

# Molecular profiling of CD34<sup>+</sup> cells identifies low expression of CD7, along with high expression of proteinase 3 or elastase, as predictors of longer survival in patients with CML

Agnes S. M. Yong, Richard M. Szydlo, John M. Goldman, Jane F. Apperley, and Junia V. Melo

Although most patients with chronic myeloid leukemia (CML) have the same initial molecular abnormality, the *BCR-ABL* fusion gene, the duration of chronic phase (CP) varies widely. To identify the possible molecular basis of this heterogeneity, we studied CD34<sup>+</sup> cells collected at diagnosis from 68 patients with CML-CP. By using oligonucleotide microarray screening, we performed gene-expression profiling on 2 subsets of patients, one comprising patients with an “aggressive disease” who developed blastic

transformation (BT) within 3 years of diagnosis (n = 10) and, at the other extreme, patients with an “indolent disease” whose BT occurred 7 or more years from diagnosis (n = 9). This screening revealed 20 genes differentially expressed in patients with aggressive and indolent disease, which were validated by quantitative reverse transcriptase/polymerase chain reaction (Q-RT/PCR). A multivariate Cox regression model identified the combination of low CD7 expression with high expression of proteinase 3 or elastase as

associated with longer survival in the complete cohort of 68 patients. This differential pattern of gene expression probably reflects the intrinsic heterogeneity of the disease; if so, assessing expression levels of selected genes at diagnosis may be valuable in predicting duration of survival in patients treated with imatinib and the newer tyrosine kinase inhibitors. (Blood. 2006;107:205-212)

© 2006 by The American Society of Hematology

## Introduction

Chronic myeloid leukemia (CML) is characterized cytogenetically by a reciprocal t(9;22)(q34;q11) chromosomal translocation which results in the formation of the *BCR-ABL* oncogene. This encodes a constitutively activated tyrosine kinase<sup>1</sup> responsible for the clinical features of the leukemia.<sup>2,3</sup> CML typically presents in chronic phase (CP), when the disease responds well to treatments such as hydroxyurea, interferon- $\alpha$ , or the new tyrosine kinase inhibitor, imatinib. Before the imatinib era, CP lasted on average 4 to 5 years before progressing to blastic transformation (BT), which usually proved fatal within a few months. Thus, until recently the median survival for patients treated predominantly with interferon- $\alpha$  was around 5 to 6 years,<sup>4</sup> but imatinib, now widely regarded as the best initial therapy for CML in CP, promises to prolong life very substantially.<sup>5</sup> However, despite the very encouraging responses to imatinib, there is good evidence that it does not eradicate all leukemic progenitor cells,<sup>6-8</sup> this in effect means that allogeneic stem cell transplantation (SCT), ideally performed in the early stage of the disease, is still the only curative treatment for CML.

Despite a consistent molecular finding, CML has always exhibited a marked heterogeneity in the duration of CP,<sup>4</sup> and preliminary experience suggests that this heterogeneity is also seen in patients treated with imatinib.<sup>5</sup> Various attempts have been made

to determine prognosis for individual patients at the time of diagnosis in CP. Thus, the Sokal and Hasford prognostic scores<sup>9,10</sup> proved moderately useful in predicting the duration of survival for individual patients treated with busulfan or interferon- $\alpha$ , respectively. The Sokal score also serves to discriminate survival without progression in patients treated with imatinib<sup>11</sup> and may still be useful to guide therapeutic decisions.<sup>12</sup> Deletions adjacent to the *ABL-BCR* junction on the derivative 9q+ chromosome may be associated with an adverse prognosis,<sup>13</sup> but only about 10% of patients with CML have this additional cytogenetic abnormality. Telomere lengths, which are already shorter in CML compared with normal cells, have been reported to be further shortened in patients in CP with early disease progression<sup>14</sup> or a more rapid onset of BT.<sup>15,16</sup>

We investigated the hypothesis that differential expression of specific genes at the time of diagnosis may underlie the intrinsic heterogeneity of CML, in which case profiling the expression of selected genes would be useful to predict the duration of survival even in patients treated with tyrosine kinase inhibitors. Using oligonucleotide microarrays, we identified 20 genes whose expression significantly differentiated between patients with aggressive versus indolent disease in an initial cohort of 19 patients diagnosed

From the Department of Haematology, Imperial College London, Hammersmith Hospital, London, United Kingdom.

Submitted May 31, 2005; accepted August 4, 2005. Prepublished online as *Blood* First Edition Paper, September 6, 2005; DOI 10.1182/blood-2005-05-2155.

Supported by grants from Malaysia Public Services Department, Lady Tata Memorial Trust, Hammersmith Hospitals Trust Research Committee (A.S.M.Y. and J.V.M.), and the Leukaemia Research Fund, United Kingdom (J.V.M.).

A.S.M.Y. collected the patient data, processed all samples, performed all experiments, analyzed the microarray data and wrote the report. R.M.S. performed the statistical analyses and helped write the report. J.F.A. and J.M.G. provided clinical care, recorded clinical data, provided advice on the

design of the study, and commented on the manuscript. J.V.M. conceived and designed the study, supervised its execution, and helped write the report. The corresponding author (J.V.M.) had full access to all the data in the study and had final responsibility for the decision to submit for publication.

**Reprints:** Junia V. Melo, Department of Haematology, Imperial College London, Hammersmith Hospital, Du Cane Rd, London W12 0NN, United Kingdom; e-mail: j.melo@imperial.ac.uk.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked “advertisement” in accordance with 18 U.S.C. section 1734.

© 2006 by The American Society of Hematology

in CP. We then confirmed these findings using quantitative real-time reverse transcription/polymerase chain reaction (Q-RT/PCR) and correlated gene expression with patient survival using a multivariate Cox regression model on 68 patients with CML. We found that the combination of a low expression level of *CD7*, with a high expression of proteinase 3 (*PR-3*) or elastase (*ELA2*), was highly predictive of a longer survival.

## Patients, materials, and methods

### Patients

Patients with Philadelphia (Ph) chromosome–positive CML whose nucleated cells were collected by leukapheresis and cryopreserved within 3 months of diagnosis before start of treatment were eligible for this study if the stored cells were still viable and if adequate clinical data were available. Informed consent for the use of these cells for research had been obtained according to the requirements of the Hammersmith, Queen Charlotte, and Chelsea and Acton Hospitals Research Ethics Committee (London, United Kingdom), and, where possible, in accordance with the Declaration of Helsinki. Diagnosis of CML-CP was based on clinical parameters and morphology of blood and bone marrow<sup>17</sup> and was confirmed by a karyotype exhibiting the t(9;22)(q34;q11) translocation with a single Ph chromosome. Patients were excluded if they (1) had variant translocations or additional cytogenetic abnormalities at diagnosis and (2) were in long-term remission after allogeneic SCT. Out of more than 400 patients, a cohort of 68 fulfilled all of these criteria; each of these 68 patients had been diagnosed with CML-CP and underwent leukapheresis within 3 months of diagnosis between October 1979 and July 2001. The majority of exclusions were due to the unavailability of cells collected and stored within 3 months of diagnosis. The median age at diagnosis of the selected patients was 45.2 years (range, 17.6–68.3 years). The male-to-female ratio was 1.8:1 (44 males, 24 females). The majority of patients were diagnosed in the era before imatinib and treated with hydroxyurea and/or interferon- $\alpha$ . Patient characteristics are shown in Table 1. Their median survival was 8.7 years. Patients who developed BT within 3 years of diagnosis were defined as having aggressive disease (n = 18), whereas those who survived for longer than 7 years prior to the onset of BT were defined as having indolent disease (n = 23). Patients who had survived between 3 and 7 years without developing BT were categorized as having intermediate disease (n = 27). The relatively low median age of patients in this series reflects the pattern of tertiary referral to the Hammersmith Hospital in London.

### Sample preparation and oligonucleotide microarray methods

Mononuclear cells from cryopreserved material were isolated by density gradient centrifugation (Lymphoprep; Nycomed, Oslo, Norway), and CD34<sup>+</sup> cells were selected by binding to immunomagnetic beads (Mini-MACS [magnetic-activated cell sorting]; Miltenyi Biotech, Bergisch-Gerbach, Germany). The mean CD34<sup>+</sup> purity was 92%  $\pm$  5.5%. Total RNA was extracted from CD34<sup>+</sup> cells using the Qiagen RNeasy kit (Qiagen, Crawley, United Kingdom). Of the 68 samples, only 19 (10 aggressive and 9 indolent CML) yielded sufficient RNA for gene expression profiling on HG-U133A GeneChip arrays (Affymetrix, Santa Clara, CA). The detailed protocol for sample preparation and microarray processing is available on line.<sup>18</sup> The 19 samples used for gene expression profiling were also analyzed for confirmation of the presence of a single Ph chromosome and for screening for a deletion of the 9q+ derivative by fluorescence in situ hybridization using the *BCR-ABL* LSI dual-color dual-fusion translocation probes (Vysis, Downers Grove, IL).

### Microarray data analysis

Microarray expression measures were generated from 3 data analysis packages: MAS 5.0 (Affymetrix, Santa Clara, CA),<sup>19</sup> dChip<sup>20</sup> (www.dchip.org), and RMA (Robust Multichip Average (www.bioconductor.org)),<sup>21</sup> which use different data extraction and normalization algorithms, and were loaded into GeneSpring 5.0 software (Silicon Genetics, Redwood City, CA). Expression measures derived from MAS 5.0 were normalized by using the median expression values of each array and each probe set in the experiment. Normalized expression measures generated from both MAS 5.0 and dChip (which use an invariant set normalization method) were then filtered by excluding probe sets with (1) expression measures within 2-fold of the local background in all arrays and (2) greater than 75% of arrays having “Absent” calls as assigned by the software. The expression measures between the indolent and aggressive CML groups were then compared using the Mann-Whitney *U* test. Expression measures from the 2 patient groups derived from RMA, which had undergone quantile normalization, were compared using a 2-tailed Student *t* test. A *P* value of less than .05 was considered significant and identified genes as differentially expressed between the 2 groups only if confirmed by all 3 analysis packages. In addition to minimizing the inclusion of probe sets identified because of an inherent bias present in a given algorithm, this methodology is also expected to allow the implementation of different measures to minimize nonbiologically relevant noise. Hierarchical clustering of the probe sets representing the most discriminatory genes was performed with HCE software (Human-Computer Interaction Lab, University of Maryland, College Park; www.cs.umd.edu/hcil/hce).

**Table 1. Patient characteristics**

Clinical characteristics	Microarray series by disease phenotype; n = 19		All patients by disease phenotype; n = 68		
	Aggressive; n = 10	Indolent; n = 9	Aggressive; n = 18	Intermediate; n = 27	Indolent; n = 23
Male/female	7/3	3/6	15/3	17/10	12/11
Age, y, median (range)	42.1 (22.3–55.9)	49.3 (34.8–61.0)	44.1 (20.6–56.0)	48.7 (27.5–68.3)	39.8 (20.8–64.1)
WBC count at diagnosis, $\times 10^9/L$ , median (range)	234.0 (45.7–361.0)	209.0 (82.8–336.0)	202.4 (45.7–393.0)	152.3 (16.1–506.0)	149.0 (25.9–379.0)
Sokal score, median (range)	0.89 (0.55–2.44)	0.94 (0.64–3.90)	0.92 (0.5–2.4)	1.09 (0.5–1.9)	1.0 (0.5–3.9)
<b>Treatment prior to BT*</b>					
BU	2	1	5	0	4
IFN	5	3	7	20	16
HU	9	4	14	7	8
Imatinib	0	2	2	19	11
Autograft	0	1	0	0	6
Allograft	2	0	3	0	0
Unknown	0	4	0	2	4

WBC indicates white blood cell; BU, busulfan; IFN, interferon- $\alpha$ ; HU, hydroxyurea; Autograft, autologous SCT; Allograft, allogeneic SCT; Unknown, not available because medical case records were irretrievable but most likely HU with or without IFN.

\*Main treatment during CP shown; most patients had more than one type of treatment over the course of their disease.

Probe sequences of each probe set on the array identified as differentially regulated between the aggressive and indolent CML groups were obtained from the Affymetrix website (<http://www.affymetrix.com>) and aligned with the most updated version of the mRNA sequence stored in the Reference Sequence (RefSeq) database curated by the National Center for Biotechnology Information (Bethesda, MD). Probe sets containing more than 2 (of 11) probe sequences mismatched with the consensus mRNA were considered unreliable for gene expression analysis.

**Q-RT/PCR amplification**

RNA was treated with DNase I (Invitrogen, Paisley, United Kingdom) to eliminate genomic DNA, and random hexamer primed cDNA was synthesized according to standard methods. Expression of the 20 genes identified by microarray screening as differentially regulated in the aggressive and indolent groups was assessed by Q-RT/PCR using the ABI PRISM 7700 sequence detection system (Applied Biosystems, Foster City, CA). All Q-RT/PCR reactions were performed on 25 µL volume (Tables 2 and 3). The ABI Assays-on-demand TaqMan probe-and-primer reagents were used according to the manufacturer’s instructions for 14 genes, *ELA2*, *PR-3*, *STARD1*, *AZU1*, *EDN*, *DF*, *MBP*, *CG*, *CLC*, *GFI1*, *CXCR4*, *HIST1H2BG*, *HTM4*, and *CEBPA*. Fluorogenic LUX primers (Invitrogen) were used for *MPO*, *CD7*, *SMA4*, *FUT4*, *ECP*, and *HM74* according to the manufacturer’s instructions. *ABL* expression was used as the endogenous cDNA quantity control for all samples.<sup>22</sup> Its expression was measured using the Q-RT/PCR core kit (Eurogentec, Romsey, United Kingdom), 300 nM primers and 200 nM probe.

**Flow cytometry**

Fluorescence-activated cell sorting (FACS) was performed with the FACS-Calibur flow cytometer and analyzed using CellQuestPRO software (Becton Dickinson, Cowley, Oxford, United Kingdom). Lineage and maturation markers were assessed on CD34+ cells from 16 of the 19 samples included in the microarray experiments. The panel of monoclonal antibodies consisted of fluorescein isothiocyanate (FITC)-conjugated lineage markers CD2, CD19, CD14, CD16, CD61, and glycophorin A (Caltag, Burlingame, CA); phycoerythrin (PE)-conjugated CD38; peridinin chlorophyll protein (PerCP)-conjugated CD34; and allophycocyanin (APC)-conjugated CD7. A minimum of 5 × 10<sup>5</sup> cells were labeled with 2 µL each of the FITC-conjugated lineage cocktail, CD38-PE and CD7-APC and 10 µL CD34-PerCP, incubated on ice for 30 minutes, and subsequently washed in phosphate-buffered saline (PBS) prior to analysis. A minimum of 100 000 events were acquired.

**Table 2. Primers and probes for TaqMan and LUX**

Gene	Designation	Sequence (5'→3')
<i>ABL</i>	TaqMan forward primer	GATACGAAGGGAGGGTGTACCA
<i>ABL</i>	TaqMan reverse primer	CTCGCCAGGGTGTGAA
<i>ABL</i>	TaqMan probe	TGCTTCTGATGGCAAGCTCTACGTCTCCT
<i>ABL</i>	LUX forward primer	caccaaATCAATGGCAGCTTCTTGGtG
<i>ABL</i>	LUX reverse primer	CACCCTCCCCTCGTATCTCAGC
<i>MPO</i>	LUX forward primer	gtaccgTCGCCTGCATCATCGGtAC
<i>MPO</i>	LUX reverse primer	ACCCTCGTTCTCCACCAAA
<i>CD7</i>	LUX forward primer	cagccTCTCCTTCTCCTCGGGctG
<i>CD7</i>	LUX reverse primer	CGCCACGAGCACAGTTTCTT
<i>ECP</i>	LUX forward primer	cacagtGCGGGTGCCTTTACTCCACTGtG
<i>ECP</i>	LUX reverse primer	ACCTCCTTCCCTGGTCTGTCTGc
<i>FUT4</i>	LUX forward primer	gactgcGAGAAAGGTGAGGAGGGCAGtC
<i>FUT4</i>	LUX reverse primer	GCACTTTGCTTCCATTACAGG
<i>HM74</i>	LUX forward primer	caacacaTGAGATTGGAGGGAATTGTGtG
<i>HM74</i>	LUX reverse primer	CACGGAAGCCAGGTGACCTTAC
<i>SMA4</i>	LUX forward primer	cacggTCTGTATGTGGATGTGATCCGtG
<i>SMA4</i>	LUX reverse primer	AATCAACTCCAGGTGCCCGtA

The 5 to 7 bases in lower case at the 5' end of the LUX forward primers are complementary to the 3' end of the primer to form a hairpin. The single base in lower case at the 3' end of the primer is the site of fluorophore attachment.

**Table 3. Assays-on-demand primer-and-probe sets**

Gene	Assays-on-demand code
<i>AZU1</i>	Hs00156049_m1
<i>CEBPA</i>	Hs00269972_s1
<i>CLC</i>	Hs00171342_m1
<i>CG</i>	Hs00175195_m1
<i>CXCR4</i>	Hs00607978_s1
<i>DF</i>	Hs00157263_m1
<i>EDN</i>	Hs00795553_s1
<i>ELA2</i>	Hs00357734_m1
<i>GFI1</i>	Hs00382207_m1
<i>HIST1H2BG</i>	Hs00374317_s1
<i>HTM4</i>	Hs00193192_m1
<i>MBP</i>	Hs00794928_m1
<i>PR-3</i>	Hs00160521_m1
<i>STARD1</i>	Hs00264912_m1

For proteinase 3 (PR-3) intracellular staining, CD34+ cells were fixed with 4% paraformaldehyde for 10 minutes at room temperature, permeabilized with 0.3% saponin in 0.5% bovine serum albumin-supplemented PBS, and stained with 3 µL murine CLB-12.8 clone (a kind gift from Dr Y. M. van der Geld, Groningen, The Netherlands) for 1 hour at room temperature. After washing in 0.3% saponin PBS, the cells were incubated for 30 minutes at room temperature with 2 µL goat-anti-mouse Ig conjugated to FITC. For identification of cells with dual CD7 and CD34 surface markers, CD34+ cells were incubated on ice for 30 minutes with 2 µL CD7-FITC (Caltag) and 10 µL CD34-PerCP (Becton Dickinson) or 2 µL CD34-PE (Caltag) conjugated monoclonal antibodies. After washing in PBS, a minimum of 30 000 events were acquired.

**Statistical methods**

The results obtained from microarray and Q-RT/PCR assays were compared by Spearman ρ. Survival curves were calculated by using the Kaplan-Meier method, and groups were compared using the log-rank test. Patients were divided into groups by using Q-RT/PCR or protein expression values (FACS measurement) delineated by the median, upper, or lower quartile cutoffs. Genes with discriminating expression values identified from the univariate analysis with *P* values of less than .20 were entered into a Cox regression analysis, and a forward and backward stepping procedure was used to find the best model to predict survival. All quoted *P* values are from 2-sided tests, with values less than .05 considered significant.

**Results**

**Standard prognostic indicators**

Sokal prognostic scores were calculated, and comparisons were made between the 3 prognostic groups, but in this group of patients there were no differences in survival probabilities (*P* = .90) (Table 1). The derivative 9q+ deletion<sup>13</sup> was found in only 1 patient of the 19 included for microarray screening. There was no difference in the proportion of the CD34+ subsets of CD38+Lin+, CD38+Lin-, or CD38-Lin- cells between patients in the indolent as compared with the aggressive CML groups (data not shown).

**Gene expression profile of CML CD34+ cells**

Microarray files are available from <http://www.ebi.ac.uk/arrayexpress/query/login>. The expression measures of 32 probe sets, from a total of 22 283 present on the HG-U133A GeneChip array, were identified as significantly different between the indolent and aggressive CML groups using the combination of MAS 5.0, dChip, and RMA software. Subsequently, the exclusion of probe sets with

**Table 4. Twenty-five most significant probe sets representing 20 genes identified by MAS 5.0, dChip, and RMA**

Greater expression by disease phenotype	Common name	Affymetrix accession no.	GenBank accession no.	Description
Agg	<i>CD7</i>	214049_x_at	AI829961	CD7 antigen (p41)
Agg	<i>CD7</i>	214551_s_at	NM_006137	CD7 antigen (p41)
Agg	<i>SMA4</i>	215599_at	NM_021652	Spinal muscular atrophy gene 4
Agg	<i>HIST1H2BG</i>	215779_s_at	NM_003518	H2B histone family, member A
Agg	<i>HIST1H2BG</i>	210387_at	BC001131	H2B histone family, member A
Ind	<i>EDN</i>	206111_at	NM_002934	Eosinophil-derived neurotoxin
Ind	<i>ECP</i>	206851_at	NM_002935	Eosinophil cationic protein
Ind	<i>MBP</i>	211743_s_at	NM_002728	Eosinophil granule major basic protein
Ind	<i>STARD1</i>	204548_at	NM_000349	Steroidogenic acute regulatory protein
Ind	<i>CXCR4</i>	209201_x_at	L01639	Chemokine (C-X-C motif), receptor 4 (fusin)
Ind	<i>CXCR4</i>	211919_s_at	AF348491	Chemokine (C-X-C motif), receptor 4 (fusin)
Ind	<i>CXCR4</i>	217028_at	AJ224869	Chemokine (C-X-C motif), receptor 4 (fusin)
Ind	<i>HTM4</i>	210254_at	NM_006138	Membrane-spanning 4-domains, subfamily A, member 3 (hematopoietic cell-specific)
Ind	<i>HM74</i>	205220_at	NM_006018	Nicotinic acid receptor; GTP-binding protein
Ind	<i>AZU1</i>	214575_s_at	NM_001700	Azurocidin 1
Ind	<i>PR-3</i>	207341_at	NM_002777	Proteinase 3
Ind	<i>ELA2</i>	206871_at	NM_001972	Elastase 2
Ind	<i>GFI1</i>	206589_at	NM_005263	Growth factor independent 1
Ind	<i>CG</i>	205653_at	NM_001911	Cathepsin G
Ind	<i>DF</i>	205382_s_at	NM_001928	D component of complement (adipsin)
Ind	<i>MPO</i>	203949_at	NM_000250	Myeloperoxidase
Ind	<i>MPO</i>	203948_s_at	J02694	Myeloperoxidase
Ind	<i>CEBPA</i>	204039_at*	NM_004364	CCAAT/enhancer binding protein, alpha
Ind	<i>CLC</i>	206207_at*	NM_001828.3	Charcot-Leyden crystal protein
Ind	<i>FUT4</i>	209893_s_at†	NM_002033	Fucosyltransferase 4

Agg indicates aggressive CML; Ind, indolent CML.

\*Probe sets with single nucleotide mismatch according to RefSeq in 2 of 11 probe sequences.

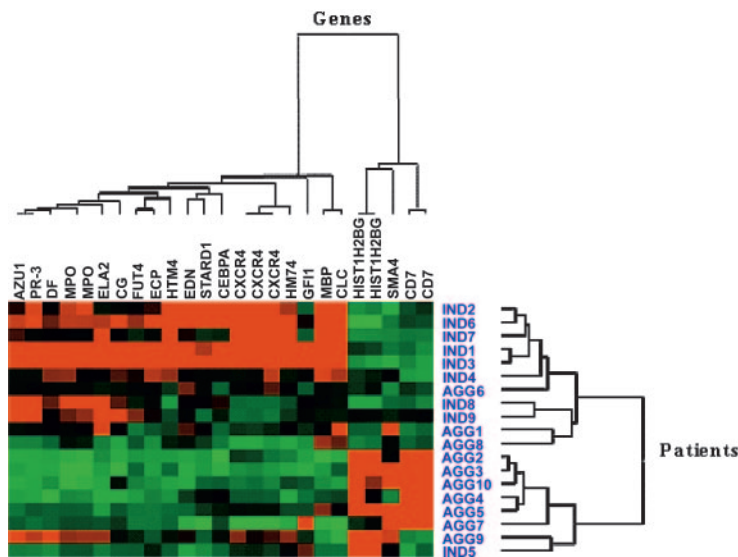
†*FUT4* has 2 probe sets in the array; the other probe set *209892\_at* was only significantly different by dChip and RMA.

sequences mismatching public mRNA databases and those bearing gene sequences which were Y-chromosome associated (excluded as there were more men in the aggressive CML group) yielded 25 probe sets, representing 20 genes (Table 4). Hierarchical clustering using the expression of these genes demonstrated the demarcation between indolent and aggressive CML groups (Figure 1).

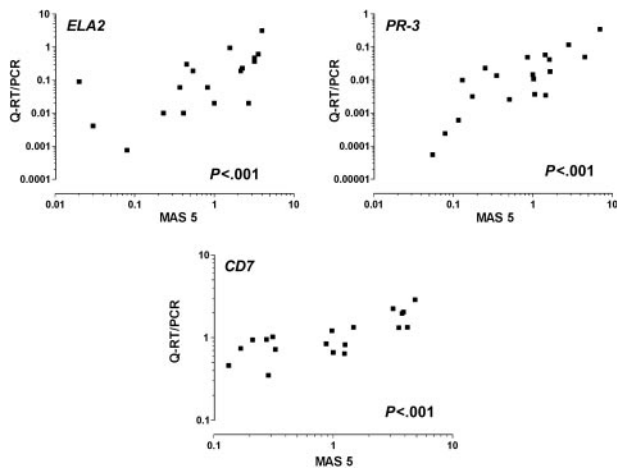
There was good correlation between the measurements of gene expression by microarray hybridization and Q-RT/PCR amplification, with 18 of the 20 genes having *P* values less than .005, *STARD1* *P* = .033, and only *FUT4* with poor correlation (*P* = .21) (Figure 2).

**Association between gene expression and survival**

Univariate analyses of survival using expression measures from Q-RT/PCR for each of the 20 genes identified *STARD1*, *MBP*, *ELA2*, *AZU1*, *PR-3*, *CD7*, and *SMA4* as being associated with differential survival in the entire cohort of 68 patients (Table 5). In Cox regression analyses, a backward-stepping approach found high expression of *CD7* (relative risk [RR], 2.42; 95% confidence interval [CI], 1.16-5.03) and low expression of *ELA2* (RR, 3.26; 95% CI, 1.28-8.32) to be associated with poor survival, whereas a forward stepping method disclosed high expression of *CD7* (RR,



**Figure 1. Hierarchical clustering using microarray expression measures on 20 most significantly different genes, represented by 25 probe sets, in 19 patient samples.** Ind indicates indolent CML; Agg, aggressive CML. Red squares indicate genes that were overexpressed, and green squares indicate genes that were underexpressed. Note that in 4 cases individual genes were studied with more than one probe set.



**Figure 2.** Representative scatter plots showing a good correlation between the expression of *ELA2*, *PR-3*, and *CD7* measured by microarray (MAS 5.0, using an arbitrary scale) and Q-RT/PCR methods (*P* values refer to Spearman  $\rho$ ).

2.28; 95% CI, 1.11-4.67) and low expression of *PR3* (RR, 2.30; 95% CI, 1.11-4.78). Combinations of these variables enabled patients with poor risk to be identified (Figure 3).

Measurement of gene expression at the protein level by FACS confirmed that the percentage of CD34+ cells with *PR-3* (n = 25) and *CD7* (n = 29) expression was also of prognostic value by a univariate analysis (*P* = .029 and *P* = .031, respectively) (Figure 4). Patients who had greater than 11% CD34+ cells expressing surface *CD7* protein at diagnosis had a poor survival. Conversely, patients with greater than 2% CD34+ cells expressing cytoplasmic *PR-3* had a superior survival.

## Discussion

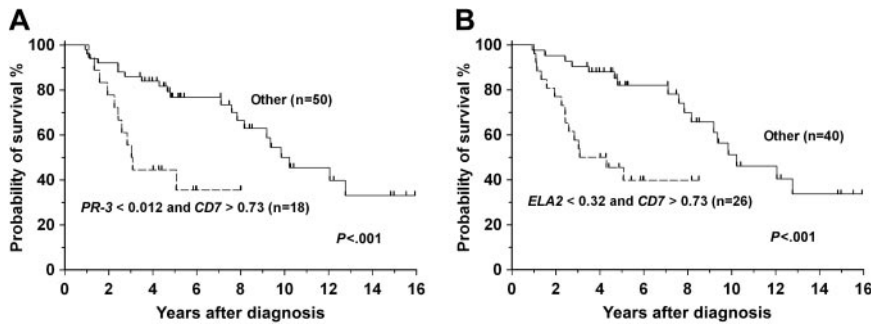
The prognosis for patients with CML depends on the stage of the disease at presentation,<sup>9</sup> but, even for patients diagnosed in CP, there is significant variability in survival.<sup>4,23</sup> Many efforts have in the past been made to predict survival for individual patients diagnosed in CP but neither the Sokal<sup>9</sup> nor the Hasford<sup>10</sup> systems, both currently in common use, provide the prognostic precision necessary to make important therapeutic decisions, notably to decide for a given patient whether to offer initial treatment with antileukemia drugs or to proceed immediately to allogeneic SCT. Moreover, the ability to characterize a patient's leukemia before starting treatment with a tyrosine kinase inhibitor would contribute to the decision whether to change therapy if the patient's subsequent response were suboptimal. Here, we compared gene expression in CD34+ cells collected at diagnosis from patients with CML who proved subsequently to have either relatively short or relatively long survivals and showed that 20 genes were differentially regulated. In particular, patients whose CD34+ cells had increased expression of *CD7* in combination with a low expression of either *ELA2* or *PR-3* had a poor survival probability. Furthermore, the prognostic information derived from the differential level of expression of these genes in patients with CML was confirmed at the protein level for the 2 genes that could be tested by FACS analysis, namely *PR-3* and *CD7*.

The adverse influence on prognosis of increased numbers of CD34+ cells with increased *CD7* expression in patients with CML in CP has been suggested previously.<sup>24,25</sup> Recently, Kosugi et al<sup>26</sup> have shown that CD34+CD7+ cells in CML may be involved in

**Table 5.** Univariate analyses of survival using Q-RT/PCR expression of 20 genes

Gene*	N	Median survival, y (95% CI)	<i>P</i>
<b>AZU1</b>			.024
Less than 0.056	16	3.5 (2.5-4.5)	
More than 0.055	52	9.4 (7.5-11.3)	
<b>CD7</b>			.028
Less than 0.73	34	9.8 (8.4-11.3)	
More than 0.72	34	7.8 (4.1-11.6)	
<b>CEBPA</b>			.56
Less than 0.074	34	12.8 (5.2-20.4)	
More than 0.073	34	9.2 (7.1-11.3)	
<b>CG</b>			.51
Less than 0.098	32	9.8 (2.3-17.4)	
More than 0.097	32	9.2 (7.3-11.1)	
<b>CLC</b>			.82
Less than 2.22	32	9.8 (2.3-17.4)	
More than 2.21	32	8.2 (5.99-10.4)	
<b>CXCR4</b>			.19
Less than 1.15	34	8.7 (2.9-14.5)	
More than 1.14	34	9.4 (6.8-12.0)	
<b>DF</b>			.13
Less than 0.073	34	5.1 (0-11.8)	
More than 0.072	34	9.4 (7.2-11.5)	
<b>ECP</b>			.14
Less than 0.067	32	8.7 (1-16.4)	
More than 0.066	32	9.4 (6.7-12.0)	
<b>EDN</b>			.15
Less than 0.035	32	8.7 (3.0-14.4)	
More than 0.034	32	9.4 (6.8-11.9)	
<b>ELA2</b>			.021
Less than 0.32	50	8.7 (4.1-13.3)	
More than 0.31	16	12.1 (8.9-14.1)	
<b>FUT4</b>			.82
Less than 3.34	32	9.2 (4.3-14.1)	
More than 3.33	32	8.2 (5.8-10.5)	
<b>GFI1</b>			.52
Less than 0.029	32	9.2 (4.3-14.1)	
More than 0.028	32	8.2 (3.2-13.3)	
<b>HIST1H2BG</b>			.89
Less than 0.56	32	8.7 (7.1-10.3)	
More than 0.55	32	3.6 (2.3-16.4)	
<b>HM74</b>			.45
Less than 0.071	32	9.8 (0-20.9)	
More than 0.070	32	9.2 (7.6-10.8)	
<b>HTM4</b>			.26
Less than 0.019	32	8.7 (3.6-13.8)	
More than 0.018	32	9.2 (7.4-11.0)	
<b>MBP</b>			.016
Less than 0.045	16	3.5 (0-7.0)	
More than 0.044	48	9.4 (7.2-11.5)	
<b>MPO</b>			.42
Less than 0.021	34	9.8 (2-17.7)	
More than 0.020	34	9.2 (7.2-11.5)	
<b>PR-3</b>			.027
Less than 0.011	34	5.1 (0.9-9.2)	
More than 0.010	34	12.1 (7.2-16.9)	
<b>SMA4</b>			.047
Less than 1.65	32	9.2 (4.5-13.8)	
More than 1.64	32	9.4 (2.3-16.4)	
<b>STARD1</b>			< .001
Less than 0.027	32	5.1 (4.4-5.7)	
More than 0.026	32	12.1 (7.1-17.0)	

\*Expression values by Q-RT/PCR are given as a ratio of amplification of the individual gene and the *ABL* control gene.

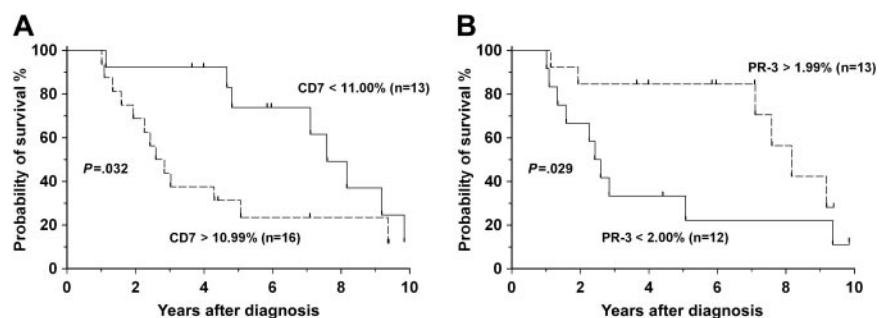


**Figure 3. Probability curves based on Cox regression analysis models.** A backward-stepping procedure (B) yielded a model with the combination of low *ELA2* and high *CD7* as predictive of poor survival ( $P < .001$ ). A forward-stepping procedure (A) yielded a model with the combination of low *PR-3* and high *CD7* as predictive of poor survival ( $P < .001$ ). Values of genes represent the Q-RT/PCR expression as a ratio of the gene of interest to the *ABL* control gene.

clonal maintenance and evolution. Furthermore, PR-3 and neutrophil elastase (NE), which are encoded by the *PR-3* and *ELA2* genes, respectively, have also been implicated as possible prognostic indicators in CML.<sup>27,28</sup> Both proteins are serine proteases that accumulate in the primary azurophilic granules of granulocytes and are overexpressed in CML as compared with normal progenitors. Patients with CML have cytotoxic T lymphocytes (CTLs) against PR1, a nonapeptide common to PR-3<sup>29</sup> and NE,<sup>28</sup> and the degree of cytotoxicity correlates with the amount of PR-3 expression.<sup>30</sup> Moreover the presence of PR1-directed CTLs has been correlated with clinical responses to interferon- $\alpha$  and allogeneic SCT.<sup>27</sup> Our study provides the first demonstration that higher levels of *ELA2* and *PR-3* expression are in fact associated with a longer survival and suggests that in some patients with CML an immune-mediated effect, possibly involving CTLs directed against overexpressed antigens such as NE and PR-3, may render the chronic phase of CML relatively indolent and thereby reduce the risk of progression to BT.

Microarray screening is currently the most powerful tool to identify differences in the transcriptome between 2 or more populations of cells/individuals. In the case of CML, the use of CD34<sup>+</sup> progenitors, which are better representatives of the clonogenic leukemia population than mixed mononuclear or total leukocyte fractions,<sup>31,32</sup> should provide a profile more biologically reflective of disease phenotype in individual patients. Our preliminary experiments, which have been confirmed by others,<sup>33</sup> demonstrated substantial differences in gene expression profile between mononuclear cells and CD34<sup>+</sup> progenitors from the same patients, which were a consequence of the different cell populations present in the mononuclear pool (data not shown). We reasoned that to avoid the potential difficulties in interpreting data from highly variable mixed cell populations in total leukocytes between individuals, the use of CD34<sup>+</sup> progenitors for gene profiling in our study was preferred. This reasoning has been strengthened by a recent report that expression profiling of *unselected* total leukocytes was unhelpful in differentiating 2 populations of patients with CML in CP with varying responses to imatinib treatment.<sup>32</sup> Although the

CD34<sup>+</sup> progenitor pool is still a relatively heterogeneous population of cells at differing stages of “stemness,” we detected no significant differences in the composition of CD38<sup>+</sup>Lin<sup>+</sup>, CD38<sup>+</sup>Lin<sup>-</sup>, or CD38<sup>-</sup>Lin<sup>-</sup> cells between the 2 groups of patients. We concede that our results may still reflect a heterogeneity in the maturation profile of the bulk of the CD34<sup>+</sup> population in patients with CML at diagnosis, as distinct from differences in equivalent leukemia cells between different patients, but even this heterogeneity may be of prognostic significance. We were not able to ascertain the proportion of *BCR-ABL*-negative CD34<sup>+</sup> cells in all the patients included in this study because of the lack of biologic material. However, it is expected that at least 70% of CD34<sup>+</sup> cells would be leukemic from previously published observations<sup>26,31</sup> of similarly unmanipulated CD34<sup>+</sup> cells from untreated patients with CML CP at diagnosis. Our choice of the 2 extremes of disease patterns defined by the duration of CP from diagnosis to the onset of BT aimed at maximizing the identification of genes underlying the extreme prognostic heterogeneity characteristic of CML. Because all patients in our cohort had the same chromosomal and molecular abnormality, the gene profiles of the groups under comparison were expected to have a greater homogeneity than, for example, those with acute leukemias, whereby a variety of chromosomal abnormalities are present.<sup>34,35</sup> We increased the stringency of our analysis by obtaining gene expression measures using 3 separate microarray data analysis packages, MAS 5.0,<sup>19</sup> dChip,<sup>20</sup> and RMA,<sup>21</sup> to identify the most consistently discriminatory genes. Our data must be interpreted with some caution because we had insufficient patient numbers to divide our cohort into “training set” versus “testing set” groups for separate microarray tests. However, even if this had been possible, the results might not have been entirely conclusive because, as recently pointed out by Michiels et al,<sup>36</sup> even in large patient series, which genes are identified as prognostically important is highly dependent on the actual samples included in the training set. Taken together, our data could therefore be used as a guide to ensure that the genes we have identified are included in future prospective studies involving larger numbers of patients.



**Figure 4. Probabilities of survival using CD7 and PR-3 protein expression.** The median protein expression by FACS is used to segregate the patients into 2 groups. The expression of both CD7 (A) ( $n = 29$ ) and PR-3 (B) ( $n = 25$ ) are significant predictors of survival ( $P = .031$  and  $P = .029$ , respectively).

The majority of patients included in this study were treated with agents in the era before imatinib, and it is possible that initial treatment of patients with CML with this tyrosine kinase inhibitor will eliminate the heterogeneity in survival observed with prior therapy. However, data from the prospective studies of imatinib<sup>5</sup> and from single center experience<sup>37</sup> reveal that significant variability still exists in both response to imatinib and rate of relapse among patients in CP with similar hematologic and clinical features.<sup>5,11</sup> These differences in response rates are shown to translate into heterogeneity in progression-free survival.<sup>11</sup> The mechanisms underlying this heterogeneity in treatment response are still unknown. Recently, the combination of the Sokal score and the concentration of imatinib producing a 50% decrease in the level of its Crkl substrate was shown to have a strong predictive value in a subset of patients.<sup>12</sup> These results, taken together with those presented here, suggest that the heterogeneity in treatment response and associated progression-free survival are related to the biology of the CML cell in individual patients. This, in turn, may be governed by transcriptional or metabolic pathways inherent to the CD34<sup>+</sup> progenitor population at diagnosis.

Previous microarray studies involving patients with CML have compared the gene signatures of normal with leukemic cells, or of cells from different stages of CML.<sup>38,39</sup> These approaches were not able to identify subtle gene expression differences within the patients in CP and did not take into account possible differences in disease tempo between the patients studied. We aimed to avoid this issue by using samples within 3 months of diagnosis and excluded patients who, despite

fulfilling morphologic CP criteria, had additional cytogenetic abnormalities. Patients in long-term remission after allogeneic SCT were also excluded to eliminate bias toward the good prognostic groups attributable to the benefits of a successful allograft. Despite the relatively small numbers of patient samples included in this study, we obtained clear-cut results. Thus, we believe that the expression levels of *ELA2*, *PR-3*, and *CD7* in progenitor cells from CML at diagnosis may constitute important new prognostic markers. For practical purposes, expression of these genes could readily be measured at the RNA or protein levels by Q-RT/PCR or immunophenotyping of CD34<sup>+</sup> cells, laboratory assays which are now widely available in clinical practice. Our results offer a timely opportunity to exploit the prognostic markers here identified in the design of prospective trials of imatinib and of the promising new tyrosine kinase inhibitors.<sup>40-44</sup> They also reinforce the rationale for treating patients who have not responded optimally to Abl tyrosine kinase inhibitors by vaccination with relevant peptides, and these studies are now being planned.

## Acknowledgments

We thank the staff of the Stem Cell Laboratory and the CSC/IC Microarray Centre, Hammersmith Hospital, for help with identification of stored cells and practical assistance with the microarray studies, and Dr Eric Hoffman and Sara Hilmer, Research Center for Genetic Medicine at the Children's National Medical Center, Washington, DC, for advice on microarray data analysis.

## References

- Deininger MW, Goldman JM, Melo JV. The molecular biology of chronic myeloid leukemia. *Blood*. 2000;96:3343-3356.
- Deininger MW, Goldman JM, Lydon NB, Melo JV. The tyrosine kinase inhibitor CGP57148B selectively inhibits the growth of BCR-ABL positive cells. *Blood*. 1997;90:3691-3698.
- Druker BJ, Talpaz M, Resta DJ, et al. Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. *N Engl J Med*. 2001;344:1031-1037.
- Kantarjian HM, O'Brien S, Cortes JE, et al. Complete cytogenetic and molecular responses to interferon-alpha-based therapy for chronic myelogenous leukemia are associated with excellent long-term prognosis. *Cancer*. 2003;97:1033-1041.
- O'Brien SG, Guilhot F, Larson RA, et al. Imatinib compared with interferon and low-dose cytarabine for newly diagnosed chronic-phase chronic myeloid leukemia. *N Engl J Med*. 2003;348:994-1004.
- Graham SM, Jorgensen HG, Allan E, et al. Primitive, quiescent, Philadelphia-positive stem cells from patients with chronic myeloid leukemia are insensitive to ST1571 in vitro. *Blood*. 2002;99:319-325.
- Bhatia R, Holtz M, Niu N, et al. Persistence of malignant hematopoietic progenitors in chronic myelogenous leukemia patients in complete cytogenetic remission following imatinib mesylate treatment. *Blood*. 2003;101:4701-4707.
- Michor F, Hughes TP, Iwasa Y, et al. Dynamics of chronic myeloid leukaemia. *Nature*. 2005;435:1267-1270.
- Sokal JE, Cox EB, Baccarani M, et al. Prognostic discrimination in "good-risk" chronic granulocytic leukemia. *Blood*. 1984;63:789-799.
- Hasford J, Pfirrmann M, Hehlmann R, et al. A new prognostic score for survival of patients with chronic myeloid leukemia treated with interferon alfa. Writing Committee for the Collaborative CML Prognostic Factors Project Group. *J Natl Cancer Inst*. 1998;90:850-858.
- Hughes TP, Kaeda J, Branford S, et al. Frequency of major molecular responses to imatinib or interferon alfa plus cytarabine in newly diagnosed chronic myeloid leukemia. *N Engl J Med*. 2003;349:1423-1432.
- White D, Saunders V, Lyons AB, et al. In vitro sensitivity to imatinib-induced inhibition of ABL kinase activity is predictive of molecular response in de novo CML. *Blood*. 2005;106:2520-2526.
- Huntly BJ, Bench A, Green AR. Double jeopardy from a single translocation: deletions of the derivative chromosome 9 in chronic myeloid leukemia. *Blood*. 2003;102:1160-1168.
- Boultonwood J, Peniket A, Watkins F, et al. Telomere length shortening in chronic myelogenous leukemia is associated with reduced time to accelerated phase. *Blood*. 2000;96:358-361.
- Brummendorf TH, Holyoake TL, Rufer N, et al. Prognostic implications of differences in telomere length between normal and malignant cells from patients with chronic myeloid leukemia measured by flow cytometry. *Blood*. 2000;95:1883-1890.
- Drummond M, Lennard A, Brummendorf T, Holyoake T. Telomere shortening correlates with prognostic score at diagnosis and proceeds rapidly during progression of chronic myeloid leukemia. *Leuk Lymphoma*. 2004;45:1775-1781.
- Vardiman JW, Pierre R, Thiele J, et al. Chronic myelogenous leukaemia. In: Jaffe ES, Harris NL, Stein H, Vardiman JW, eds. *World Health Organization Classification of Tumours. Pathology and Genetics of Tumours of Haematopoietic and Lymphoid Tissues*. Lyon, France: IARC Press; 2001: 20-26.
- CSC/IC Microarray Centre. Protocol for sample preparation and microarray processing. [http://microarray.csc.mrc.ac.uk/\\_private/Support/affy\\_protocols\\_text.htm](http://microarray.csc.mrc.ac.uk/_private/Support/affy_protocols_text.htm). Accessed December 20, 2004.
- Irizarry RA, Bolstad BM, Collin F, et al. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res*. 2003;31:e15.
- Li C, Wong WH. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc Natl Acad Sci U S A*. 2001;98:31-36.
- Irizarry RA, Hobbs B, Collin F, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*. 2003;4:249-264.
- Beillard E, Pallisgaard N, Van DV, et al. Evaluation of candidate control genes for diagnosis and residual disease detection in leukemic patients using "real-time" quantitative reverse-transcriptase polymerase chain reaction (RQ-PCR): a Europe Against Cancer program. *Leukemia*. 2003;17:2474-2486.
- Cortes J. Natural history and staging of chronic myelogenous leukemia. *Hematol Oncol Clin North Am*. 2004;18:569-84, viii.
- Normann AP, Egeland T, Madhus IH, Heim S, Tjonnfjord GE. CD7 expression by CD34<sup>+</sup> cells in CML patients, of prognostic significance? *Eur J Haematol*. 2003;71:266-275.
- Martin-Henao GA, Quiroga R, Sureda A, Garcia J. CD7 expression on CD34<sup>+</sup> cells from chronic myeloid leukaemia in chronic phase. *Am J Hematol*. 1999;61:178-186.
- Kosugi N, Ebihara Y, Nakahata T, et al. CD34<sup>+</sup> CD7<sup>+</sup> leukemic progenitor cells may be involved in maintenance and clonal evolution of chronic myeloid leukemia. *Clin Cancer Res*. 2005;11:505-511.
- Moldrem JJ, Lee PP, Wang C, et al. Evidence that specific T lymphocytes may participate in the

- elimination of chronic myelogenous leukemia. *Nat Med*. 2000;6:1018-1023.
28. Fujiwara H, El Ouriaghli F, Grube M, et al. Identification and in vitro expansion of CD4<sup>+</sup> and CD8<sup>+</sup> T cells specific for human neutrophil elastase. *Blood*. 2004;103:3076-3083.
  29. Molldrem JJ, Clave E, Jiang YZ, et al. Cytotoxic T lymphocytes specific for a nonpolymorphic proteinase 3 peptide preferentially inhibit chronic myeloid leukemia colony-forming units. *Blood*. 1997;90:2529-2534.
  30. Molldrem J, Dermime S, Parker K, et al. Targeted T-cell therapy for human leukemia: cytotoxic T lymphocytes specific for a peptide derived from proteinase 3 preferentially lyse human myeloid leukemia cells. *Blood*. 1996;88:2450-2457.
  31. Takahashi N, Miura I, Saitoh K, Miura AB. Lineage involvement of stem cells bearing the Philadelphia chromosome in chronic myeloid leukemia in the chronic phase as shown by a combination of fluorescence-activated cell sorting and fluorescence in situ hybridization. *Blood*. 1998;92:4758-4763.
  32. Crossman LC, Mori M, Hsieh YC, et al. In chronic myeloid leukemia white cells from cytogenetic responders and non-responders to imatinib have very similar gene expression signatures. *Haematologica*. 2005;90:459-464.
  33. Szaniszló P, Wang N, Sinha M, et al. Getting the right cells to the array: gene expression microarray analysis of cell mixtures and sorted cells. *Cytometry A*. 2004;59:191-202.
  34. Yeoh EJ, Ross ME, Shurtleff SA, et al. Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell*. 2002;1:133-143.
  35. Chiaretti S, Li X, Gentleman R, et al. Gene expression profile of adult T-cell acute lymphocytic leukemia identifies distinct subsets of patients with different response to therapy and survival. *Blood*. 2004;103:2771-2778.
  36. Michiels S, Koscielny S, Hill C. Prediction of cancer outcome with microarrays: a multiple random validation strategy. *Lancet*. 2005;365:488-492.
  37. Marin D, Kaeda J, Szydlo R, et al. Monitoring patients in complete cytogenetic remission after treatment of CML in chronic phase with imatinib: patterns of residual leukaemia and prognostic factors for cytogenetic relapse. *Leukemia*. 2005;19:507-512.
  38. Nowicki MO, Pawlowski P, Fischer T, et al. Chronic myelogenous leukemia molecular signature. *Oncogene*. 2003;22:3952-3963.
  39. Ohmine K, Ota J, Ueda M, et al. Characterization of stage progression in chronic myeloid leukemia by DNA microarray with purified hematopoietic stem cells. *Oncogene*. 2001;20:8249-8257.
  40. von Bubnoff N, Veach DR, Miller WT, et al. Inhibition of wild-type and mutant Bcr-Abl by pyridopyrimidine-type small molecule kinase inhibitors. *Cancer Res*. 2003;63:6395-6404.
  41. Shah NP, Tran C, Lee FY, et al. Overriding imatinib resistance with a novel ABL kinase inhibitor. *Science*. 2004;305:399-401.
  42. O'Hare T, Pollock R, Stoffregen EP, et al. Inhibition of wild-type and mutant Bcr-Abl by AP23464, a potent ATP-based oncogenic protein kinase inhibitor: implications for CML. *Blood*. 2004;104:2532-2539.
  43. Gumireddy K, Baker SJ, Cosenza SC, et al. A non-ATP-competitive inhibitor of BCR-ABL overrides imatinib resistance. *Proc Natl Acad Sci U S A*. 2005;102:1992-1997.
  44. Weisberg E, Manley PW, Breitenstein W, et al. Characterization of AMN107, a selective inhibitor of native and mutant Bcr-Abl. *Cancer Cell*. 2005;7:129-141.