
TEL-AML1-C10	T13A	CCR	TEL-AML1-C50	T13B	CCR
TEL-AML1-C11	T13A	CCR	TEL-AML1-C51	T13B	CCR
TEL-AML1-C12	T13A	CCR	TEL-AML1-C52	T13B	CCR
TEL-AML1-C13	T13A	CCR	TEL-AML1-C53	T13B	CCR
TEL-AML1-C14	T13A	CCR	TEL-AML1-C54	T13B	CCR
TEL-AML1-C15	T13A	CCR	TEL-AML1-C55	T13B	CCR
TEL-AML1-C16	T13A	CCR	TEL-AML1-C56	T13B	CCR
TEL-AML1-C17	T13A	CCR	TEL-AML1-C57	T13B	CCR
TEL-AML1-C18	T13A	CCR	TEL-AML1-R1	T13A	Heme Relapse
TEL-AML1-C19	T13A	CCR	TEL-AML1-R2	T13A	Heme Relapse
TEL-AML1-C20	T13A	CCR	TEL-AML1-R3	T13B	Heme Relapse
TEL-AML1-C21	T13A	CCR	TEL-AML1-2M#1	T13A	2nd AML
TEL-AML1-C22	T13A	CCR	TEL-AML1-2M#2	T13A	2nd AML
TEL-AML1-C23	T13A	CCR	TEL-AML1-2M#3	T13A	2nd AML
TEL-AML1-C24	T13A	CCR	TEL-AML1-2M#4	T13B	2nd AML
TEL-AML1-C25	T13A	CCR	TEL-AML1-2M#5	T13B	2nd AML
TEL-AML1-C26	T13A	CCR	TEL-AML1-#1	T13B	Other Relapse
TEL-AML1-C27	T13A	CCR	TEL-AML1-#2	T13A	Censored
TEL-AML1-C28	T13A	CCR	TEL-AML1-#3	T13A	Censored
TEL-AML1-C29	T13B	CCR	TEL-AML1-#4	T13B	Censored
TEL-AML1-C30	T13B	CCR	TEL-AML1-#5	T15	NA
TEL-AML1-C31	T13B	CCR	TEL-AML1-#6	T15	NA
TEL-AML1-C32	T13B	CCR	TEL-AML1-#7	T15	NA
TEL-AML1-C33	T13B	CCR	TEL-AML1-#8	T15	NA
TEL-AML1-C34	T13B	CCR	TEL-AML1-#9	T15	NA
TEL-AML1-C35	T13B	CCR	TEL-AML1-#10	T15	NA
TEL-AML1-C36	T13B	CCR	TEL-AML1-#11	T15	NA
TEL-AML1-C37	T13B	CCR	TEL-AML1-#12	T15	NA
TEL-AML1-C38	T13B	CCR	TEL-AML1-#13	T15	NA
TEL-AML1-C39	T13B	CCR	TEL-AML1-#14	T15	NA
TEL-AML1-C40	T13B	CCR			

**@Label key-**

Subtype Name-C#	Dx Sample of patient in CCR
Subtype Name-R#	Dx Sample of patient who developed a hematologic relapse
Subtype Name-#	Dx Sample used for subgroup classification only
Subtype Name-2M#	Dx Sample of patient who later developed 2 <sup>nd</sup> AML
Subtype Name-N	Dx Sample in novel group

**#Protocol-** Protocol that patient was treated on

**%Outcome-**

CCR	Continuous complete remission
Heme Relapse	Hematologic relapse
Other Relapse	Extramedullary relapse
2nd AML	Diagnostic samples of patients who later developed 2 <sup>nd</sup> AML
Censored	Censored due to BM transplant, treated off protocol, or died in CR
NA	Not applicable, primarily because the patient was not treated on Total 13, and thus is excluded from the analysis used to identify gene expression profiles predictive of outcome

### **Diagnostic samples used for prediction of prognosis**

In addition to the 201 CCR and 27 Heme Relapse cases listed in Table 10, five additional relapse cases were also included in the prognostic analysis. This gives a total of 233 cases for this analysis. These additional cases were not included in the subgroup prediction dataset because they did not meet the established criteria for the reasons listed below.

<b>Label</b>	<b>Protocol</b>	<b>Comment</b>
BCR-ABL-R4	T13B	Did not meet QC criteria because contained 70% blasts
MLL-R5	T13A	Peripheral Blood Sample (90% blasts)
Normal-R4	T13B	Molecular studies not performed
T-ALL-R7	T13A	Peripheral Blood Sample (90% blasts)
T-ALL-R8	T13B	Peripheral Blood Sample (90% blasts)

### **Diagnostic Samples used for prediction of 2<sup>nd</sup> AML**

In addition to the 201 CCR and 13 secondary AML cases listed in Table 10, three additional diagnostic marrow samples from patients who developed secondary AML were also included in the prognostic analysis. This gives a total of 217 cases used for this analysis. These additional cases were not included in the diagnostic data set because they did not meet the established criteria for the reasons listed below.

<b>Label</b>	<b>Protocol</b>	<b>Comment</b>
Hyperdip>50-2M#3	T12	Non Total 13 diagnostic sample
Hypodip-2M#2	T13B	No molecular studies performed
Hypodip-2M#3	T12	Non Total 13 diagnostic sample

### **Relapsed Samples (n=25)**

Twenty-five relapse samples were analyzed, 17 samples which were paired to the diagnostic samples listed in Table 10 and 8 additional non-paired relapse samples.

## **Section III: Detailed Analysis**

### **Hierarchical cluster analysis of diagnostic cases using all genes that passed the variation filter**

Two-dimensional hierarchical clustering was performed using Pearson correlation coefficient and an unweighted pair group method using arithmetic averages (GeneMaths, version 1.5). The results of hierarchical clustering of the 327 diagnostic samples using genes selected by a variety of metrics are shown below.

### **Methods for gene selection**

Discriminating genes for the various leukemia subtypes were selected using a variety of statistical metrics. The individual metrics used and the list of selected probe sets and corresponding genes are given below. Genes selected by Chi-square, T-statistics, Wilkins' and CFS, were chosen using the decision tree format shown in Figure 19 below. In this process, genes were selected that distinguished the class for all classes listed below it in the decision tree structure. For the selection of genes using SOM/DAV genes were selected that distinguished the class from all others. The degree of overlap between the lists of genes selected for each genetic subtype by the various metrics is discussed below.

### **Chi-square**

The Chi-square method evaluates each gene individually by measuring the Chi-square statistics with respect to the classes. The method first discretizes the observed expression values of the gene into several intervals using an entropy-based discretization method<sup>3</sup>. The Chi-square statistics of a gene is then

calculated as  $X^2 = \sum \sum (A_{ij} - E_{ij})^2 / E_{ij}$ , summing over intervals  $i = 1..m$  and classes  $j = 1..k$ .  $A_{ij}$  is the number of samples in the  $i^{\text{th}}$  interval that are of the  $j^{\text{th}}$  class.  $E_{ij}$  is the expected frequency of  $A_{ij}$  and is calculated as  $E_{ij} = R_i * C_j / N$ , where  $R_i$  is the number of samples in the  $i^{\text{th}}$  interval,  $C_j$  is the number of samples in the  $j^{\text{th}}$  class, and  $N$  is the total number of samples. The genes are then sorted according to their Chi-square statistics---the larger the Chi-square statistics, the more important the gene. The 40 genes with the highest Chi-square statistics in each subtype are listed in Table 11. Generally, using anywhere from the top 20 to 40 genes did not result in significant differences in subtype prediction accuracy. Therefore, we used only the top 20 genes in subtype prediction, unless noted otherwise.

**Table 11. Genes selected by Chi-square**

	<b>Affymetrix number</b>	<b>Gene Name</b>	<b>GeneSymbol</b>	<b>Reference number</b>	<b>Chi- square value</b>	<b>Above/ Below Mean</b>
<b><i>BCR-ABL</i></b>						
1	1637_at	mitogen-activated protein kinase- activated protein kinase 3	MAPKAPK3	U09578	62.75	Above
2	36650_at	cyclin D2	CCND2	D13639	59.79	Above
3	40196_at	HYA22 protein	HYA22	D88153	54.79	Above
4	1635_at	proto-oncogene tyrosine-protein kinase ABL gene	ABL	U07563	54.77	Above
5	33775_s_at	caspase 8 apoptosis-related cysteine protease	CASP8	X98176	49.70	Above
6	1636_g_at	proto-oncogene tyrosine-protein kinase ABL gene	ABL	U07563	48.29	Above
7	41295_at	GTT1 protein	GTT1	AL041780	42.60	Above
8	37600_at	extracellular matrix protein 1	ECM1	U68186	42.60	Above
9	37012_at	capping protein actin filament muscle Z- line beta	CAPZB	U03271	38.46	Above
10	39225_at	alkylglycerone phosphate synthase	AGPS	Y09443	38.46	Above
11	1326_at	caspase 10 apoptosis-related cysteine protease	CASP10	U60519	37.83	Above
12	34362_at	solute carrier family 2 facilitated glucose transporter member 5	SLC2A5	M55531	37.54	Above
13	33150_at	disrupter of silencing 10	SAS10	AI126004	36.95	Above
14	40051_at	TRAM-like protein	KIAA0057	D31762	36.95	Above
15	39061_at	bone marrow stromal cell antigen 2	BST2	D28137	36.95	Above
16	33172_at	hypothetical protein FLJ10849	FLJ10849	T75292	36.95	Above
17	37399_at	aldo-keto reductase family 1 member C3 3-alpha hydroxysteroid dehydrogenase type II	AKR1C3	D17793	36.95	Above
18	317_at	protease cysteine 1 legumain	PRSC1	D55696	36.95	Above
19	40953_at	calponin 3 acidic	CNN3	S80562	33.94	Above
20	330_s_at	tubulin, alpha 1, isoform 44	TUBA1	HG2259- HT2348	33.32	Above
21	40504_at	paraoxonase 2	PON2	AF001601	31.46	Above
22	38578_at	tumor necrosis factor receptor superfamily member 7	TNFRSF7	M63928	30.47	Above
23	39044_s_at	diacylglycerol kinase delta 130kD	DGKD	D73409	29.59	Below

24	36634_at	BTG family member 2	BTG2	U72649	29.16	Below
25	38119_at	glycophorin C Gerbich blood group	GYPC	X12496	29.16	Above
26	32562_at	endoglin Osler-Rendu-Weber syndrome 1	ENG	X72012	27.96	Above
27	33228_g_at	interleukin 10 receptor beta	IL10RB	AI984234	27.70	Below
28	37006_at	step II splicing factor SLU7	SLU7	AI660656	27.15	Above
29	38641_at	Homo sapiens mRNA for TSC-22-like protein		AJ133115	27.15	Above
30	38220_at	dihydropyrimidine dehydrogenase	DPYD	U20938	27.15	Above
31	1211_s_at	CASP2 and RIPK1 domain containing adaptor with death domain	CRADD	U84388	26.46	Above
32	39730_at	v-abl Abelson murine leukemia viral oncogene homolog 1	ABL1	X16416	25.90	Above
33	36591_at	tubulin alpha 1 testis specific	TUBA1	X06956	25.90	Above
34	36035_at	anchor attachment protein 1 Gaa1p yeast homolog	GPAA1	AB002135	25.34	Above
35	980_at	Niemann-Pick disease type C1	NPC1	AF002020	25.29	Above
36	671_at	secreted protein acidic cysteine-rich osteonectin	SPARC	J03040	25.29	Above
37	40698_at	C-type calcium dependent carbohydrate-recognition domain lectin superfamily member 2 activation-induced	CLECSF2	X96719	23.80	Above
38	39330_s_at	actinin alpha 1	ACTN1	M95178	23.70	Above
39	1983_at	cyclin D2	CCND2	X68452	23.70	Above
40	2001_g_at	ataxia telangiectasia mutated	ATM	U26455	22.60	Above
<b><i>E2A-PBX1</i></b>						
1	41146_at	ADP-ribosyltransferase NAD poly ADP-ribose polymerase	ADPRT	J03473	187.00	Above
2	1287_at	ADP-ribosyltransferase NAD poly ADP-ribose polymerase	ADPRT	J03473	187.00	Above
3	32063_at	pre-B-cell leukemia transcription factor 1	PBX1	M86546	187.00	Above
4	33355_at	Homo sapiens cDNA FLJ12900 fis clone NT2RP2004321 (by CELERA search of target sequence = PBX1)	PBX1	AL049381	187.00	Above
5	430_at	nucleoside phosphorylase	NP	X00737	187.00	Above
6	40454_at	FAT tumor suppressor Drosophila homolog	FAT	X87241	176.11	Above
7	753_at	nidogen 2	NID2	D86425	164.28	Above
8	33821_at	Human DNA sequence from clone RP3-483K16 on chromosome 6p12.1-21.1	HELO1	AL034374	155.00	Above
9	39614_at	KIAA0802 protein	KIAA0802	AB018345	153.46	Above
10	38340_at	huntingtin interacting protein-1-related	KIAA0655	AB014555	143.85	Above
11	1786_at	c-mer proto-oncogene tyrosine kinase	MERTK	U08023	142.34	Above
12	39929_at	KIAA0922 protein	KIAA0922	AB023139	139.97	Above
13	39379_at	Homo sapiens mRNA cDNA DKFZp586C1019 from clone DKFZn586C1019		AL049397	139.49	Above

## DKFZp586C1019

14	717_at	GS3955 protein	GS3955	D87119	135.24	Above
15	362_at	protein kinase C zeta	PRKCZ	Z15108	131.36	Above
16	33513_at	signaling lymphocytic activation molecule	SLAM	U33017	131.36	Above
17	37225_at	KIAA0172 protein	KIAA0172	D79994	131.36	Above
18	854_at	B lymphoid tyrosine kinase	BLK	S76617	130.95	Above
19	35974_at	lymphoid-restricted membrane protein	LRMP	U10485	123.33	Above
20	36452_at	synaptopodin	KIAA1029	AB028952	123.33	Above
21	40648_at	c-mer proto-oncogene tyrosine kinase	MERTK	U08023	120.51	Above
22	38393_at	KIAA0247 gene product	KIAA0247	D87434	120.51	Above
23	38994_at	STAT induced STAT inhibitor-2	STAT2	AF037989	118.58	Below
24	34861_at	golgi autoantigen golgin subfamily a 3	GOLGA3	D63997	116.80	Above
25	38748_at	adenosine deaminase RNA-specific B1 homolog of rat RED1	ADARB1	U76421	114.13	Above
26	40113_at	GS3955 protein	GS3955	D87119	114.13	Above
27	36179_at	mitogen-activated protein kinase-activated protein kinase 2	MAPKAPK2	U12779	113.43	Above
28	37493_at	colony stimulating factor 2 receptor beta low-affinity granulocyte-macrophage	CSF2RB	H04668	113.04	Above
29	578_at	Human recombination activating protein (RAG2) gene	RAG2	M94633	111.32	Above
30	41017_at	myosin-binding protein H	MYBPH	U27266	109.73	Above
31	37625_at	interferon regulatory factor 4	IRF4	U52682	108.51	Above
32	38679_g_at	small nuclear ribonucleoprotein polypeptide E	SNRPE	AA73305	106.02	Above
33	1389_at	membrane metallo-endopeptidase neutral endopeptidase enkephalinase CALLA CD10	MME	J03779	105.65	Below
34	34783_s_at	BUB3 budding uninhibited by benzimidazoles 3 yeast homolog	BUB3	AF047473	103.87	Above
35	36959_at	ubiquitin-conjugating enzyme E2 variant1	UBE2V1	U49278	103.87	Above
36	39864_at	cold inducible RNA-binding protein	CIRBP	D78134	99.76	Below
37	41862_at	KIAA0056 protein	KIAA0056	D29954	99.76	Above
38	41425_at	Friend leukemia virus integration 1	FLI1	M98833	96.47	Above
39	37177_at	CD58 antigen lymphocyte function-associated antigen 3	CD58	Y00636	93.84	Above
40	37485_at	fatty-acid-Coenzyme A ligase very long-chain 1	FACVL1	D88308	93.17	Above

**Hyperdiploid >50**

1	36620_at	superoxide dismutase 1 soluble amyotrophic lateral sclerosis 1 adult	SOD1	X02317	52.43	Above
2	37350_at	Human DNA sequence from clone 889N15 on chromosome Xq22.1-22.3.	PSMD10	AL031177	48.71	Above
3	171_at	von Hippel-Lindau binding protein 1	VBP1	U56833	45.80	Above

4	37677_at	phosphoglycerate kinase 1	PGK1	V00572	45.80	Above
5	41724_at	accessory proteins BAP31/BAP29	DXS1357E	X81109	45.58	Above
6	32207_at	membrane protein palmitoylated 1 55kD	MPP1	M64925	44.07	Above
7	38738_at	SMT3 suppressor of mif two 3 yeast homolog 1	SMT3H1	X99584	43.57	Above
8	40480_s_at	FYN oncogene related to SRC FGR YES	FYN	M14333	43.57	Above
9	38518_at	sex comb on midleg Drosophila like 2	SCML2	Y18004	43.20	Above
10	41132_r_at	heterogeneous nuclear ribonucleoprotein H2 H	HNRPH2	U01923	43.15	Above
11	31492_at	muscle specific gene	M9	AB019392	43.01	Below
12	38317_at	transcription elongation factor A SII like 1	TCEAL1	M99701	41.10	Above
13	40998_at	trinucleotide repeat containing 11 THR-associated protein 230 kDa subunit	TNRC11	AF071309	40.88	Above
14	35688_g_at	mature T-cell proliferation 1	MTCP1	Z24459	40.52	Above
15	40903_at	ATPase H transporting lysosomal vacuolar proton pump membrane sector associated protein M8-9	APT6M8-9	AL049929	40.33	Above
16	36489_at	phosphoribosyl pyrophosphate synthetase 1	PRPS1	D00860	40.33	Above
17	1520_s_at	interleukin 1 beta	IL1B	X04500	40.29	Above
18	35939_s_at	POU domain class 4 transcription factor 1	POU4F1	L20433	38.74	Above
19	38604_at	neuropeptide Y	NPY	AI198311	38.26	Above
20	31863_at	KIAA0179 protein	KIAA0179	D80001	38.26	Above
21	890_at	ubiquitin-conjugating enzyme E2A RAD6 homolog	UBE2A	M74524	37.99	Above
22	39402_at	interleukin 1 beta	IL1B	M15330	37.92	Above
23	41490_at	phosphoribosyl pyrophosphate synthetase 2	PRPS2	Y00971	37.72	Above
24	34753_at	synaptobrevin-like 1	SYBL1	X92396	37.72	Above
25	40891_f_at	DNA segment on chromosome X unique 9879 expressed sequence	DXS9879E	X92896	37.15	Above
26	306_s_at	high-mobility group nonhistone chromosomal protein 14	HMG14	J02621	37.15	Above
27	37640_at	hypoxanthine phosphoribosyltransferase 1 Lesch-Nyhan syndrome	HPRT1	M31642	37.15	Above
28	34829_at	dyskeratosis congenita 1 dyskerin	DKC1	U59151	36.48	Above
29	36169_at	NADH dehydrogenase ubiquinone 1 alpha subcomplex 1 7.5kD MWFE	NDUFA1	N47307	36.48	Above
30	38968_at	SH3-domain binding protein 5 BTK-associated	SH3BP5	AB005047	35.95	Above
31	36128_at	transmembrane trafficking protein	TMP21	L40397	35.88	Above
32	37014_at	myxovirus influenza resistance 1 homolog of murine interferon-inducible protein p78	MX1	M33882	35.65	Above
33	34374_g_at	upstream regulatory element binding protein 1	UREB1	Z97054	35.55	Above
34	36542_at	solute carrier family 9 sodium/hydrogen exchanger isoform 6	SLC9A6	AF030409	35.55	Above

35	688_at	proteasome prosome macropain 26S subunit ATPase 1	PSMC1	L02426	35.55	Above
36	955_at	calmodulin type I		HG1862-HT1897	35.55	Above
37	35816_at	cystatin B stefin B	CSTB	U46692	35.27	Above
38	38459_g_at	Human cytochrome b5 (CYB5) gene	CYB5	L39945	35.18	Above
39	41288_at	matrix Gla protein	MGP	AL036744	35.18	Above
40	32251_at	hypothetical protein FLJ21174	FLJ21174	AA149307	35.14	Above

### **MLL**

1	34306_at	muscleblind Drosophila like	MBNL	AB007888	64.07	Above
2	40797_at	a disintegrin and metalloproteinase domain 10	ADAM10	AF009615	62.85	Above
3	33412_at	LGALS1 Lectin, galactoside-binding, soluble, 1	LGALS1	AI535946	57.97	Above
4	39338_at	S100 calcium-binding protein A10 annexin II ligand calpactin I light polypeptide p11	S100A10	AI201310	57.97	Above
5	2062_at	insulin-like growth factor binding protein 7	IGFBP7	L19182	55.22	Above
6	32193_at	plexin C1	PLXNC1	AF030339	53.59	Above
7	40518_at	protein tyrosine phosphatase receptor type C	PTPRC	Y00062	53.40	Above
8	36777_at	DNA segment on chromosome 12 unique 2489 expressed sequence	D12S2489E	AJ001687	51.47	Above
9	32207_at	membrane protein palmitoylated 1 55kD	MPP1	M64925	50.73	Below
10	33859_at	sin3-associated polypeptide 18kD	SAP18	U96915	50.48	Above
11	38391_at	capping protein actin filament gelsolin-like	CAPG	M94345	50.26	Above
12	40763_at	Meis1 mouse homolog	MEIS1	U85707	50.26	Above
13	1126_s_at	cell surface glycoprotein CD44 gene	CD44	L05424	50.17	Above
14	34721_at	FK506-binding protein 5	FKBP5	U42031	50.17	Above
15	37809_at	homeo box A9	HOXA9	U41813	50.17	Above
16	34861_at	golgi autoantigen golgin subfamily a 3	GOLGA3	D63997	47.58	Below
17	38194_s_at	immunoglobulin kappa constant	IGKC	M63438	46.18	Below
18	657_at	protocadherin gamma subfamily C 3	PCDHGC3	L11373	46.05	Above
19	36918_at	guanylate cyclase 1 soluble alpha 3	GUCY1A3	Y15723	43.90	Above
20	32215_i_at	KIAA0878 protein	KIAA0878	AB020685	43.90	Above
21	38160_at	lymphocyte antigen 75	LY75	AF011333	43.90	Above
22	38413_at	defender against cell death 1	DAD1	D15057	43.90	Above
23	1389_at	membrane metallo-endopeptidase neutral endopeptidase enkephalinase CALLA CD10	MME	J03779	43.82	Below
24	34168_at	deoxynucleotidyltransferase terminal	DNTT	M11722	43.82	Below
25	2036_s_at	CD44 antigen homing function and Indian blood group system	CD44	M59040	42.55	Above
26	40522_at	glutamate-ammonia ligase glutamine	GLUL	X59834	42.55	Above

		synthase				
27	854_at	B lymphoid tyrosine kinase	BLK	S76617	42.34	Above
28	40067_at	E74-like factor 1 ets domain transcription factor	ELF1	M82882	40.85	Above
29	39756_g_at	X-box binding protein 1	XBP1	Z93930	39.95	Below
30	36940_at	TGFB1-induced anti-apoptotic factor 1	TIAF1	D86970	39.82	Below
31	36935_at	RAS p21 protein activator GTPase activating protein 1	RASA1	M23379	38.77	Above
32	32134_at	testin	DKFZP586B2 022	AL050162	38.77	Above
33	39379_at	Homo sapiens mRNA cDNA DKFZp586C1019 from clone DKFZp586C1019		AL049397	38.77	Above
34	40493_at	Human cell surface glycoprotein CD44	CD44	L05424	38.44	Above
35	769_s_at	annexin A2	ANXA2	D00017	37.61	Above
36	40415_at	acetyl-Coenzyme A acyltransferase 1 peroxisomal 3-oxoacyl-Coenzyme A thiolase	ACAA1	X14813	37.55	Above
37	35983_at	hypothetical protein R32184_1	R32184_1	AC004528	37.55	Above
38	40519_at	protein tyrosine phosphatase receptor type C	PTPRC	Y00638	36.56	Above
39	794_at	protein tyrosine phosphatase non-receptor type 6	PTPN6	X62055	36.56	Above
40	41234_at	DnaJ Hsp40 homolog subfamily B member 6	DNAJB6	AI540318	36.56	Above
<b>Novel</b>						
1	37960_at	carbohydrate chondroitin 6/keratan sulfotransferase 2	CHST2	AB014679	175.82	Above
2	31892_at	protein tyrosine phosphatase receptor type M	PTPRM	X58288	172.85	Above
3	994_at	protein tyrosine phosphatase receptor type M	PTPRM	X58288	172.85	Above
4	995_g_at	protein tyrosine phosphatase receptor type M	PTPRM	X58288	172.85	Above
5	41074_at	G protein-coupled receptor 49	GPR49	AF062006	139.36	Above
6	41073_at	G protein-coupled receptor 49	GPR49	AI743745	139.36	Above
7	34676_at	KIAA1099 protein	KIAA1099	AB029022	137.71	Above
8	36139_at	DKFZP586G0522 protein	DKFZP586G0 522	AL050289	127.05	Above
9	37542_at	lipoma HMGIC fusion partner-like 2	LHFPL2	D86961	120.79	Above
10	41159_at	clathrin heavy polypeptide Hc	CLTC	D21260	115.15	Above
11	40081_at	phospholipid transfer protein	PLTP	L26232	108.33	Above
12	32800_at	Human retinoid X receptor alpha mRNA, 3' UTR, partial sequence	RXR	U66306	107.39	Above
13	36906_at	cannabinoid receptor 1 brain	CNR1	U73304	107.39	Above
14	39878_at	protocadherin 9	PCDH9	AI524125	99.20	Above



15	41747_s_at	Human myocyte-specific enhancer factor 2A (MEF2A) gene, last coding exon, and complete cds.	MEF2A	U49020	99.20	Above
16	33410_at	integrin alpha 6	ITGA6	S66213	96.17	Above
17	34947_at	phorbolin-like protein MDS019	MDS019	AA442560	93.59	Above
18	36029_at	chromosome 11 open reading frame 8	C11ORF8	U57911	93.59	Above
19	41708_at	KIAA1034 protein	KIAA1034	AB028957	92.60	Above
20	1664_at	insulin-like growth factor 2	IGF2	HG3543-HT3739	92.60	Above
21	32736_at	HSPC022 protein	HSPC022	W68830	91.62	Below
22	41266_at	integrin alpha 6	ITGA6	X53586	86.95	Above
23	36566_at	cystinosis nephropathic	CTNS	AJ222967	82.89	Above
24	1825_at	IQ motif containing GTPase activating protein 1	IQGAP1	L33075	81.20	Below
25	1731_at	platelet-derived growth factor receptor alpha polypeptide	PDGFRA	M21574	78.22	Above
26	37023_at	lymphocyte cytosolic protein 1 L-plastin	LCP1	J02923	78.22	Below
27	33037_at	carbohydrate N-acetylglucosamine 6-O sulfotransferase 7	CHST7	AL022165	76.00	Above
28	33411_g_at	integrin alpha 6	ITGA6	S66213	75.47	Above
29	538_at	CD34 antigen	CD34	S53911	74.86	Above
30	39108_at	lanosterol synthase 2 3-oxidosqualene-lanosterol cyclase	LSS	U22526	71.90	Above
31	38364_at	BCE-1 protein	BCE-1	AF068197	71.90	Above
32	40423_at	KIAA0903 protein	KIAA0903	AB020710	71.29	Above
33	35192_at	glycine dehydrogenase decarboxylating glycine decarboxylase glycine cleavage system protein P	GLDC	D90239	71.29	Above
34	39037_at	myeloid/lymphoid or mixed-lineage leukemia trithorax Drosophila homolog translocated to 2	MLLT2	L13773	71.29	Above
35	38747_at	Human CD34 gene, exon 8.	CD34	M81945	69.45	Above
36	37687_i_at	Fc fragment of IgG low affinity IIa receptor for CD32	FCGR2A	M31932	67.75	Above
37	1857_at	MAD mothers against decapentaplegic Drosophila homolog 7	MADH7	AF010193	66.28	Above
38	38618_at	Human PAC clone RP3-515N1 from 22q11.2-q22	LIMK2	AC002073	64.03	Above
39	31782_at	prostaglandin D2 receptor DP	PTGDR	U31099	61.92	Above
40	32842_at	B-cell CLL/lymphoma 7A	BCL7A	X89984	61.57	Above

#### T-ALL

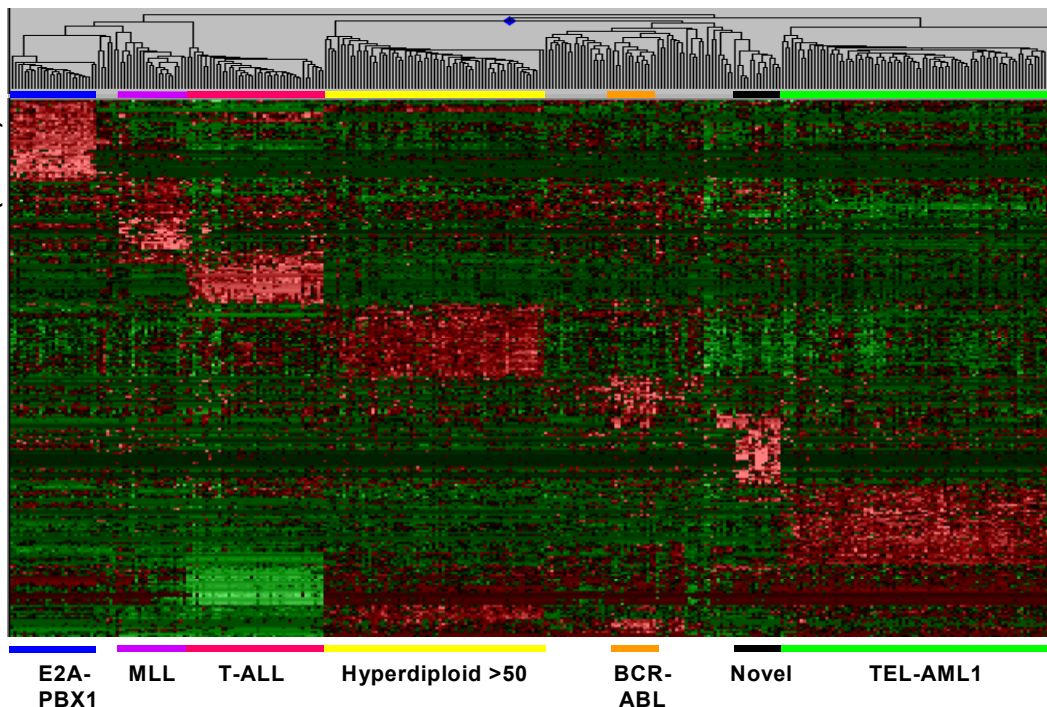
1	38319_at	CD3D antigen delta polypeptide TIT3 complex	CD3D	AA919102	215.00	Above
2	1096_g_at	CD19 antigen	CD19	M28170	206.48	Below
3	38242_at	B cell linker protein	SLP65	AF068180	198.52	Below
4	32794_g_at	T cell receptor beta locus	TRB	X00437	197.71	Above

5	37988_at	CD79B antigen immunoglobulin-associated beta	CD79B	M89957	197.71	Below
6	38017_at	CD79A antigen immunoglobulin-associated alpha	CD79A	U05259	197.53	Below
7	35016_at	Human Ia-associated invariant gamma-chain gene, exon 8, clones lambda-y(1,2,3).		M13560	M13560	Below
8	36277_at	Human membran protein (CD3-epsilon) gene, exon 9.	CD3E	M23323	197.53	Above
9	38095_i_at	major histocompatibility complex class II DP beta 1	HLA-DPB1	M83664	191.09	Below
10	39318_at	T-cell leukemia/lymphoma 1A	TCL1A	X82240	189.78	Below
11	38147_at	SH2 domain protein 1A Duncan s disease lymphoproliferative syndrome	SH2D1A	AL023657	189.78	Above
12	41723_s_at	major histocompatibility complex class II DR beta 1	HLA-DRB1	M32578	189.25	Below
13	38833_at	Human mRNA for SB classII histocompatibility antigen alpha-chain		X00457	189.03	Below
14	33238_at	Human T-lymphocyte specific protein tyrosine kinase p56lck (lck) abberant mRNA	lck	U23852	189.03	Above
15	37039_at	major histocompatibility complex class II DR alpha	HLA-DRA	J00194	188.93	Below
16	38051_at	mal T-cell differentiation protein	MAL	X76220	188.93	Above
17	37344_at	major histocompatibility complex class II DM alpha	HLA-DMA	X62744	187.25	Below
18	38096_f_at	major histocompatibility complex class II DP beta 1	HLA-DPB1	M83664	182.38	Below
19	2059_s_at	lymphocyte-specific protein tyrosine kinase	LCK	M36881	182.38	Above
20	1105_s_at	T cell receptor beta locus	TRB	M12886	180.45	Above
21	32649_at	transcription factor 7 T-cell specific HMG-box	TCF7	X59871	177.84	Above
22	38949_at	protein kinase C theta	PRKCQ	L01087	172.59	Below
23	39709_at	selenoprotein W 1	SEPW1	U67171	171.96	Above
24	41165_g_at	immunoglobulin heavy constant mu	IGHM	X67301	171.96	Below
25	36473_at	ubiquitin specific protease 20	USP20	AB023220	167.27	Above
26	266_s_at	CD24 antigen small cell lung carcinoma cluster 4 antigen	CD24	L33930	165.56	Below
27	40570_at	forkhead box O1A rhabdomyosarcoma	FOXO1A	AF032885	165.29	Below
28	40775_at	integral membrane protein 2A	ITM2A	AL021786	164.14	Above
29	37420_i_at	Human DNA sequence from clone RP3-377H14 on chromosome 6p21.32-22.1.		AL022723	164.14	Below
30	1085_s_at	phospholipase C gamma 2 phosphatidylinositol-specific	PLCG2	M37238	161.30	Below
31	38018_g_at	CD79A antigen immunoglobulin-associated alpha	CD79A	U05259	160.51	Below
32	35643_at	nucleobindin 2	NUCB2	X76732	160.07	Above
33	41166_at	immunoglobulin heavy constant mu	IGHM	X58529	158.50	Below

34	38415_at	protein tyrosine phosphatase type IVA member 2	PTP4A2	U14603	155.78	Above
35	38893_at	neutrophil cytosolic factor 4 40kD	NCF4	AL008637	155.78	Below
36	1241_at	protein tyrosine phosphatase type IVA member 2	PTP4A2	U14603	155.78	Above
37	32793_at	T cell receptor beta locus	TRB	X00437	155.43	Above
38	36571_at	topoisomerase DNA II beta 180kD	TOP2B	X68060	152.16	Below
39	37399_at	aldo-keto reductase family 1 member C3 3-alpha hydroxysteroid dehydrogenase type II	AKR1C3	D17793	151.93	Above
40	41097_at	telomeric repeat binding factor 2	TERF2	AF002999	151.86	Below
<b><i>TEL-AML1</i></b>						
1	38652_at	hypothetical protein FLJ20154	FLJ20154	AF070644	137.92	Above
2	36239_at	POU domain class 2 associating factor 1	POU2AF1	Z49194	131.43	Above
3	41442_at	core-binding factor runt domain alpha subunit 2 translocated to 3	CBFA2T3	AB010419	130.17	Above
4	37780_at	piccolo presynaptic cytomatrix protein	PCLO	AB011131	126.79	Above
5	36985_at	isopentenyl-diphosphate delta isomerase	IDI1	X17025	125.47	Above
6	38578_at	tumor necrosis factor receptor superfamily member 7	TNFRSF7	M63928	115.72	Above
7	38203_at	potassium intermediate/small conductance calcium-activated channel subfamily N member 1	KCNN1	U69883	112.87	Above
8	35614_at	transcription factor-like 5 basic helix-loop-helix	TCFL5	AB012124	108.45	Above
9	32224_at	KIAA0769 gene product	KIAA0769	AB018312	107.08	Above
10	32730_at	Homo sapiens mRNA for KIAA1750 protein partial cds		AL080059	104.93	Above
11	35665_at	phosphoinositide-3-kinase class 3	PIK3C3	Z46973	104.83	Above
12	1077_at	recombination activating gene 1	RAG1	M29474	102.90	Above
13	36524_at	Rho guanine nucleotide exchange factor GEF 4	ARHGEF4	AB029035	100.67	Above
14	34194_at	Homo sapiens cDNA FLJ21697 fis clone COL09740		AL049313	98.31	Above
15	36937_s_at	PDZ and LIM domain 1 elfin	PDLIM1	U90878	96.91	Below
16	36008_at	protein tyrosine phosphatase type IVA member 3	PTP4A3	AF041434	96.68	Above
17	1299_at	telomeric repeat binding factor 2	TERF2	X93512	93.08	Above
18	41814_at	fucosidase alpha-L- 1 tissue	FUCA1	M29877	92.77	Above
19	41200_at	CD36 antigen collagen type I receptor thrombospondin receptor like 1	CD36L1	Z22555	90.86	Above
20	35238_at	TNF receptor-associated factor 5	TRAF5	AB000509	90.81	Above
21	880_at	FK506-binding protein 1A 12kD	FKBP1A	M34539	86.69	Above
22	33690_at	Homo sapiens mRNA cDNA DKFZp434A202 from clone DKFZp434A202		AL080190	86.69	Above
23	40272_at	collapsin response mediator protein 1	CRMP1	D78012	85.44	Above

24	35362_at	myosin X	MYO10	AB018342	83.60	Above
25	41819_at	FYN-binding protein FYB-120/130	FYB	U93049	83.25	Above
26	40279_at	KIAA0121 gene product	KIAA0121	D50911	81.66	Above
27	1488_at	protein tyrosine phosphatase receptor type K	PTPRK	L77886	81.66	Above
28	1325_at	MAD mothers against decapentaplegic Drosophila homolog 1	MADH1	U59423	81.17	Above
29	37908_at	guanine nucleotide binding protein 11	GNG11	U31384	80.37	Above
30	769_s_at	annexin A2	ANXA2	D00017	78.68	Below
31	33415_at	non-metastatic cells 2 protein NM23B expressed in	NME2	X58965	77.04	Below
32	1980_s_at	non-metastatic cells 2 protein NM23B expressed in	NME2	X58965	76.35	Below
33	32579_at	SWI/SNF related matrix associated actin dependent regulator of chromatin subfamily a member 4	SMARCA4	D26156	76.35	Above
34	39425_at	thioredoxin reductase 1	TXNRD1	X91247	75.97	Above
35	755_at	inositol 1 4 5-triphosphate receptor type 1	ITPR1	D26070	75.56	Above
36	37343_at	inositol 1 4 5-triphosphate receptor type 3	ITPR3	U01062	75.11	Above
37	1336_s_at	protein kinase C beta 1	PRKCB1	X06318	73.96	Above
38	41097_at	telomeric repeat binding factor 2	TERF2	AF002999	73.84	Above
39	31786_at	Sam68-like phosphotyrosine protein T-STAR	T-STAR	AF051321	73.72	Above
40	160029_at	protein kinase C beta 1	PRKCB1	X07109	73.66	Above

Illustrated below are the results of a two-dimensional hierarchical clustering algorithm of the 327 diagnostic ALL cases using the top 40 probe sets for each of the 7 groups selected by the Chi-square metric. This represents 271 unique probe sets as some probe sets were selected for more than one diagnostic group.



**Figure 14.** Hierarchical cluster of 327 diagnostic ALL bone marrow samples and genes chosen by Chi-square metric.

### Correlation-based Feature Selection (CFS)

The Correlation-based Feature Selection (CFS) is a method that evaluates subsets of genes rather than individual genes.<sup>4</sup> The core of the algorithm is a subset evaluation heuristic that takes into account the usefulness of individual features for predicting the class along with the level of intercorrelation among them with the belief that “good feature subsets contain features highly correlated with the class, yet uncorrelated with each other”. The heuristic assigns a score  $Merit_s$  to a subset  $S$  containing  $k$  genes, defined as  $Merit_s = (k * r_{cf}) / \sqrt{k + k * (k - 1) * r_{ff}}$ , where  $r_{cf}$  is the average gene-class correlation and  $r_{ff}$  is the average gene-gene correlation. Like the Chi-square method, CFS first discretizes the gene expressions into intervals and then calculates a matrix of gene-class and gene-gene correlations from the training data for merit calculation. The correlation between two genes or a gene and a class is calculated as  $r_{xy} = 2 * [H(X) + H(Y) - H(X,Y)] / [H(X) + H(Y)]$ , where  $H(X)$  is the entropy of a gene  $X$ . CFS starts from an empty set of genes and uses the best-first search technique with a stopping criterion of 5 consecutive fully expanded non-improving subsets. The subset with the highest merit found during the search is selected. Table 12 lists the top gene subsets chosen by CFS for each subtype. For subtype prediction, each gene subset must be used in its entirety, as within each subset, all genes are equally ranked. A two-dimensional clustering of these genes is shown in Figure 15. Recall that in CFS, genes in a subset are chosen so that they have high correlation with the corresponding class and low correlation with each other. This characteristic is clearly visible in Figure 15.

**Table 12. Genes selected by CFS**

	Affymetrix number	Gene Name	GeneSymbol	Reference number	Above/ Below Mean
<b><i>BCR-ABL</i></b>					
1	36650_at	cyclin D2	CCND2	D13639	Above
2	40196_at	HYA22 protein	HYA22	D88153	Above
3	1635_at	proto-oncogene tyrosine-protein kinase (ABL) gene	ABL	U07563	Above
4	33775_s_at	caspase 8 apoptosis-related cysteine protease	CASP8	X98176	Above
5	1636_g_at	proto-oncogene tyrosine-protein kinase (ABL) gene	ABL	U07563	Above
6	41295_at	GTT1 protein	GTT1	AL041780	Above
7	1326_at	caspase 10 apoptosis-related cysteine protease	CASP10	U60519	Above
8	33150_at	disrupter of silencing 10	SAS10	AI126004	Above
9	40051_at	TRAM-like protein	KIAA0057	D31762	Above
10	39061_at	bone marrow stromal cell antigen 2	BST2	D28137	Above
11	33172_at	hypothetical protein FLJ10849	FLJ10849	T75292	Above
12	37399_at	aldo-keto reductase family 1 member C3 3-alpha hydroxysteroid dehydrogenase type II	AKR1C3	D17793	Above
13	317_at	protease cysteine 1 legumain	PRSC1	D55696	Above
14	330_s_at	tubulin, alpha 1, isoform 44	TUBA1	HG2259- HT2348	Above

15	38578_at	tumor necrosis factor receptor superfamily member 7	TNFRSF7	M63928	Above
16	39044_s_at	diacylglycerol kinase delta 130kD	DGKD	D73409	Below
17	32562_at	endoglin Osler-Rendu-Weber syndrome 1	ENG	X72012	Above
18	38641_at	Homo sapiens mRNA for TSC-22-like protein		AJ133115	Above
19	1211_s_at	CASP2 and RIPK1 domain containing adaptor with death domain	CRADD	U84388	Above
20	39730_at	v-abl Abelson murine leukemia viral oncogene homolog 1	ABL1	X16416	Above
21	36591_at	tubulin alpha 1 testis specific	TUBA1	X06956	Above
22	36035_at	anchor attachment protein 1 Gaa1p yeast homolog	GPAA1	AB002135	Above
23	980_at	Niemann-Pick disease type C1	NPC1	AF002020	Above
24	40698_at	C-type calcium dependent carbohydrate-recognition domain lectin superfamily member 2 activation-induced	CLECSF2	X96719	Above
25	39330_s_at	actinin alpha 1	ACTN1	M95178	Above
26	2001_g_at	ataxia telangiectasia mutated includes complementation groups A C and D	ATM	U26455	Above
27	39319_at	lymphocyte cytosolic protein 2 SH2 domain-containing leukocyte protein of 76kD	LCP2	U20158	Above
28	37685_at	Clathrin assembly lymphoid-myeloid leukemia gene	CLTH	U45976	Above
29	33813_at	tumor necrosis factor receptor superfamily member 1B	TNFRSF1B	AI813532	Above
30	33134_at	adenylate cyclase 3	ADCY3	AB011083	Above
31	36536_at	schwannomin interacting protein 1	SCHIP-1	AF070614	Above
32	36985_at	isopenentenyl-diphosphate delta isomerase	IDI1	X17025	Below
33	35991_at	Sm protein F	LSM6	AA917945	Above
34	33774_at	caspase 8 apoptosis-related cysteine protease	CASP8	X98172	Above
35	37470_at	leukocyte-associated Ig-like receptor 1	LAIR1	AF013249	Above
36	39245_at	Human 40871 mRNA partial sequence		U72507	Above
37	40076_at	tumor protein D52-like 2	TPD52L2	AF004430	Below
38	39370_at	Microtubule-associated proteins 1A and 1B light chain 3	MAP1ALC3	W28807	Below
39	41594_at	Janus kinase 1 a protein tyrosine kinase	JAK1	M64174	Above
40	41338_at	amino-terminal enhancer of split	AES	AI969192	Below
41	32319_at	tumor necrosis factor ligand superfamily member 4 tax-transcriptionally activated glycoprotein 1 34kD	TNFSF4	AL022310	Above
42	33924_at	KIAA1091 protein	KIAA1091	AB029014	Above
43	37397_at	platelet/endothelial cell adhesion molecule-1 (PECAM-1) gene	PECAM	L34657	Above
44	37190_at	WAS protein family member 1	WASF1	D87459	Below
45	39070_at	singed Drosophila like sea urchin fascin homolog like	SNL	U03057	Above
46	38994_at	STAT induced STAT inhibitor-2	STAT2	AF037989	Above

47	32621_at	down-regulator of transcription 1 TBP-binding negative cofactor 2	DR1	M97388	Above
48	40108_at	KIAA0005 gene product	KIAA0005	D13630	Below
49	35238_at	TNF receptor-associated factor 5	TRAF5	AB000509	Above
50	1558_g_at	p21/Cdc42/Rac1-activated kinase 1 yeast Ste20-related	PAK1	U24152	Above
51	1373_at	transcription factor 3 E2A immunoglobulin enhancer binding factors E12/E47	TCF3	M31523	Below
52	35731_at	integrin alpha 4 antigen CD49D alpha 4 subunit of VLA-4 receptor	ITGA4	X16983	Above
53	38659_at	suppressor of clear C. elegans homolog of	SHOC2	AB020669	Below

### ***E2A-PBX1***

1	33355_at	Homo sapiens cDNA FLJ12900 fis clone NT2RP2004321 (by CELERA search of target sequence = PBX1)	PBX1	AL049381	Above
---	----------	--	------	----------	-------

### **Hyperdiploid >50**

1	36620_at	superoxide dismutase 1 soluble amyotrophic lateral sclerosis 1 adult	SOD1	X02317	Above
2	37350_at	clone 889N15 on chromosome Xq22.1-22.3. Contains part of the gene for a novel protein similar to X. laevis Cortical Thymocyte Marker CTX	PSMD10	AL031177	Above
3	41724_at	accessory proteins BAP31/BAP29	DXS1357E	X81109	Above
4	38738_at	SMT3 suppressor of mif two 3 yeast homolog 1	SMT3H1	X99584	Above
5	40480_s_at	FYN oncogene related to SRC FGR YES	FYN	M14333	Above
6	38518_at	sex comb on midleg Drosophila like 2	SCML2	Y18004	Above
7	31492_at	muscle specific gene	M9	AB019392	Below
8	35688_g_at	mature T-cell proliferation 1	MTCP1	Z24459	Above
9	35939_s_at	POU domain class 4 transcription factor 1	POU4F1	L20433	Above
10	36128_at	transmembrane trafficking protein	TMP21	L40397	Above
11	37014_at	myxovirus influenza resistance 1 homolog of murine interferon-inducible protein p78	MX1	M33882	Above
12	34374_g_at	upstream regulatory element binding protein 1	UREB1	Z97054	Above
13	688_at	proteasome prosome macropain 26S subunit ATPase 1	PSMC1	L02426	Above
14	39878_at	protocadherin 9	PCDH9	AI524125	Below
15	38771_at	histone deacetylase 1	HDAC1	D50405	Below
16	865_at	ribosomal protein S6 kinase 90kD polypeptide 3	RPS6KA3	U08316	Above
17	41143_at	calmodulin (CALM1) gene	CALM1	U12022	Above
18	39867_at	Tu translation elongation factor mitochondrial	TUFM	S75463	Below
19	41470_at	prominin mouse like 1	PROML1	AF027208	Above

20	41503_at	KIAA0854 protein	KIAA0854	AB020661	Below
21	2039_s_at	FYN oncogene related to SRC FGR YES	FYN	M14333	Above
22	36845_at	KIAA0136 protein	KIAA0136	D50926	Above
23	36940_at	TGFB1-induced anti-apoptotic factor 1	TIAF1	D86970	Above
24	32236_at	ubiquitin-conjugating enzyme E2G 2 homologous to yeast UBC7	UBE2G2	AF032456	Above
25	36885_at	spleen tyrosine kinase	SYK	L28824	Below
26	40200_at	heat shock transcription factor 1	HSF1	M64673	Below
27	40842_at	U1 snRNP-specific protein A gene	SNRPA	M60784	Below
28	40514_at	hypothetical 43.2 Kd protein	LOC51614	AF091085	Below
29	41222_at	signal transducer and activator of transcription 6 (STAT6) gene	STAT6	AF067575	Below
30	1294_at	ubiquitin-activating enzyme E1-like	UBE1L	L13852	Below
31	34315_at	AFG3 ATPase family gene 3 yeast like 2	AFG3L2	Y18314	Above
32	39806_at	DKFZP547E2110 protein	DKFZP547E2110	AL050261	Above
33	40875_s_at	small nuclear ribonucleoprotein 70kD polypeptide RNP antigen	SNRP70	X06815	Below
34	38458_at	cytochrome b5 (CYB5) gene	CYB5	L39945	Above
35	1817_at	prefoldin 5	PFDN5	D89667	Below
36	34709_r_at	stromal antigen 2	STAG2	Z75331	Above
37	33447_at	myosin light polypeptide regulatory non-sarcomeric 20kD	MLCB	X54304	Above
38	1077_at	recombination activating gene 1	RAG1	M29474	Below
39	1915_s_at	v-fos FBJ murine osteosarcoma viral oncogene homolog	FOS	V01512	Above
40	38854_at	KIAA0635 gene product	KIAA0635	AB014535	Above
41	37732_at	RING1 and YY1 binding protein	RYBP	AL049940	Above
42	35940_at	POU domain class 4 transcription factor 1	POU4F1	X64624	Above
43	34733_at	splicing factor 3a subunit 1 120kD	SF3A1	X85237	Below
44	245_at	selectin L lymphocyte adhesion molecule 1	SELL	M25280	Below
45	40146_at	RAP1B member of RAS oncogene family	RAP1B	AL080212	Below
46	40104_at	serine/threonine kinase 25 Ste20 yeast homolog	STK25	D63780	Below
47	430_at	nucleoside phosphorylase	NP	X00737	Above
48	36899_at	special AT-rich sequence binding protein 1 binds to nuclear matrix/scaffold-associating DNA s	SATB1	M97287	Below
49	35727_at	hypothetical protein FLJ20517	FLJ20517	AI249721	Below
50	38649_at	KIAA0970 protein	KIAA0970	AB023187	Below
51	36107_at	ATP synthase H transporting mitochondrial F0 complex subunit F6	ATP5J	AA845575	Above
52	38789_at	transketolase Wernicke-Korsakoff syndrome	TKT	L12711	Below
53	39301_at	calpain 3 p94	CAPN3	X85030	Below
54	41278_at	BAF53	BAF53A	AF041474	Below
55	41162_at	protein phosphatase 1G formerly 2C magnesium-dependent gamma isoform	PPM1G	Y13936	Below



56	37819_at	hypothetical protein	LOC54104	AF007130	Below
57	38717_at	DKFZP586A0522 protein	DKFZP586A0522	AL050159	Below
58	40019_at	ecotropic viral integration site 2B	EVI2B	M60830	Above
59	39489_g_at	protocadherin 9	PCDH9	W27720	Below
60	857_at	protein phosphatase 1A formerly 2C magnesium-dependent alpha isoform	PPM1A	S87759	Above
61	32804_at	RNA binding motif protein 5	RBM5	AF091263	Below
62	37676_at	phosphodiesterase 8A	PDE8A	AF056490	Below
63	1519_at	v-ets avian erythroblastosis virus E26 oncogene homolog 2	ETS2	J04102	Above
64	37680_at	A kinase PRKA anchor protein gravin 12	AKAP12	U81607	Below
65	548_s_at	spleen tyrosine kinase	SYK	S80267	Below
66	39797_at	KIAA0349 protein	KIAA0349	AB002347	Above
67	32789_at	nuclear cap binding protein subunit 2 20kD	NCBP2	AA149428	Below
68	38091_at	lectin galactoside-binding soluble 9 galectin 9	LGALS9	Z49107	Below
69	41223_at	cytochrome c oxidase subunit Va	COX5A	M22760	Below
70	933_f_at	zinc finger protein 91 HPF7 HTF10	ZNF91	L11672	Below
71	37012_at	capping protein actin filament muscle Z-line beta	CAPZB	U03271	Below
72	35214_at	UDP-glucose dehydrogenase	UGDH	AF061016	Above
73	32434_at	myristoylated alanine-rich protein kinase C substrate MARCKS 80K-L	MACS	D10522	Above
74	38345_at	centrosomal protein 1	CEP1	AF083322	Below
75	40404_s_at	CDC16 cell division cycle 16 S. cerevisiae homolog	CDC16	U18291	Below
76	39096_at	SON DNA binding protein	SON	AB028942	Above
77	33429_at	DKFZP586M1523 protein	DKFZP586M1523	AL050225	Above
78	40641_at	TBP-associated factor 172	TAF-172	AF038362	Above
79	41381_at	KIAA0308 protein	KIAA0308	AB002306	Below
80	35135_at	Homo sapiens Similar to CG15084 gene product clone MGC 10471 mRNA complete cds		X13956	Below
81	39421_at	runt-related transcription factor 1 acute myeloid leukemia 1 aml1 oncogene	RUNX1	D43969	Below
82	195_s_at	caspase 4 apoptosis-related cysteine protease	CASP4	U28014	Below
83	36898_r_at	primase polypeptide 2A 58kD	PRIM2A	X74331	Above
84	38792_at	spermine synthase	SMS	AD001528	Above
85	32643_at	glucan 1 4-alpha- branching enzyme 1 glycogen branching enzyme Andersen disease glycogen storage disease type IV	GBE1	L07956	Below
86	38808_at	cell membrane glycoprotein 110000M r surface antigen	GP110	D64154	Below
87	36062_at	Leupaxin	LPXN	AF062075	Below
88	300_f_at	transcription factor BTF3 homolog (GB:M90355)		HG4518-HT4921	Below
89	1979_s_at	nucleolar protein 1 120kD	NOL1	X55504	Below

90	32230_at	eukaryotic translation initiation factor 3 subunit 2 beta 36kD	EIF3S2	U39067	Below
91	39893_at	guanine nucleotide binding protein G protein gamma 7	GNG7	AB010414	Below
92	34651_at	catechol-O-methyltransferase	COMT	M58525	Above
93	1052_s_at	CCAAT/enhancer binding protein C/EBP delta	CEBPD	M83667	Below
94	36272_r_at	peripheral myelin protein 2	PMP2	X62167	Below
95	2044_s_at	retinoblastoma 1 including osteosarcoma	RB1	M15400	Below
96	32135_at	sterol regulatory element binding transcription factor 1	SREBF1	U00968	Below

### **MLL**

1	34306_at	muscleblind Drosophila like	MBNL	AB007888	Above
2	40797_at	a disintegrin and metalloproteinase domain 10	ADAM10	AF009615	Above
3	33412_at	LGALS1 Lectin, galactoside-binding, soluble, 1 (galectin 1)	LGALS1	AI535946	Above
4	39338_at	S100 calcium-binding protein A10 annexin II ligand calpactin I light polypeptide p11	S100A10	AI201310	Above
5	2062_at	insulin-like growth factor binding protein 7	IGFBP7	L19182	Above
6	32193_at	plexin C1	PLXNC1	AF030339	Above
7	40518_at	protein tyrosine phosphatase receptor type C	PTPRC	Y00062	Above
8	36777_at	DNA segment on chromosome 12 unique 2489 expressed sequence	D12S2489E	AJ001687	Above
9	38391_at	capping protein actin filament gelsolin-like	CAPG	M94345	Above
10	40763_at	Meis1 mouse homolog	MEIS1	U85707	Above
11	34721_at	FK506-binding protein 5	FKBP5	U42031	Above
12	37809_at	homeo box A9	HOXA9	U41813	Above
13	32215_i_at	KIAA0878 protein	KIAA0878	AB020685	Above
14	38160_at	lymphocyte antigen 75	LY75	AF011333	Above
15	1389_at	membrane metallo-endopeptidase neutral endopeptidase enkephalinase CALLA CD10	MME	J03779	Below
16	34168_at	deoxynucleotidyltransferase terminal	DNTT	M11722	Below
17	40522_at	glutamate-ammonia ligase glutamine synthase	GLUL	X59834	Above
18	854_at	B lymphoid tyrosine kinase	BLK	S76617	Above
19	40067_at	E74-like factor 1 ets domain transcription factor	ELF1	M82882	Above
20	39756_g_at	X-box binding protein 1	XBP1	Z93930	Below
21	32134_at	Testing	DKFZP586B20	AL050162	Above
22	39379_at	Homo sapiens mRNA cDNA DKFZp586C1019 from clone DKFZp586C1019		AL049397	Above
23	40415_at	acetyl-Coenzyme A acyltransferase 1 peroxisomal 3-oxoacyl-Coenzyme A thiolase	ACAA1	X14813	Above

24	40519_at	protein tyrosine phosphatase receptor type C	PTPRC	Y00638	Above
25	33847_s_at	cyclin-dependent kinase inhibitor 1B p27 Kip1	CDKN1B	U10906	Above
26	32696_at	pre-B-cell leukemia transcription factor 3	PBX3	X59841	Above
27	40417_at	KIAA0098 protein		D43950	Above
28	1644_at	eukaryotic translation initiation factor 3 subunit 2 beta 36kD	EIF3S2	U36764	Above
29	948_s_at	peptidylprolyl isomerase D cyclophilin D	PPID	D63861	Above
30	34337_s_at	putative DNA binding protein	M96	AJ010014	Below
31	41747_s_at	myocyte-specific enhancer factor 2A (MEF2A) gene	MEF2A	U49020	Above
32	39516_at	hypothetical protein	HSPC004	AI827793	Above
33	31820_at	hematopoietic cell-specific Lyn substrate 1	HCLS1	X16663	Above
34	33305_at	serine or cysteine proteinase inhibitor clade B ovalbumin member 1	SERPINB1	M93056	Above
35	40520_g_at	protein tyrosine phosphatase receptor type C	PTPRC	Y00638	Above
36	41222_at	signal transducer and activator of transcription 6 (STAT6) gene	STAT6	AF067575	Above
37	1718_at	actin related protein 2/3 complex subunit 2 kD	ARPC2	U50523	Above
38	38342_at	KIAA0239 protein	KIAA0239	D87076	Below
39	38805_at	TG-interacting factor TALE family homeobox	TGIF	X89750	Below
40	32089_at	sperm associated antigen 6	SPAG6	AF079363	Above
41	1950_s_at	Smad 3, exon 1		AB004922	Above
42	39410_at	development and differentiation enhancing factor 2	DDEF2	AB007860	Above
43	37280_at	MAD mothers against decapentaplegic Drosophila homolog 1	MADH1	U59912	Below
44	32607_at	brain acid-soluble protein 1	BASP1	AF039656	Above
45	39389_at	CD9 antigen p24	CD9	M38690	Below
46	40913_at	ATPase Ca transporting plasma membrane 4	ATP2B4	W28589	Below
47	1039_s_at	hypoxia-inducible factor 1 alpha subunit basic helix-loop-helix transcription factor	HIF1A	U22431	Below
48	35939_s_at	POU domain class 4 transcription factor 1	POU4F1	L20433	Below
49	963_at	ligase IV DNA ATP-dependent	LIG4	X83441	Below
50	39628_at	RAB9 member RAS oncogene family	RAB9	U44103	Below
51	38242_at	B cell linker protein	SLP65	AF068180	Below
52	37692_at	diazepam binding inhibitor GABA receptor modulator acyl-Coenzyme A binding protein	DBI	AI557240	Above
53	32166_at	KIAA1027 protein	KIAA1027	AB028950	Above
54	34800_at	DKFZP586O1624 protein	DKFZP586O16 24	AL039458	Below
55	34386_at	methyl-CpG binding domain protein 4	MBD4	AF072250	Below
56	40296_at	hypothetical protein	753P9	AL023653	Below
57	40456_at	up-regulated by BCG-CWS	LOC64116	AL049963	Above
58	33943_at	ferritin heavy polypeptide 1	FTH1	L20941	Below

59	39049_at	G18.1a and G18.1b proteins (G18.1a and G18.1b genes, located in the class III region of the major histocompatibility complex)		AJ243937	Below
60	38075_at	synaptophysin-like protein	SYPL	X68194	Above
61	932_i_at	zinc finger protein 91 HPF7 HTF10	ZNF91	L11672	Below
62	1825_at	IQ motif containing GTPase activating protein 1	IQGAP1	L33075	Above
63	34210_at	CDW52 antigen CAMPATH-1 antigen	CDW52	N90866	Below
64	39778_at	mannosyl alpha-1 3- glycoprotein beta-1 2-N-acetylglucosaminyltransferase	MGAT1	M55621	Below
65	34699_at	CD2-associated protein	CD2AP	AL050105	Below
66	40066_at	ubiquitin-activating enzyme E1C homologous to yeast UBA3	UBE1C	AF046024	Above
67	41177_at	hypothetical protein FLJ12443	FLJ12443	AW024285	Above
68	32736_at	HSPC022 protein	HSPC022	W68830	Above
69	1928_s_at	mad protein homolog Smad2 gene	Smad2	U78733	Below
70	1081_at	ornithine decarboxylase 1	ODC1	M33764	Above
71	37345_at	Calumenin	CALU	AF013759	Above
72	34099_f_at	nucleosome assembly protein 1-like 1	NAP1L1	W26056	Above
73	933_f_at	zinc finger protein 91 HPF7 HTF10	ZNF91	L11672	Below
74	32214_at	thioredoxin-like 32kD	TXNL	AF003938	Below
75	33501_r_at	SNC73 protein SNC73 mRNA complete cds		S71043	Below
76	950_at	translocation protein 1	TLOC1	D87127	Below
77	41161_at	death-associated protein 6	DAXX	AB015051	Below
78	41381_at	KIAA0308 protein	KIAA0308	AB002306	Below
79	38705_at	ubiquitin-conjugating enzyme E2D 2 homologous to yeast UBC4/5	UBE2D2	AI310002	Above
80	38617_at	LIM domain kinase 2	LIMK2	D45906	Below
81	34305_at	poly rC binding protein 1	PCBP1	Z29505	Above
82	40436_g_at	solute carrier family 25 mitochondrial carrier adenine nucleotide translocator member 6	SLC25A6	J03592	Above
83	1827_s_at	c-myc-P64 mRNA, initiating from promoter P0		M13929	Above
84	38479_at	acidic protein rich in leucines	SSP29	Y07969	Below
85	33207_at	DnaJ Hsp40 homolog subfamily C member 3	DNAJC3	AI095508	Below
86	39039_s_at	CGI-76 protein	LOC51632	AI557497	Below
87	32157_at	protein phosphatase 1 catalytic subunit alpha isoform	PPP1CA	S57501	Above
88	905_at	guanylate kinase 1	GUK1	L76200	Below
89	35794_at	KIAA0942 protein	KIAA0942	AB023159	Below
90	1007_s_at	discoidin domain receptor family member 1	DDR1	U48705	Below
91	39424_at	tumor necrosis factor receptor superfamily member 14 herpesvirus entry mediator	TNFRSF14	U70321	Below
92	36634_at	BTG family member 2	BTG2	U72649	Below
93	38760_f_at	butyrophilin subfamily 3 member A2	BTN3A2	U90546	Below

**Novel**

1	37960_at	carbohydrate chondroitin 6/keratan sulfotransferase 2	CHST2	AB014679	Above
2	31892_at	protein tyrosine phosphatase receptor type M	PTPRM	X58288	Above
3	994_at	protein tyrosine phosphatase receptor type M	PTPRM	X58288	Above
4	995_g_at	protein tyrosine phosphatase receptor type M	PTPRM	X58288	Above
5	41074_at	G protein-coupled receptor 49	GPR49	AF062006	Above
6	41073_at	G protein-coupled receptor 49	GPR49	AI743745	Above
7	34676_at	KIAA1099 protein	KIAA1099	AB029022	Above
8	36139_at	DKFZP586G0522 protein	DKFZP586G0522	AL050289	Above
9	37542_at	lipoma HMGIC fusion partner-like 2	LHFPL2	D86961	Above
10	41159_at	clathrin heavy polypeptide Hc	CLTC	D21260	Above
11	32800_at	retinoid X receptor alpha mRNA		U66306	Above
12	1664_at	insulin-like growth factor 2	IGF2	HG3543-HT3739	Above
13	36566_at	cystinosis nephropathic	CTNS	AJ222967	Above

**T-ALL**

1	38319_at	CD3D antigen delta polypeptide TiT3 complex	CD3D	AA919102	Above
---	----------	---	------	----------	-------

**TEL-AML1**

1	38652_at	hypothetical protein FLJ20154	FLJ20154	AF070644	Above
2	36239_at	POU domain class 2 associating factor 1	POU2AF1	Z49194	Above
3	41442_at	core-binding factor runt domain alpha subunit 2 translocated to 3	CBFA2T3	AB010419	Above
4	37780_at	piccolo presynaptic cytomatrix protein	PCLO	AB011131	Above
5	36985_at	isopenentenyl-diphosphate delta isomerase	IDI1	X17025	Above
6	38578_at	tumor necrosis factor receptor superfamily member 7	TNFRSF7	M63928	Above
7	35614_at	transcription factor-like 5 basic helix-loop-helix	TCFL5	AB012124	Above
8	32224_at	KIAA0769 gene product	KIAA0769	AB018312	Above
9	32730_at	KIAA1750 protein		AL080059	Above
10	36937_s_at	PDZ and LIM domain 1 elfin	PDLIM1	U90878	Below
11	36008_at	protein tyrosine phosphatase type IVA member 3	PTP4A3	AF041434	Above
12	41200_at	CD36 antigen collagen type I receptor thrombospondin receptor like 1	CD36L1	Z22555	Above
13	33690_at	DKFZp434A202 from clone DKFZp434A202		AL080190	Above
14	755_at	inositol 1 4 5-triphosphate receptor type 1	ITPR1	D26070	Above
15	41097_at	telomeric repeat binding factor 2	TERF2	AF002999	Above
16	160029_at	protein kinase C beta 1	PRKCB1	X07109	Above
17	34481_at	vav proto-oncogene	Vav	AF030227	Above
18	41498_at	KIAA0911 protein	KIAA0911	AB020718	Above

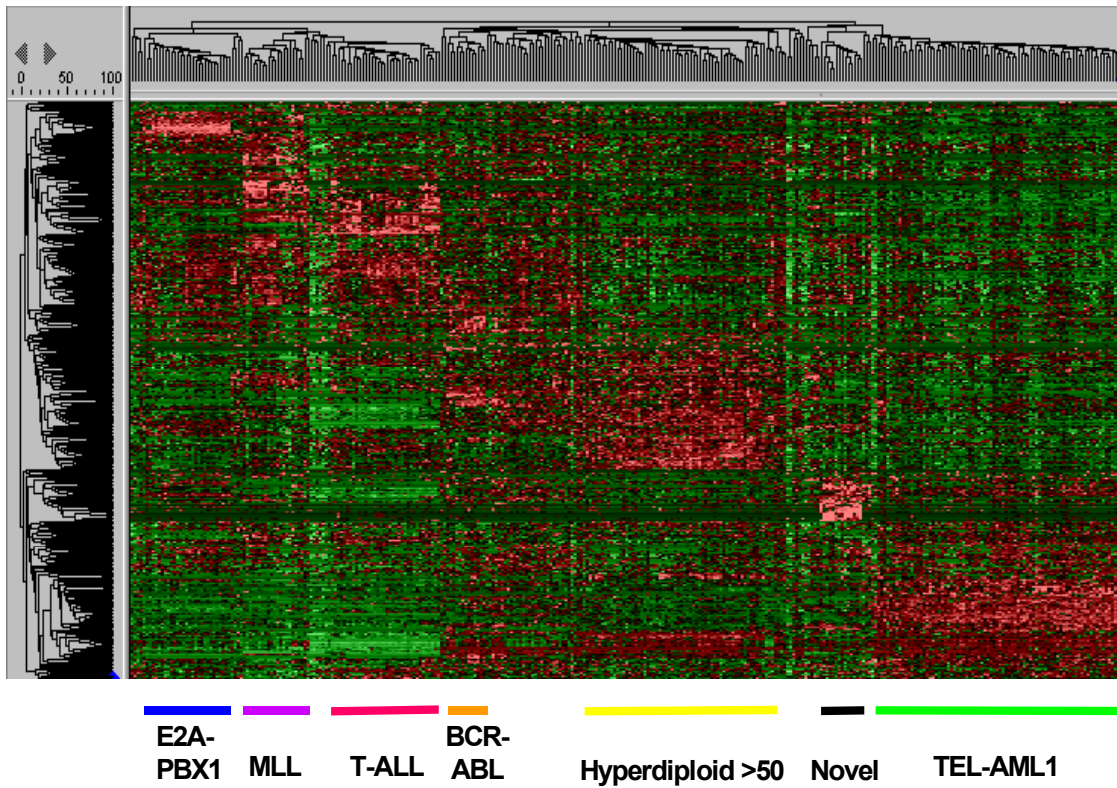
19	37280_at	MAD mothers against decapentaplegic Drosophila homolog 1	MADH1	U59912	Above
20	1647_at	IQ motif containing GTPase activating protein 2	IQGAP2	U51903	Below
21	37724_at	v-myc avian myelocytomatosis viral oncogene homolog	MYC	V00568	Below
22	37981_at	drebrin 1	DBN1	U00802	Above
23	37326_at	proteolipid protein 2 colonic epithelium-enriched	PLP2	U93305	Below
24	37344_at	major histocompatibility complex class II DM alpha	HLA-DMA	X62744	Above
25	38666_at	pleckstrin homology Sec7 and coiled/coil domains 1 cytohesin 1	PSCD1	M85169	Below
26	39039_s_at	CGI-76 protein	LOC51632	AI557497	Below
27	34819_at	CD164 antigen sialomucin	CD164	D14043	Below
28	40729_s_at	nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor-like 1	NFKBIL1	Y14768	Above
29	34224_at	fatty acid desaturase 3	FADS3	AC004770	Above
30	39827_at	hypothetical protein	FLJ20500	AA522530	Below
31	32157_at	protein phosphatase 1 catalytic subunit alpha isoform	PPP1CA	S57501	Below
32	34183_at	DKFZP434C171 protein	DKFZP434C171	AL080169	Below
33	39329_at	actinin alpha 1	ACTN1	X15804	Below
34	38124_at	midkine neurite growth-promoting factor 2	MDK	X55110	Above
35	33304_at	interferon stimulated gene 20kD	ISG20	U88964	Above
36	41295_at	GTT1 protein	GTT1	AL041780	Below
37	40745_at	adaptor-related protein complex 1 beta 1 subunit	AP1B1	L13939	Above
38	38906_at	spectrin alpha erythrocytic 1 elliptocytosis 2	SPTA1	M61877	Above
39	263_g_at	S-adenosylmethionine decarboxylase 1	AMD1	M21154	Below
40	41609_at	major histocompatibility complex class II DM beta	HLA-DMB	U15085	Above
41	39045_at	hypothetical protein FLJ21432	FLJ21432	W26655	Below
42	39421_at	runt-related transcription factor 1 acute myeloid leukemia 1 aml1 oncogene	RUNX1	D43969	Above
43	34210_at	CDW52 antigen CAMPATH-1 antigen	CDW52	N90866	Above
44	37276_at	IQ motif containing GTPase activating protein 2	IQGAP2	U51903	Below
45	38763_at	L-iditol-2 dehydrogenase gene		L29254	Below
46	40960_at	UDP-Gal betaGlcNAc beta 1 4-galactosyltransferase polypeptide 1	B4GALT1	D29805	Below
47	1127_at	ribosomal protein S6 kinase 90kD polypeptide 1	RPS6KA1	L07597	Below
48	37359_at	KIAA0102 gene product	KIAA0102	D14658	Below
49	38968_at	SH3-domain binding protein 5 BTK-associated	SH3BP5	AB005047	Below

50	39135_at	KIAA0767 protein	KIAA0767	AB018310	Below
51	36128_at	transmembrane trafficking protein	TMP21	L40397	Below
52	1158_s_at	calmodulin 3 phosphorylase kinase delta	CALM3	J04046	Above
53	34782_at	jumonji mouse homolog	JMJ	AL021938	Below
54	37893_at	protein tyrosine phosphatase non-receptor type 2	PTPN2	AI828880	Below
55	39758_f_at	Lysosomal-associated membrane protein 1	LAMP1	J04182	Below
56	35151_at	tumor suppressor deleted in oral cancer-related 1	DOC-1R	AF089814	Below
57	38096_f_at	major histocompatibility complex class II DP beta 1	HLA-DPB1	M83664	Above
58	40467_at	succinate dehydrogenase complex subunit D integral membrane protein	SDHD	AB006202	Below
59	39712_at	S100 calcium-binding protein A13	S100A13	AI541308	Below
60	41812_s_at	KIAA0906 protein	KIAA0906	AB020713	Below
61	34336_at	lysyl-tRNA synthetase	KARS	D32053	Below
62	38336_at	KIAA1013 protein	KIAA1013	AB023230	Below
63	32253_at	arginine-glutamic acid dipeptide RE repeats	RERE	AB007927	Below
64	35731_at	integrin alpha 4 antigen CD49D alpha 4 subunit of VLA-4 receptor	ITGA4	X16983	Below
65	40698_at	C-type calcium dependent carbohydrate-recognition domain lectin superfamily member 2 activation-induced	CLECSF2	X96719	Below
66	840_at	zinc finger protein 220	ZNF220	U47742	Above
67	41171_at	proteasome prosome macropain activator subunit 2 PA28 beta	PSME2	D45248	Above
68	34877_at	Janus kinase 1 a protein tyrosine kinase	JAK1	AL039831	Above
69	37190_at	WAS protein family member 1	WASF1	D87459	Below
70	31690_at	Glutamate dehydrogenase-2	GLUD2	U08997	Below
71	40961_at	SWI/SNF related matrix associated actin dependent regulator of chromatin subfamily a member 2	SMARCA2	X72889	Below
72	38149_at	KIAA0053 gene product	KIAA0053	D29642	Above
73	2061_at	integrin alpha 4 antigen CD49D alpha 4 subunit of VLA-4 receptor	ITGA4	L12002	Below
74	2012_s_at	protein kinase DNA-activated catalytic polypeptide	PRKDC	U34994	Below
75	36878_f_at	major histocompatibility complex class II DQ beta 1	HLA-DQB1	M60028	Above
76	34821_at	DKFZP586D0623 protein	DKFZP586D0623	AL050197	Below
77	36980_at	proline-rich protein with nuclear targeting signal	B4-2	U03105	Below
78	853_at	nuclear factor erythroid-derived 2 like 2	NFE2L2	S74017	Below
79	39320_at	caspase 1 apoptosis-related cysteine protease interleukin 1 beta convertase	CASP1	U13697	Below

80	32572_at	ubiquitin specific protease 9 X chromosome Drosophila fat facets related	USP9X	X98296	Below
81	387_at	cyclin-dependent kinase 9 CDC2-related kinase	CDK9	X80230	Below
82	35300_at	glutamyl-prolyl-tRNA synthetase	EPRS	X54326	Below
83	36155_at	KIAA0275 gene product	KIAA0275	D87465	Below
84	37625_at	Interferon regulatory factor 4	IRF4	U52682	Below
85	35763_at	KIAA0540 protein	KIAA0540	AB011112	Below
86	39077_at	DR1-associated protein 1 negative cofactor 2 alpha	DRAP1	U41843	Below
87	40132_g_at	Follistatin-like 1	FSTL1	D89937	Below
88	32615_at	aspartyl-tRNA synthetase	DARS	J05032	Below
89	38357_at	Homo sapiens mRNA cDNA DKFZp564D156 from clone DKFZp564D156		AL049321	Above
90	34817_s_at	ataxin 2 related protein	A2LP	U70671	Above
91	40856_at	serine or cysteine proteinase inhibitor clade F alpha-2 antiplasmin pigment epithelium derived factor member 1	SERPINF1	U29953	Below
92	39784_at	eukaryotic translation initiation factor 2 subunit 1 alpha 35kD	EIF2S1	U26032	Below
93	37600_at	extracellular matrix protein 1	ECM1	U68186	Below
94	40839_at	ubiquitin-like 3	UBL3	AL080177	Below
95	34832_s_at	KIAA0763 gene product	KIAA0763	AB018306	Below
96	33244_at	chimerin chimaerin 2	CHN2	U07223	Below
97	31516_f_at	basic transcription factor 3 like 1	BTF3L1	M90354	Below
98	35266_at	bladder cancer associated protein	BLCAP	AL049288	Above
99	253_g_at	(clone GPCR W) G protein-linked receptor gene (GPCR) gene		L42324	Below
100	35227_at	retinoblastoma-binding protein 8	RBBP8	U72066	Below
101	41073_at	G protein-coupled receptor 49	GPR49	AI743745	Below
102	38084_at	chromobox homolog 3 Drosophila HP1 gamma	CBX3	AI797801	Below
103	39025_at	6.2 kd protein	LOC54543	AI557912	Below
104	32085_at	KIAA0981 protein	KIAA0981	AB023198	Above
105	38902_r_at	Activating transcription factor 2	ATF2	X15875	Below

Illustrated below are the results of a two-dimensional hierarchical clustering algorithm of the 327 diagnostic ALL cases using the top 362 probe sets selected by the CFS metric. This represents 345 unique probe sets as some probe sets were selected for more than one diagnostic group.





**Figure 15.** Hierarchical cluster of 327 Diagnostic ALL samples with genes chosen by the CFS metric

### T-statistics

T-statistics is a classical feature selection approach. The t-statistics of a gene is defined as  $T = \frac{|\mu_1 - \mu_2|}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$ , where  $\mu_i$  is the mean expression of that gene in the  $i^{\text{th}}$  class,  $\sigma_i^2$  is the variance of that gene in the  $i^{\text{th}}$  class and  $n_i$  is the size of the  $i^{\text{th}}$  class. This formula assigns higher value to a gene that has larger mean difference between two classes and has smaller variance within both classes. Data was log transformed prior to the application of the metric. The 40 top ranked genes for each diagnostic group are listed in Table 13. Generally, using the top 20-40 genes did not result in significant changes to subtype prediction accuracy. So we used only the top 20 genes for subtype prediction, unless noted otherwise.

**Table 13. Genes Selected by T statistics**

	Affymetrix number	Gene Name	Gene Symbol	Reference number	T-stat value	Above/ Below Mean
<b><i>BCR-ABL</i></b>						
1	32319_at	tumor necrosis factor ligand superfamily member 4 tax-transcriptionally activated glycoprotein 1 34kD	TNFSF4	AL022310	12.0346	Above
2	36194_at	low density lipoprotein-related protein-associated protein 1 alpha-2-macroglobulin receptor-associated protein 1	LRPAP1	M63959	-11.3077	Below
3	1211_s_at	CASP2 and RIPK1 domain containing adaptor with death domain	CRADD	U84388	10.6627	Above

4	37397_at	Homo sapiens platelet/endothelial cell adhesion molecule-1 (PECAM-1) gene, exon 16 and complete cds.	PECAM	L34657	10.2460	Above
5	330_s_at	tubulin, alpha 1, isoform 44	TUBA1	HG2259-HT2348	10.0540	Above
6	33774_at	caspase 8 apoptosis-related cysteine protease	CASP8	X98172	9.9147	Above
7	202_at	heat shock transcription factor 2	HSF2	M65217	-9.7639	Below
8	1558_g_at	p21/Cdc42/Rac1-activated kinase 1 yeast Ste20-related	PAK1	U24152	9.6562	Above
9	39691_at	SH3-containing protein SH3GLB1	SH3GLB1	AB007960	9.5307	Above
10	2045_s_at	hemopoietic cell kinase	HCK	M16592	-9.3898	Below
11	36591_at	tubulin alpha 1 testis specific	TUBA1	X06956	9.3382	Above
12	1386_at	protein tyrosine phosphatase non-receptor type 9	PTPN9	M83738	-9.2414	Below
13	35991_at	Sm protein F	LSM6	AA917945	9.0298	Above
14	41273_at	FK506 binding protein 12-rapamycin associated protein 1	FRAP1	AL046940	8.9732	Above
15	35970_g_at	M-phase phosphoprotein 9	MPHOSPH9	N23137	8.6474	Above
16	38636_at	immunoglobulin superfamily containing leucine-rich repeat	ISLR	AB003184	8.4291	Above
17	36683_at	matrix Gla protein	MGP	AI953789	-8.3872	Below
18	39070_at	singed Drosophila like sea urchin fascin homolog like	SNL	U03057	8.2583	Above
19	40798_s_at	a disintegrin and metalloproteinase domain 10	ADAM10	Z48579	8.2283	Above
20	41649_at	FOXJ2 forkhead factor	LOC55810	AF038177	8.2275	Above
21	38966_at	glycoprotein synaptic 2	GPSN2	AF038958	8.2080	Above
22	34759_at	Human hbc647 mRNA sequence		U68494	8.1863	Above
23	1434_at	phosphatase and tensin homolog mutated in multiple advanced cancers 1	PTEN	U92436	8.1671	Above
24	40167_s_at	CS box-containing WD protein	LOC55884	AF038187	8.1655	Above
25	40264_g_at	zinc finger protein-like 1	ZFPL1	AF001891	8.1384	Above
26	36129_at	KIAA0397 gene product	KIAA0397	AB007857	8.0041	Above
27	551_at	E1A binding protein p300	EP300	U01877	-7.7578	Below
28	38345_at	centrosomal protein 1	CEP1	AF083322	-7.7431	Below
29	41137_at	myosin phosphatase target subunit 2	MYPT2	AB007972	-7.7301	Below
30	39068_at	protein phosphatase 2 regulatory subunit B B56 delta isoform	PPP2R5D	L76702	-7.6161	Below
31	38160_at	lymphocyte antigen 75	LY75	AF011333	7.5830	Above
32	34314_at	ribonucleotide reductase M1 polypeptide	RRM1	X59543	7.5778	Above
33	39519_at	KIAA0692 protein	KIAA0692	AB014592	7.4662	Above
34	32788_at	RAN binding protein 2	RANBP2	D42063	7.4114	Above
35	34882_at	nucleolar protein KKE/D repeat	NOP56	Y12065	7.3622	Above
36	2064_g_at	excision repair cross-complementing rodent repair deficiency	ERCC5	L20046	7.3597	Above

37	41836_at	complementation group 5 protein with polyglutamine repeat calcium ca2 homeostasis endoplasmic reticulum protein	ERPROT21	U94836	7.3350	Above
38	1563_s_at	tumor necrosis factor receptor superfamily member 1A	TNFRSF1A	M58286	7.3039	Above
39	37047_at	Niemann-Pick disease type C1	NPC1	AF002020	7.2357	Above
40	32724_at	phytanoyl-CoA hydroxylase Refsum disease	PHYH	AF023462	-7.2252	Below

### ***E2A-PBX1***

1	32063_at	pre-B-cell leukemia transcription factor 1	PBX1	M86546	126.7442	Above
2	33355_at	Homo sapiens cDNA FLJ12900 fis clone NT2RP2004321 (by CELERA search of target sequence = PBX1)	PBX1	AL049381	36.6116	Above
3	40454_at	FAT tumor suppressor Drosophila homolog	FAT	X87241	30.7577	Above
4	717_at	GS3955 protein	GS3955	D87119	23.7813	Above
5	39070_at	singed Drosophila like sea urchin fascin homolog like	SNL	U03057	-22.8956	Below
6	33641_g_at	nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor-like 1	NFKBIL1	Y14768	-20.4637	Below
7	36536_at	schwannomin interacting protein 1	SCHIP-1	AF070614	-20.1554	Below
8	854_at	B lymphoid tyrosine kinase	BLK	S76617	19.6467	Above
9	37625_at	interferon regulatory factor 4	IRF4	U52682	18.8419	Above
10	39614_at	KIAA0802 protein	KIAA0802	AB018345	17.8214	Above
11	37099_at	arachidonate 5-lipoxygenase-activating protein	ALOX5AP	AI806222	-17.7944	Below
12	38994_at	STAT induced STAT inhibitor-2	STATI2	AF037989	-17.6553	Below
13	37641_at	Human gene for hepatitis C-associated microtubular aggregate protein p44, exon 9 and complete cds.		D28915	-17.3074	Below
14	40113_at	GS3955 protein	GS3955	D87119	16.7288	Above
15	2031_s_at	cyclin-dependent kinase inhibitor 1A p21 Cip1	CDKN1A	U03106	-14.9826	Below
16	330_s_at	tubulin, alpha 1, isoform 44	TUBA1	HG2259-HT2348	-14.8016	Below
17	38340_at	huntingtin interacting protein-1-related	KIAA0655	AB014555	14.7180	Above
18	38510_at	Homo sapiens mRNA cDNA DKFZp586B0220		AL049435	-14.4522	Below
19	268_at	Homo sapiens platelet/endothelial cell adhesion molecule-1 (PECAM-1) gene, exon 16 and complete cds.	PECAM	L34657	-13.7540	Below
20	2062_at	insulin-like growth factor binding protein 7	IGFBP7	L19182	13.6403	Above

21	37893_at	protein tyrosine phosphatase non-receptor type 2	PTPN2	AI828880	13.5099	Above
22	38580_at	guanine nucleotide binding protein G protein q polypeptide	GNAQ	U43083	-12.8525	Below
23	40049_at	death-associated protein kinase 1	DAPK1	X76104	-12.3837	Below
24	38393_at	KIAA0247 gene product	KIAA0247	D87434	12.3436	Above
25	39379_at	Homo sapiens mRNA cDNA DKFZp586C1019		AL049397	12.2102	Above
26	430_at	nucleoside phosphorylase	NP	X00737	12.1307	Above
27	37975_at	cytochrome b-245 beta polypeptide chronic granulomatous disease	CYBB	X04011	-12.0743	Below
28	34862_at	CGI-49 protein	LOC51097	AA005018	12.0264	Above
29	39756_g_at	X-box binding protein 1	XBP1	Z93930	-11.9796	Below
30	307_at	arachidonate 5-lipoxygenase	ALOX5	J03600	-11.9492	Below
31	37304_at	chromobox homolog 1 Drosophila HP1 beta	CBX1	U35451	11.9422	Above
32	1287_at	ADP-ribosyltransferase NAD poly ADP-ribose polymerase	ADPRT	J03473	11.9051	Above
33	1520_s_at	interleukin 1 beta	IL1B	X04500	11.7327	Above
34	596_s_at	colony stimulating factor 3 receptor granulocyte	CSF3R	M59820	-11.6814	Below
35	37493_at	colony stimulating factor 2 receptor beta low-affinity granulocyte-macrophage	CSF2RB	H04668	11.6620	Above
36	36452_at	synaptopodin	KIAA1029	AB028952	11.4021	Above
37	1081_at	ornithine decarboxylase 1	ODC1	M33764	11.2865	Above
38	1563_s_at	tumor necrosis factor receptor superfamily member 1A	TNFRSF1A	M58286	-11.1361	Below
39	39069_at	AE-binding protein 1	AEBP1	AF053944	11.0984	Above
40	36203_at	ornithine decarboxylase 1	ODC1	X16277	10.9475	Above

### Hyperdiploid >50

1	36620_at	superoxide dismutase 1 soluble amyotrophic lateral sclerosis 1 adult	SOD1	X02317	9.1574	Above
2	39878_at	protocadherin 9	PCDH9	AI524125	-6.9008	Below
3	37543_at	Rac/Cdc42 guanine exchange factor GEF 6	ARHGEF6	D25304	6.8366	Above
4	41470_at	prominin mouse like 1	PROML1	AF027208	6.7290	Above
5	31492_at	muscle specific gene	M9	AB019392	-6.6885	Below
6	38968_at	SH3-domain binding protein 5 BTK-associated	SH3BP5	AB005047	6.4051	Above
7	1915_s_at	v-fos FBJ murine osteosarcoma viral oncogene homolog	FOS	V01512	6.4008	Above
8	37677_at	phosphoglycerate kinase 1	PGK1	V00572	6.2865	Above
9	39867_at	Tu translation elongation factor mitochondrial	TUFM	S75463	-6.2299	Below
10	36795_at	prosaposin variant Gaucher disease and variant metachromatic leukodystrophy	PSAP	J03077	6.1812	Above

11	40875_s_at	small nuclear ribonucleoprotein 70kD polypeptide RNP antigen	SNRP70	X06815	-6.0877	Below
12	306_s_at	high-mobility group nonhistone chromosomal protein 14	HMG14	J02621	6.0804	Above
13	41724_at	accessory proteins BAP31/BAP29	DXS1357E	X81109	6.0244	Above
14	39168_at	Ac-like transposable element	ALTE	AB018328	5.9336	Above
15	955_at	calmodulin type I	CALM1	HG1862-HT1897	5.8650	Above
16	38604_at	neuropeptide Y	NPY	AI198311	5.8313	Above
17	39147_g_at	alpha thalassemia/mental retardation syndrome X-linked RAD54 <i>S. cerevisiae</i> homolog	ATRX	U72936	5.8181	Above
18	39069_at	AE-binding protein 1	AEBP1	AF053944	-5.6901	Below
19	37014_at	myxovirus influenza resistance 1 homolog of murine interferon-inducible protein p78	MX1	M33882	5.6688	Above
20	1520_s_at	interleukin 1 beta	IL1B	X04500	5.6605	Above
21	1488_at	protein tyrosine phosphatase receptor type K	PTPRK	L77886	-5.5877	Below
22	32553_at	MYC-associated zinc finger protein purine-binding transcription factor	MAZ	M94046	-5.5000	Below
23	36169_at	NADH dehydrogenase ubiquinone 1 alpha subcomplex 1 7.5kD MWFE	NDUFA1	N47307	5.4376	Above
24	1817_at	prefoldin 5	PFDN5	D89667	-5.4110	Below
25	578_at	Human recombination activating protein (RAG2) gene, last exon	RAG2	M94633	-5.4026	Below
26	1556_at	RNA binding motif protein 5	RBM5	U23946	-5.3032	Below
27	40998_at	trinucleotide repeat containing 11 THR-associated protein 230 kDa subunit	TNRC11	AF071309	5.2349	Above
28	37294_at	B-cell translocation gene 1 anti-proliferative	BTG1	X61123	-5.1877	Below
29	1447_at	proteasome prosome macropain subunit beta type 1	PSMB1	D00761	5.1699	Above
30	35940_at	POU domain class 4 transcription factor 1	POU4F1	X64624	5.1200	Above
31	33307_at	kraken-like	BK126B4.1	AL022316	-5.0984	Below
32	1081_at	ornithine decarboxylase 1	ODC1	M33764	-5.0822	Below
33	34336_at	lysyl-tRNA synthetase	KARS	D32053	-5.0692	Below
34	41143_at	Human calmodulin (CALM1) gene, exons 2,3,4,5 and 6, and complete cds	CALM1	U12022	5.0543	Above
35	32251_at	hypothetical protein FLJ21174	FLJ21174	AA149307	5.0373	Above
36	35298_at	eukaryotic translation initiation factor 3 subunit 7 zeta 66/67kD	EIF3S7	U54558	-4.9499	Below
37	38649_at	KIAA0970 protein	KIAA0970	AB023187	-4.9228	Below
38	36629_at	glucocorticoid-induced leucine zipper	GILZ	AI635895	4.8061	Above
39	39721_at	ephrin-B1	EFNB1	U09303	4.7968	Above

40	2094_s_at	v-fos FBJ murine osteosarcoma viral oncogene homolog	FOS	K00650	4.7446	Above
<b>MLL</b>						
1	307_at	arachidonate 5-lipoxygenase	ALOX5	J03600	-16.8244	Below
2	37280_at	MAD mothers against decapentaplegic Drosophila homolog 1	MADH1	U59912	-15.4460	Below
3	1520_s_at	interleukin 1 beta	IL1B	X04500	-13.6764	Below
4	36908_at	Human macrophage mannose receptor (MRC1) gene, exon 30.	MRC1	M93221	-11.8629	Below
5	33412_at	LGALS1 Lectin, galactoside-binding, soluble, 1 (galectin 1)	LGALS1	AI535946	11.0223	Above
6	2062_at	insulin-like growth factor binding protein 7	IGFBP7	L19182	10.4318	Above
7	35940_at	POU domain class 4 transcription factor 1	POU4F1	X64624	-10.1815	Below
8	39721_at	ephrin-B1	EFNB1	U09303	-9.6158	Below
9	39402_at	interleukin 1 beta	IL1B	M15330	-9.5998	Below
10	1737_s_at	insulin-like growth factor-binding protein 4	IGFBP4	M62403	-9.4119	Below
11	37413_at	dipeptidase 1 renal	DPEP1	J05257	-9.4101	Below
12	40519_at	protein tyrosine phosphatase receptor type C	PTPRC	Y00638	9.3163	Above
13	1971_g_at	fragile histidine triad gene	FHIT	U46922	-9.2257	Below
14	1983_at	cyclin D2	CCND2	X68452	-9.2213	Below
15	38869_at	KIAA1069 protein	KIAA1069	AB028992	-9.1951	Below
16	40520_g_at	protein tyrosine phosphatase receptor type C	PTPRC	Y00638	9.1099	Above
17	1718_at	actin related protein 2/3 complex subunit 2 34 kD	ARPC2	U50523	9.0435	Above
18	34237_at	HBS1 S. cerevisiae like	HBS1L	AB028961	-8.8208	Below
19	1726_at	DNA polymerase, epsilon, catalytic subunit		HG919-HT919	-8.4664	Below
20	36643_at	discoidin domain receptor family member 1	DDR1	L20817	-8.4627	Below
21	1325_at	MAD mothers against decapentaplegic Drosophila homolog 1	MADH1	U59423	-8.3762	Below
22	39379_at	Homo sapiens mRNA cDNA DKFZp586C1019		AL049397	8.2974	Above
23	36536_at	schwannomin interacting protein 1	SCHIP-1	AF070614	-8.1177	Below
24	564_at	guanine nucleotide binding protein G protein alpha 11 Gq class	GNA11	M69013	-8.1107	Below
25	39705_at	KIAA0700 protein	KIAA0700	AB014600	-7.9334	Below
26	36105_at	Human nonspecific crossreacting antigen mRNA, complete cds.	NCA	M18728	-7.6911	Below
27	174_s_at	intersectin 2	ITSN2	U61167	7.5752	Above
28	39114_at	decidual protein induced by progesterone	DEPP	AB022718	-7.4767	Below

29	40436_g_at	solute carrier family 25 mitochondrial carrier adenine nucleotide translocator member 6	SLC25A6	J03592	7.3952	Above
30	794_at	protein tyrosine phosphatase non-receptor type 6	PTPN6	X62055	7.2192	Above
31	38032_at	KIAA0736 gene product	KIAA0736	AB018279	-7.0718	Below
32	40518_at	protein tyrosine phosphatase receptor type C	PTPRC	Y00062	6.9829	Above
33	41762_at	TIA1 cytotoxic granule-associated RNA-binding protein-like 1	TIAL1	D64015	-6.9118	Below
34	1389_at	membrane metallo-endopeptidase neutral endopeptidase enkephalinase CALLA CD10	MME	J03779	-6.7734	Below
35	39967_at	leucine zipper down-regulated in cancer 1	LDOC1	AB019527	-6.7415	Below
36	188_at	ephrin-B1	EFNB1	U09303	-6.5964	Below
37	160033_s_at	X-ray repair complementing defective repair in Chinese hamster cells 1	XRCC1	NM_006297	-6.5936	Below
38	40913_at	ATPase Ca transporting plasma membrane 4	ATP2B4	W28589	-6.5774	Below
39	37398_at	platelet/endothelial cell adhesion molecule CD31 antigen	PECAM1	AA100961	-6.5675	Below
40	1488_at	protein tyrosine phosphatase receptor type K	PTPRK	L77886	-6.5584	Below

### Novel

1	41734_at	KIAA0870 protein	KIAA0870	AB020677	-40.5168	Below
2	31892_at	protein tyrosine phosphatase receptor type M	PTPRM	X58288	33.4654	Above
3	995_g_at	protein tyrosine phosphatase receptor type M	PTPRM	X58288	24.7557	Above
4	34676_at	KIAA1099 protein	KIAA1099	AB029022	14.0491	Above
5	37908_at	guanine nucleotide binding protein 11	GNG11	U31384	11.4548	Above
6	37960_at	carbohydrate chondroitin 6/keratan sulfotransferase 2	CHST2	AB014679	10.9971	Above
7	33410_at	integrin alpha 6	ITGA6	S66213	10.0370	Above
8	40585_at	adenylate cyclase 7	ADCY7	D25538	-9.5897	Below
9	33284_at	myeloperoxidase	MPO	M19507	-9.4724	Below
10	41159_at	clathrin heavy polypeptide Hc	CLTC	D21260	9.4489	Above
11	36591_at	tubulin alpha 1 testis specific	TUBA1	X06956	-9.1387	Below
12	37712_g_at	MADS box transcription enhancer factor 2 polypeptide C myocyte enhancer factor 2C	MEF2C	S57212	-9.1225	Below
13	38576_at	H2B histone family member B	H2BFB	AJ223353	-9.0869	Below
14	38408_at	transmembrane 4 superfamily member 2	TM4SF2	L10373	-8.7026	Below
15	33907_at	eukaryotic translation initiation factor 4 gamma 3	EIF4G3	AF012072	-8.3540	Below

16	41273_at	FK506 binding protein 12-rapamycin associated protein 1	FRAP1	AL046940	-8.3212	Below
17	402_s_at	intercellular adhesion molecule 3	ICAM3	X69819	-7.9741	Below
18	35112_at	regulator of G-protein signalling 9	RGS9	AF071476	7.8348	Above
19	34850_at	ubiquitin-conjugating enzyme E2E 3 homologous to yeast UBC4/5	UBE2E3	AB017644	7.8197	Above
20	37030_at	KIAA0887 protein	KIAA0887	AB020694	-7.6343	Below
21	36322_at	fucosyltransferase 7 alpha 1 3 fucosyltransferase	FUT7	AB012668	-7.6240	Below
22	39509_at	Homo sapiens cDNA FLJ22071		AI692348	-7.6232	Below
23	40091_at	B-cell CLL/lymphoma 6 zinc finger protein 51	BCL6	U00115	-7.6171	Below
24	37280_at	MAD mothers against decapentaplegic Drosophila homolog 1	MADH1	U59912	7.5991	Above
25	1325_at	MAD mothers against decapentaplegic Drosophila homolog 1	MADH1	U59423	7.5824	Above
26	831_at	DEAD/H Asp-Glu-Ala-Asp/His box polypeptide 10 RNA helicase	DDX10	U28042	7.4276	Above
27	37600_at	extracellular matrix protein 1	ECM1	U68186	-7.2991	Below
28	41266_at	integrin alpha 6	ITGA6	X53586	7.2985	Above
29	36958_at	zyxin	ZYX	X95735	-7.2889	Below
30	36564_at	Human DNA sequence from clone RP5-1174N9 on chromosome 1p34.1-35.3		W27419	-7.2848	Below
31	32174_at	solute carrier family 9 sodium/hydrogen exchanger isoform 3 regulatory factor 1	SLC9A3R1	AF015926	-7.2749	Below
32	619_s_at	membrane-spanning 4-domains subfamily A member 2 Fc fragment of IgE high affinity I receptor for beta polypeptide	MS4A2	M27394	-7.2325	Below
33	40749_at	membrane-spanning 4-domains subfamily A member 2 Fc fragment of IgE high affinity I receptor for beta polypeptide	MS4A2	X07203	-7.2063	Below
34	31894_at	centromere protein C 1	CENPC1	M95724	6.9679	Above
35	32319_at	tumor necrosis factor ligand superfamily member 4 tax-transcriptionally activated glycoprotein 1 34kD	TNFSF4	AL022310	6.8225	Above
36	38259_at	syntaxin binding protein 2	STXBP2	AB002559	-6.6992	Below
37	35629_at	hypothetical protein	DJ1042K10	AL022238	-6.6968	Below
38	38700_at	cysteine and glycine-rich protein 1	CSRP1	M33146	-6.6962	Below
39	37397_at	Homo sapiens platelet/endothelial cell adhesion molecule-1 (PECAM-1) gene, exon 16 and complete cds.	PECAM	L34657	-6.6934	Below
40	41127_at	solute carrier family 1 glutamate/neutral amino acid transporter member 4	SLC1A4	L14595	-6.6892	Below



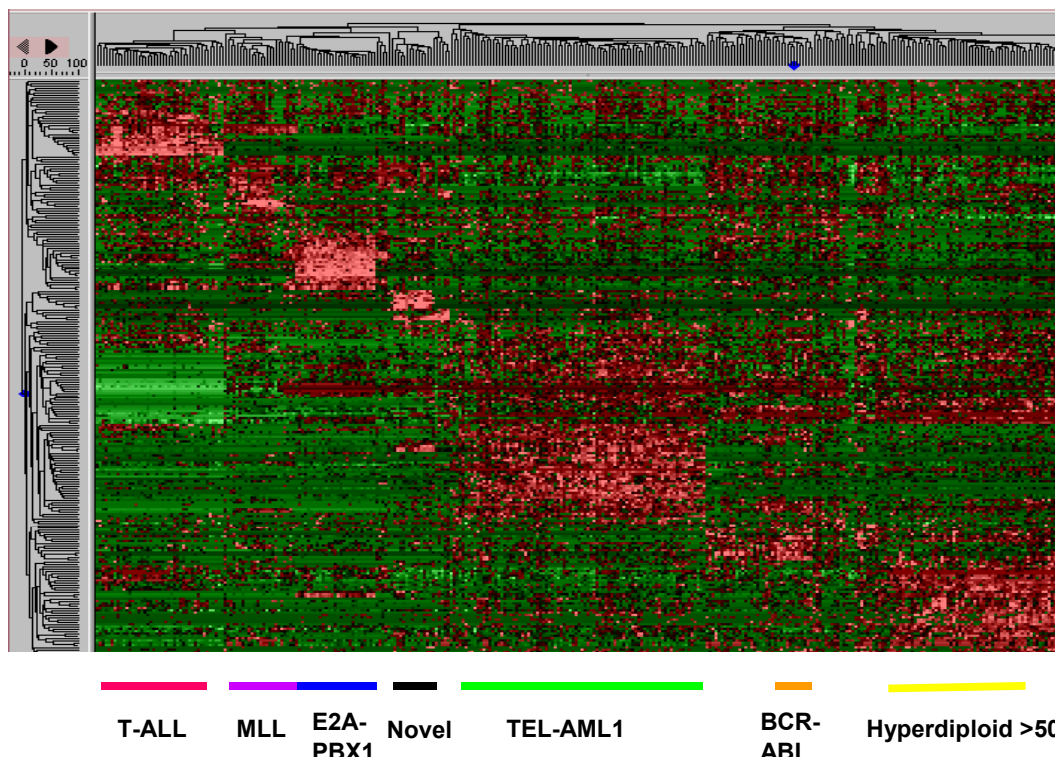
**T -ALL**

1	38242_at	B cell linker protein	SLP65	AF068180	-115.8362	Below
2	38319_at	CD3D antigen delta polypeptide TiT3 complex	CD3D	AA919102	27.6995	Above
3	37988_at	CD79B antigen immunoglobulin-associated beta	CD79B	M89957	-23.7294	Below
4	38147_at	SH2 domain protein 1A Duncan s disease lymphoproliferative syndrome	SH2D1A	AL023657	22.4501	Above
5	38522_s_at	CD22 antigen	CD22	X52785	-21.2795	Below
6	35350_at	B cell RAG associated protein	BRAG	AB011170	-19.1460	Below
7	36277_at	Human membran protein (CD3-epsilon) gene, exon 9.	CD3E	M23323	19.0859	Above
8	38604_at	neuropeptide Y	NPY	AI198311	-18.8194	Below
9	33705_at	phosphodiesterase 4B cAMP-specific dunce Drosophila homolog phosphodiesterase E4	PDE4B	L20971	-18.6383	Below
10	36878_f_at	major histocompatibility complex class II DQ beta 1	HLA-DQB1	M60028	-18.5620	Below
11	36638_at	connective tissue growth factor	CTGF	X78947	-18.2772	Below
12	32794_g_at	T cell receptor beta locus	TRB	X00437	17.9081	Above
13	32174_at	solute carrier family 9 sodium/hydrogen exchanger isoform 3 regulatory factor 1	SLC9A3R1	AF015926	17.4427	Above
14	160041_at	protein tyrosine phosphatase non-receptor type 18 brain-derived	PTPN18	X79568	-17.3412	Below
15	38521_at	CD22 antigen	CD22	X59350	-17.0388	Below
16	38018_g_at	CD79A antigen immunoglobulin-associated alpha	CD79A	U05259	-16.7948	Below
17	36571_at	topoisomerase DNA II beta 180kD	TOP2B	X68060	-16.7508	Below
18	1096_g_at	CD19 antigen	CD19	M28170	-16.4583	Below
19	39318_at	T-cell leukemia/lymphoma 1A	TCL1A	X82240	-16.2017	Below
20	41710_at	hypothetical protein	LOC54103	AL079277	-15.9099	Below
21	599_at	H2.0 Drosophila like homeo box 1	HLX1	M60721	-15.5425	Below
22	266_s_at	CD24 antigen small cell lung carcinoma cluster 4 antigen	CD24	L33930	-15.0123	Below
23	36502_at	PFTAIRE protein kinase 1	PFTK1	AB020641	-14.9972	Below
24	39114_at	decidual protein induced by progesterone	DEPP	AB022718	-14.9886	Below
25	37539_at	RalGDS-like gene KIAA0959 protein	KIAA0959	AB023176	-14.6872	Below
26	40775_at	integral membrane protein 2A	ITM2A	AL021786	14.5666	Above
27	34033_s_at	leukocyte immunoglobulin-like receptor subfamily A with TM domain member 2	LILRA2	AF025531	-14.3809	Below
28	2031_s_at	cyclin-dependent kinase inhibitor 1A p21 Cip1	CDKN1A	U03106	-14.1071	Below
29	38051_at	mal T-cell differentiation protein	MAL	X76220	14.0743	Above
30	35794_at	KIAA0942 protein	KIAA0942	AB023159	-13.9659	Below
31	41156_g_at	catenin cadherin-associated protein alpha 1 102kD	CTNNA1	U03100	-13.8135	Below

32	32979_at	GRB2-associated binding protein 1	GAB1	U43885	-13.5842	Below
33	32562_at	endoglin Osler-Rendu-Weber syndrome 1	ENG	X72012	-13.4209	Below
34	36536_at	schwannomin interacting protein 1	SCHIP-1	AF070614	-13.4172	Below
35	36108_at	major histocompatibility complex class II DQ beta 1	HLA-DQB1	M16276	-13.3518	Below
36	41734_at	KIAA0870 protein	KIAA0870	AB020677	-13.2672	Below
37	41153_f_at	Homo sapiens alphaE-catenin (CTNNA1) gene, exon 18 and complete cds.	CTNNA1	AF102803	-12.7927	Below
38	37710_at	MADS box transcription enhancer factor 2 polypeptide C myocyte enhancer factor 2C	MEF2C	L08895	-12.7716	Below
39	39893_at	guanine nucleotide binding protein G protein gamma 7	GNG7	AB010414	-12.7696	Below
40	37908_at	guanine nucleotide binding protein 11	GNG11	U31384	-12.7353	Below
<b><i>TEL-AML1</i></b>						
1	38578_at	tumor necrosis factor receptor superfamily member 7	TNFRSF7	M63928	15.2209	Above
2	38203_at	potassium intermediate/small conductance calcium-activated channel subfamily N member 1	KCNN1	U69883	15.0804	Above
3	36524_at	Rho guanine nucleotide exchange factor GEF 4	ARHGEF4	AB029035	14.9774	Above
4	37780_at	piccolo presynaptic cytomatrix protein	PCLO	AB011131	14.1405	Above
5	35614_at	transcription factor-like 5 basic helix-loop-helix	TCFL5	AB012124	12.9369	Above
6	160029_at	protein kinase C beta 1	PRKCB1	X07109	12.5429	Above
7	1980_s_at	non-metastatic cells 2 protein NM23B expressed in	NME2	X58965	-12.5035	Below
8	1488_at	protein tyrosine phosphatase receptor type K	PTPRK	L77886	12.3871	Above
9	34194_at	Homo sapiens cDNA FLJ21697		AL049313	12.1089	Above
10	37908_at	guanine nucleotide binding protein 11	GNG11	U31384	11.4322	Above
11	40272_at	collapsin response mediator protein 1	CRMP1	D78012	11.0625	Above
12	41097_at	telomeric repeat binding factor 2	TERF2	AF002999	11.0133	Above
13	33690_at	Homo sapiens mRNA cDNA DKFZp434A202		AL080190	10.8763	Above
14	32730_at	Homo sapiens mRNA for KIAA1750		AL080059	10.7439	Above
15	1325_at	MAD mothers against decapentaplegic Drosophila homolog 1	MADH1	U59423	10.5332	Above
16	41819_at	FYN-binding protein FYB-120/130	FYB	U93049	10.3692	Above
17	1299_at	telomeric repeat binding factor 2	TERF2	X93512	10.2921	Above
18	35665_at	phosphoinositide-3-kinase class 3	PIK3C3	Z46973	10.0568	Above
19	36537_at	Rho-specific guanine nucleotide exchange factor p114	P114-RHO-GEF	AB011093	9.8824	Above

20	37280_at	MAD mothers against decapentaplegic Drosophila homolog 1	MADH1	U59912	9.8662	Above
21	1936_s_at	proto-oncogene c-myc, alt. transcript 3, ORF 114		HG3523-HT4899	-9.6621	Below
22	1077_at	recombination activating gene 1	RAG1	M29474	9.4563	Above
23	38763_at	Human (clone D21-1) L-Iditol-2 dehydrogenase gene, exon 9 and complete cds.		L29254	-9.2719	Below
24	41295_at	GTT1 protein	GTT1	AL041780	-9.1813	Below
25	36008_at	protein tyrosine phosphatase type IVA member 3	PTP4A3	AF041434	9.1682	Above
26	38570_at	major histocompatibility complex class II DO beta	HLA-DOB	X03066	9.0394	Above
27	32163_f_at	EST		AA216639	9.0392	Above
28	40570_at	forkhead box O1A rhabdomyosarcoma	FOXO1A	AF032885	8.9931	Above
29	32724_at	phytanoyl-CoA hydroxylase Refsum disease	PHYH	AF023462	8.9571	Above
30	932_i_at	zinc finger protein 91 HPF7 HTF10	ZNF91	L11672	8.8075	Above
31	37343_at	inositol 1 4 5-triphosphate receptor type 3	ITPR3	U01062	8.7321	Above
32	33447_at	myosin light polypeptide regulatory non-sarcomeric 20kD	MLCB	X54304	-8.6848	Below
33	35362_at	myosin X	MYO10	AB018342	8.6700	Above
34	38906_at	spectrin alpha erythrocytic 1 elliptocytosis 2	SPTA1	M61877	8.5010	Above
35	324_f_at	basic transcription factor 3	BTF3	HG1515-HT1515	-8.4705	Below
36	39329_at	actinin alpha 1	ACTN1	X15804	-8.3219	Below
37	577_at	midkine neurite growth-promoting factor 2	MDK	M94250	8.2693	Above
38	40729_s_at	nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor-like 1	NFKBIL1	Y14768	8.2000	Above
39	41442_at	core-binding factor runt domain alpha subunit 2 translocated to 3	CBFA2T3	AB010419	8.0604	Above
40	36275_at	Homo sapiens mRNA from chromosome 5q21-22 clone FBR89		AB002438	7.8550	Above

Illustrated below are the results of a two-dimensional hierarchical clustering algorithm of the 327 diagnostic ALL cases using the top 280 probe sets selected by the T-statistics method. Some probe sets were chosen for more than one diagnostic group, and thus there are only 248 unique probe sets represented.



**Figure 16.** Hierarchical cluster of 327 Diagnostic ALL samples with genes selected by the T-statistics method.

### Wilkins'

This method of selecting genes uses the weighted sum of three components to estimate the discriminative value of each gene. The higher the score, the better the gene is at discriminating between the two classes. The input to the scoring method is preprocessed and normalized data. The idea of the metric is that a gene is a good discriminator if (1) it is expressed in one class and not in the other, or if the gene is expressed in both classes, but significantly more so in one than the other, or (2) the gene is present in most samples, and the data are pure, in the sense that there is a threshold expression value for the gene where the gene generally has expression levels larger than the threshold in one class, and smaller than the threshold in the other class. The components of the metric were quantified as follows. For a gene, assume  $PR_1$  is the ratio of "present" samples to all samples in class 1, where present means that the gene's expression value was not preprocessed to a constant (1). Assume  $PR_2$  is defined similarly for class 2. The first component of the metric,  $M_1$ , is estimated as the absolute difference between  $PR_1$  and  $PR_2$ . This value is between 0 (when the gene is equally present in both classes) and 1 (when the gene is expressed in one class and not in the other). The second component of the metric,  $M_2$ , measures the extent to which the gene is present overall, and is defined as the average of  $PR_1$  and  $PR_2$ . The final component,  $M_3$ , estimates the "purity", or existence of a threshold value. The gene expression values for the present samples are sorted into ascending order and a vector of their class labels is built, for example  $\{+, +, +, -, -, -, +, -, -, +, -\}$ . The next step is to find the best place to partition the samples so that the expression values for one class (maybe  $+$ ) are less than the partition point, and the values from the other class are larger. Let  $L_{C1}$  and  $L_{C2}$  be the number of class 1 and class 2 samples on the left side of the partition, respectively. Assume  $R_{C1}$  and  $R_{C2}$  are defined similarly for the right side of the partition. Then the purity is estimated as:  $\max\{L_{C1} - L_{C2} + R_{C2} - R_{C1}, L_{C2} - L_{C1} + R_{C1} - R_{C2}\} / \text{number of total present samples}$ . Each possible partition is checked. In the example above, the partition  $\{+, +, +, || -, -, -, +, -, -, +, -\}$  is the best partition, with a purity value of  $M_3 = 7 / 11 = 0.64$ . The score for the gene is the weighted sum of  $0.5 \cdot M_1 + 0.25 \cdot M_2 +$

0.25\*M<sub>3</sub>. The top 50 genes for each subgroup selected by this metric are listed in Table 14. For class prediction all 50 genes were used, unless otherwise stated.

**Table 14. Genes Selected by Wilkins'**

	<b>Affymetrix number</b>	<b>Gene Name</b>	<b>Gene Symbol</b>	<b>Reference number</b>	<b>Train set score</b>	<b>Above/ Below Mean</b>
<b><i>BCR-ABL</i></b>						
1	32319_at	tumor necrosis factor ligand superfamily member 4 tax-transcriptionally activated glycoprotein 1 34kD	TNFSF4	AL022310	0.6354	Above
2	37479_at	CD72 antigen	CD72	M54992	0.6352	Below
3	1211_s_at	CASP2 and RIPK1 domain containing adaptor with death domain	CRADD	U84388	0.6265	Above
4	37397_at	platelet/endothelial cell adhesion molecule-1 (PECAM-1) gene	PECAM	L34657	0.6161	Above
5	33162_at	insulin receptor	INSR	X02160	0.6118	Below
6	39691_at	SH3-containing protein SH3GLB1	SH3GLB1	AB007960	0.6089	Above
7	1558_g_at	p21/Cdc42/Rac1-activated kinase 1 yeast Ste20-related	PAK1	U24152	0.6087	Above
8	34759_at	Human hbc647 mRNA sequence		U68494	0.6061	Above
9	33774_at	caspase 8 apoptosis-related cysteine protease	CASP8	X98172	0.6040	Above
10	1326_at	caspase 10 apoptosis-related cysteine protease	CASP10	U60519	0.6021	Above
11	38312_at	DKFZp564O222 from clone DKFZp564O222		AL050002	0.6010	Above
12	35970_g_at	M-phase phosphoprotein 9	MPHOSPH9	N23137	0.5989	Above
13	41273_at	FK506 binding protein 12-rapamycin associated protein 1	FRAP1	AL046940	0.5989	Above
14	40798_s_at	a disintegrin and metalloproteinase domain 10	ADAM10	Z48579	0.5980	Above
15	40953_at	calponin 3 acidic	CNN3	S80562	0.5972	Above
16	1434_at	phosphatase and tensin homolog mutated in multiple advanced cancers 1	PTEN	U92436	0.5963	Below
17	38966_at	glycoprotein synaptic 2	GPSN2	AF038958	0.5953	Above
18	35991_at	Sm protein F	LSM6	AA917945	0.5938	Above
19	330_s_at	tubulin, alpha 1, isoform 44	TUBA1	HG2259-HT2348	0.5938	Above
20	38032_at	KIAA0736 gene product	KIAA0736	AB018279	0.5934	Above
21	1983_at	cyclin D2	CCND2	X68452	0.5927	Above
22	36194_at	low density lipoprotein-related protein-associated protein 1 alpha-2-macroglobulin receptor-associated protein 1	LRPAP1	M63959	0.5914	Below

23	34460_at	peripheral benzodiazepine receptor-associated protein 1	PRAX-1	AB014512	0.5911	Above
24	2001_g_at	ataxia telangiectasia mutated includes complementation groups A C and D	ATM	U26455	0.5910	Above
25	31443_at	AML1	AML1	S76346	0.5896	Above
26	33410_at	integrin alpha 6	ITGA6	S66213	0.5896	Above
27	37472_at	mannosidase beta A lysosomal	MANBA	U60337	0.5887	Below
28	36099_at	splicing factor arginine/serine-rich 1 splicing factor 2 alternate splicing factor	SFRS1	M69040	0.5877	Below
29	38636_at	immunoglobulin superfamily containing leucine-rich repeat	ISLR	AB003184	0.5858	Above
30	34314_at	ribonucleotide reductase M1 polypeptide	RRM1	X59543	0.5858	Below
31	36129_at	KIAA0397 gene product	KIAA0397	AB007857	0.5858	Above
32	40264_g_at	zinc finger protein-like 1	ZFPL1	AF001891	0.5858	Above
33	37399_at	aldo-keto reductase family 1 member C3 3-alpha hydroxysteroid dehydrogenase type II	AKR1C3	D17793	0.5852	Above
34	38160_at	lymphocyte antigen 75	LY75	AF011333	0.5832	Above
35	41649_at	FOXJ2 forkhead factor	LOC55810	AF038177	0.5832	Above
36	36591_at	tubulin alpha 1 testis specific	TUBA1	X06956	0.5832	Above
37	40167_s_at	CS box-containing WD protein	LOC55884	AF038187	0.5832	Above
38	2064_g_at	excision repair cross-complementing rodent repair deficiency complementation group	ERCC5	L20046	0.5832	Above
39	39729_at	Human natural killer cell enhancing factor (NKEFB) mRNA, complete cds.	NKEFB	L19185	0.5829	Below
40	38270_at	poly ADP-ribose glycohydrolase	PARG	AF005043	0.5828	Below
41	40613_at	uncharacterized hypothalamus protein HT012	HT012	AL031775	0.5819	Below
42	39070_at	singed Drosophila like sea urchin fascin homolog like	SNL	U03057	0.5813	Above
43	40782_at	short-chain dehydrogenase/reductase 1	SDR1	AF061741	0.5813	Above
44	34256_at	sialyltransferase 9 CMP-NeuAc lactosylceramide alpha-2 3-sialyltransferase GM3 synthase	SIAT9	AB018356	0.5797	Above
45	41836_at	protein with polyglutamine repeat calcium ca2 homeostasis endoplasmic reticulum protein	ERPROT21 3-21	U94836	0.5777	Above
46	35681_r_at	zinc finger homeobox 1B	ZFHX1B	AB011141	0.5759	Below
47	37190_at	WAS protein family member 1	WASF1	D87459	0.5759	Below
48	32788_at	RAN binding protein 2	RANBP2	D42063	0.5756	Above
49	828_at	prostaglandin E receptor 2 subtype EP2 53kD	PTGER2	U19487	0.5740	Above
50	38220_at	dihydropyrimidine dehydrogenase	DPYD	U20938	0.5737	Above

**E2A-PBX1**

1	32063_at	pre-B-cell leukemia transcription factor 1	PBX1	M86546	0.8750	Above
2	38994_at	STAT induced STAT inhibitor-2	STATI2	AF037989	0.8252	Below
3	33355_at	Homo sapiens cDNA FLJ12900 fis clone NT2RP2004321 (by CELERA serach of target sequence = PBX1)	PBX1	AL049381	0.8040	Above
4	40454_at	FAT tumor suppressor Drosophila homolog	FAT	X87241	0.7899	Above
5	753_at	nidogen 2	NID2	D86425	0.7368	Above
6	717_at	GS3955 protein	GS3955	D87119	0.7306	Above
7	1786_at	c-mer proto-oncogene tyrosine kinase	MERTK	U08023	0.7300	Above
8	39070_at	singed Drosophila like sea urchin fascin homolog like	SNL	U03057	0.7271	Below
9	1065_at	fms-related tyrosine kinase 3	FLT3	U02687	0.7160	Below
10	36650_at	cyclin D2	CCND2	D13639	0.7151	Below
11	33513_at	signaling lymphocytic activation molecule	SLAM	U33017	0.7096	Above
12	33748_at	minor histocompatibility antigen HA-1	KIAA0223	D86976	0.7084	Below
13	37225_at	KIAA0172 protein	KIAA0172	D79994	0.7033	Above
14	38717_at	DKFZP586A0522 protein	DKFZP586A0522	AL050159	0.7003	Below
15	854_at	B lymphoid tyrosine kinase	BLK	S76617	0.6982	Above
16	33641_g_at	nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor-like 1	NFKBIL1	Y14768	0.6975	Below
17	40468_at	KIAA0554 protein	KIAA0554	AB011126	0.6971	Below
18	41266_at	integrin alpha 6	ITGA6	X53586	0.6965	Below
19	36536_at	schwannomin interacting protein 1	SCHIP-1	AF070614	0.6938	Below
20	362_at	protein kinase C zeta	PRKCZ	Z15108	0.6904	Above
21	755_at	inositol 1 4 5-triphosphate receptor type 1	ITPR1	D26070	0.6877	Below
22	307_at	arachidonate 5-lipoxygenase	ALOX5	J03600	0.6875	Below
23	39614_at	KIAA0802 protein	KIAA0802	AB018345	0.6863	Above
24	1563_s_at	tumor necrosis factor receptor superfamily member 1A	TNFRSF1A	M58286	0.6837	Below
25	38748_at	adenosine deaminase RNA-specific B1 homolog of rat RED1	ADARB1	U76421	0.6763	Above
26	41409_at	basement membrane-induced gene	ICB-1	AF044896	0.6757	Below
27	34892_at	tumor necrosis factor receptor superfamily member 10b	TNFRSF10B	AF016266	0.6726	Below
28	40648_at	c-mer proto-oncogene tyrosine kinase	MERTK	U08023	0.6710	Above
29	38408_at	transmembrane 4 superfamily member 2	TM4SF2	L10373	0.6667	Below
30	34583_at	fms-related tyrosine kinase 3	FLT3	U02687	0.6665	Below
31	36900_at	stromal interaction molecule 1	STIM1	U52426	0.6650	Below
32	37625_at	interferon regulatory factor 4	IRF4	U52682	0.6636	Above
33	38340_at	huntingtin interacting protein-1-related	KIAA0655	AB014555	0.6609	Above

34	1830_s_at	transforming growth factor beta 1	TGFB1	M38449	0.6608	Below
35	37099_at	arachidonate 5-lipoxygenase-activating protein	ALOX5AP	AI806222	0.6605	Below
36	38254_at	KIAA0882 protein	KIAA0882	AB020689	0.6539	Below
37	37641_at	Human gene for hepatitis C-associated microtubular aggregate protein p44, exon 9 and complete cds.		D28915	0.6531	Below
38	33865_at	adenovirus 5 E1A binding protein	BS69	AA127624	0.6515	Below
39	40729_s_at	nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor-like 1	NFKBIL1	Y14768	0.6502	Below
40	40113_at	GS3955 protein	GS3955	D87119	0.6476	Above
41	32979_at	GRB2-associated binding protein 1	GAB1	U43885	0.6457	Below
42	36591_at	tubulin alpha 1 testis specific	TUBA1	X06956	0.6427	Below
43	38739_at	v-ets avian erythroblastosis virus E26 oncogene homolog 2	ETS2	AF017257	0.6424	Below
44	37485_at	fatty-acid-Coenzyme A ligase very long-chain 1	FACVL1	D88308	0.6363	Above
45	538_at	CD34 antigen	CD34	S53911	0.6326	Below
46	37893_at	protein tyrosine phosphatase non-receptor type 2	PTPN2	AI828880	0.6318	Above
47	41017_at	myosin-binding protein H	MYBPH	U27266	0.6297	Above
48	37967_at	lymphocyte antigen 117	LY117	AF000424	0.6260	Below
49	37281_at	KIAA0233 gene product	KIAA0233	D87071	0.6250	Below
50	35675_at	vinexin beta SH3-containing adaptor molecule-1	SCAM-1	AF037261	0.6229	Below

### Hyperdiploid > 50

1	39878_at	protocadherin 9	PCDH9	AI524125	0.5838	Below
2	41470_at	prominin mouse like 1	PROML1	AF027208	0.5616	Above
3	39069_at	AE-binding protein 1	AEBP1	AF053944	0.5423	Below
4	1520_s_at	interleukin 1 beta	IL1B	X04500	0.5399	Above
5	578_at	Human recombination activating protein (RAG2) gene, last exon	RAG2	M94633	0.5208	Below
6	32251_at	hypothetical protein FLJ21174	FLJ21174	AA149307	0.5164	Above
7	40480_s_at	FYN oncogene related to SRC FGR YES	FYN	M14333	0.5090	Above
8	38604_at	neuropeptide Y	NPY	AI198311	0.5083	Above
9	40903_at	ATPase H transporting lysosomal vacuolar proton pump membrane sector associated protein M8-9	APT6M8-9	AL049929	0.5080	Above
10	38968_at	SH3-domain binding protein 5 BTK-associated	SH3BP5	AB005047	0.5057	Above
11	37272_at	inositol 1 4 5-trisphosphate 3-kinase B	ITPKB	X57206	0.5025	Below
12	35688_g_at	mature T-cell proliferation 1	MTCP1	Z24459	0.5018	Above
13	1488_at	protein tyrosine phosphatase receptor type K	PTPRK	L77886	0.4977	Below



14	36885_at	spleen tyrosine kinase	SYK	L28824	0.4964	Below
15	1630_s_at	tyrosine kinase syk	syk	HG3730- HT4000	0.4913	Below
16	38317_at	transcription elongation factor A SII like 1	TCEAL1	M99701	0.4901	Above
17	38649_at	KIAA0970 protein	KIAA0970	AB023187	0.4898	Below
18	39721_at	ephrin-B1	EFNB1	U09303	0.4895	Above
19	33307_at	kraken-like	BK126B4.1	AL022316	0.4880	Below
20	38518_at	sex comb on midleg Drosophila like 2	SCML2	Y18004	0.4879	Above
21	39402_at	interleukin 1 beta	IL1B	M15330	0.4750	Above
22	36489_at	phosphoribosyl pyrophosphate synthetase 1	PRPS1	D00860	0.4718	Above
23	37747_at	Human annexin V (ANX5) gene, exon 13.	(ANX5	U05770	0.4717	Above
24	40200_at	heat shock transcription factor 1	HSF1	M64673	0.4689	Below
25	35940_at	POU domain class 4 transcription factor 1	POU4F1	X64624	0.4685	Above
26	35727_at	hypothetical protein FLJ20517	FLJ20517	AI249721	0.4675	Below
27	1357_at	ubiquitin specific protease 4 proto- oncogene	USP4	U20657	0.4670	Below
28	36592_at	prohibitin	PHB	S85655	0.4668	Above
29	37014_at	myxovirus influenza resistance 1 homolog of murine interferon-inducible protein p78	MX1	M33882	0.4635	Above
30	40891_f_at	DNA segment on chromosome X unique 9879 expressed sequence	DXS9879E	X92896	0.4608	Above
31	40846_g_at	interleukin enhancer binding factor 3 90kD	ILF3	U10324	0.4605	Below
32	41132_r_at	heterogeneous nuclear ribonucleoprotein H2 H	HNRPH2	U01923	0.4605	Above
33	37280_at	MAD mothers against decapentaplegic Drosophila homolog 1	MADH1	U59912	0.4595	Below
34	35939_s_at	POU domain class 4 transcription factor 1	POU4F1	L20433	0.4594	Above
35	890_at	ubiquitin-conjugating enzyme E2A RAD6 homolog	UBE2A	M74524	0.4570	Above
36	38738_at	SMT3 suppressor of mif two 3 yeast homolog 1	SMT3H1	X99584	0.4568	Above
37	38458_at	Human cytochrome b5 (CYB5) gene, exon 6 and complete cds.	CYB5	L39945	0.4552	Above
38	38869_at	KIAA1069 protein	KIAA1069	AB028992	0.4549	Above
39	915_at	interferon-induced protein with tetra-trico-peptide repeats 1	IFIT1	M24594	0.4544	Above
40	38408_at	transmembrane 4 superfamily member 2	TM4SF2	L10373	0.4535	Above
41	39301_at	calpain 3 p94	CAPN3	X85030	0.4533	Below
42	41425_at	Friend leukemia virus integration 1	FLI1	M98833	0.4519	Below
43	2094_s_at	v-fos FBJ murine osteosarcoma viral oncogene homolog	FOS	K00650	0.4514	Above

44	36605_at	transcription factor 4	TCF4	M74719	0.4497	Above
45	37709_at	DNA segment numerous copies expressed probes GS1 gene	DXF68S1E	M86934	0.4493	Above
46	36128_at	transmembrane trafficking protein	TMP21	L40397	0.4488	Above
47	171_at	von Hippel-Lindau binding protein 1	VBP1	U56833	0.4473	Above
48	41490_at	phosphoribosyl pyrophosphate synthetase 2	PRPS2	Y00971	0.4466	Above
49	36536_at	schwannomin interacting protein 1	SCHIP-1	AF070614	0.4448	Above
50	35843_at	Homo sapiens mRNA cDNA DKFZp434D0935		L40402	0.4443	Above

### **MLL**

1	39402_at	interleukin 1 beta	IL1B	M15330	0.7355	Below
2	307_at	arachidonate 5-lipoxygenase	ALOX5	J03600	0.7221	Below
3	1389_at	membrane metallo-endopeptidase neutral endopeptidase enkephalinase CALLA CD10	MME	J03779	0.7178	Below
4	37280_at	MAD mothers against decapentaplegic Drosophila homolog 1	MADH1	U59912	0.7021	Below
5	36650_at	cyclin D2	CCND2	D13639	0.6759	Below
6	37043_at	inhibitor of DNA binding 3 dominant negative helix-loop-helix protein	ID3	AL021154	0.6743	Below
7	1520_s_at	interleukin 1 beta	IL1B	X04500	0.6689	Below
8	40913_at	ATPase Ca transporting plasma membrane 4	ATP2B4	W28589	0.6684	Below
9	36536_at	schwannomin interacting protein 1	SCHIP-1	AF070614	0.6554	Below
10	37398_at	platelet/endothelial cell adhesion molecule CD31 antigen	PECAM1	AA100961	0.6548	Below
11	39114_at	decidual protein induced by progesterone	DEPP	AB022718	0.6478	Below
12	37967_at	lymphocyte antigen 117	LY117	AF000424	0.6432	Below
13	1325_at	MAD mothers against decapentaplegic Drosophila homolog 1	MADH1	U59423	0.6421	Below
14	38336_at	KIAA1013 protein	KIAA1013	AB023230	0.6395	Below
15	577_at	midkine neurite growth-promoting factor 2	MDK	M94250	0.6363	Below
16	38671_at	KIAA0620 protein	KIAA0620	AB014520	0.6353	Below
17	33412_at	LGALS1 Lectin, galactoside-binding, soluble, 1	LGALS1	AI535946	0.6351	Above
18	40451_at	hypothetical protein FLJ21434	FLJ21434	AL080203	0.6350	Below
19	36908_at	Human macrophage mannose receptor (MRC1) gene, exon 30.	MRC1	M93221	0.6290	Below
20	963_at	ligase IV DNA ATP-dependent	LIG4	X83441	0.6282	Below
21	41346_at	like-glycosyltransferase	LARGE	AJ007583	0.6214	Below
22	32207_at	membrane protein palmitoylated 1 55kD	MPP1	M64925	0.6155	Below
23	2062_at	insulin-like growth factor binding protein 7	IGFBP7	L19182	0.6145	Above

24	38408_at	transmembrane 4 superfamily member 2	TM4SF2	L10373	0.6137	Below
25	854_at	B lymphoid tyrosine kinase	BLK	S76617	0.6075	Above
26	32193_at	plexin C1	PLXNC1	AF030339	0.6065	Above
27	35939_s_at	POU domain class 4 transcription factor 1	POU4F1	L20433	0.6046	Below
28	33705_at	phosphodiesterase 4B cAMP-specific duncce Drosophila homolog phosphodiesterase E4	PDE4B	L20971	0.5991	Below
29	34168_at	deoxynucleotidyltransferase terminal	DNTT	M11722	0.5979	Below
30	36383_at	v-ets avian erythroblastosis virus E26 oncogene related	ERG	M17254	0.5976	Below
31	38968_at	SH3-domain binding protein 5 BTK-associated	SH3BP5	AB005047	0.5976	Below
32	39263_at	2 5 oligoadenylate synthetase 2	OAS2	M87434	0.5967	Below
33	39329_at	actinin alpha 1	ACTN1	X15804	0.5953	Below
34	34699_at	CD2-associated protein	CD2AP	AL050105	0.5945	Below
35	1267_at	protein kinase C eta	PRKCH	M55284	0.5941	Below
36	35172_at	tyrosylprotein sulfotransferase 2	TPST2	AF049891	0.5937	Below
37	38124_at	midkine neurite growth-promoting factor 2	MDK	X55110	0.5936	Below
38	33813_at	tumor necrosis factor receptor superfamily member 1B	TNFRSF1B	AI813532	0.5934	Below
39	34176_at	hypothetical protein from clone 643	LOC57228	AF091087	0.5930	Below
40	39424_at	tumor necrosis factor receptor superfamily member 14 herpesvirus entry mediator	TNFRSF14	U70321	0.5930	Below
41	40729_s_at	nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor-like 1	NFKBIL1	Y14768	0.5905	Below
42	32607_at	brain acid-soluble protein 1	BASP1	AF039656	0.5905	Above
43	38342_at	KIAA0239 protein	KIAA0239	D87076	0.5896	Below
44	32533_s_at	vesicle-associated membrane protein 5 myobrevin	VAMP5	AF054825	0.5880	Below
45	39330_s_at	actinin alpha 1	ACTN1	M95178	0.5867	Below
46	40519_at	protein tyrosine phosphatase receptor type C	PTPRC	Y00638	0.5848	Above
47	39338_at	S100 calcium-binding protein A10 annexin II ligand calpactin I light polypeptide p11	S100A10	AI201310	0.5844	Above
48	35940_at	POU domain class 4 transcription factor 1	POU4F1	X64624	0.5824	Below
49	39712_at	S100 calcium-binding protein A13	S100A13	AI541308	0.5818	Below
50	39379_at	Homo sapiens mRNA cDNA DKFZp586C1019 from clone DKFZp586C1019		AL049397	0.5811	Above

## Novel

1	31892_at	protein tyrosine phosphatase receptor type M	PTPRM	X58288	0.8668	Above
2	41734_at	KIAA0870 protein	KIAA0870	AB020677	0.8614	Below
3	995_g_at	protein tyrosine phosphatase receptor type M	PTPRM	X58288	0.8505	Above
4	994_at	protein tyrosine phosphatase receptor type M	PTPRM	X58288	0.7694	Above
5	37967_at	lymphocyte antigen 117	LY117	AF000424	0.7399	Below
6	34676_at	KIAA1099 protein	KIAA1099	AB029022	0.7298	Above
7	41159_at	clathrin heavy polypeptide Hc	CLTC	D21260	0.7283	Above
8	39728_at	interferon gamma-inducible protein 30	IFI30	J03909	0.7138	Below
9	37542_at	lipoma HMGIC fusion partner-like 2	LHFPL2	D86961	0.7069	Above
10	35350_at	B cell RAG associated protein	BRAG	AB011170	0.7049	Below
11	41438_at	KIAA1451 protein	KIAA1451	AL049923	0.6999	Below
12	34370_at	archain 1	ARCN1	X81198	0.6999	Below
13	36029_at	chromosome 11 open reading frame 8	C11ORF8	U57911	0.6964	Above
14	37960_at	carbohydrate chondroitin 6/keratan sulfotransferase 2	CHST2	AB014679	0.6947	Above
15	35869_at	MD-1 RP105-associated	MD-1	AB020499	0.6908	Below
16	36601_at	vinculin	VCL	M33308	0.6908	Below
17	40775_at	integral membrane protein 2A	ITM2A	AL021786	0.6879	Above
18	37281_at	KIAA0233 gene product	KIAA0233	D87071	0.6837	Below
19	957_at	arrestin, beta 2	ARRB2	HG2059- HT2114	0.6744	Below
20	33284_at	myeloperoxidase	MPO	M19507	0.6712	Below
21	40585_at	adenylate cyclase 7	ADCY7	D25538	0.6712	Below
22	37908_at	guanine nucleotide binding protein 11	GNG11	U31384	0.6656	Above
23	40167_s_at	CS box-containing WD protein	LOC55884	AF038187	0.6581	Below
24	38576_at	H2B histone family member B	H2BFB	AJ223353	0.6576	Below
25	36591_at	tubulin alpha 1 testis specific	TUBA1	X06956	0.6576	Below
26	37712_g_at	MADS box transcription enhancer factor 2 polypeptide C myocyte enhancer factor 2C	MEF2C	S57212	0.6576	Below
27	33924_at	KIAA1091 protein	KIAA1091	AB029014	0.6484	Below
28	32724_at	phytanoyl-CoA hydroxylase Refsum disease	PHYH	AF023462	0.6466	Above
29	33358_at	EST (retina)		W29087	0.6457	Above
30	33740_at	chromosome 1 open reading frame 2	C1ORF2	AF023268	0.6441	Below
31	36588_at	KIAA0810 protein	KIAA0810	AB018353	0.6441	Below
32	38802_at	progesterone binding protein	HPR6.6	Y12711	0.6441	Below
33	38408_at	transmembrane 4 superfamily member 2	TM4SF2	L10373	0.6440	Below
34	32227_at	proteoglycan 1 secretory granule	PRG1	X17042	0.6409	Below
35	34840_at	Homo sapiens cDNA FLJ22642 fis clone HSI06970		AI700633	0.6409	Below
36	1131_at	mitogen-activated protein kinase kinase 2	MAP2K2	L11285	0.6409	Below
37	33410_at	integrin alpha 6	ITGA6	S66213	0.6391	Above

38	38006_at	CD48 antigen B-cell membrane protein	CD48	M37766	0.6342	Below
39	33907_at	eukaryotic translation initiation factor 4 gamma 3	EIF4G3	AF012072	0.6304	Below
40	41273_at	FK506 binding protein 12-rapamycin associated protein 1	FRAP1	AL046940	0.6304	Below
41	39781_at	insulin-like growth factor-binding protein 4	IGFBP4	U20982	0.6301	Below
42	39893_at	guanine nucleotide binding protein G protein gamma 7	GNG7	AB010414	0.6301	Below
43	37326_at	proteolipid protein 2 colonic epithelium-enriched	PLP2	U93305	0.6267	Below
44	36687_at	cytochrome c oxidase subunit VIIb	COX7B	N50520	0.6266	Below
45	40423_at	KIAA0903 protein	KIAA0903	AB020710	0.6254	Above
46	32542_at	four and a half LIM domains 1	FHL1	AF063002	0.6236	Below
47	33232_at	cysteine-rich protein 1 intestinal	CRIP1	AI017574	0.6211	Below
48	37280_at	MAD mothers against decapentaplegic Drosophila homolog 1	MADH1	U59912	0.6208	Above
49	1325_at	MAD mothers against decapentaplegic Drosophila homolog 1	MADH1	U59423	0.6208	Above
50	40729_s_at	nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor-like 1	NFKBIL1	Y14768	0.6199	Below

#### T-ALL

1	38242_at	B cell linker protein	SLP65	AF068180	0.8683	Below
2	37988_at	CD79B antigen immunoglobulin-associated beta	CD79B	M89957	0.8422	Below
3	1096_g_at	CD19 antigen	CD19	M28170	0.8181	Below
4	39318_at	T-cell leukemia/lymphoma 1A	TCL1A	X82240	0.8128	Below
5	38018_g_at	CD79A antigen immunoglobulin-associated alpha	CD79A	U05259	0.8127	Below
6	36878_f_at	major histocompatibility complex class II DQ beta 1	HLA-DQB1	M60028	0.8053	Below
7	38147_at	SH2 domain protein 1A Duncan s disease lymphoproliferative syndrome	SH2D1A	AL023657	0.8016	Above
8	35350_at	B cell RAG associated protein	BRAG	AB011170	0.7914	Below
9	38051_at	mal T-cell differentiation protein	MAL	X76220	0.7900	Above
10	266_s_at	CD24 antigen small cell lung carcinoma cluster 4 antigen	CD24	L33930	0.7867	Below
11	38521_at	CD22 antigen	CD22	X59350	0.7856	Below
12	37344_at	major histocompatibility complex class II DM alpha	HLA-DMA	X62744	0.7835	Below
13	34033_s_at	leukocyte immunoglobulin-like receptor subfamily A with TM domain member 2	LILRA2	AF025531	0.7761	Below
14	36638_at	connective tissue growth factor	CTGF	X78947	0.7755	Below
15	38213_at	galactosidase alpha	GLA	U78027	0.7701	Below

16	41734_at	KIAA0870 protein	KIAA0870	AB020677	0.7693	Below
17	37711_at	MADS box transcription enhancer factor 2 polypeptide C myocyte enhancer factor 2C	MEF2C	S57212	0.7560	Below
18	36239_at	POU domain class 2 associating factor 1	POU2AF1	Z49194	0.7440	Below
19	38319_at	CD3D antigen delta polypeptide TiT3 complex	CD3D	AA919102	0.7426	Above
20	38894_g_at	neutrophil cytosolic factor 4 40kD	NCF4	AL008637	0.7422	Below
21	33705_at	phosphodiesterase 4B cAMP-specific dunce Drosophila homolog phosphodiesterase E4	PDE4B	L20971	0.7414	Below
22	38017_at	CD79A antigen immunoglobulin-associated alpha	CD79A	U05259	0.7360	Below
23	41156_g_at	catenin cadherin-associated protein alpha 1 102kD	CTNNA1	U03100	0.7315	Below
24	38994_at	STAT induced STAT inhibitor-2	STAT2	AF037989	0.7292	Below
25	37710_at	MADS box transcription enhancer factor 2 polypeptide C myocyte enhancer factor 2C	MEF2C	L08895	0.7283	Below
26	41155_at	catenin cadherin-associated protein alpha 1 102kD	CTNNA1	U03100	0.7278	Below
27	40570_at	forkhead box O1A rhabdomyosarcoma	FOXO1A	AF032885	0.7258	Below
28	34224_at	fatty acid desaturase 3	FADS3	AC004770	0.7254	Below
29	38604_at	neuropeptide Y	NPY	AI198311	0.7212	Below
30	36773_f_at	major histocompatibility complex class II DQ beta 1	HLA-DQB1	M81141	0.7197	Below
31	32562_at	endoglin Osler-Rendu-Weber syndrome 1	ENG	X72012	0.7180	Below
32	36502_at	PFTAIRE protein kinase 1	PFTK1	AB020641	0.7179	Below
33	37180_at	phospholipase C gamma 2 phosphatidylinositol-specific	PLCG2	X14034	0.7114	Below
34	38893_at	neutrophil cytosolic factor 4 40kD	NCF4	AL008637	0.7100	Below
35	387_at	cyclin-dependent kinase 9 CDC2-related kinase	CDK9	X80230	0.7024	Below
36	32035_at	Human MHC class II HLA-DRw53-associated glycoprotein beta- chain mRNA complete cds		M16942	0.6992	Below
37	41153_f_at	Homo sapiens alphaE-catenin (CTNNA1) gene	CTNNA1	AF102803	0.6976	Below
38	40780_at	C-terminal binding protein 2	CTBP2	AF016507	0.6976	Below
39	40775_at	integral membrane protein 2A	ITM2A	AL021786	0.6952	Above
40	39402_at	interleukin 1 beta	IL1B	M15330	0.6945	Below
41	38522_s_at	CD22 antigen	CD22	X52785	0.6945	Below
42	41166_at	immunoglobulin heavy constant mu	IGHM	X58529	0.6941	Below
43	36937_s_at	PDZ and LIM domain 1 elfin	PDLIM1	U90878	0.6937	Below
44	38833_at	Human mRNA for SB classII histocompatibility antigen alpha-chain		X00457	0.6925	Below

45	2047_s_at	junction plakoglobin	JUP	M23410	0.6920	Below
46	36277_at	Human membran protein (CD3-epsilon) gene, exon 9.	CD3E	M23323	0.6899	Above
47	40688_at	linker for activation of T cells	LAT	AJ223280	0.6898	Above
48	39389_at	CD9 antigen p24	CD9	M38690	0.6879	Below
49	33162_at	insulin receptor	INSR	X02160	0.6879	Below
50	31891_at	chitinase 3-like 2	CHI3L2	U58515	0.6872	Above

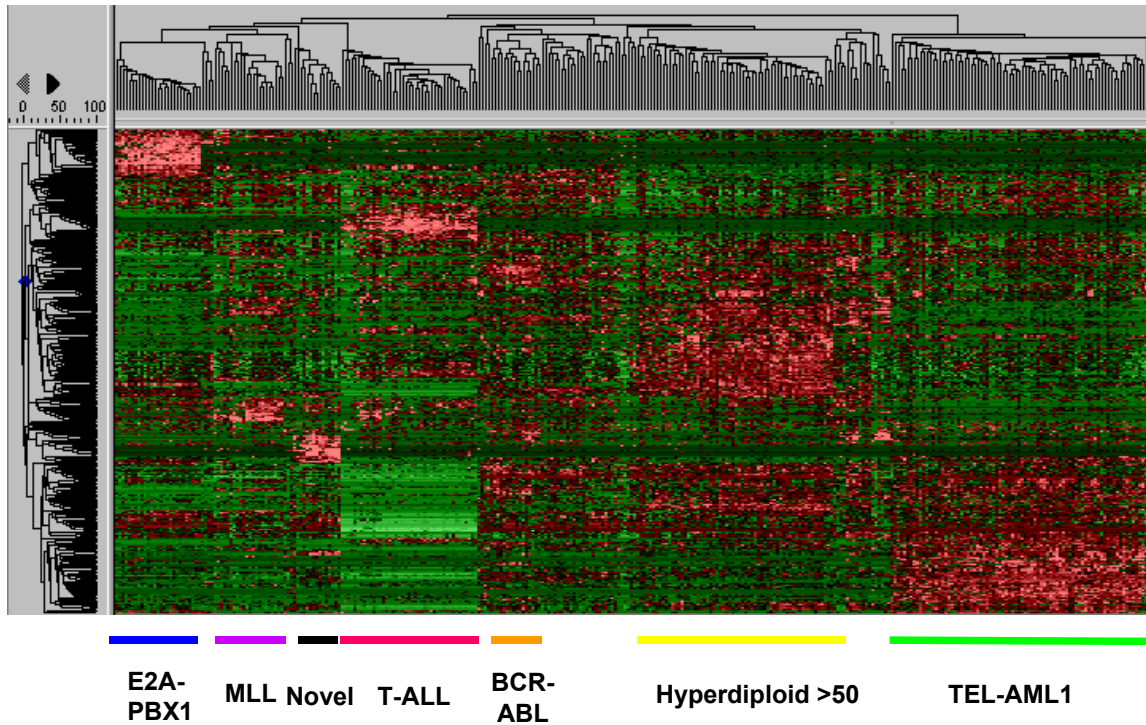
### **TEL-AML1**

1	37780_at	piccolo presynaptic cytomatrix protein	PCLO	AB011131	0.7121	Above
2	38203_at	potassium intermediate/small conductance calcium-activated channel subfamily N member 1	KCNN1	U69883	0.7086	Above
3	36524_at	Rho guanine nucleotide exchange factor GEF 4	ARHGEF4	AB029035	0.6782	Above
4	38578_at	tumor necrosis factor receptor superfamily member 7	TNFRSF7	M63928	0.6718	Above
5	32730_at	Homo sapiens mRNA for KIAA1750 protein partial cds		AL080059	0.6616	Above
6	34194_at	Homo sapiens cDNA FLJ21697 fis clone COL09740		AL049313	0.6518	Above
7	40272_at	collapsin response mediator protein 1	CRMP1	D78012	0.6160	Above
8	41819_at	FYN-binding protein FYB-120/130	FYB	U93049	0.6058	Above
9	1488_at	protein tyrosine phosphatase receptor type K	PTPRK	L77886	0.6056	Above
10	35665_at	phosphoinositide-3-kinase class 3	PIK3C3	Z46973	0.6022	Above
11	35614_at	transcription factor-like 5 basic helix-loop-helix	TCFL5	AB012124	0.5983	Above
12	36008_at	protein tyrosine phosphatase type IVA member 3	PTP4A3	AF041434	0.5976	Above
13	35362_at	myosin X	MYO10	AB018342	0.5964	Above
14	37908_at	guanine nucleotide binding protein 11	GNG11	U31384	0.5888	Above
15	39329_at	actinin alpha 1	ACTN1	X15804	0.5840	Below
16	1936_s_at	proto-oncogene c-myc, alt. transcript 3, ORF 114		HG3523-HT4899	0.5761	Below
17	33690_at	Homo sapiens mRNA cDNA DKFZp434A202	DKFZp434A202	AL080190	0.5725	Above
18	39389_at	CD9 antigen p24	CD9	M38690	0.5684	Below
19	37343_at	inositol 1 4 5-triphosphate receptor type 3	ITPR3	U01062	0.5642	Above
20	1299_at	telomeric repeat binding factor 2	TERF2	X93512	0.5585	Above
21	38652_at	hypothetical protein FLJ20154	FLJ20154	AF070644	0.5563	Above
22	38763_at	(clone D21-1) L-iditol-2 dehydrogenase gene		L29254	0.5535	Below
23	37724_at	v-myc avian myelocytomatosis viral oncogene homolog	MYC	V00568	0.5506	Below

24	36937_s_at	PDZ and LIM domain 1 elfin	PDLIM1	U90878	0.5506	Below
25	1325_at	MAD mothers against decapentaplegic Drosophila homolog 1	MADH1	U59423	0.5482	Above
26	41549_s_at	adaptor-related protein complex 1 sigma 2 subunit	AP1S2	AF091077	0.5474	Below
27	39827_at	hypothetical protein	FLJ20500	AA522530	0.5471	Below
28	32724_at	phytanoyl-CoA hydroxylase Refsum disease	PHYH	AF023462	0.5459	Above
29	31786_at	Sam68-like phosphotyrosine protein T-STAR	T-STAR	AF051321	0.5403	Above
30	38570_at	major histocompatibility complex class II DO beta	HLA-DOB	X03066	0.5384	Above
31	39330_s_at	actinin alpha 1	ACTN1	M95178	0.5375	Below
32	36493_at	lymphocyte-specific protein 1	LSP1	M33552	0.5356	Below
33	574_s_at	caspase 1 apoptosis-related cysteine protease interleukin 1 beta convertase	CASP1	M87507	0.5336	Below
34	32224_at	KIAA0769 gene product	KIAA0769	AB018312	0.5326	Above
35	1077_at	recombination activating gene 1	RAG1	M29474	0.5302	Above
36	37280_at	MAD mothers against decapentaplegic Drosophila homolog 1	MADH1	U59912	0.5283	Above
37	41200_at	CD36 antigen collagen type I receptor thrombospondin receptor like 1	CD36L1	Z22555	0.5261	Above
38	36009_at	hypothetical protein	CL683	AF091092	0.5259	Below
39	36933_at	N-myc downstream regulated	NDRG1	D87953	0.5254	Below
40	1126_s_at	Human cell surface glycoprotein CD44 (CD44) gene, 3' end of long tailed isoform.	CD44	L05424	0.5232	Below
41	39824_at	ESTs		AI391564	0.5231	Above
42	38078_at	filamin B beta actin-binding protein-278	FLNB	AF042166	0.5208	Below
43	38127_at	syndecan 1	SDC1	Z48199	0.5199	Above
44	32941_at	interferon consensus sequence binding protein 1	ICSBP1	M91196	0.5195	Below
45	37276_at	IQ motif containing GTPase activating protein 2	IQGAP2	U51903	0.5191	Below
46	34768_at	DKFZP564E1962 protein	DKFZP564E1962	AL080080	0.5184	Below
47	39781_at	insulin-like growth factor-binding protein 4	IGFBP4	U20982	0.5173	Below
48	37918_at	integrin beta 2 antigen CD18 p95 lymphocyte function-associated antigen 1 macrophage antigen 1 mac-1 beta subunit	ITGB2	M15395	0.5162	Below
49	41490_at	phosphoribosyl pyrophosphate synthetase 2	PRPS2	Y00971	0.5155	Below
50	41814_at	fucosidase alpha-L- 1 tissue	FUCA1	M29877	0.5101	Above



Illustrated below are the results of a two-dimensional hierarchical clustering algorithm of the 327 diagnostic ALL cases using the top 50 probe sets for each of the 7 groups chosen by the Wilkins' metric. As some genes are chosen for more than one group, there are 304 unique probe sets represented in Figure 17.



**Figure 17.** Hierarchical cluster of 327 Diagnostic ALL samples using genes selected by Wilkins' metric

### SOM/DAV

The 10,991 probe sets that passed the variation filter were used for subsequent selection of discriminating genes using the self-organizing map (SOM) and discriminant analysis with variance (DAV) programs in the GeneMaths software package (version 1.5, Applied Maths, Belgium). The subgroups for which genes were selected included T-lineage ALL, *TEL-AML1*, *E2A-PBX1*, *MLL* rearrangement, *BCR-ABL*, hyperdiploid ALL (chromosomal number > 50) and the novel subgroup described in the text of the paper. The target number of total genes chosen by each algorithm was 500.

The SOM analysis was performed using 30 X 18 node format to enable an optimal number of genes per node (~20 genes per node). Nodes that contained genes whose expression varied more than 2-fold from the mean in more than 70% of the samples in a particular subgroup were chosen. A total of 451 genes were chosen using the SOM algorithm and 443 genes using the DAV algorithm. The combined gene sets contained 755 unique genes, of which 185 were present in both subsets. 2-D hierarchical clustering of the genes and samples were performed using Pearson's correlation coefficient as the metric and unweighted pair group method using arithmetic averages (UPGMA). Approximately 10% of the genes that were found to have correlation coefficients less than 0.7 in each branch of the dendrogram were removed and the process was repeated iteratively until the correlation coefficient for all genes within a branch was > 0.7, or until the removal of additional gene resulted in a deterioration of the class distinction as indicated by inappropriate clustering of cases. Through this approach a subset of 215 genes were selected that optimally separated the 7 subgroups. These genes are listed in Table 15. The selection of genes by this approach does not provide for a ranking. For class prediction between 20 and 30 genes were used for each genetic subgroup, unless otherwise stated. The two-dimensional Hierarchical analysis of the cases using these selected genes are illustrated in Figure 18.

**Table 15. Genes selected by DAV-SOM**

	<b>Affymetrix number</b>	<b>Gene Name</b>	<b>GeneSymbol</b>	<b>Reference number</b>	<b>Above/ Below Mean</b>
<b><i>BCR-ABL</i></b>					
1	39250_at	nephroblastoma overexpressed gene	NOV	X96584	Above
2	37600_at	extracellular matrix protein 1	ECM1	U68186	Above
3	38312_at	DKFZp564O222 from clone DKFZp564O222		AL050002	Above
4	38342_at	KIAA0239 protein	KIAA0239	D87076	Above
5	39712_at	S100 calcium-binding protein A13	S100A13	AI541308	Above
6	39730_at	v-abl Abelson murine leukemia viral oncogene homolog 1	ABL1	X16416	Above
7	39781_at	Insulin-like growth factor-binding protein 4	IGFBP4	U20982	Above
8	40051_at	TRAM-like protein	KIAA0057	D31762	Above
9	40504_at	paraoxonase 2	PON2	AF001601	Above
10	33362_at	Cdc42 effector protein 3	CEP3	AF094521	Above
11	33404_at	adenylyl cyclase-associated protein 2	CAP2	U02390	Above
12	34362_at	solute carrier family 2 facilitated glucose transporter member 5	SLC2A5	M55531	Above
13	36591_at	Tubulin alpha 1 testis specific	TUBA1	X06956	Above
14	38077_at	collagen type VI alpha 3	COL6A3	X52022	Above
15	40196_at	HYA22 protein	HYA22	D88153	Above
16	1911_s_at	Growth arrest and DNA-damage-inducible alpha	GADD45A	M60974	Above
17	1702_at	interleukin 2 receptor alpha	IL2RA	X01057	Above
18	1635_at	Human proto-oncogene tyrosine-protein kinase (ABL) gene, exon 1a and exons 2-10, complete cds.	ABL	U07563	Above
19	1636_g_at	Human proto-oncogene tyrosine-protein kinase (ABL) gene, exon 1a and exons 2-10, complete cds.	ABL	U07563	Above
20	1326_at	Caspase 10 apoptosis-related cysteine protease	CASP10	U60519	Above
21	330_s_at	Tubulin, alpha 1, isoform 44	TUBA1	HG2259- HT2348	Above
<b><i>E2A-PBX1</i></b>					
1	33513_at	signaling lymphocytic activation molecule	SLAM	U33017	Above
2	37479_at	CD72 antigen	CD72	M54992	Above
3	37485_at	fatty-acid-Coenzyme A ligase very long- chain 1	FACVL1	D88308	Above
4	39614_at	KIAA0802 protein	KIAA0802	AB018345	Above
5	39929_at	KIAA0922 protein	KIAA0922	AB023139	Above
6	40648_at	c-mer proto-oncogene tyrosine kinase	MERTK	U08023	Above

7	41017_at	Myosin-binding protein H	MYBPH	U27266	Above
8	41425_at	Friend leukemia virus integration 1	FLI1	M98833	Above
9	41862_at	KIAA0056 protein	KIAA0056	D29954	Above
10	32063_at	pre-B-cell leukemia transcription factor 1	PBX1	M86546	Above
11	37225_at	KIAA0172 protein	KIAA0172	D79994	Above
12	38285_at	mu-crystallin gene		AF039397	Above
13	38286_at	KIAA1071 protein	KIAA1071	AB028994	Above
14	38340_at	huntingtin interacting protein-1-related	KIAA0655	AB014555	Above
15	39379_at	cDNA DKFZp586C1019 from clone DKFZp586C1019		AL049397	Above
16	39402_at	interleukin 1 beta	IL1B	M15330	Above
17	40454_at	FAT tumor suppressor Drosophila homolog	FAT	X87241	Above
18	41139_at	melanoma antigen family D 1	MAGED1	W26633	Above
19	41146_at	ADP-ribosyltransferase NAD poly ADP-ribose polymerase	ADPRT	J03473	Above
20	33355_at	Homo sapiens cDNA FLJ12900 fis clone NT2RP2004321		AL049381	Above
21	34783_s_at	BUB3 budding uninhibited by benzimidazoles 3 yeast homolog	BUB3	AF047473	Above
22	36179_at	mitogen-activated protein kinase-activated protein kinase 2	MAPKAPK2	U12779	Above
23	36589_at	aldo-keto reductase family 1 member B1 aldose reductase	AKR1B1	X15414	Above
24	38393_at	KIAA0247 gene product	KIAA0247	D87434	Above
25	38438_at	Nuclear factor of kappa light polypeptide gene enhancer in B-cells 1 p105	NFKB1	M58603	Above
26	1786_at	c-mer proto-oncogene tyrosine kinase	MERTK	U08023	Above
27	1520_s_at	interleukin 1 beta	IL1B	X04500	Above
28	1287_at	ADP-ribosyltransferase NAD poly ADP-ribose polymerase	ADPRT	J03473	Above
29	854_at	B lymphoid tyrosine kinase	BLK	S76617	Above
30	753_at	Nidogen 2	NID2	D86425	Above
31	430_at	nucleoside phosphorylase	NP	X00737	Above
32	362_at	Protein kinase C zeta	PRKCZ	Z15108	Above

### Hyperdiploid >50

1	36795_at	prosaposin variant Gaucher disease and variant metachromatic leukodystrophy	PSAP	J03077	Above
2	38242_at	B cell linker protein	SLP65	AF068180	Above
3	38518_at	sex comb on midleg Drosophila like 2	SCML2	Y18004	Above
4	39628_at	RAB9 member RAS oncogene family	RAB9	U44103	Above
5	31863_at	KIAA0179 protein	KIAA0179	D80001	Above
6	33228_g_at	interleukin 10 receptor beta	IL10RB	AI984234	Above
7	33753_at	KIAA0666 protein	KIAA0666	AB014566	Above
8	37543_at	Rac/Cdc42 guanine exchange factor GEF 6	ARHGEF6	D25304	Above

9	38968_at	SH3-domain binding protein 5 BTK-associated	SH3BP5	AB005047	Above
10	39039_s_at	CGI-76 protein	LOC51632	AI557497	Above
11	39329_at	Actinin alpha 1	ACTN1	X15804	Above
12	39389_at	CD9 antigen p24	CD9	M38690	Above
13	32207_at	membrane protein palmitoylated 1 55kD	MPP1	M64925	Above
14	32236_at	ubiquitin-conjugating enzyme E2G 2 homologous to yeast UBC7	UBE2G2	AF032456	Above
15	32251_at	hypothetical protein FLJ21174	FLJ21174	AA149307	Above
16	35764_at	chromosome X open reading frame 5	OFD1	Y15164	Above
17	36620_at	superoxide dismutase 1 soluble amyotrophic lateral sclerosis 1 adult	SOD1	X02317	Above
18	36937_s_at	PDZ and LIM domain 1 elfin	PDLIM1	U90878	Above
19	37326_at	proteolipid protein 2 colonic epithelium-enriched	PLP2	U93305	Above
20	37350_at	clone 889N15 on chromosome Xq22.1-22.3. Contains part of the gene for a novel protein similar to X. laevis Cortical Thymocyte Marker CTX	PSMD10	AL031177	Above
21	38738_at	SMT3 suppressor of mif two 3 yeast homolog 1	SMT3H1	X99584	Above
22	39168_at	Ac-like transposable element	ALTE	AB018328	Above
23	40903_at	ATPase H transporting lysosomal vacuolar proton pump membrane sector associated protein M8-9	APT6M8-9	AL049929	Above
24	32572_at	ubiquitin specific protease 9 X chromosome Drosophila fat facets related	USP9X	X98296	Above
25	1065_at	fms-related tyrosine kinase 3	FLT3	U02687	Above
26	306_s_at	high-mobility group nonhistone chromosomal protein 14	HMG14	J02621	Above

### **MLL**

1	31492_at	Muscle specific gene	M9	AB019392	Above
2	36777_at	DNA segment on chromosome 12 unique 2489 expressed sequence	D12S2489E	AJ001687	Above
3	39301_at	Calpain 3 p94	CAPN3	X85030	Below
4	41448_at	Homeo box A4	HOXA4	AC004080	Above
5	39424_at	tumor necrosis factor receptor superfamily member 14 herpesvirus entry mediator	TNFRSF14	U70321	Below
6	40076_at	Tumor protein D52-like 2	TPD52L2	AF004430	Above
7	40493_at	Human cell surface glycoprotein CD44 (CD44) gene, 3' end of long tailed isoform.	CD44	L05424	Above
8	40506_s_at	Homo sapiens polyadenylate binding protein mRNA, complete cds.		U75686	Above
9	40514_at	hypothetical 43.2 Kd protein	LOC51614	AF091085	Above
10	40763_at	Meis1 mouse homolog	MEIS1	U85707	Above

11	40797_at	a disintegrin and metalloproteinase domain 10	ADAM10	AF009615	Above
12	40798_s_at	a disintegrin and metalloproteinase domain 10	ADAM10	Z48579	Above
13	41747_s_at	myocyte-specific enhancer factor 2A (MEF2A) gene	MEF2A	U49020	Above
14	32193_at	Plexin C1	PLXNC1	AF030339	Above
15	32215_i_at	KIAA0878 protein	KIAA0878	AB020685	Above
16	33412_at	LGALS1 Lectin, galactoside-binding, soluble, 1 (galectin 1)	LGALS1	AI535946	Above
17	34306_at	muscleblind Drosophila like	MBNL	AB007888	Above
18	34785_at	KIAA1025 protein	KIAA1025	AB028948	Above
19	35298_at	eukaryotic translation initiation factor 3 subunit 7 zeta 66/67kD	EIF3S7	U54558	Above
20	36690_at	Nuclear receptor subfamily 3 group C member 1	NR3C1	M10901	Above
21	37675_at	solute carrier family 25 mitochondrial carrier phosphate carrier member 3	SLC25A3	X60036	Above
22	38391_at	capping protein actin filament gelsolin-like	CAPG	M94345	Above
23	38413_at	defender against cell death 1	DAD1	D15057	Above
24	39110_at	eukaryotic translation initiation factor 4B	EIF4B	X55733	Above
25	39867_at	Tu translation elongation factor mitochondrial	TUFM	S75463	Above
26	2062_at	Insulin-like growth factor binding protein 7	IGFBP7	L19182	Above
27	2036_s_at	CD44 antigen homing function and Indian blood group system	CD44	M59040	Above
28	1914_at	Cyclin A1	CCNA1	U66838	Above
29	1327_s_at	mitogen-activated protein kinase kinase kinase 5	MAP3K5	U67156	Above
30	1126_s_at	Human cell surface glycoprotein CD44 (CD44) gene, 3' end of long tailed isoform.	CD44	L05424	Above
31	1102_s_at	Nuclear receptor subfamily 3 group C member 1	NR3C1	M10901	Above
32	873_at	homeo box A5	HOXA5	M26679	Above
33	706_at	Glucocorticoid receptor, beta		HG4582- HT4987	Above
34	657_at	protocadherin gamma subfamily C 3	PCDHGC3	L11373	Above
<b>Novel</b>					
1	33137_at	latent transforming growth factor beta binding protein 4	LTBP4	Y13622	Above
2	38081_at	leukotriene A4 hydrolase	LTA4H	J03459	Above
3	38661_at	seb4D	HSRNASEB	X75314	Above
4	39878_at	protocadherin 9	PCDH9	AI524125	Above
5	35260_at	KIAA0867 protein	MONDOA	AB020674	Above
6	1373_at	transcription factor 3 E2A immunoglobulin enhancer binding factors E12/E47	TCF3	M31523	Above

7	35177_at	KIAA0725 protein	KIAA0725	AB018268	Above
8	38618_at	Human PAC clone RP3-515N1 from 22q11.2-q22	LIMK2	AC002073	Above
9	34947_at	phorbolin-like protein MDS019	MDS019	AA442560	Above
10	40692_at	transducin-like enhancer of split 4 homolog of Drosophila E sp1	TLE4	M99439	Above
11	38364_at	BCE-1 protein	BCE-1	AF068197	Above
12	37960_at	carbohydrate chondroitin 6/keratan sulfotransferase 2	CHST2	AB014679	Above
13	994_at	Protein tyrosine phosphatase receptor type M	PTPRM	X58288	Above
14	31892_at	Protein tyrosine phosphatase receptor type M	PTPRM	X58288	Above
15	995_g_at	Protein tyrosine phosphatase receptor type M	PTPRM	X58288	Above
16	41073_at	G protein-coupled receptor 49	GPR49	AI743745	Above
17	41708_at	KIAA1034 protein	KIAA1034	AB028957	Above
18	34376_at	protein kinase cAMP-dependent catalytic inhibitor gamma	PKIG	AB019517	Below
19	37978_at	quinolinate phosphoribosyltransferase nicotinate-nucleotide pyrophosphorylase carboxylating	QPRT	D78177	Below
20	38717_at	DKFZP586A0522 protein	DKFZP586A0522	AL050159	Below
21	33999_f_at	Human L2-9 transcript of unrearranged immunoglobulin V H 5 pseudogene		X58398	Above
22	36181_at	LIM and SH3 protein 1	LASP1	X82456	Below
23	41202_s_at	conserved gene amplified in osteosarcoma	OS4	AF000152	Above
24	41138_at	Antigen identified by monoclonal antibodies 12E7 F21 and O13	MIC2	M16279	Below
25	40771_at	Moesin	MSN	Z98946	Above
26	39070_at	singed Drosophila like sea urchin fascin homolog like	SNL	U03057	Below
27	32562_at	endoglin Osler-Rendu-Weber syndrome 1	ENG	X72012	Below
28	36536_at	schwannomin interacting protein 1	SCHIP-1	AF070614	Below
29	36650_at	cyclin D2	CCND2	D13639	Below
30	39756_g_at	X-box binding protein 1	XBP1	Z93930	Above
31	34168_at	deoxynucleotidyltransferase terminal	DNTT	M11722	Above
32	1389_at	membrane metallo-endopeptidase neutral endopeptidase enkephalinase CALLA CD10	MME	J03779	Below
33	41213_at	peroxiredoxin 1	PRDX1	X67951	Above
34	36571_at	Topoisomerase DNA II beta 180kD	TOP2B	X68060	Above
35	253_g_at	clone GPCR W G protein-linked receptor gene (GPCR) gene, 5' end of cds.		L42324	Below
36	252_at	clone GPCR W G protein-linked receptor gene (GPCR) gene, 5' end of cds.		L42324	Above
37	2087_s_at	cadherin 11 type 2 OB-cadherin osteoblast	CDH11	D21254	Above

38	36976_at	cadherin 11 type 2 OB-cadherin osteoblast	CDH11	D21255	Above
----	----------	---	-------	--------	-------

**T-ALL**

1	35016_at	Human Ia-associated invariant gamma-chain gene, exon 8, clones lambda-y(1,2,3).		M13560	Below
2	36277_at	membrane protein (CD3-epsilon) gene	CD3E	M23323	Above
3	38147_at	SH2 domain protein 1A Duncan s disease lymphoproliferative syndrome	SH2D1A	AL023657	Above
4	38949_at	protein kinase C theta	PRKCQ	L01087	Above
5	32649_at	transcription factor 7 T-cell specific HMG-box	TCF7	X59871	Above
6	33238_at	Human T-lymphocyte specific protein tyrosine kinase p56lck (LCK) aberrant mRNA, complete cds.	LCK	U23852	Above
7	35643_at	nucleobindin 2	NUCB2	X76732	Above
8	36473_at	ubiquitin specific protease 20	USP20	AB023220	Above
9	38319_at	CD3D antigen delta polypeptide TiT3 complex	CD3D	AA919102	Above
10	39709_at	selenoprotein W 1	SEPW1	U67171	Above
11	40775_at	integral membrane protein 2A	ITM2A	AL021786	Above
12	32794_g_at	T cell receptor beta locus	TRB	X00437	Above
13	37039_at	major histocompatibility complex class II DR alpha	HLA-DRA	J00194	Below
14	38051_at	mal T-cell differentiation protein	MAL	X76220	Above
15	38095_i_at	major histocompatibility complex class II DP beta 1	HLA-DPB1	M83664	Below
16	38096_f_at	major histocompatibility complex class II DP beta 1	HLA-DPB1	M83664	Below
17	38415_at	protein tyrosine phosphatase type IVA member 2	PTP4A2	U14603	Above
18	38833_at	Human mRNA for SB classII histocompatibility antigen alpha-chain		X00457	Below
19	2059_s_at	lymphocyte-specific protein tyrosine kinase	LCK	M36881	Above
20	1241_at	protein tyrosine phosphatase type IVA member 2	PTP4A2	U14603	Above
21	1105_s_at	T cell receptor beta locus	TRB	M12886	Above

**TEL-AML1**

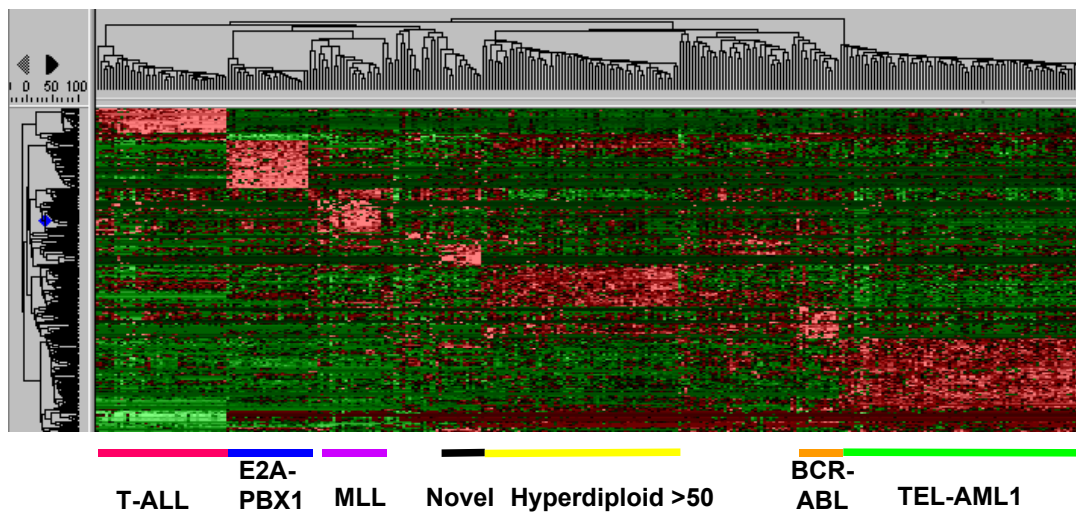
1	31508_at	upregulated by 1, 25-dihydroxyvitamin D-3	VDUP1	S73591	Above
2	33690_at	cDNA DKFZp434A202 from clone DKFZp434A202		AL080190	Above
3	34481_at	vav proto-oncogene, exon 27, and complete cds.	VAV	AF030227	Above
4	36239_at	POU domain class 2 associating factor 1	POU2AF1	Z49194	Above
5	37470_at	Leukocyte-associated Ig-like receptor 1	LAIR1	AF013249	Above

6	38203_at	Potassium intermediate/small conductance calcium-activated channel subfamily N member 1	KCNN1	U69883	Above
7	38570_at	major histocompatibility complex class II DO beta	HLA-DOB	X03066	Above
8	38578_at	tumor necrosis factor receptor superfamily member 7	TNFRSF7	M63928	Above
9	38906_at	spectrin alpha erythrocytic 1 elliptocytosis 2	SPTA1	M61877	Above
10	40729_s_at	nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor-like 1	NFKBIL1	Y14768	Above
11	40745_at	adaptor-related protein complex 1 beta 1 subunit	AP1B1	L13939	Above
12	41097_at	telomeric repeat binding factor 2	TERF2	AF002999	Above
13	41381_at	KIAA0308 protein	KIAA0308	AB002306	Above
14	41442_at	core-binding factor runt domain alpha subunit 2 translocated to 3	CBFA2T3	AB010419	Above
15	31898_at	KIAA0212 gene product	KIAA0212	D86967	Above
16	32660_at	KIAA0342 gene product	KIAA0342	AB002340	Above
17	34194_at	cDNA FLJ21697 fis clone COL09740		AL049313	Above
18	35614_at	transcription factor-like 5 basic helix-loop-helix	TCFL5	AB012124	Above
19	35665_at	Phosphoinositide-3-kinase class 3	PIK3C3	Z46973	Above
20	36008_at	protein tyrosine phosphatase type IVA member 3	PTP4A3	AF041434	Above
21	36524_at	Rho guanine nucleotide exchange factor GEF 4	ARHGEF4	AB029035	Above
22	36537_at	Rho-specific guanine nucleotide exchange factor p114	P114-RHO-GEF	AB011093	Above
23	37280_at	MAD mothers against decapentaplegic Drosophila homolog 1	MADH1	U59912	Above
24	38652_at	hypothetical protein FLJ20154	FLJ20154	AF070644	Above
25	41200_at	CD36 antigen collagen type I receptor thrombospondin receptor like 1	CD36L1	Z22555	Above
26	32224_at	KIAA0769 gene product	KIAA0769	AB018312	Above
27	36985_at	isopentenyl-diphosphate delta isomerase	IDI1	X17025	Above
28	38124_at	midkine neurite growth-promoting factor 2	MDK	X55110	Above
29	39824_at	ESTs		AI391564	Above
30	40570_at	forkhead box O1A rhabdomyosarcoma	FOXO1A	AF032885	Above
31	41498_at	KIAA0911 protein	KIAA0911	AB020718	Above
32	41814_at	fucosidase alpha-L- 1 tissue	FUCA1	M29877	Above
33	32579_at	SWI/SNF related matrix associated actin dependent regulator of chromatin subfamily a member 4	SMARCA4	D26156	Above
34	33162_at	insulin receptor	INSR	X02160	Above
35	1779_s_at	pim-1 oncogene	PIM1	M16750	Above
36	1488_at	protein tyrosine phosphatase receptor type K	PTPRK	L77886	Above



37	1325_at	MAD mothers against decapentaplegic Drosophila homolog 1	MADH1	U59423	Above
38	1336_s_at	protein kinase C beta 1	PRKCB1	X06318	Above
39	1299_at	Telomeric repeat binding factor 2	TERF2	X93512	Above
40	1217_g_at	protein kinase C beta 1	PRKCB1	X07109	Above
41	1077_at	recombination activating gene 1	RAG1	M29474	Above
42	932_i_at	zinc finger protein 91 HPF7 HTF10	ZNF91	L11672	Above
43	880_at	FK506-binding protein 1A 12kD	FKBP1A	M34539	Above
44	755_at	inositol 1 4 5-triphosphate receptor type 1	ITPR1	D26070	Above
45	577_at	midkine neurite growth-promoting factor 2	MDK	M94250	Above
46	160029_at	protein kinase C beta 1	PRKCB1	X07109	Above

Illustrated below are the results of a two-dimensional hierarchical clustering algorithm of the 327 diagnostic ALL cases using the 215 genes selected by a combination of SOM and DAV.



**Figure 18.** Hierarchical cluster of 327 Diagnostic ALL samples using genes chosen by SOM and DAV.

### Comparison of genes selected by the different metrics

To address the question of overlap among the different lists of genes, it is important to note that the selection of genes in this paper was for the purpose of defining discriminators of the different genetic or prognostic subgroups. We chose to select the top 20-50 genes for each class because we felt this was the largest number of genes we could choose without a significant risk of overtraining with the supervised learning algorithms. Thus, only the top 40 to 50 genes are listed for each metric. To draw conclusions from a comparison of these “truncated” lists would be inappropriate. The metrics use very different criteria to rank genes, and thus the top ranked genes differ significantly between metrics. If one wants to know the overlap of all statistically significant genes selected by the different metrics for each genetic subtype, you would need to compare very large lists. For example, the total number of statistically significant ( $p < 0.05$ ) genes selected by Chi-square for each genetic subtype is T-ALL (1309 genes), *E2A-PBX1* (827 genes), *TEL-AML1* (1156 genes), *BCR-ABL* (85 genes), *MLL* (358 genes), Hyperdiploid >50 (626 genes). If we then ask, what is the percentage of the top 20 genes selected by the other metrics that are contained within the significant Chi-square genes, the answer is 100% for every subgroup except for *BCR-ABL* (data not shown). The lower percentage of overlap for *BCR-ABL* results from the smaller number of genes that distinguish this genetic subgroup. Thus, there is a very high degree of overlap

between the genes chosen by the various metrics. At the top of the lists however, the ranking is quite different. Despite this the top genes selected by the various metrics are all able to accurately identify the genetic subtypes as detailed below. As a result, a limited number of genes can be used to accurately identify the genetic subtypes and one can use non-overlapping lists and still achieve high prediction accuracy. Thus, there are many genes that are distinct discriminators of these seven genetic subtypes, and one need only to use a small subset of these in a supervised learning algorithm to accurately identify a case as belonging to the genetic subtype.

### Decision tree for the diagnosis of genetic subtypes

Classification was approached using a decision tree format (Figure 19), in which the first decision was T-ALL versus B-lineage (non-T-ALL). Within the B-lineage subset, cases were then sequentially classified into the known risk groups characterized by the presence of E2A-PBX1, TEL-AML1, BCR-ABL, MLL chimeric genes, and lastly hyperdiploid >50 chromosomes. Cases not assigned to one of these classes were left unassigned. Classification was performed using the supervised learning algorithms described below.

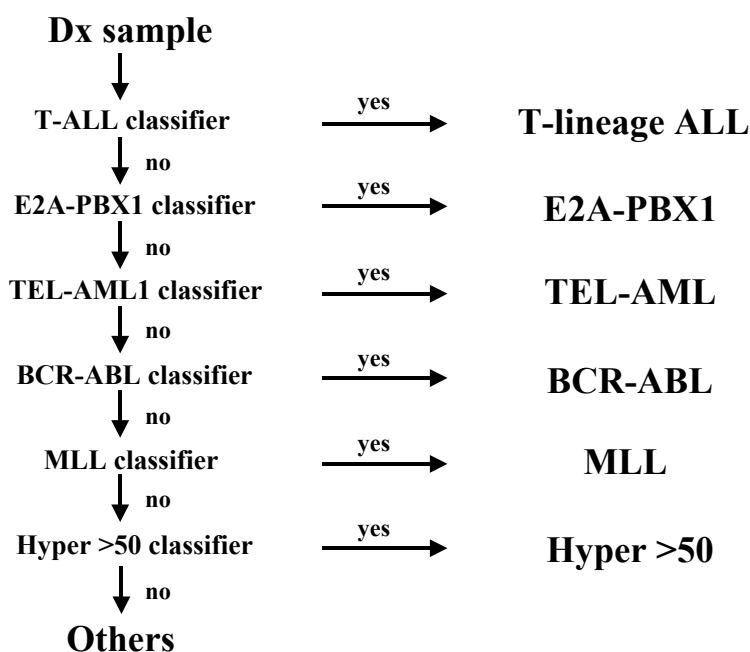


Figure 19. Diagnostic decision tree

### Description of Supervised Learning Algorithms

We performed an analysis using what is considered a sophisticated linear classifier, C4.5, and a variety of different non-linear classifiers. The non-linear classifiers consistently outperformed the linear classifier. We therefore only included the description and data from non-linear classifiers below.

### Support Vector Machine (SVM)

Support vector machine (SVM) selects a small number of critical boundary instances from each class and builds a linear discriminant function that separates them as widely as possible.<sup>5</sup> In the case where no linear separation is possible, the technique of “kernel” is used to automatically inject the training instances into a higher dimensional space and a separator is learned in that space. We used the Weka version of SVM developed at the University of Waikato of New Zealand (<http://www.cs.waikato.ac.nz/ml/weka>), which implements Platt’s sequence minimal optimization algorithm for training a support vector classifier using polynomial kernels.<sup>6</sup>

## Prediction by Collective Likelihood of Emerging Patterns (PCL)

Emerging patterns (EPs) are a notion used in data mining to discover sharp differences between two classes of data.<sup>7</sup> An EP is a pattern---the expression level of several genes in our case---whose frequency increases significantly from one class of samples to another class. In particular, we looked for the most general patterns that have infinite growth in the sense that their frequency in one class is 0% and in another class is greater than 0% and none of their proper subpatterns are EPs. These EPs can then be combined into reliable rules for subtype prediction. Three earlier methods for classification based on EPs are JEP<sup>8</sup>, DeEPs<sup>9</sup>, and CAEP.<sup>10</sup>

Here we used an original variation in the spirit of JEP but with a different manner of aggregating EPs: Given two training data sets  $D_p$  and  $D_n$  and a testing sample  $T$ , the first phase was to discover EPs from  $D_p$  and  $D_n$ . Denote the EPs of  $D_p$ , in descending order of frequency, as  $TopEP^p_1, \dots, TopEP^p_i$ , and those of  $D_n$  as  $TopEP^n_1, \dots, TopEP^n_j$ . Suppose  $T$  contains the following EPs of  $D_p$ :  $TopEP^p_{i_1}, \dots, TopEP^p_{i_x}$ , where  $i_1 < i_2 < \dots < i_x \leq i$ ; and the following EPs of  $D_n$ :  $TopEP^n_{j_1}, \dots, TopEP^n_{j_y}$ , where  $j_1 < j_2 < \dots < j_y \leq j$ . In the next step, two scores were calculated for  $T$ :  $score_p = \Sigma[\text{frequency}(TopEP^p_{i_m})/\text{frequency}(TopEP^p_m)]$  and  $score_n = \Sigma[\text{frequency}(TopEP^n_{j_m})/\text{frequency}(TopEP^n_m)]$ , summing over  $m = 1..k$ , where  $k \ll i$  and  $k \ll j$ . In our case,  $k$  is chosen to be 25. Finally, a prediction is made on  $T$  as follows: If  $score_p > score_n$ , then  $T$  is predicted to be in class  $D_p$ ; otherwise, it is predicted as class  $D_n$ .

The spirit of this variation is to measure how far the top  $k$  EPs contained in  $T$  are away from the top  $k$  EPs of a class. For example, if  $k = 1$ , then  $score_p$  indicates whether the number-one EP contained in  $T$  is far from the most frequent EP of  $D_p$ . If the score is the maximum value 1, then the “distance” is very close, namely the most common property of  $D_p$  is also present in this testing sample. With smaller scores, the distance becomes further and the likelihood of  $T$  belonging to  $D_p$  becomes weaker. Using more than one top-ranked EPs in this way leads to very reliable predictions. We call this variation of EP-based classification method “prediction by collective likelihood of EPs” or PCL for short.

## $k$ -Nearest Neighbor ( $k$ -NN)

$k$ -NN is a typical instance-based learner where the class of a new instance is decided by the majority class of its  $k$  closest neighbors.<sup>11</sup> This method was used with the Euclidean distance metric. Conceptually, this is one of the most straightforward methods and is often used as a baseline for comparison purposes. The data were normalized using the z-score method, then the “best” few genes were chosen using one of the statistical gene selection methods. For these experiments, the “top  $n$ ” genes, where  $n = 1-50$ , were used. The expression values of the top genes from each diagnostic sample were treated as a vector in  $n$ -dimensional space. To classify a new sample, the same top  $n$  genes were chosen, and the Euclidean distance was computed between this new vector and each vector in the training data. The prediction was made by a majority vote of the  $k$  nearest samples, where  $k=1$  or  $k=3$ . In our experiment,  $k$  was set to 1.

## Artificial Neural Network (ANN)

The artificial neural network (ANN) learning models built are all feed-forward, fully connected, and non-recurrent. The input layer of each ANN contains 50 units, which correspond to the 50 input values (the “top 50” scoring genes). Each ANN has one hidden layer with 4 units, and an output layer that contains two units, which represent the two class labels. In a preprocessing step all input data was normalized using the z-score method. The apparent error was estimated using 3-fold cross-validation. That is, for each training procedure, the training samples were randomly shuffled and divided into three groups of approximately equal size. A model was built with two of the groups and the third group was set aside for validation. This step was repeated three times, each time with a different group for validation. This shuffling-training process was repeated ten times, resulting in 30 ANN models. Each test sample was fed into each of the 30 ANN models, and the output was the average of the 30 outputs. The class predicted was the one that was represented by the output unit with the larger average output value.

**Table of results using the different algorithms to predict the genetic subgroups**

A summary of the true prediction accuracies on the blinded test set of 112 cases are presented in Tables 16-18. Sensitivity was calculated as the number of positive samples predicted /the number of true positives. Specificity was calculated as the number of negative samples predicted/the number of true negatives.

**Table 16. True Prediction Accuracy Results on Test Set using SVM and ANN algorithms**

		<b>SVM</b>				<b>ANN</b>
		Chi Sq	CFS	T-stats	SOM/DAV	Wilkins'
T-ALL	True Accuracy	100	100	100	100	100
	Sensitivity	100	100	100	100	100
	Specificity	100	100	100	100	100
<i>E2A-PBX1</i>	True Accuracy	100	100	100	100	100
	Sensitivity	100	100	100	100	100
	Specificity	100	100	100	100	100
<i>TEL-AML1</i>	True Accuracy	99	99	98	97	100
	Sensitivity	100	100	100	100	100
	Specificity	98	98	97	97	100
<i>BCR-ABL</i>	True Accuracy	95	97	94	97	97
	Sensitivity	50	67	33	83	83
	Specificity	100	100	100	98	98
<i>MLL</i>	True Accuracy	100	98	100	97	100
	Sensitivity	100	100	100	86	100
	Specificity	100	98	100	100	100
H>50	True Accuracy	96	96	96	95	94
	Sensitivity	100	100	100	95	100
	Specificity	93	93	93	93	89

**Table 17. True Prediction Accuracy Results on Test Set using *k*-NN**

		<b><i>k</i>-NN</b>			
		Chi Sq	CFS	T-stats	Wilkins'
T-ALL	True Accuracy	100	100	100	100
	Sensitivity	100	100	100	100
	Specificity	100	100	100	100
<i>E2A-PBX1</i>	True Accuracy	100	100	100	100
	Sensitivity	100	100	100	100
	Specificity	100	100	100	100
<i>TEL-AML1</i>	True Accuracy	98	98	99	100
	Sensitivity	100	96	96	100
	Specificity	97	98	100	100
<i>BCR-ABL</i>	True Accuracy	94	97	95	93
	Sensitivity	33	67	50	67
	Specificity	100	100	100	96
<i>MLL</i>	True Accuracy	100	98	95	100
	Sensitivity	100	83	100	100
	Specificity	100	100	94	100
H>50	True Accuracy	98	96	94	98
	Sensitivity	100	100	95	100
	Specificity	96	93	93	96

**Table 18. True Prediction Accuracy Results on Test Set using PCL**

		PCL	
		Chi Sq	CFS
T-ALL	True Accuracy	100	100
	Sensitivity	100	100
	Specificity	100	100
E2A-PBX1	True Accuracy	ND	100
	Sensitivity	ND	100
	Specificity	ND	100
TEL-AML1	True Accuracy	99	ND
	Sensitivity	96	ND
	Specificity	100	ND
BCR-ABL	True Accuracy	97	ND
	Sensitivity	67	ND
	Specificity	100	ND
MLL	True Accuracy	100	ND
	Sensitivity	100	ND
	Specificity	100	ND
H>50	True Accuracy	98	ND
	Sensitivity	100	ND
	Specificity	96	ND

**Absence of correlation of expression data for genetic subtypes with stage of B-cell differentiation**

In trying to address the issue of whether the expression profile of the different genetic subtypes of B-cell leukemias might simply correspond to markers of different stages of B-cell differentiation, we have performed a large number of experiments. The first issue is defining the stage of B-cell differentiation. The defined stages of BM derived B-cells relevant to pediatric ALL are outlined below in Table 19, along with their frequency in pediatric ALL<sup>12</sup>. As can be seen three stages of differentiation are defined by a limited number of markers. The use of additional markers is not relevant for these distinctions. In Table 20 below, the distribution of our ALL cases into these B-cell differentiation stages is shown. As can be seen, none of the genetic subtypes is specifically associated with one of these three stages of differentiation. Thus, this simple analysis clearly shows that the majority of the chromosomal translocation subgroups in pediatric ALL do not correspond to a specific stage of B-cell differentiation. This is a well-known fact in the field of pediatric ALL and differs from the relationship typically seen between chromosomal translocations and other genetic lesions, and the stage of differentiation seen in B-cell lymphomas.

**Table 19. Immunophenotyping of acute lymphoblastic leukemias<sup>12</sup>**

Subtype	Leukocyte antigen expression (% of cases positive)					Frequency (%)
	CD19	CD22	cIgμ	sIgμ	sIg κ or λ	
Early Pre-B	100	>95	0	0	0	60-65
Pre-B	100	100	100	0	0	20-25
Transitional	100	100	100	100	0	1-3

Abbreviations: cIg μ, cytoplasmic immunoglobulin μ chain; sIg μ, surface immunoglobulin μ chain; sIg κ or λ, surface immunoglobulin κ or λ chains

<sup>12</sup>D.Campana and F.G.Behm, "Immunophenotyping of leukemia", Journal of Immunological Methods 243: 59-75, 2000.

**Table 20. Distribution of genetic subtypes by immunophenotype<sup>a</sup>**

	EARLY PRE-B	PRE-B	TRANSITIONAL PRE B
E2A	0	17	6
TEL	55	23	0

BCR	11	3	0
MLL	12	6	1
Hyperdip>50	49	9	5
Novel	8	4	1
Total	172	77	24

<sup>a</sup>For this analysis, samples with other immunophenotypes (NOS or mature B-cell) were not included

We next tried to see if we could define a set of genes that could accurately identify cases by their stage of differentiation, irrespective of what genetic subgroup they belong to. To accomplish this, we assigned cases into one of three classes, early pre-B, pre-B, or transitional pre-B based on their immunophenotype. We then choose the top 50 genes that distinguished each group from the other two using the Wilkins' metric. These genes were then used in an ANN analysis to assess their performance in correctly classifying the 273 diagnostic B-lineage ALL samples, for which a stage of differentiation could be determined, through a process of cross validation. The results of this analysis are included below.

**Table 21. Accuracy Results for immunophenotype discrimination using Wilkins' metric and ANN algorithm**

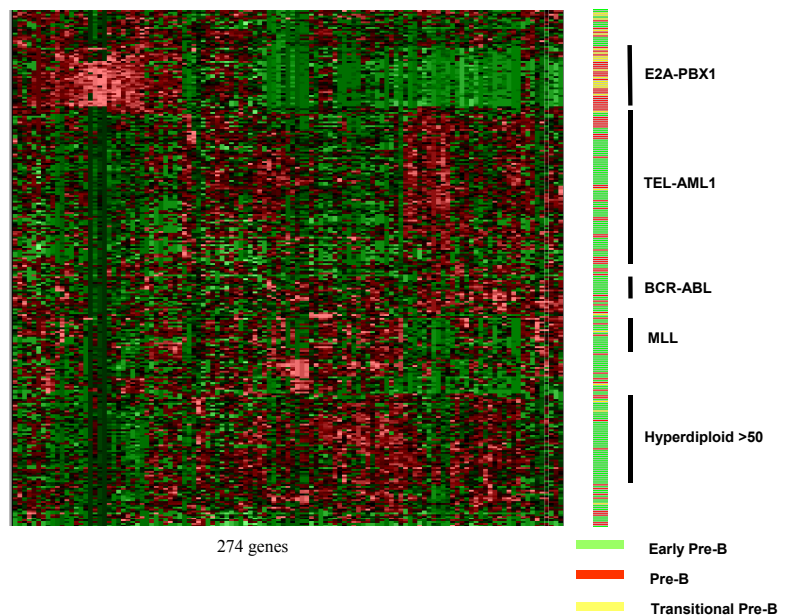
	Accuracy	Sensitivity	Specificity
Early Pre-B <sup>a</sup>	78.39%	85.47%	66.34%
Pre-B <sup>b</sup>	71.79%	38.96%	84.69%
Transitional Pre-B <sup>c</sup>	91.24%	33.33%	96.79%

<sup>a</sup>Cells with CD19+, CD22+, cytoplasmic I $\mu$ -, surface I $\mu$ - immunophenotype

<sup>b</sup>Cells with CD19+, CD22+, cytoplasmic I $\mu$ +, surface I $\mu$ - immunophenotype

<sup>c</sup>Cells with CD19+, CD22+, cytoplasmic I $\mu$ +, surface I $\mu$ + immunophenotype

As you can see, the selected genes perform rather poorly in correctly assigning cases to specific B-cell differentiation stages, with accuracies well below those achieved for prediction of the genetic subgroups. When these genes are used in a two-dimensional hierarchical clustering algorithm they failed to cluster cases by immunophenotype, but instead, resulted in the loose clustering of some of the genetic subgroups, including *E2A-PBX1*, *TEL-AML1*, *BCR-ABL*, *MLL*, and hyperdiploid >50 (Figure 20). We have repeated the analysis using genes selected by DAV and again, we fail to see clustering of the immunophenotypically defined stages, but instead see clustering of the genetic subgroups. Thus, we are unable to identify expression profiles that can accurately identify the immunophenotypically-defined differentiation stages of pediatric B-cell ALL. Moreover, the expression profiles that we have defined for the genetic subtypes are not profiles that correspond to specific stages of B-cell differentiation. Although some of the genes that define specific genetic subtypes can be associated with a particular stage of B-cell differentiation, the majority of the discriminating genes show no correlation with differentiation. These results do not conflict with the recently published data from Armstrong (Nat. Gen., 2002) As shown in



**Figure 20.** Two-dimensional hierarchical cluster of Wilkins' genes selected as discriminators of early pre-B, pre-B, and transitional pre-B.

Figure 20. Two-dimensional hierarchical cluster of Wilkins' genes selected as discriminators of early pre-B, pre-B, and transitional pre-B. As shown in

Table 20 above, we see a slight increase of *MLL* cases in the early pre-B immunophenotype. However, not all cases have this early differentiation stage.

### Results for relapse prediction

In the prediction of whether a patient would go into continuous complete remission or would relapse, we adopted a subtype-specific approach. That is, we constructed individual classifier for each subtype of ALL. Given a sample, we first predicted its subtype, and then invoked the corresponding subtype-specific prognostic classifier to predict whether the patient would relapse. This subtype-specific approach was required because an expression profile predictive of relapse for the entire group could not be defined.

In the construction of the type-specific classifiers, genes were selected by CFS unless it returned >20 genes, in which case the top 20 ranked genes by T-statistics were used. When the T-statistics method was used, we decided how many among the top 20 T-statistics genes were to be used by performing cross validation experiments---that is, we tried the top n genes for n = 1..20 and picked the n that gave the best cross validation results. The genes that were chosen for use in subtype-specific relapse predictions are summarized in Table 22. A permutation test was used to calculate whether the selected genes were statistically significant discriminators of relapse versus CCR. This was accomplished by performing 1000 random permutations of the dataset and then defining the T-statistic score for all ranked genes in each permutation. The 1% and 5%, top score for each level of ranking were then defined. The observed T-statistic score for the genes selected as discriminator of relapse were then compared to these experimentally determined 1% and 5% significance levels. This permutation analysis was performed and the top 7 ranked genes for T-ALL and the top 20 ranked genes for hyperdiploid >50 are shown, since statistically significant predictors of relapse were obtained only for these two genetic subgroups. The results of this permutation test are presented in Table 23 and Table 24. The result of the supervised learning algorithm for the prediction of relapse for each of the genetic subgroups are summarized in Table 25 below.

**Table 22. Genes Selected by Tstats/CFS for Relapse**

	<b>Affymetrix number</b>	<b>Gene Name</b>	<b>GeneSymbol</b>	<b>Reference number</b>	<b>Above/ Below mean</b>
<b>T ALL</b>					
1	33777_at	Human TBXAS1 gene for thromboxane synthase	TBXAS1	D34625	Above
2	41853_at	Homo sapiens mRNA for 41-kDa phosphoribosylpyrophosphate synthetase-associated protein		AB007851	Above
3	38866_at	Human DNA sequence from PAC 370M22		Z82206	Above
4	41643_at	Human spinal muscular atrophy gene	SMA5	X83301	Above
5	1126_s_at	Human cell surface glycoprotein CD44	CD44	L05424	Above
6	41862_at	Human mRNA for KIAA0056 gene	KIAA0056	D29954	Above
7	41131_f_at	Human BTK region clone ftp-3 mRNA		U01923	Above
<b>Hyperdiploid &gt;50</b>					
1	37721_at	deoxyhypusine synthase	DHPS	U79262	Above
2	38721_at	KIAA1536 protein	KIAA1536	W72733	Above
3	40120_at	hydroxyacyl glutathione hydrolase	HAGH	X90999	Above
4	41386_i_at	KIAA0346 protein	KIAA0346	AB002344	Above

5	38677_at	stress 70 protein chaperone microsome-associated 60kD	STCH	U04735	Above
6	37620_at	Human TFIID subunits TAF20 and TAF15 mRNA, complete cds.		U57693	Above
7	34703_f_at	EST		AA151971	Above
8	38355_at	DEAD/H Asp-Glu-Ala-Asp/His box polypeptide Y chromosome	DBY	AF000984	Above
9	41214_at	ribosomal protein S4 Y-linked	RPS4Y	M58459	Above
10	34530_at	Homo sapiens cDNA FLJ22448 fis clone HRC09541		W73822	Above
11	603_at	nuclear receptor subfamily 2 group C member 1	NR2C1	M29960	Above
12	32697_at	inositol myo 1 or 4 monophosphatase 1	IMPA1	AF042729	Above
13	41129_at	KIAA0033 protein	KIAA0033	D26067	Above
14	33333_at	KIAA0403 protein	KIAA0403	AB007863	Above
15	37078_at	CD3Z antigen zeta polypeptide TiT3 complex	CD3Z	J04132	Above
16	38148_at	cryptochrome 1 photolyase-like	CRY1	D83702	Above
17	39150_at	ring finger protein 11	RNF11	U69559	Above
18	33869_at	DKFZp586N1323 from clone DKFZp586N1323		AL080218	Above
19	41447_at	KIAA0990 protein	KIAA0990	AB023207	Above
20	39369_at	KIAA0935 protein	KIAA0935	AB023152	Above

### **TEL-AML1**

1	35797_at	Human interleukin-13 gene	IL-13Ra	Y10659	Above
2	37524_at	Human death-associated protein kinase	DRAK2	AB011421	Above
3	34243_i_at	Human l(3)mbt protein homolog mRNA		U89358	Above
4	41398_at	Homo sapiens mRNA. CDNA DKFZp564A186		AL049305	Above
5	35195_at	H. sapiens mRNA for phosphate cyclase		Y11651	Above
6	32393_s_at	Homo sapiens cDNA		W27466	Above
7	31909_at	Homo sapiens mRNA for KIAA0754 protein	KIAA0754	AB018297	Above

### **MLL**

1	294_s_at	Protein Kinase Pitslre, Alpha, Alt. Splice 1-Feb			Below
2	38226_at	23h11 Homo sapiens cDNA		W27152	Below
3	1398_g_at	Human protein kinase (MLK-3) mRNA	HUMMLK3A	L32976	Bbove
4	409_at	Human mRNA for 14.3.3 protein, a protein kinase regulator		X56468	Below

### **Others**

1	33782_r_at	nn82f03.s1 Homo sapiens cDNA, 3 end /clone=IMAGE-1090397		AA587372	Above
2	33338_at	Human transcription factor ISGF-3 mRNA		M97936	Above



3	40242_at	Human (clone N5-4) protein p84 mRNA		L36529	Above
4	37018_at	qd05c04.x1 Homo sapiens cDNA, 3 end /clone=IMAGE-1722822		AI189287	Above
5	38337_at	Homo sapiens zinc finger protein mRNA		U62392	Above
6	41464_at	Human mRNA for KIAA0339 gene	KIAA0339	AB002337	Above
7	38064_at	H.sapiens lrp mRNA	LRP	X79882	Above
8	33173_g_at	yc89b05.r1 Homo sapiens cDNA, 5 end /clone=IMAGE-23231		T75292	Below
9	33365_at	Homo sapiens mRNA for KIAA0945 protein	KIAA0945	AB023162	Above
10	39367_at	ni38e08.s1 Homo sapiens cDNA, 3 end /clone=IMAGE-979142		AA522537	Above
11	41108_at	Homo sapiens mRNA for putative GTP- binding protein	PGPL	Y14391	Above
12	37304_at	Homo sapiens heterochromatin protein p25 mRNA	P25beta	U35451	Below
13	40359_at	Human DNA-binding protein (HRC1) mRNA	HRC1	M91083	Above
14	32792_at	Human DNA sequence from clone 465N24 on chromosome 1p35.1-36.13. Contains two novel genes, ESTs, GSSs and CpG islands		AL031432	Above
15	34726_at	Human voltage-gated calcium channel beta subunit mRNA		U07139	Above
16	40299_at	Homo sapiens G-protein coupled receptor RE2 mRNA,		AF091890	Above
17	40704_at	H.sapiens mRNA for phosphatidylinositol 3- kinase		Z29090	Above
18	38568_at	Homo sapiens p53 binding protein mRNA		U82939	Above
19	32038_s_at	wi30c12.x1 Homo sapiens cDNA, 3 end /clone=IMAGE-2391766		AI739308	Above
20	39613_at	H.sapiens HUMM9 mRNA		X74837	Above

**Table 23. Permutation test results for predictors of T-ALL relapse**

Rank	Affymetrix number	t-statistic value	t-statistic		neighbors
			Perm 1%	Perm 5%	
1	33777_at	7.8337	7.3774	5.4783	6
2	41853_at	6.1727	6.5948	4.8117	16
3	38866_at	5.9890	6.0293	4.5611	12
4	41643_at	5.6106	5.6815	4.3877	12
5	1126_s_at	5.4777	5.5162	4.2375	11
6	41862_at	5.3734	5.3759	4.1208	11
7	41131_f_at	4.9134	5.2280	4.0295	17

**Table 24. Permutation test results for predictors of Hyperdiploid > 50 relapse**

Rank	Affymetrix number	t-statistics		Perm 5%	neighbors
		value	Perm 1%		
1	37721_at	8.7160	12.7358	9.9506	75
2	38721_at	8.4162	10.7256	8.8438	59
3	40120_at	7.2736	9.9837	8.0383	73
4	41386_i_at	6.3436	9.0552	7.5579	88
5	38677_at	6.2698	8.8633	7.2466	88
6	37620_at	6.2174	8.4154	6.9604	82
7	34703_f_at	6.0770	8.0982	6.8835	83
8	38355_at	5.5120	7.8657	6.7434	92
9	41214_at	5.4262	7.6583	6.6094	90
10	34530_at	5.4013	7.5991	6.5109	87
11	603_at	5.3142	7.5903	6.4409	87
12	32697_at	5.1785	7.5146	6.3265	90
13	41129_at	5.1450	7.3939	6.2121	88
14	33333_at	5.1061	7.2601	6.1389	87
15	37078_at	5.0738	7.1484	6.0308	86
16	38148_at	4.9256	6.9688	5.9230	93
17	39150_at	4.9061	6.9273	5.9015	93
18	33869_at	4.8256	6.8900	5.8367	93
19	41447_at	4.7919	6.8135	5.7621	93
20	39369_at	4.7790	6.7731	5.7391	92

Individually, the discriminating genes for relapse in T-ALL are significant at either the 1% or 5% level, while those for hyperdiploid >50 fall at approximately the 7% level.

**Table 25. Results of relapse prediction on indicated subgroups**

	Relapse	CCR	# genes	metric	Accuracy	P value by permutation test
T-ALL	8	26	7	t-stats	97	0.034
H>50	5	43	13	t-stats	100	0.018
<i>TEL-AML1</i>	3	56	7	CFS	100	0.145
<i>MLL</i>	5	7	4	t-stats	100	0.104
Others	4	56	20	t-stats	98.3	0.079

As the number of relapse samples were small, in addition to the usual cross validation experiments, we also performed 1000 permutation experiments for each subtype-specific relapse study. In each permutation experiment, we re-partitioned the samples in a manner that preserved class size by randomly swapping the class labels (“relapse” or “continuous complete remission”). Then we employed the same metric to pick the same number of genes as in the original partitioning of the samples given by the original class labels. SVM was then used to obtain a prediction accuracy by cross validation for this random partition using these freshly selected genes. The percentage of these 1000 permutation experiments was taken as a p-value that gave an indication on how many random partitions of the original

samples could achieve the same accuracy as the original samples. The results of these permutation experiments are summarized in the last column of Table 25 above. These results show that the high accuracy obtained on the predictability of relapse in T-lineage ALL, Hyperdiploid>50, and others are unlikely to be a random event. The p-values for the subtypes of *TEL-AML1* and *MLL* are weaker than the other subtypes. However, in the case of *TEL-AML1* the number of relapse samples were exceedingly small (3) and in the case of *MLL* the number of relapse and non-relapse samples were both very small.

### Results for secondary AML prediction

For the secondary AML prediction we also adopted the same subtype-specific approach as described earlier in relapse prediction. This time only the *TEL-AML1* subtype had sufficient number of samples for a secondary AML prediction model to be developed. For this model, we used the MIT score<sup>13</sup> to select genes and SVM to perform classification using these genes. The MIT score of a gene is defined as  $T = |\mu_1 - \mu_2| / (\sigma_1 + \sigma_2)$ , where  $\mu_i$  is the mean expression of that gene in the  $i^{\text{th}}$  class and  $\sigma_i$  is the standard deviation of that gene in the  $i^{\text{th}}$  class. This formula assigns higher value to a gene that has larger mean difference between two classes and has smaller variance within both classes. The 20 genes with the highest MIT scores in *TEL-AML1* patients that went into continuous complete remission vs those *TEL-AML1* samples that developed secondary AML are listed in Table 26 below. 100% accuracy for secondary AML prediction accuracy was achieved on *TEL-AML1* specific subtype samples using these 20 genes. We also performed a permutation test, in the same manner as described earlier in the subtype-specific relapse prediction, and obtained a p-value of 0.031 which suggested that the predictability of the development of secondary AML in *TEL-AML1* -specific patients was unlikely to be a random event.

**Table 26. Genes selected by MIT score for secondary AML**

	Affymetrix number	Gene Name	GeneSymbol	Reference number	Above/ Below mean
<i>TEL-AML1</i>					
1	34890_at	ATPase H transporting lysosomal vacuolar proton pump alpha polypeptide 70kD isoform 1	ATP6A1	L09235	Above
2	40925_at	hypothetical protein FLJ10803	FLJ10803	AA554945	Above
3	1719_at	mutS E. coli homolog 3	MSH3	U61981	Above
4	32877_i_at	EST IMAGE:954213		AA524802	Above
5	32650_at	neuronal protein	NP25	Z78388	Above
6	33173_g_at	hypothetical protein FLJ10849	FLJ10849	T75292	Above
7	32545_r_at	RSU-1/RSP-1	RSU-1	L12535	Above
8	34889_at	ATPase H transporting lysosomal vacuolar proton pump alpha polypeptide 70kD isoform 1	ATP6A1	AA056747	Above
9	35180_at	cDNA DKFZp586F1323 from clone DKFZp586F1323		AL050205	Above
10	34274_at	KIAA1116 protein	KIAA1116	AB029039	Above
11	35727_at	hypothetical protein FLJ20517	FLJ20517	AI249721	Above
12	1627_at	tyrosine kinase (GB:Z25437)		HG2715- HT2811	Above

13	1461_at	nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor alpha	NFKBIA	M69043	Below
14	36023_at	lacrimal proline rich protein	LPRP	AI864120	Above
15	39167_r_at	serine or cysteine proteinase inhibitor clade H heat shock protein 47 member 2	SERPINH2	D83174	Above
16	39969_at	H4 histone family member G	H4FG	AA255502	Above
17	38692_at	NGFI-A binding protein 1 ERG1 binding protein 1	NAB1	AF045451	Above
18	1594_at	polymerase RNA II DNA directed polypeptide C 33kD	POLR2C	J05448	Above
19	33234_at	RBP1-like protein	LOC51742	AA887480	Above
20	34739_at	hypothetical protein FLJ20275	FLJ20275	W26023	Above

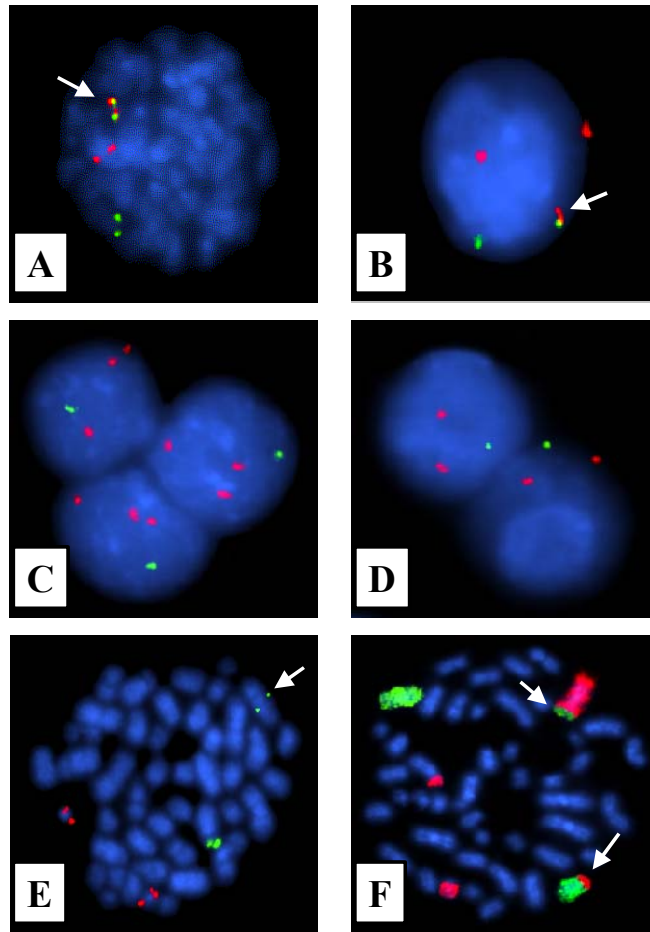
**Table 27. Permutation test results for secondary AML**

Rank	Affymetrix	t-statistics				
	number	number	Perm 1%	Perm 5%	Perm median	neighbors
1	34890_at	1.2204	2.7933	2.2138	1.4712	822
2	40925_at	1.0712	2.0006	1.7607	1.2884	859
3	1719_at	1.0599	1.8536	1.6272	1.1894	767
4	32877_i_at	1.0364	1.7125	1.5218	1.1200	715
5	32650_at	1.0217	1.6580	1.4584	1.0776	646
6	33173_g_at	1.0126	1.5868	1.4132	1.0416	595
7	32545_r_at	1.0097	1.5536	1.3630	1.0223	536
8	34889_at	0.9959	1.5164	1.3241	1.0009	512
9	35180_at	0.9854	1.4838	1.2938	0.9777	477
10	34274_at	0.9420	1.4759	1.2721	0.9600	550
11	35727_at	0.8493	1.4482	1.2507	0.9415	809
12	1627_at	0.8471	1.4207	1.2398	0.9254	782
13	1461_at	0.8312	1.4012	1.2260	0.9114	801
14	36023_at	0.8177	1.3551	1.2012	0.8995	813
15	39167_r_at	0.8136	1.3462	1.1806	0.8894	790
16	39969_at	0.8122	1.3395	1.1702	0.8785	759
17	38692_at	0.8109	1.3333	1.1565	0.8696	729
18	1594_at	0.8103	1.3142	1.1503	0.8626	696

Individually, no gene exceeds a significant level as a predictor of secondary AML. However, the combination of genes can be used in a supervised learning algorithm to accurately identify patients that eventually go on to develop secondary AML.

## FISH analysis

Below are the results of interphase and metaphase FISH analysis on 4 cases that were classified as *TEL-AML1* by microarray analysis, but were negative for a *TEL-AML1* chimeric transcript by RT-PCR.



**Figure 21.** Results from FISH analysis of the four cases that lacked a *TEL-AML1* chimeric transcript by RT-PCR but were found to have abnormalities of *TEL* by FISH. The *AML1* probe is red and the *TEL* probe is green (Vysis, Downer's Grove, IL). **(A)** Metaphase analysis of case 1 demonstrating a *TEL-AML1* fusion indicated by the arrow. **(B)** Interphase analysis of case 1 showing the *TEL-AML1* fusion. **(C)** Interphase FISH of case 2 showing trisomy of chromosome 21 and deletion of one allele of *TEL*. **(D)** Interphase FISH of case 3 showing loss of one *TEL* allele. **(E)** Metaphase analysis of case 4 showing a partial deletion of one *TEL* allele, indicated by the arrow. **(F)** Metaphase FISH of case 4 with painting probes for chromosome 7 (green) and 12 (red) showing a complex translocation as indicated by the arrows.

## References

---

- <sup>1</sup> Pui C-H, Rivera GK, Hancock ML et al. Risk-adapted treatment for acute lymphoblastic leukemia: findings from St. Jude Children's Research Hospital. In: Buchner T, Hiddemann W, Wormann B et al. eds. *Acute Leukemias. VI: Prognostic Factors and Treatment Strategies*. Haematology and Blood Transfusions. Springer-Verlag, Berlin, Heidelberg, 1997; 629-637.
- <sup>2</sup> Pui C-H, Boyett JM, Rivera GK, Hancock ML, Sandlund JT, Ribieiro RC et al. Long-term results of Total Therapy Studies 11, 12, and 13A for childhood acute lymphoblastic leukemia at St. Jude Children's Research Hospital. *Leukemia* 2000; 14:2286-94.
- <sup>3</sup> U.M.Fayyad and K.B.Irani, "Multi-interval Discretization of Continuous-valued Attributes", Proc. 19<sup>th</sup> International Joint Conference on Artificial Intelligence, Morgan Kaufmann, 1993, pages 1022-1027.
- <sup>4</sup> M.A.Hall and G.Holmes, "Benchmarking Attribute Selection Techniques for Data Mining", Working Paper 00/10, Department of Computer Science, University of Waikato, New Zealand, 2000.
- <sup>5</sup> H.Witten and E.Frank, "Data Mining: Practical Machine Learning Tools and Techniques with Java Implementation", Morgan Kaufmann, 1999.
- <sup>6</sup> J.Platt, "Fast Training of Support Vector Machines Using Sequential Minimal Optimization", in B.Scholkopf, C. Burges, and A.Smola (eds), "Advances in Kernel Methods---Support Vector Learning", MIT Press, 1998.
- <sup>7</sup> G.Dong and J.Li, "Efficient Mining of Emerging Patterns: Discovering Trends and Differences", Proc. 5<sup>th</sup> ACM SIGKDD International Conference on Knowledge Discovery and Data Mining", pages 43-52, 1999.
- <sup>8</sup> J.Li, G.Dong, and K.Ramamohanarao, "Making Use of the Most Expressive Jumping Emerging Patterns for Classification", *Knowledge and Information System*, 3:131-145, 2001.
- <sup>9</sup> J.Li, G.Dong, and K.Ramamohanarao, "DeEPs: Instance-based Classification by Emerging Patterns", Proc. 4<sup>th</sup> European Conference on Principles and Practice of Knowledge Discovery in Databases, pages 191-200, 2000.
- <sup>10</sup> G.Dong, X.Zhang, J.Li, and L.Wong, "CAEP: Classification by Aggregating Emerging Patterns", Proc. 2<sup>nd</sup> International Conference on Discovery Science, pages 30—42, 1999.
- <sup>11</sup> T.M.Cover and P.E.Hart, "Nearest Neighbor Pattern Classification", *IEEE Transactions on Information Theory*, 13:21—27, 1967.
- <sup>12</sup> D. Campana and F.G. Behm, "Immunophenotyping of Leukemia", *J of Immunologic Methods*, 243(1-2):59-75, 2000.
- <sup>13</sup> T.R.Golub, D.K.Slonim, P.Tamayo, C.Huard, M.Gaasenbeek, J.P.Mesirov, H.Coller, M.L.Loh, J.R.Downing, M.A. Caligiuri, C.D. Bloomfield, E.S. Lander, "Molecular classification of cancer: class discovery and class prediction by gene expression monitoring", *Science*, 286, pp531-537