Int. J. Cancer: **91,** 474–480 (2001)

Publication of the International Union Against Cancer

CDNA MICROARRAY GENE EXPRESSION ANALYSIS OF B-CELL CHRONIC LYMPHOCYTIC LEUKEMIA PROPOSES POTENTIAL NEW PROGNOSTIC MARKERS INVOLVED IN LYMPHOCYTE TRAFFICKING

Christian STRATOWA[1]*, Gerald LÖFFLER[1], Peter LICHTER[2], Stephan STILGENBAUER[3], Peter HABERL[4], Norbert SCHWEIFER[1], Hartmut DÖHNER[3] and Klaus K. WILGENBUS[1]

[1]*Boehringer Ingelheim Austria, Vienna, Austria*
[2]*Abteilung "Organisation komplexer Genome," Deutsches Krebsforschungszentrum, Heidelberg, Germany*
[3]*Medizinische Klinik und Poliklinik V, University of Heidelberg, Heidelberg, Germany*
[4]*GeneData AG, Basel, Switzerland*

Human cancer is characterized by complex molecular perturbations leading to variable clinical behavior, often even in single-disease entities. We performed a feasibility study systematically comparing large-scale gene expression profiles with clinical features in human B-cell chronic lymphocytic leukemia (B-CLL). cDNA microarrays were employed to determine the expression levels of 1,024 selected genes in 54 peripheral blood lymphocyte samples obtained from patients with B-CLL. Statistical analyses were applied to correlate the expression profiles with a number of clinical parameters including patient survival and disease staging. We were able to identify genes whose expression levels significantly correlated with patient survival and/or with clinical staging. Most of these genes code either for cell adhesion molecules (L-selectin, integrin-β2) or for factors inducing cell adhesion molecules (IL-1β, IL-8, EGR1), suggesting that prognosis of this disease may be related to a defect in lymphocyte trafficking. This report demonstrates the feasibility of a systematic integration of large-scale gene expression profiles with clinical data as a general approach for dissecting human diseases.

Key words: *B-CLL; L-selectin; CD18; TCL1; cDNA microarrays*

Cancer is a diverse family of complex diseases characterized by genotypic and phenotypic changes that result in a tremendous variability in clinical behavior. The understanding of the complex changes taking place at the molecular level should lead to more precise prognostic markers and to the identification of novel drug intervention sites. Until recently, limitations in conventional techniques only allowed focus on few biological parameters. Advances in nucleic acid microarray technology now permit study of gene expression on a genome-wide scale. Correlation of these results with clinical data should allow the systematic identification of molecular changes contributing to clinical phenotype. B-cell chronic lymphocytic leukemia (B-CLL) was selected as a model for a study aiming to test the feasibility of such an approach.

B-CLL, the most common form of leukemia in the Western world in adults, is characterized by a highly variable clinical course including varying survival rates, which can range from a few months to more than 20 years. Reliable individual prognostic tools are still limited and clinical stage remains the strongest predictor of survival.[1] More recently, genetic alterations, *i.e.*, 17p13 and 11q22.3-11q23.1 deletions, have been shown to provide independent prognostic information for disease progression and survival.[2–4]

Nucleic acid microarray technology, which is based on hybridization of fluorescence-labeled probes to either cDNA clones or oligonucleotides arrayed on solid surfaces, allows the simultaneous analysis of the expression of thousands of genes in a single experiment.[5,6] However, because of the quantity and complexity of the data produced, data mining tools including advanced statistical algorithms are necessary to identify patterns in these complex data sets.[7] Nucleic acid microarrays have been successfully used for the identification of inflammatory disease-related genes.[8] cDNA microarrays containing virtually every gene of the yeast *Saccharomyces cerevisiae* were used to investigate the temporal gene expression pattern accompanying the shift from anaerobic to aerobic metabolism.[9] Oligonucleotide arrays were used to monitor the expression levels of nearly all yeast genes[10] and to investigate the mitotic cell cycle.[11] Recently, it was demonstrated that gene expression profiling allowed the molecular classification of tumors and thus the identification of previously undetected and clinically significant subtypes of cancer.[12,13]

We describe that comprehensive expression profiles, as determined by cDNA microarrays, can be successfully correlated with clinical parameters for cancer such as patient survival and disease staging. An additional common reference sample was introduced in order to allow a cross-sectional analysis over the entire range of experiments. In addition, a rigorous quality assessment was performed to prove the technical feasibility of this approach.

MATERIAL AND METHODS

Microarray design and preparation

Clones that code for genes belonging to different categories (*e.g.*, cell cycle, apoptosis) were identified by a nucleotide sequence search in dBEST[14] utilizing an 800 bp long sequence that codes for the 3′-end of the gene of interest (using WorkBench, GeneData, Basel, Switzerland). The BLAST search yielded a number of expressed sequence tags (EST) for the gene of interest. Respective IMAGE clones[15] were purchased from Ressourcen-Zentrum/PrimärDatenbank (RZPD; Berlin, Germany) and from Genome Systems (St. Louis, MO). All 1,237 IMAGE clones (representing 1,024 distinct known genes) were partially sequenced using the standard primers M13 forward or M13 reverse, respectively, in order to verify the correct sequence. Sequences were assigned to a human gene by a BLAST search in the GenBank. When verified sequences did not match with the database entry (only 78.9% were identified as correct sequences), different ESTs for the same gene were ordered and subsequently verified by sequencing.

Microtiter plates containing plasmid DNA for all 1,237 sequence-verified EST clones were sent to Incyte Pharmaceuticals (Palo Alto, CA), where 60 custom microarrays were fabricated.

C. Stratowa and G. Löffler contributed equally to the work.

*Correspondence to: Exploratory Research, Boehringer Ingelheim Austria, Dr. Boehringergasse 5-11, A-1121 Vienna, Austria. Fax: +43-1-80105-2683.
E-mail: christian.stratowa@vie.boehringer-ingelheim.com

Received 10 May 2000; Revised 1 August 2000; Accepted 29 September 2000

Patient selection and tissue specimens

Frozen peripheral blood samples from 54 patients with chronic B-CLL were studied. All patients included in our study were seen at the University of Heidelberg, a secondary and tertiary care referral center for patients with B-CLL. They were diagnosed for B-CLL and subjected to interphase cytogenetic analysis. The diagnosis of B-CLL required a persistent lymphocytosis of >5,000/μl. Mononuclear cells were separated from peripheral blood samples by Ficoll-Hypaque density gradient centrifugation. Phenotype was determined by immunofluorescent flow cytometry analysis and/or by immunohistochemical staining. Immunophenotypic data were available for 51 patients: all samples were $CD19^+$, 49 of 51 tested were $CD5^+$, 49 of 51 tested were $CD23^+$ (48 of 51 were $CD5^+$ $CD23^+$); 36 of 36 tested were both $CD4^-$ and $CD8^-$; leukemic cells from 12 patients strongly expressed the CD22 antigen. The disease was staged according to Rai *et al.*[16] at diagnosis and at the time of sampling.

Sample preparation, hybridization and scanning

Total RNA from all 54 B-CLL samples (about 2×10^7 cells per sample) was extracted using the Triazol reagent (Life Technologies, Bethesda, MD). Poly-A^+-RNA was prepared with the use of Oligotex-dT resin (Qiagen, Hilden, Germany). Aliquots from random samples were reverse transcribed into P^{32}-labeled cDNA and size separated on agarose gels for quality control. The total amount of poly-A^+-RNA obtained varied from 250 to 1,000 ng. About 200–500 ng of each mRNA were sent to Incyte where it was reverse transcribed into Cy3-labeled cDNA.

A standard control hybridization pool was prepared by Cy5 labeling Δ, a mixture of equal amounts of human mRNAs from placenta, bone marrow, heart, kidney, liver, lung, pancreas, skeletal muscle, spleen and thymus (poly-A^+-RNA was purchased from Clontech, Palo Alto, CA). At Incyte, each Cy3-labeled test sample was mixed with an equal amount of Cy5-labeled reference cDNA and co-hybridized onto 1 microarray. For control purposes, aliquots of the reference mRNA mixture were Cy3 or Cy5 labeled and mixed before hybridization. After, arrays were scanned using a fluorescence laser-scanning device. Raw data were obtained from Incyte in tabular form for further analysis.

RNA quantification by TaqMan PCR analysis

Primers were designed with Tm of 58° to 60°C, whereas the Tms of the probes were 5° to 7°C higher. Probes containing either a FAM (6-carboxyfluorescein) or TET (tetrachloro-6-carboxy-fluorescein) reporter dye and TAMRA (6-carboxy-tetramethyl-rhodamine) as quencher were purchased from Perkin Elmer (PE; Norwalk, CT). The following primer and probes were used for the analysis:

L-selectin: 5′-TCACGTCGTCTTCTGTATACTGTGG-3′, 5′-TGCAGCTAGCATTTCAGTGATG-3′ and 5′FAM-ACTCCAG-TGAAGTAATGGGGTCCTGC-TAMRA 3′.

IL-8: 5′-ACTTCATGTATTGTGTGGGTCTGTTG-3′, 5′-TC-CCGTGCAATATCTAGGAAAATCC-3′ and 5′FAM-TGCCA-GATGCAATACAAGATTCCTGGTT-TAMRA 3′.

GAPDH: 5′- ACGGGAAGCTTGTCATCAATG -3′, 5′- TAC-TCAGCGCCAGCATCG -3′ and 5′TET- CCATCACCATCTTC-CAGGAGCGAGA-TAMRA 3′.

Approximately 100 ng poly-A^+-RNA was reverse transcribed using an oligo-dT primer and the Superscript II reverse transcriptase (200 U/μl; Gibco BRL, Gaithersburg, MD). ss-cDNA (1 μl) was directly subjected to TaqMan PCR in a total volume of 25 μl, including 10× Ampli-Taq Gold PCR-buffer A, 200 μM dATP, dCTP, dGTP and 400 μM dUTP, 3.5 mM $MgCl_2$, 0.025 U/μl Ampli-Taq Gold DNA polymerase, 0.005 U/μl AmpErase uracil N-glycosylase, 400 nM of each primer and 100 nM probe. The following thermal cycling conditions were used: 2 min at 50°C, 10 min at 95°C and 45 cycles with 15 sec of 95°C and 1 min of 60°C for the respective gene or 68°C for GAPDH. All reactions were performed in the model 7700 sequence detector (PE). TaqMan readings were analyzed using the SDS 1.6.3 software, data were exported to Excel and normalized using GAPDH TaqMan signals. To be able to correlate the results, microarray expression values for the corresponding genes were also normalized to the expression values for GAPDH.

Determination of confidence limits of sample hybridizations

To prove the correctness and reproducibility of a single DNA chip measurement (direct approach), 2 identical samples were co-hybridized in a single measurement. Therefore, the relative mRNA level l_i of every gene i on the chip should be exactly 1. The logarithms $\log(l_i)$ of the relative mRNA levels did approximately follow a normal distribution centered at 0 (corresponding to $l_i = 1$). Thus, the number of genes i with $\log(l_i) = x$ and hence $l_i = \exp(x) = f$ is approximately the same as the number of genes j with $\log(l_j) = -x$ and hence $l_j = 1/\exp(x) = 1/f$ for every value x. The factor f_p is defined as the measure at which, for p percent of the measured relative mRNA levels l_i, the true relative mRNA level of 1 falls within the interval $[l_i/f_p, l_i x f_p]$. Thus, f_p is called the confidence factor at the confidence level p. Comparisons with TaqMan PCR studies (where the "correct" relative mRNA level can be determined even if it is not equal to 1) have shown that f_p is independent of the relative mRNA levels l_i.

To determine the correctness and reproducibility of 2 identical DNA microarray measurements (indirect approach), 2 identical samples A_1 and A_2 were co-hybridized with a common reference sample R in 2 separate measurements. Again, the relative mRNA level l_i of every gene i on the microarray in sample A_1 relative to sample A_2 (or vice versa) should be exactly 1. $\text{Log}(l_i)$ of the relative mRNA levels followed approximately a normal distribution centered at 0 (corresponding to $l_i = 1$). A confidence factor f_p was determined that defines a confidence interval $[l_i/f_p, l_i x f_p]$ around every (indirectly measured) relative mRNA level l_i such that the probability is p percent that the true relative mRNA level falls within the confidence interval.

Data analysis

To identify genes that correlate with patient survival, an algorithm was developed for the identification of a "significant" mRNA threshold level that can be used to divide patients into a low- and a high-expression group. Ranked expression values were divided into 2 groups while successively increasing the number of patients in the low-expressing group. For each case, the ratio of the mean (median) expression values of the patients in the high- and low-expression groups and the *p* value using the Mantel-Haenszel test were calculated. An mRNA threshold level is considered to be significant if (1) the mRNA levels for the resulting high-expression group are sufficiently higher than those for the low-expression group (factor 3), (2) the survival behavior of the resulting high- and low-expression groups of patients is sufficiently different (*p* value for the null hypothesis of identical survival behavior in the Mantel-Haenszel test ≤ 0.03), and (3) the minimal *p* value is found for a separation between the 2 groups that is sufficiently close to the median.

Genes for which mRNA levels correlate with any single continuous (numerical) clinical variable were identified as follows: The mRNA levels of a given gene were correlated with a defined clinical variable by calculating a (rank-based) Spearman correlation coefficient. In general, neither mRNA levels nor the clinical variables fulfill the requirements for applying a standard Pearson correlation measure. Therefore, a rank-based correlation coefficient was calculated. Additionally, the *p* value for the null-hypothesis was calculated, which showed that the mRNA levels and the clinical variable are not correlated. A gene was considered to correlate with a clinical parameter if (1) *p* value < 0.05 and (2) the absolute value of Spearman's rank correlation rho >0.5. All statistical calculations were performed using the software package S-Plus (Math-Soft, Seattle, WA).

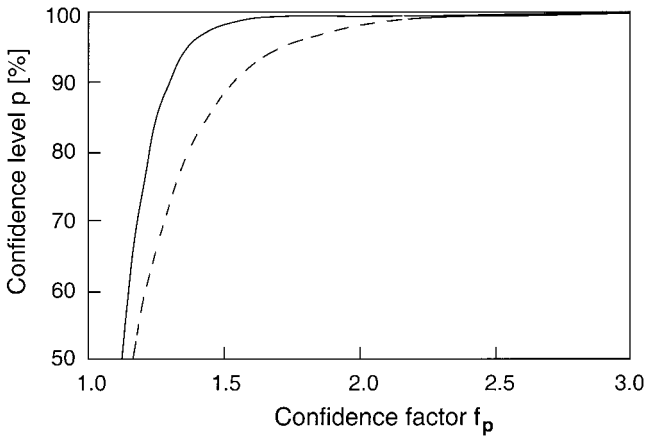

FIGURE 1 – Percentages of expression levels that lie in the interval $[1/f_p, f_p]$, shown as a function of f_p. The plots represent the mean variation of the confidence levels with the confidence factor f_p for direct comparisons of 2 samples (solid curve) or for indirect comparisons of 2 samples using a common reference sample (dotted curve), respectively, as described in the Material and Methods.

RESULTS

Evaluation of hybridization reproducibility and reliability

The cDNA microarray technology requires the hybridization of 2 samples per DNA chip measurement, each sample being marked with a different fluorescent dye.[6] The primary fluorescence intensities of the genes in sample A are divided by the primary fluorescence intensities of the corresponding genes in sample B. Comparing every sample in a set of N samples to every other sample in the same set would require $N(N - 1)/2$ measurements (direct measurement). Alternatively, each of the N samples can be compared with a reference sample in a separate DNA chip measurement (indirect measurement). However, the accuracy of the relative mRNA levels is lower in the indirect approach than in the direct approach. Each of the 54 B-CLL cDNA samples was mixed with a differently labeled common reference cDNA sample and co-hybridized onto 54 microarrays. By carefully selecting this reference sample (an mRNA mixture of 10 human tissues was used as reported in the Material and Methods), up to 98% of the spots on the microarray could deliver valuable fluorescence signals.

The reproducibility was determined for both the direct and the indirect approach as described in the Material and Methods. The results are presented in Figure 1. The confidence interval of a measured relative mRNA level l_i at a confidence level of 99% was defined by a factor f_{99} to be $[l_i/f_{99}, l_i \times f_{99}]$. For the direct and indirect measuring approach, f_{99} was found to be 1.6 and 2.3, respectively. Thus, for a measured relative expression level larger than 1.6 (2.3) or smaller than 1/1.6 (1/2.3), the probability is less than 1% that the actual expression is 1.0. Because a number of clones were spotted on the microarray in duplicate, the "intra-chip" reproducibility could also be tested. At the 99% confidence level, a factor f_{99} of 1.5 was determined. This is in the same range as the value 1.6 obtained for the direct measurement and is consistent with data calculated at Incyte for significance in ratios. Moreover, several genes were represented by different IMAGE clones. In this case, an overall confidence factor, including inter-chip and intra-chip variability, of 2.5 at a confidence level of 99% was determined (data not shown).

Finally, to validate the microarray results, mRNA levels in different patient samples were measured by quantitative PCR analysis for a few selected genes, which were found to correlate with patient survival rates (see below). The correlation between the expression levels as determined by DNA microarray hybridization and by TaqMan PCR analysis, respectively, is shown in Figure 2 for L-selectin and IL-8. The IL-8 data are shown for 2 distinct clones, each of which is spotted on different locations onto the microarrays. As shown in Figure 2*b*, the correlation coefficients were in consent.

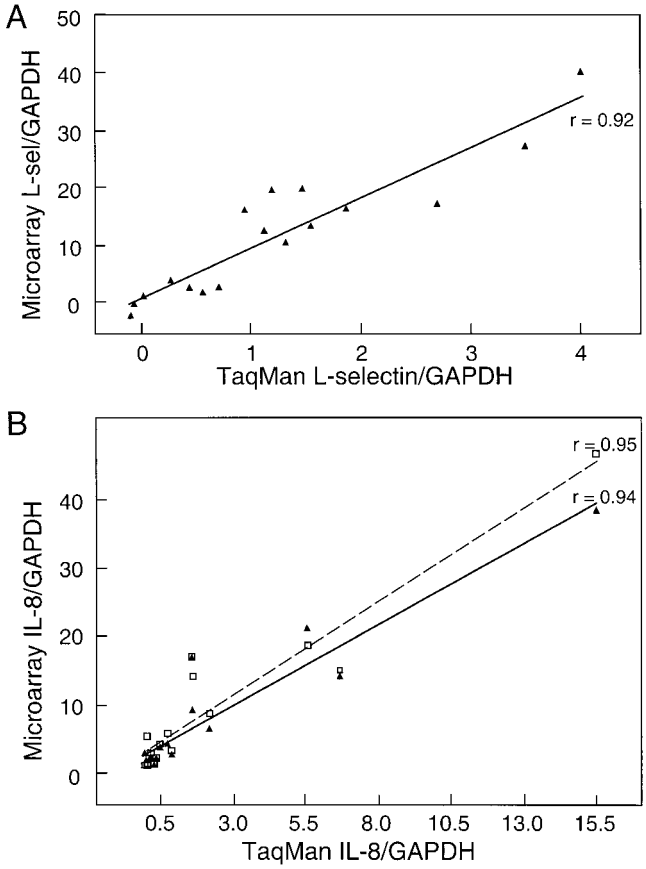

FIGURE 2 – Validation of microarray expression profiles by TaqMan PCR analysis. Expression values for L-selectin (*a*) and IL-8 (*b*), using GAPDH as control, were measured by TaqMan PCR analysis for different patient samples. Correlation of these data, normalized to GAPDH, to data obtained with cDNA microarrays was determined. The correlation coefficients, r, are indicated in the figures. (*a*) Data are shown for 1 distinct clone for L-selectin. (*b*) Data are shown for 2 distinct clones for IL-8, which were spotted onto the microarray.

Genes abundantly expressed in B-CLL samples

Microarrays were co-hybridized with Cy3-cDNA from the B-CLL cells from the 54 selected patients (Table I) and the common reference Cy5-cDNA mixture (see Material and Methods). In the initial analysis, we identified genes that were abundantly expressed in the B-CLL samples compared with the mixture of human normal tissues used as the reference sample. The results for all genes that are expressed in mean more than 4-fold higher are shown in Table II.

Most genes listed in Table II are known to be expressed in B lymphocytes, indicating that cDNA microarray hybridization is able to identify relevant genes. Many of these genes are involved in B-cell antigen receptor signaling, *i.e.*, tyrosine kinase Lyn, phospholipase PLC-gamma-2, SH2-domain-containing inositol 5-phosphatase SHIP, growth factor receptor binding protein Grb2, MAP-kinase ERK1 and the transcription factors Jun-B, ELF-1 and Spi-B. Genes that are at least 20-fold less expressed in the B-CLL samples compared with the reference mRNA sample include serum albumin (ratio = 0.028), human growth hormone (ratio = 0.029), placental lactogen hormone PL-4 (ratio = 0.032), haptoglobin Hp-alpha-2 (ratio = 0.036) and tissue factor pathway inhibitor-2 (ratio = 0.041).

TABLE I – CLINICAL DATA FOR THE SELECTED PATIENTS

	Diagnosis	Sampling
No. of CLL patients[1]		
Sex		
Male	33 (61%)	33 (61%)
Female	21 (39%)	21 (39%)
Age (years)		
Mean ± SEM	57.8 ± 10.2	60.5 ± 9.7
Range	39–85	39–85
Stage at		
Rai 0	14 (29%)	12 (22%)
Rai 1–2	29 (59%)	28 (52%)
Rai 3–4	6 (12%)	14 (26%)
Time from diagnosis to sampling (months)		
Mean number of patients (range)	33.7 (0–187)	
Survival time since (months)		
Mean ± SD	71.2 ± 56.0	34.3 ± 21.8
Range	4–262	1–86

[1]N = 54 (100%).

Correlation of gene expression profiles with patient survival

In order to correlate the mRNA levels of a given gene with the survival rate of the patients, we divided the patients into a low-expression group and a high-expression group based on the mRNA level of that gene in a patient. The significance of any difference in survival behavior of these 2 groups was then statistically evaluated. Genes obeying the statistical criteria described in the Material and Methods are listed in Table III. Survival data were calculated from both the date of tumor sampling and the date of diagnosis of the disease. In most cases, low patient survival was correlated with decreased gene expression. However, TCL1 was expressed at high levels in B-CLL cells from patients with low survival probability. Furthermore, genes that were found to be associated with survival because either diagnosis or sampling with $p < 0.03$, also showed correlation in the other survival measure. Kaplan-Meier survival curves are presented in Figure 3 for 4 selected genes at both time points. Patients showing reduced expression of the genes coding for IL-1β, IL-8 and L-selectin in B-CLL cells had a significantly reduced overall survival time compared with patients showing higher expression rates. In contrast, high expression of TCL1 correlated with low patient survival.

Investigating whether the expression levels of those genes were positively correlated in individual patients belonging to the group with either high or low survival probability, respectively, revealed that IL-1β and IL-8 showed the same tendency in 80% of the patients. This suggests a relationship between these 2 genes. Expression of the transcription factor EGR1 was positively associated with expression of IL-8 in 66% and of IL-1β in 63% of the patients.

Correlation of gene expression profiles with clinical staging of the patients

Because clinical stage remains the strongest predictor of survival for patients with B-CLL, we tested whether specific genes can be identified for which the expression levels correlate with different stages. Demanding an at least 3-fold difference in the expression for high *vs.* low Rai stages, a clear correlation ($p < 0.001$) between gene expression and Rai staging could be detected for only 3 genes. Higher clinical stage correlated with a decreased expression of the genes for IL-1β, IL-8 and EGR1, whereas no correlation was found for other genes listed in Table III such as L-selectin. Data in Figure 4 are presented for Rai stage at sampling. The results could be confirmed by comparing the expression profiles from patients with Rai stage 0, who are still alive, with the profiles obtained from deceased patients with Rai stages 3 or 4 (data not shown).

TABLE II – GENES ABUNDANTLY EXPRESSED IN B-CLL CELLS COMPARED WITH A MIXTURE OF HUMAN NORMAL TISSUES

Acc No.[1]	Gene name	Relative mRNA levels[2]
X07109	Protein kinase C gene beta-II	8.9 ± 3.4
M16650	ODC1 ornithine decarboxylase-1	7.9 ± 4.0
M60333	MHC class II HLA-DR alpha-chain	7.9 ± 2.1
X61123	BTG1 B-cell translocation gene-1	7.6 ± 2.8
X52425	IL-4 receptor	7.0 ± 3.2
J03571	5-lipoxygenase	6.4 ± 2.7
X14034	Phospholipase C-gamma-2	6.0 ± 2.2
M16038	Lyn tyrosine kinase	5.9 ± 1.9
M82882	ELF-1 Ets-related transcription factor	5.6 ± 2.4
U82972	IL-16	5.3 ± 2.2
X06318	Protein kinase C gene beta-I	5.2 ± 2.7
X60188	ERK1 Ser/Thr-kinase	5.2 ± 1.5
L13463	G0S8/RGS2 helix-loop-helix protein	4.9 ± 3.7
X16150	L-selectin	4.5 ± 2.9
U18297	MST1/Krs-2 stress-responsive Ser/Thr-kinase	4.5 ± 2.0
U50040	SHIP-110 signaling IP-5 phosphatase	4.3 ± 1.5
X51345	Jun-B	4.3 ± 1.8
X96998	Spi-B Ets-related transcription factor	4.2 ± 1.8
L29511	Grb2 growth factor receptor binding protein-2	4.2 ± 1.1

[1]GenBank accession number.–[2]Mean levels in B-CLL cells relative to a mixture of human normal tissues with standard deviation.

Correlation of gene expression profiles with other molecular and clinical criteria

The correlation between expression profiles and a number of different molecular and clinical parameters was also tested. Statistical analyses were performed to identify associations between gene expression and the chromosomal aberrations 11q22.3-q23.1 deletion and 17p13 deletion. Both deletions have been recently shown to represent independent markers for poor survival. Results are depicted in Figure 5. Applying $p < 0.05$ in the Kruskal-Wallis rank sum test, the only gene exhibiting a significant correlation between its expression level and chromosomal aberration was L-selectin (17p13 deletion; p value: 0.0337). The purity of the tumor samples can be estimated from the percentage of $CD19^+$ cells which is at mean 89.2% (range 73% to 98%). However, no correlation to the percentage of $CD19^+$ cells or to the percentage of $CD5^+CD23^+$ cells, respectively, could be detected (data not shown). No further correlation between gene expression levels and other molecular or clinical parameters could be found.

DISCUSSION

The goal of our study was to demonstrate that gene expression profiles of tumor samples can be correlated with clinical data such as disease staging and patient survival. To this aim, cDNA microarray methodology was adapted to profile B-CLL samples from 54 patients, thereby allowing extensive cross-sectional data analysis. However, this experimental setup requires a high accuracy of the microarray measurements. For this reason, a rigorous quality assessment was performed beforehand.

Because the cDNA microarray technology used is based on 2-color differential expression analysis, a reference cDNA sample had to be used as common control for 1 channel, and the ability to correlate the ratio Cy3/Cy5 of the fluorescent signals across numerous samples had to be proven first. The problem of choosing the "correct" reference sample remains open to debate. We selected an mRNA mixture of 10 human tissues as reference sample in order to assure that most of the spots could deliver valuable fluorescence signals in the reference channel. The robustness and validation of the data obtained by microarray hybridization were tested for both the direct and the indirect approach (Fig.1). The differences in the confidence factor at a confidence level of 99% reflect the fact that the reproducibility for direct A to B comparison

TABLE III – SURVIVAL ANALYSIS RESULTS FOR GENES CORRELATED WITH PATIENT SURVIVAL

Acc. No.[1]	Gene name	Time since sampling		Time since diagnosis		Expression correlated with low survival
		p value[2]	High/low[3]	*p* value[2]	High/low[3]	
X16150	L-selectin	0.0005	3.3 (2.7)	0.0129	3.1 (2.7)	Low
M15330	IL-1β	0.0034	3.7 (2.4)	0.0153	4.3 (3.1)	Low
M15395	Integrin-β2 (CD18, LFA-1 beta)	0.0036	3.2 (2.2)	0.0487	3.0 (2.1)	Low
M17017	IL-8	0.0044	4.4 (3.4)	0.0030	4.4 (3.4)	Low
X52541	EGR1 early growth response protein-1	0.0356	5.4 (3.5)	0.0021	5.4 (3.5)	Low
X82240	TCL1 T-cell leukemia/lymphoma-1	0.0426	3.2 (3.1)	0.0079	3.1 (3.0)	High

[1]GenBank accession number.–[2]Calculated using the Mantel-Haenszel (log-rank) test.–[3]Ratio of the mean (median) expression values of the patients in the high-expression and low-expression groups.

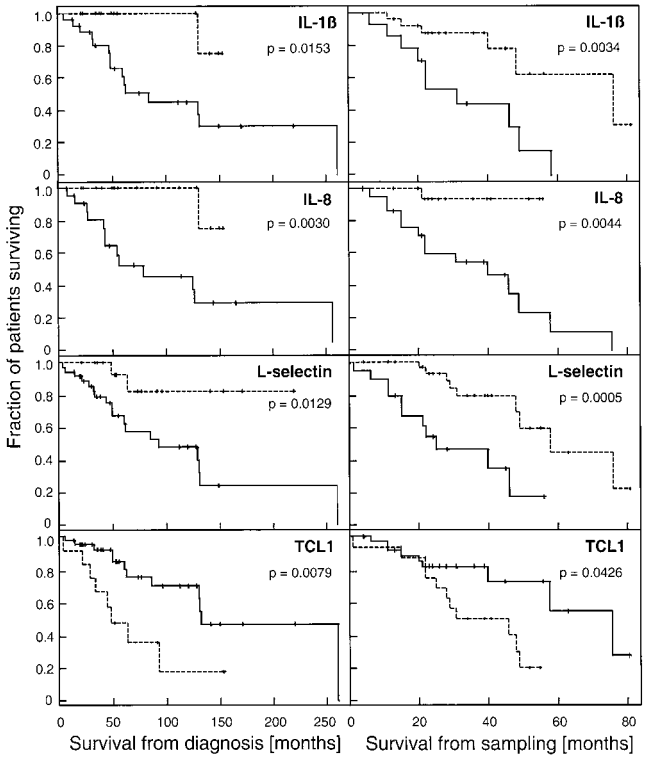

FIGURE 3 – Survival probabilities from the time of diagnosis and the time of sampling in 54 patients separated according to the expression of IL-1β, IL-8, L-selectin or TCL1 in B-CLL cells, respectively. Kaplan-Meier curves for the survival of patients separated into low- and high-expression groups, respectively, are presented. Solid (dotted) lines correspond to low (high) expression values of the respective gene as compared to a threshold level (see Material and Methods). The significance (*p* value) of the identity of the 2 curves is indicated in the figures.

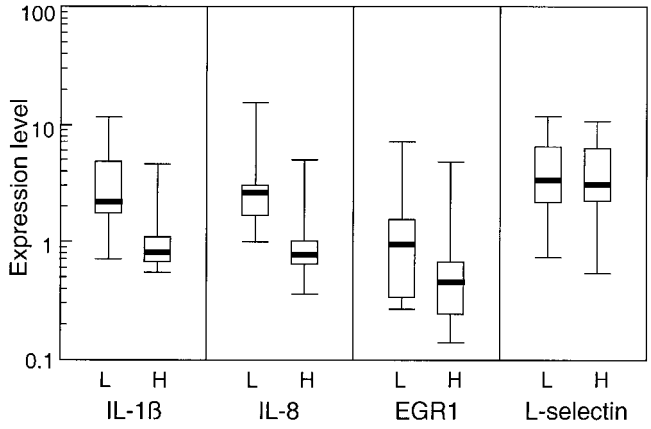

FIGURE 4 – Expression patterns of IL-1β, IL-8, EGR1 and L-selectin by B-CLL cells of patients with low *vs.* high disease progression state. L denotes patients with Rai stages 0 or 1 and H denotes patients with Rai stages of 2, 3 or 4, respectively. Marked by the boxplots are the minimum, the 1st quartile, the median, the 3rd quartile and the maximum value. The Kruskal-Wallis rank sum test computes *p* values of 0.0007 (IL-1β), 0.0002 (IL-8), 0.0536 (EGR1) and 0.3602 (L-selectin).

on 1 microarray is higher than for the indirect comparison of different samples using a reference cDNA sample as common control. The measurement of a number of genes on the microarray in duplicate allowed testing of the intra-chip reproducibility. Interestingly, the reproducibility for 2 identical clones was found to be higher than for 2 different clones of the same gene. This suggests that the selection of a suitable clone for a certain gene is a critical step for the quality of the data. However, the results obtained when using distinct ESTs coding for the same gene were comparable (Fig. 2*b*). A differential expression value of at least 2.5 could be considered as factual with a confidence of 99%.

Correlation analysis with clinical data identified genes, the cytokines IL-1β and IL-8 and the transcription factor EGR1, for which low expression was found to be correlated with both low survival probability and clinical stage, and genes, the cell adhesion molecules integrin-β2 (CD18) and L-selectin (CD62L), for which low expression was found to correlate with low survival probability only. The portion of tumor cells in the samples (≥80% $CD19^+$ cells) did not influence these results.

L-selectin is the only selectin expressed on lymphocytes and mediates a number of leukocyte-endothelial interactions, including leukocyte rolling, the first step of the cell adhesion cascade[17] and the attachment of lymphocytes to high endothelial venules. Therefore, L-selectin induces the passing of lymphocytes from blood into the peripheral lymph nodes.[18] Reduced expression of L-selectin might therefore prevent lymphocyte homing.

The cytokines IL-1β and IL-8 are involved in the triggering step preceding strong cell adhesion. IL-1β induces ICAM-1 expression[19] on the endothelium whereas IL-8 signals conversion of integrins into a functionally active state. The reduced expression of IL-1β found in B-CLL cells from patients with low survival and at higher Rai stages is in accordance with published data that show a loss of IL-1β production in B-CLL cells from progressive disease.[20] The role of IL-8 in B-CLL remains unclear although it has been reported that *in vitro* IL-8 is capable of inducing prolonged survival of B-CLL cells.[21] However, IL-8, a chemotactic cytokine, is induced by IL-1β[22] and both IL-1β and IL-8 are independently able to induce expression of the Egr-1 gene.[23,24] This is in accordance with our observation that the expression profiles for IL-1β, IL-8 and EGR1 correlate with survival in individual patients. The immediate early gene Egr-1 encodes the zinc-finger transcription factor EGR1 that is normally induced during B-cell maturation and activation.[25] Recently identified downstream targets of EGR1 in B cells include the cell adhesion molecule ICAM-1.[26]

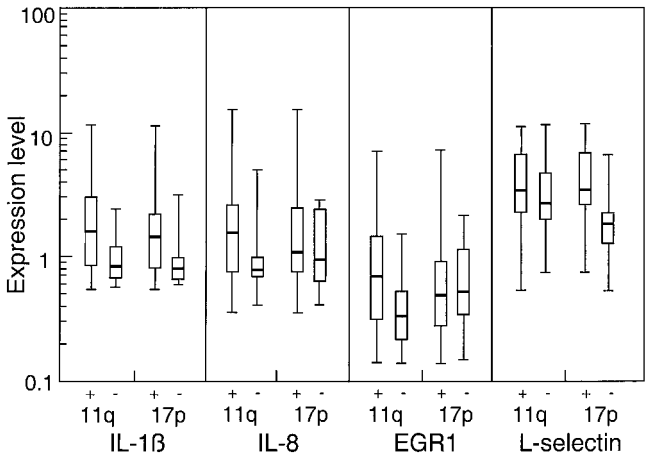

FIGURE 5 – Expression patterns of IL-1β, IL-8, EGR1 and L-selectin for B-CLL cells of patients with different karyotypes. Grey boxes denote patients with 11q22.3-q23.1 or 17p13 deletions, respectively. Marked by the boxplots are the minimum, the 1st quartile, the median, the 3rd quartile and the maximum value. The Kruskal-Wallis rank sum test computes the following *p* values: IL-1β: 0.0567 (11q), 0.2198 (17p); IL-8: 0.2618 (11q), 0.6975 (17p); EGR1: 0.0516 (11q), 0.7884 (17p); L-selectin: 0.6752 (11q), 0.0337 (17p).

The counter-receptor for ICAM-1 on lymphocytes is LFA-1, which consists of CD11a and CD18. CD18 is the common beta-subunit of the LEU-CAMs LFA-1 (CD11a), CR3 (CD11b) and CR4 (CD11c) involved in leukocyte-leukocyte adhesion, adherence of leukocytes to the endothelium and complement binding. Down-regulation of CD18 impairs all these functions. Furthermore, it has been demonstrated that LFA-1/ICAM-1 interactions are crucial for Th1 cell-mediated B-cell apoptosis[27] and may thus contribute to the maintenance of B-cell homeostasis. In accordance with our data, it has been shown previously that low levels of CD18 on B-CLL cells were associated with higher mortality rates.[28,29]

Although the expression of all genes described so far is decreased with low survival, increased expression of TCL1 correlates with low survival (Table III). Several breakpoints at the TCL1 (T-cell lymphoma/leukemia-1) locus 14q32.1 have been classified in a large proportion of patients with T-cell prolymphocytic leukemia and T-cell leukemia developing from ataxia telangiectasia.[30,31] It has also been demonstrated that deregulated expression of the human TCL1 gene in transgenic mice causes T-cell leukemia.[32] The function of this gene remains unknown although a role as a novel anti-apoptotic protein has been suggested.[33]

The comparative analysis of gene expression with survival times revealed that genes shown to correlate significantly with survival time since sampling also correlated with survival time since diagnosis and vice versa. If expression of these genes were regulated according to elapsed time from disease onset or disease progression, this finding could not be explained easily. In contrast, our data suggest that the expression levels of these genes resemble a disease-specific feature already present at early stages of the disease. Experiments with samples obtained from the same patient at different time points will allow testing of this hypothesis.

Our study demonstrates how comprehensive gene expression profiling can be correlated with clinical data to systematically identify genes whose expression levels are associated with clinical behavior of a human malignant disease entity. Most genes, for which decreased expression in B-CLL cells was found to be correlated with low patient survival, code either for cell adhesion molecules or for factors inducing cell adhesion molecules. This indicates that one factor contributing to the prognosis of B-CLL might be a defect in lymphocyte trafficking, resulting in accumulation of leukemic B cells in the blood.

In an accompanying study, Voss *et al.*[34] described the correlation of clinical data with proteomics profiles in 24 patients with B-CLL. However, due to the limited number of genes spotted onto the cDNA microarrays, distinct sets of genes were identified by the 2 approaches.

As the recent developments of large-scale expression profiling technologies such as cDNA microarrays now allow the acquisition of robust biological data in a previously unsurmountable throughput, both the availability of well-characterized clinical samples and the interpretation of this information become increasingly important. Thus, more sophisticated bioinformatics tools to browse, analyze and visualize such data sets are urgently needed. However, given the foreseeable rapid developments in data acquisition and computational tools, one can envision the analysis of all human genes in parallel and their association to complex clinical phenotypes. It can be anticipated that this information will foster our knowledge on the complex nature of human cancer and its clinical consequences.

ACKNOWLEDGEMENTS

We thank H. Dauwa-Stummer, A. Gupta and Sandra Orth for excellent technical assistance, C. Schaffner for technical advice and help in respect to RNA preparation and Uta Francke for critically reading the manuscript. Generation of custom arrays and hybridization were performed at Incyte Pharmaceuticals (Fremont, CA).

REFERENCES

1. Zwiebel JA, Cheson BD. Chronic lymphocytic leukemia: staging and prognostic factors. Semin Oncol 1998;25:42–59.
2. Juliusson G, Merup M. Cytogenetics in chronic lymphocytic leukemia. Semin Oncol 1998;25:19–26.
3. Döhner H, Fischer K, Bentz M, Hansen K, Benner A, Cabot G, et al. p53 gene deletion predicts for poor survival and non-response to therapy with purine analogs in chronic B-cell leukemias. Blood 1995;85:1580–89.
4. Döhner H, Stilgenbauer S, Fischer K, Bentz M, Lichter P. Cytogenetic and molecular cytogenetic analysis of B cell chronic lymphocytic leukemia: specific chromosome aberrations identify prognostic subgroups of patients and point to loci of candidate genes. Leukemia 1997;11:S19–S24.
5. Ramsay G. DNA chips: state-of-the art. Nature Biotech 1998;16: 40–4.
6. Schena M, Heller RA, Theriault TP, Konrad K, Lachenmeier E, Davis RW. Microarrays: biotechnology's discovery platform for functional genomics. TibTech 1998;16:301–6.
7. Ermolaeva O, Rastogi M, Pruitt KD, Schuler GD, Bittner ML, Chen Y, et al. Data management and analysis for gene expression arrays. Nature (Genet) 1998;20:19–23.
8. Heller RA, Schena M, Chai A, Shalon D, Bedilion T, Gilmore J, et al. Discovery and analysis of inflammatory disease-related genes using cDNA microarrays. Proc nat Acad Sci Wash 1997;94:2150–55.
9. DeRisi J, Iyer VR, Brown PO. Exploring the metabolic and genetic control of gene expression on a genomic scale. Science 1997;278: 680–86.
10. Wodicka L, Dong H, Mittmann M, Ho M-H, Lockhart DJ. Genome-wide expression monitoring in *Saccharomyces cerevisiae*. Nature Biotech 1997;15:1359–67.
11. Cho RJ, Campbell MJ, Winzeler EA, Steinmetz L, Conway A, Wodicka L, et al. A genome-wide transcriptional analysis of the mitotic cell cycle. Mol Cell 1998;2:65–73.
12. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 1999;286:531–37.
13. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature (Lond) 2000;403:503–11.
14. Benson DA, Boguski MS, Lipman DJ, Ostell J, Ouellette BFF. GenBank. Nucleic Acids Res 1988;26:1–7.

15. Lennon G, Auffray C, Polymeropoulos M, Soares MB. The I.M-.A.G.E. consortium: an integrated molecular analysis of genomes and their expression. Genomics 1996;33:151–52.
16. Rai KR, Sawitsky A, Cronkite EP, Chanana AD, Levy RN, Pasternack BS. Clinical staging of chronic lymphocytic leukemia. Blood 1975; 46:219–34.
17. Kansas GS. Structure and function of L-selectin. APMIS 1992;100: 287–293.
18. Csanaky G, Vass JA, Milosevits J, Ocsovszki I, Szomor A, Schmelczer M. Expression and function of L-selectin molecules (LECAM-1) in B-cell chronic lymphocytic leukemia. Haematology 1994;79:132–36.
19. Stratowa C, Audette M. Transcriptional regulation of the human intercellular adhesion molecule-1 gene: a short overview. Immunobiology 1995;193:293–304.
20. Aguilar-Santelises M, Magnusson R, Svenson SB, Loftenius A, Andersson B, Mellstedt H, et al. Expression of interleukin-1α, interleukin-1β and interleukin-6 in chronic B lymphocytic leukemia (B-CLL) cells from patients at different stages of disease progression. Clin Exp Immunol 1991;84:422–28.
21. DiCelle PF, Mariani S, Riera L, Stacchini A, Reato G, Foa R. Interleukin-8 induces the accumulation of B-cell chronic lymphocytic leukemia cells by prolonging survival in an autocrine fashion. Blood 1996;87:4382–89.
22. Matsushima K, Oppenheim JJ. Interleukin 8 and MCAF: novel inflammatory cytokines inducible by IL 1 and TNF. Cytokine 1989;1: 2–13.
23. Sells SF, Muthukumar S, Sukhatme VP, Christ SA, Rangnekar VM. The zinc finger transcription factor EGR-1 impedes interleukin-1-inducible tumor growth arrest. Mol Cell Biol 1995;15:682–92.
24. Yue T-T, Wang X, Sung C-P, Olson B, McKenna PJ, Gu J-L, et al. Interleukin-8: a mitogen and chemoattractant for vascular smooth muscle cells. Circ Res 1994;75:1–7.
25. Watson DK, Robinson L, Hodge DR, Kola I, Papas TS, Seth A. FLI1 and EWS-FLI1 function as ternary complex factors and ELK1 and SAP1a function as ternary and quaternary complex factors on the Egr1 promoter serum response elements. Oncogene 1997;14:213–21.
26. Maltzman JS, Carman JA, Monroe JG. Transcriptional regulation of the Icam-1 gene in antigen receptor and phorbol ester stimulated B lymphocytes: role for transcription factor EGR1. J exp Med 1996; 183:1747–59.
27. Wang J, Lenardo MJ. Essential lymphocyte function associated 1 (LFA-1): intercellular adhesion molecule interactions for T cell-mediated B cell apoptosis by Fas/APO-1/CD95. J exp Med 1997;186: 1171–76.
28. Domingo A, Gonzales-Barca E, Castellsague X, Fernandez-Sevilla A, Granena A, Crespo N, et al. Expression of adhesion molecules in 113 patients with B-cell chronic lymphocytic leukemia: relationship with clinico-prognostic features. Leukemia Res 1997;21:67–73.
29. Erikstein BK, Holte H, Kvaldy S, Anderson KB, Steen HB, Hannisdal E, et al. Expression of CD18 (integrin beta 2 chain) correlates with prognosis in malignant B cell lymphomas. Brit J Haematol 1993;83: 392–98.
30. Virgilio L, Narducci MG, Isobe M, Bilips L, Cooper M, Croce CM, et al. Identification of the TCL1 gene involved in T-cell malignancies. Proc nat Acad Sci Wash 1994;91:12530–34.
31. Stilgenbauer S, Schaffner C, Litterst A, Liebisch P, Gilad S, Bar-Shira A, et al. Biallelic mutations in the ATM gene in T-prolymphocytic leukemia. Nature (Med) 1997;3:1155–59.
32. Virgilio L, Lazzeri C, Bichi R, Nibu K-I, Narducci MG, Russo G, et al. Deregulated expression of TCL1 causes T cell leukemia in mice. Proc nat Acad Sci Wash 1998;95:3885–89.
33. Fu TB, Virgilio L, Narducci MG, Facchiano A, Russo G, Croce CM. Characterization and localization of the TCL-1 oncogene product. Cancer Res 1994;54: 6297–6301.
34. Voss T, Ahorn H, Haberl P, Doehner H, Wilgenbus K. Correlation of clinical data with proteomics profiles in 24 patients with B-cell chronic lymphocytic leukemia. Int J Cancer 2001;91:180–186.