Relation of Gene Expression Phenotype to Immunoglobulin Mutation Genotype in B Cell Chronic Lymphocytic Leukemia

Andreas Rosenwald,[1] Ash A. Alizadeh,[2] George Widhopf,[5]
Richard Simon,[6] R. Eric Davis,[1] Xin Yu,[1] Liming Yang,[1]
Oxana K. Pickeral,[1] Laura Z. Rassenti,[5] John Powell,[7] David Botstein,[3]
John C. Byrd,[8] Michael R. Grever,[9] Bruce D. Cheson,[10]
Nicholas Chiorazzi,[11] Wyndham H. Wilson,[12] Thomas J. Kipps,[5]
Patrick O. Brown,[2, 4] and Louis M. Staudt[1]

[1]*Metabolism Branch, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892*

[2]*Department of Biochemistry, the* [3]*Department of Genetics and* [4]*The Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA 94305*

[5]*University of California at San Diego, Department of Medicine, La Jolla, CA 92093*

[6]*Biometric Research Branch, Division of Cancer Treatment and Diagnosis, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892*

[7]*Bioinformatics and Molecular Analysis Section, CBEL, CIT, National Institutes of Health, Bethesda, MD 20892*

[8]*Department of Medicine, Walter Reed Army Medical Center, Washington, D.C. 20307*

[9]*Department of Internal Medicine, Ohio State University, Columbus, OH 43214*

[10]*CTEP, Division of Cancer Treatment and Diagnosis, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892*

[11]*North Shore-Long Island Jewish Research Institute, Manhasset, NY 11030*

[12]*Medicine Branch, Division of Clinical Sciences, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892*

Abstract

The most common human leukemia is B cell chronic lymphocytic leukemia (CLL), a malignancy of mature B cells with a characteristic clinical presentation but a variable clinical course. The rearranged immunoglobulin (Ig) genes of CLL cells may be either germ-line in sequence or somatically mutated. Lack of Ig mutations defined a distinctly worse prognostic group of CLL patients raising the possibility that CLL comprises two distinct diseases. Using genomic-scale gene expression profiling, we show that CLL is characterized by a common gene expression "signature," irrespective of Ig mutational status, suggesting that CLL cases share a common mechanism of transformation and/or cell of origin. Nonetheless, the expression of hundreds of other genes correlated with the Ig mutational status, including many genes that are modulated in expression during mitogenic B cell receptor signaling. These genes were used to build a CLL subtype predictor that may help in the clinical classification of patients with this disease.

Key words: cDNA microarrays • gene expression profiling • leukemia • lymphocytic • chronic

Introduction

The observation that the rearranged Ig variable genes in chronic lymphocytic leukemia (CLL)* cells can either be unmutated or mutated suggested that CLL might comprise two different diseases that have been lumped together using

---

Abbreviations used in this paper: BCR, B cell receptor; CLL, chronic lymphocytic leukemia; DLBCL, diffuse large B cell lymphoma; PKC, protein kinase C.

standard diagnostic methods (1–3). Somatic hypermutation of Ig genes is a specialized diversification mechanism that is activated in B cells at the germinal center stage of differentiation (4, 5). Thus, it was suggested that CLL might include two disparate malignancies, one derived from an Ig-unmutated, pregerminal center B cell, and the other from an Ig-mutated B cell that has passed through the germinal center. This "two disease" model of CLL was further supported by the observation that Ig-unmutated and Ig-mutated CLL patients had distinctly different clinical courses (2, 3). This model predicts that Ig-unmutated and Ig-mutated CLL would not be highly related to each other in gene expression. A precedent for this model is found in the recent demonstration that another lymphoid malignancy, diffuse large B cell lymphoma (DLBCL), actually includes two distinct diseases that are morphologically indistinguishable but which have largely nonoverlapping gene expression profiles (6). An alternative hypothesis is that all cases of CLL have a common cellular origin and/or a common mechanism of malignant transformation. This model predicts that Ig-mutated and Ig-unmutated CLL cases should share a gene expression signature that is characteristic of CLL.

To test these two models, and to identify molecular differences between CLL patients that might influence their clinical course, we determined the gene expression phenotype of CLL on a genomic scale using Lymphochip cDNA microarrays (6, 7). Our data demonstrate that CLL, irrespective of the Ig mutational status, is defined by a characteristic gene expression signature, thus favoring the notion that all cases share some aspects of pathogenesis. Nonetheless, we found hundreds of genes differentially expressed between Ig-unmutated and Ig-mutated CLL providing the first molecular insight into the biological mechanisms that lead to the divergent clinical behaviors of these subgroups of CLL patients. The unexpected finding that B cell activation genes were differentially expressed between the two Ig-mutational subgroups in CLL suggests the intriguing possibility that signaling pathways downstream of the B cell receptor (BCR) contribute to the more aggressive clinical behavior of the Ig-unmutated subtype.

Materials and Methods

Microarray Procedures. Peripheral blood samples from CLL patients diagnosed according to National Cancer Institute guidelines (8) were obtained after informed consent and were treated anonymously during microarray analysis. 33 CLL patients studied had not received chemotherapy at the time of sample acquisition and four patients had received prior treatment. Ig mutational status was only studied in untreated patients. Leukemic cells from CLL blood samples were purified by magnetic selection for CD19+ (Miltenyi Biotec) at 4°C before mRNA extraction and microarray analysis. Other mRNA samples from normal and malignant lymphoid populations have been described previously as have cell purification methods and array methods (6). All microarray experiments used the Cy5 dye to generate the experimental cDNA probe from mRNA of normal and malignant lymphocytes, and the Cy3 dye to generate the reference cDNA probe from mRNA pooled from nine lymphoma cell lines as described previously (6).

Expression data presented in Figs. 1, 4, and 5 are available at http://llmpp.nih.gov/cll.

Initial microarray data selection was based on fluorescence signal intensity. Each selected data point either had 100 relative fluorescent units (RFU's) above background in both the Cy3 and Cy5 channels, or 500 RFU's above background in either channel alone. A supervised selection of genes preferentially expressed in CLL cells (see Fig. 1 A) was performed as follows. First, we used the fact that the majority of cell lines that were used to construct the reference pool of mRNA were derived from DLBCL. The percentage of CLL samples with expression ratio >3 relative to the reference cell line pool was calculated, and the same calculation was also performed for the DLBCL samples. Genes were selected for which >50% of the CLL samples, and <25% of the DLBCL samples, had ratios >3. Additionally, genes were selected if the average CLL ratio was greater than the average DLBCL ratio by greater than threefold. For Fig. 1 B, representative genes were chosen from Fig. 1 A by computing the average expression in CLL samples and the average expression in resting B cell samples (adult and cord blood B cells). CLL signature genes were chosen to be at least twofold more highly expressed in CLL than in resting B cells and CLL/resting B cell genes were chosen to be expressed equivalently (within twofold) in the two sample sets. Duplicate array elements representing the same genes were removed. Germinal center genes were chosen from a previous analysis (6).

RT-PCR. 500 ng poly-A+ mRNA was used to generate first strand cDNA using Superscript (Life Technologies) together with random hexamers and oligo-dT primers. ZAP-70 oligonucleotide primers (5′ TCTCCAAAGCACTGGGTG 3′, 5′ AGCT-GTGTGTGGAGACAACCAAG 3′) were then used for PCR amplification for 27 cycles.

Statistical Analysis. A two-group t-statistic on log2 expression ratios was used to measure the ability of each array element to discriminate between the two CLL mutational subtypes univariately. For multivariate subtype prediction, we used a linear combination of log2 expression ratios for array elements that were significant at the $P < 0.001$ significance level in the univariate analysis. The expression ratios were weighted in the linear combination by the univariate t statistics. The linear combination was computed for each sample and the average linear combination was computed for each CLL subtype. The midpoint of the two CLL subtype means was used as a cut-point for subtype prediction. For the cross-validation analysis, the subtype predictor was calculated by sequentially omitting one sample from the test set of cases, and using the remaining cases to generate the predictor. In Fig. 4 B, calculation of the P value from the permutation distribution of the t-statistic also demonstrated the high statistical significance of the differential gene expression between the CLL subtypes (data not shown). Classification was determined on all CLL cases with the exception of CLL-60 (Ig-unmutated) and CLL-21 and CLL-51 (minimally mutated cases).

In Fig. 5, the choice of B cell activation genes was made as follows. The B cell activation series of microarray experiments included several different stimulations with anti-IgM for 6, 24, and 48 h for each Lymphochip array element, we averaged the data at each activation time point, and then selected those elements that gave a twofold induction compared with the resting B cell average for at least one time point.

Results

The Gene Expression Signature of CLL. We profiled gene expression in CLL samples ($n = 37$) using Lymphochip

cDNA microarrays containing 17,856 human cDNAs (7). To facilitate comparison of each CLL mRNA sample with the others and with previously generated data sets, we compared gene expression in each CLL mRNA sample to a common reference mRNA pool prepared from lymphoid cell lines (6, 7). Using this strategy, the relative gene expression in the CLL cases could be compared with other B cell malignancies (DLBCL and follicular lymphoma) and of normal B cell and T cell subpopulations. Fig. 1 A presents expression data from 328 Lymphochip array elements representing ~247 genes that were selected in a supervised fashion (see Materials and Methods) to be more highly expressed in the majority of CLL samples than in DLBCL samples ($n = 40$). These genes fall into two broad categories, which are highlighted by representative genes in Fig. 1 B. Genes in the first category define a CLL gene expression "signature" that distinguishes CLL from various normal B cell subsets and from other B cell malignancies. The CLL signature genes were not expressed highly in resting blood B cells or in germinal center B cells. This group of genes includes several named genes not previously suspected to be expressed in CLL (e.g., Wnt3, titin, Ror1) as well as a number of novel genes from various normal and malignant B cell cDNA libraries. By contrast, CLL cells lacked expression of most genes that are preferentially expressed in germinal cen-

ter B cells (Fig. 1 C). In addition to this set of CLL signature genes, CLL preferentially expressed a set of genes that distinguish resting, G0 stage blood B cells from mitogenically activated blood B cells and germinal center B cells that are traversing the cell cycle (Fig. 1 B). The expression of these resting B cell genes by CLL cells is consistent with the indolent, slowly proliferating character of this malignancy.

One of these resting B cell samples was prepared from human cord blood that is enriched for B cells bearing the CD5 surface marker, a B cell subpopulation that has been proposed to be the normal counterpart of CLL. The cord blood B cells were >80% CD5+ by FACS® analysis (data not shown) whereas resting B cells from adult blood are 10–20% CD5+ (9). We did not observe notably higher expression of the CLL signature genes in the cord blood B cell sample than in the adult B cell sample (Fig. 1) and no overall correlation in the expression of genes in Fig. 1 was observed between CLL and either adult or cord blood B cells (Pearson correlation coefficients −0.27 and −0.21, respectively). Thus, our gene expression profiling analysis does not provide support for the hypothesis that the CD5+ B cell is a CLL precursor. It is certainly possible, however, that the expression of the CLL signature genes might be due to the oncogenic mechanisms of CLL and therefore might not be a feature of any normal B cell subpopulation.
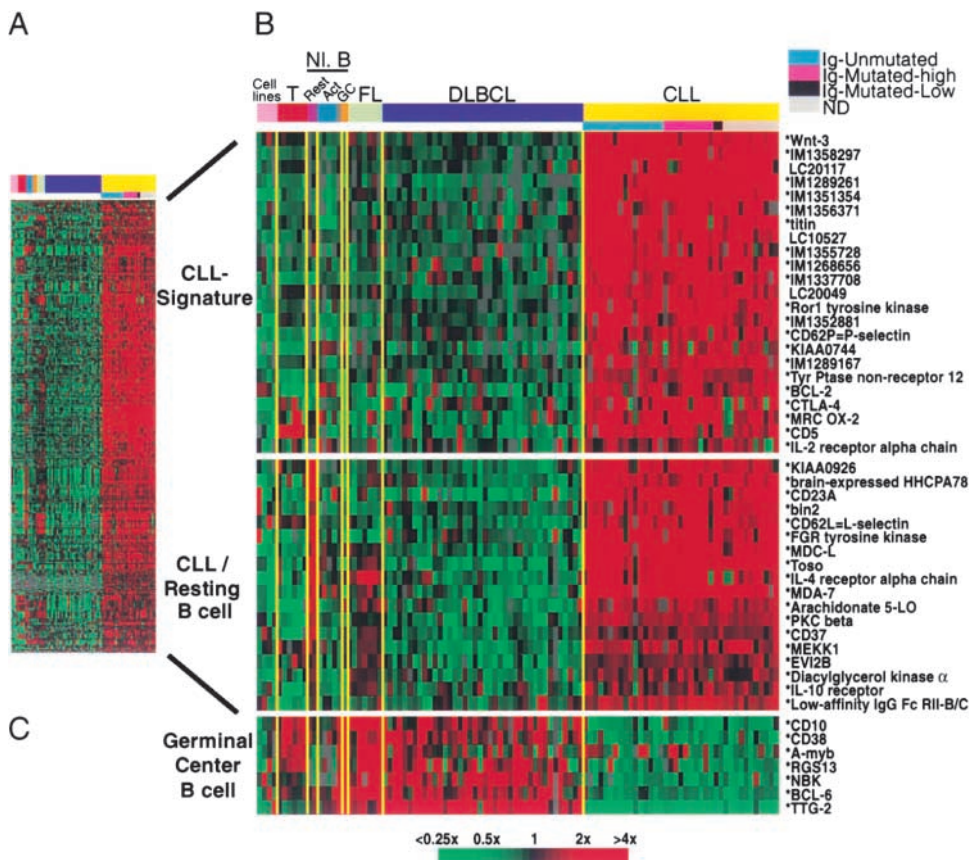
Figure 1. Discovery of a common gene expression phenotype in CLL. (A) 328 Lymphochip array elements representing ~247 genes that were more highly expressed as mRNA in the majority of CLL samples relative to DLBCL samples. The expression data are presented as a matrix in which the rows represent individual genes, and the columns represent individual mRNA samples. The relative level of gene expression is depicted according to the color scale shown at the bottom. Gray squares indicate missing or excluded data. (B) Relative expression levels of selected genes from A in mRNA samples presented in the following order, from left to right: cell lines (JVM-HH, OCI-Ly10, OCI-Ly3, U937); T cells (adult, CD4+, unstimulated; neonatal, CD4+, unstimulated; fetal, CD4+, unstimulated; adult, CD4+, + PMA [P] and ionomycin [I]; neonatal cord blood T cells, + P and I; fetal, CD4+, + P and I), resting B-cells (cord blood CD19+ B cells; adult blood CD19+ B cells), activated B cells (adult blood B cells, anti-IgM + CD40L 6 h; adult blood B cells, anti-IgM 24 h; adult blood B cells, anti-IgM + CD40L 24 h; adult blood B cells, anti-IgM + IL-4 24 h), adult blood memory B-cells (CD27+), tonsil germinal center B cells, follicular lymphomas ($n = 7$), DLBCLs ($n = 40$), and CLL ($n = 37$). The CLL samples were grouped according to Ig mutational status as indicated. In the gene names, an asterisk denotes a sequence verified gene, IM indicates an IMAGE consortium (reference 29) clone identification number, and LC indicates an unsequenced Lymphochip clone identification number. (C) Relative expression levels of selected genes characteristic of the germinal center B cell differentiation stage.

1641   Rosenwald et al.

Ig Mutational Status. The expressed Ig heavy chain genes were sequenced from 28 CLL cases and compared with known germ-line encoded Ig V_H segments as described previously (10) (Fig. 2 A). By convention, V_H sequences that matched known germ line sequences with >98% identity were considered unmutated, as any minor differences observed in this group were assumed to reflect genetic polymorphism (1–3). By this criterion, 16 CLL cases in our study set were unmutated. The remaining cases were further separated into a group of 10 highly mutated cases (<97% identity with any germ-line V_H segment) and a group of two cases that were minimally mutated (>97% but <98% identity with known germ-line V_H genes). CLL cases were grouped in Fig. 1 according to Ig mutational status as indicated. Although some variation in expression of the CLL signature and CLL/resting B cell genes was evident between CLL patients, most patients in each Ig mutational subtype highly expressed these genes at comparable levels. Furthermore, an unsupervised hierarchical clustering of the CLL cases using 10,249 Lymphochip array elements resulted in a clustering dendrogram in which the Ig-unmutated and Ig-mutated CLL cases were extensively intermingled (data not shown). Thus, the overall gene expression profiles of the two CLL subtypes were largely overlapping.

Segregation of our patients according to Ig mutational status revealed that Ig-unmutated CLL patients had a significantly worse clinical course, requiring earlier treatment, than the Ig-mutated CLL patients (Fig. 2 B), in keeping with previous reports (2, 3).

CLL Subtype Distinction Genes. Given the dramatically different clinical behavior of the Ig-unmutated and Ig-mutated CLL patients, it was evident that gene expression differences should be discernible between these groups. To both discover such genes and statistically validate their relationship to the Ig-mutational subgroups, we conducted the Ig mutational analysis independently and sequentially in two random subsets of our CLL patients (Fig. 3). The "training" set consisted of 10 Ig-unmutated cases and eight Ig-mutated cases. In this gene discovery phase, we assigned the minimally mutated CLL cases to the mutated class. The mean expression of each gene was then calculated for both mutational subgroups and the statistical significance of the difference of these means was determined. All genes that discriminated between the mutational subgroups at a significance of $P < 0.001$ ($n = 56$) were used to form a "predictor" that could be used to assign a CLL sample to a mutational subgroup based on gene expression (see Methods).

The performance of this CLL subtype predictor was initially tested using a cross-validation strategy (Fig. 3 A). One of the 18 CLL samples in the training set was omitted, the statistically significant genes were determined, and a predictor was calculated based on the remaining 17 samples. The omitted sample was then assigned to a CLL subtype based
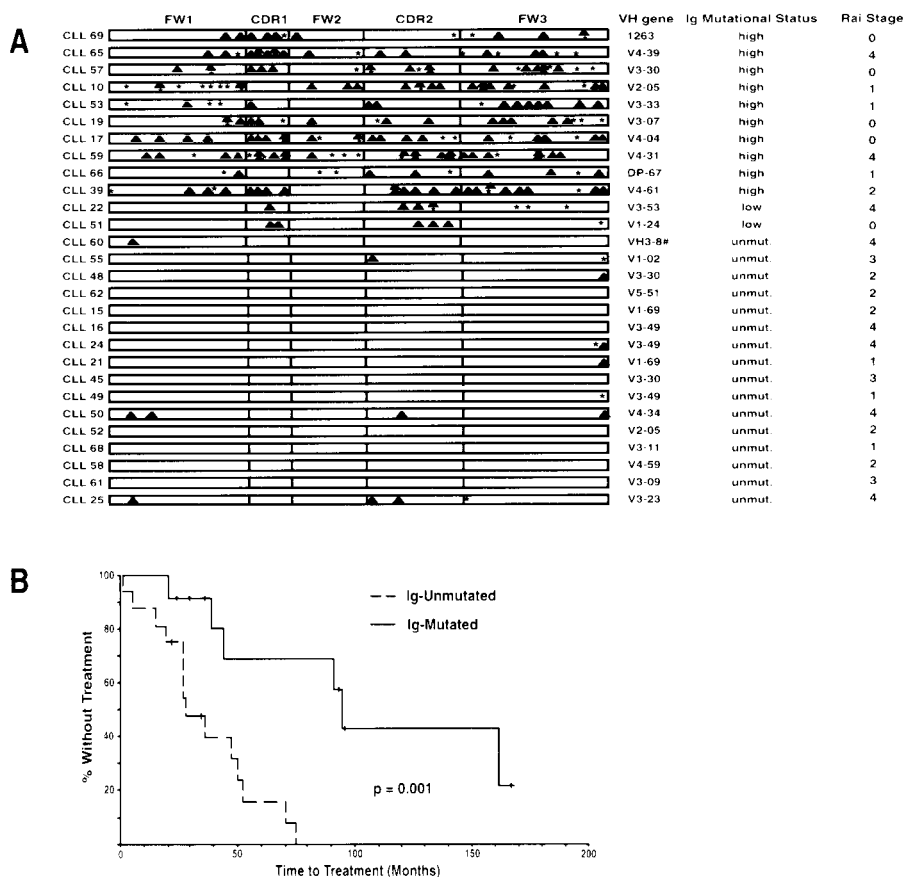
		FW1	CDR1	FW2	CDR2	FW3	VH gene	Ig Mutational Status	Rai Stage
A	CLL 69						1263	high	0
	CLL 65						V4-39	high	4
	CLL 57						V3-30	high	0
	CLL 10						V2-05	high	1
	CLL 53						V3-33	high	1
	CLL 19						V3-07	high	0
	CLL 17						V4-04	high	0
	CLL 59						V4-31	high	4
	CLL 66						DP-67	high	1
	CLL 39						V4-61	high	2
	CLL 22						V3-53	low	4
	CLL 51						V1-24	low	0
	CLL 60						VH3-8#	unmut.	4
	CLL 55						V1-02	unmut.	3
	CLL 48						V3-30	unmut.	2
	CLL 62						V5-51	unmut.	2
	CLL 15						V1-69	unmut.	2
	CLL 16						V3-49	unmut.	4
	CLL 24						V3-49	unmut.	4
	CLL 21						V1-69	unmut.	1
	CLL 45						V3-30	unmut.	3
	CLL 49						V3-49	unmut.	1
	CLL 50						V4-34	unmut.	4
	CLL 52						V2-05	unmut.	2
	CLL 68						V3-11	unmut.	1
	CLL 58						V4-59	unmut.	2
	CLL 61						V3-09	unmut.	3
	CLL 25						V3-23	unmut.	4

Figure 2. Analysis of somatic mutations in the Ig V_H genes in 28 CLL patients and correlation with their clinical courses. (A) V_H gene usage and distribution of replacement (▲) and silent (*) mutations in the complementarity determining regions (CDRs) and framework regions (FWs). #, V_H sequence most homologous to multiple cDNA sequences. (B) Kaplan-Meier curve comparing the time from diagnosis to treatment between CLL patients with mutated and unmutated V_H genes. Median time to treatment in Ig-mutated CLL: 95 mo; median time to treatment in Ig-unmutated CLL: 28 mo. The difference is significant at the $P = 0.001$ level (log-rank test).
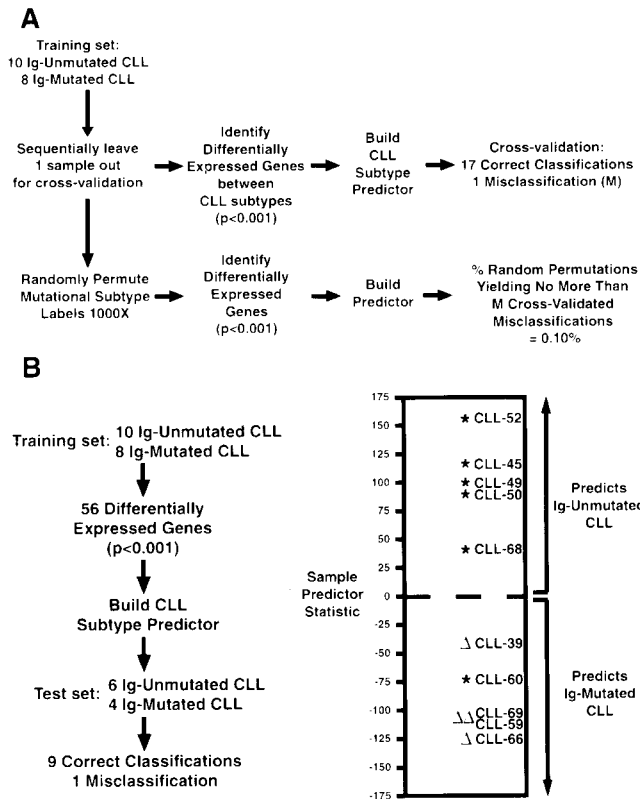
1642   Gene Expression Profiling in Chronic Lymphocytic Leukemia

A

Training set:
10 Ig-Unmutated CLL
8 Ig-Mutated CLL

Sequentially leave 1 sample out for cross-validation → Identify Differentially Expressed Genes between CLL subtypes (p<0.001) → Build CLL Subtype Predictor → Cross-validation: 17 Correct Classifications 1 Misclassification (M)

Randomly Permute Mutational Subtype Labels 1000X → Identify Differentially Expressed Genes (p<0.001) → Build Predictor → % Random Permutations Yielding No More Than M Cross-Validated Misclassifications = 0.10%

B

Training set: 10 Ig-Unmutated CLL 8 Ig-Mutated CLL

↓

56 Differentially Expressed Genes (p<0.001)

↓

Build CLL Subtype Predictor

↓

Test set: 6 Ig-Unmutated CLL 4 Ig-Mutated CLL

↓

9 Correct Classifications 1 Misclassification

Sample Predictor Statistic

★ CLL-52
★ CLL-45
★ CLL-49
★ CLL-50
★ CLL-68

Predicts Ig-Unmutated CLL

Δ CLL-39
★ CLL-60
ΔΔ CLL-69
CLL-59
Δ CLL-66

Predicts Ig-Mutated CLL

Figure 3. Statistical methodology for the creation and validation of an Ig-mutational status predictor in CLL. (A) Performance of the predictor using a cross-validation strategy. (B) Performance of the Ig-mutational subtype predictor in a test set of six unmutated (★) and four mutated CLL (Δ) samples.

on gene expression using this predictor. The Ig mutational status of 17 CLL samples was correctly assigned by this procedure with one misassignment. To test the statistical significance of this result, we created 1,000 random permutations of the assignments of CLL samples to the Ig mutation subgroups. For each permutation, the cross-validation process described above was repeated. Only one of the 1,000 random permutations generated a predictor that performed as well as the predictor based on the unpermuted data, demonstrating that the significance of the gene expression difference between the CLL subtypes was $P = 0.001$.

As a final test of the CLL subtype predictor, we determined the Ig mutational status of a "test" set of 10 additional CLL cases and used the predictor derived from the training set to assign the cases in this test set to a CLL subtype based on gene expression in a blinded fashion (Fig. 3 B). Nine out of ten of the test cases were correctly assigned, showing the ability of the CLL subtype predictor to correctly assign new CLL cases based on gene expression data that was not used to generate the predictor. The one misclassified CLL case (CLL-60) clearly was an outlier in gene expression (see below). Taken together with the cross-validation results, these data demonstrate that gene expression can define CLL subtypes that have different degrees of Ig mutation.

An important practical benefit of these findings would be to create a diagnostic test for the CLL subtypes based upon gene expression. In this regard, one of the most differentially expressed genes from the analysis of the training set of cases, ZAP-70, could classify all of the cases in both the training and the test set with 100% accuracy. Likewise, predictors based on two genes (ZAP-70 and IM1286077) or three genes (ZAP-70, IM1286077, activation-induced C-type lectin) discovered using the training set formed CLL subtype predictors that performed with 100% accuracy on the training set and test set of CLL cases.

We next expanded our search for CLL subtype distinction genes using data from both the training set and test set of CLL cases. The two CLL cases with minimal Ig mutations (CLL-22 and CLL-51) were excluded based on the possibility that their Ig sequences might actually represent as yet undescribed polymorphic V_H alleles. CLL-60 was excluded based on its unusual gene expression characteristics that led to its misclassification by the CLL subtype predictor. Fig. 4 A presents 205 Lymphochip array elements (~175 genes) that were differentially expressed between the CLL subtypes with a statistical significance of $P < 0.001$. Hierarchical clustering of the CLL cases based on expression of these genes placed the majority of Ig-unmutated CLL cases in one cluster and the Ig-highly mutated CLL cases in another. As expected, CLL-60 was more closely aligned with the Ig-mutated CLL cases, though it was an outlier from the major cluster of Ig-mutated CLL cases. Interestingly, both of the CLL cases with a low Ig mutational load were also outliers, though they were more closely related to the Ig-mutated CLL subtype than to the Ig-unmutated CLL subtype. These data define two predominant CLL subtypes that differ in the expression of hundreds of genes but also demonstrate that additional minor CLL subtypes may exist that have distinct gene expression profiles. Fig. 4 B highlights some of the genes that most strongly differentiate between the CLL subtypes. ZAP-70 was the most tightly discriminating gene, with an average 4.3-fold higher expression in Ig-unmutated CLL than in Ig-mutated CLL ($P < 10^{-6}$). RT-PCR analysis confirmed ZAP-70 expression in two Ig-unmutated CLL cases (CLL-48 and CLL-49), in contrast to CLL-66 and CLL-69 that were Ig-mutated (Fig. 4 C). Surprisingly, ZAP-70 expression was also observed in several B cell lines (LILA, LK-6, OCI-Ly2), but not in many others (Raji; Fig. 4 C, and data not shown).

Relationship between B Cell Activation and the CLL Subtype Distinction. Several of the CLL subtype distinction genes are known or suspected to be induced by protein kinase C (PKC) signaling, including activation-induced C-type lectin (11), MDS019, a very close paralogue of phorbolin 1 (12), and gravin, a scaffold protein that binds PKC and may regulate its activity (13). One mechanism by which PKC is activated in B cells is through BCR signaling (14). Therefore, we investigated whether the CLL subtype distinction genes are regulated during activation of blood B cells, using a gene expression database generated previously using Lymphochip microarrays (6). Strikingly, many of the
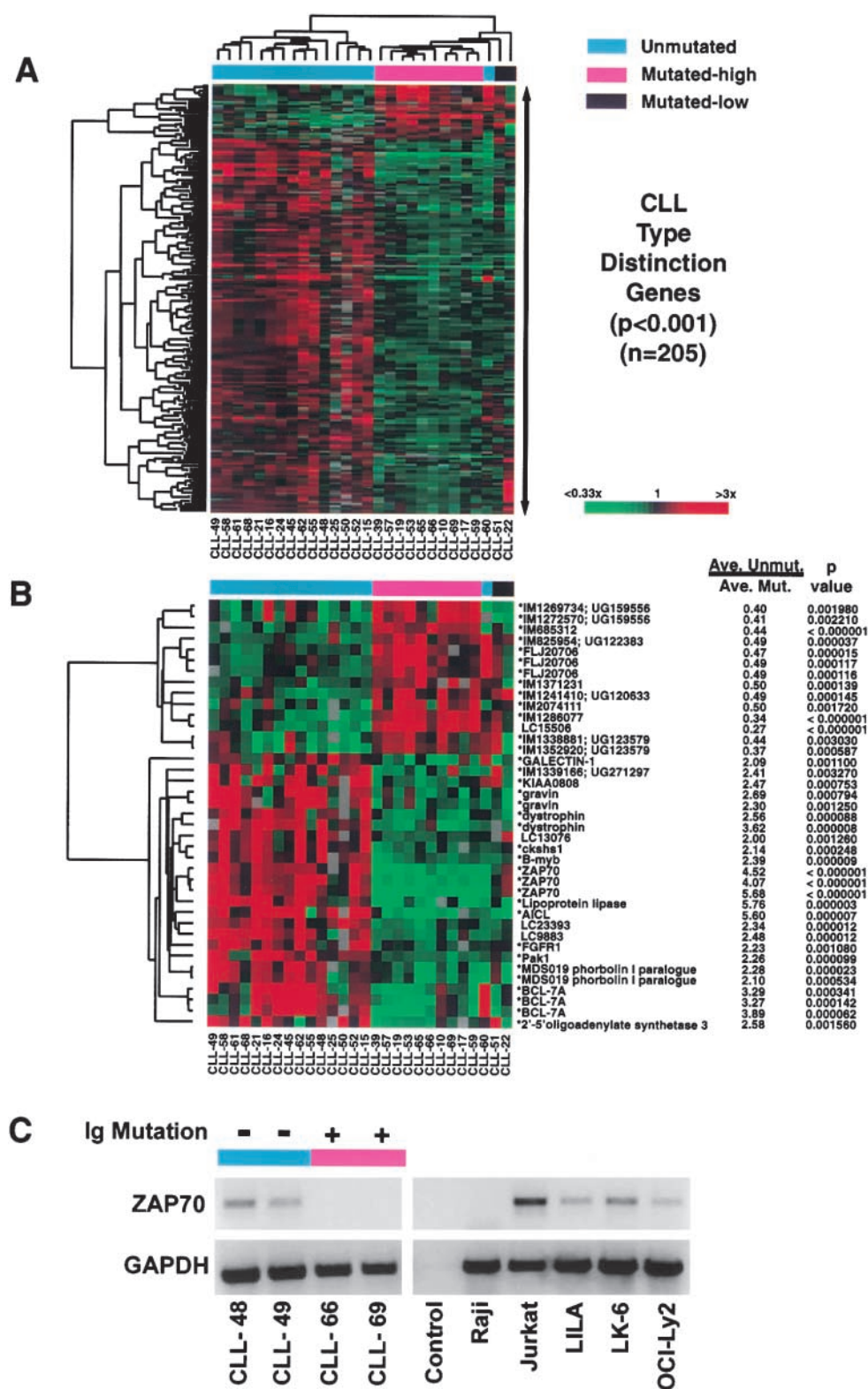
Figure 4. Relative gene expression levels of CLL subtype distinction genes. (A) Hierarchical clustering of gene expression data for 205 array elements representing ~175 genes that were differentially expressed between mutated and unmutated CLL samples ($P < 0.001$). (B) Hierarchical clustering of genes that most strongly discriminated between the CLL subtypes. Also shown for each gene is the ratio of mean expression of the gene in Ig-unmutated CLL samples (excluding CLL-60) versus mean expression in Ig-mutated (high) CLL samples, together with the P values (Student's t test) that quantitate the significance of the difference in mean expression between the two CLL subtypes. (C) RT-PCR analysis of ZAP-70 expression. Shown are data from two Ig-unmutated and two Ig-mutated CLL cases, a T cell line (Jurkat), various B cell lines found by microarray analysis to express ZAP-70 (LILA, LK6, OCI-Ly2), and a B cell line not expressing ZAP-70 (Raji; C, and data not shown). The control lane represents a reaction in which the reverse transcriptase was omitted.

genes that were more highly expressed in Ig-unmutated CLL were induced during activation of blood B cells (Fig. 5 A). Many of these genes encode proteins involved in cell cycle control (e.g., cyclin D2) or in cellular metabolism required for cell cycle progression (e.g., HPRT and other nucleotide modifying enzymes). Conversely, the majority of the genes that were expressed at lower levels in Ig-unmutated CLL were strongly downmodulated during B cell activation (Fig. 5 B). These results demonstrate that the CLL subtype distinction genes are enriched for genes that
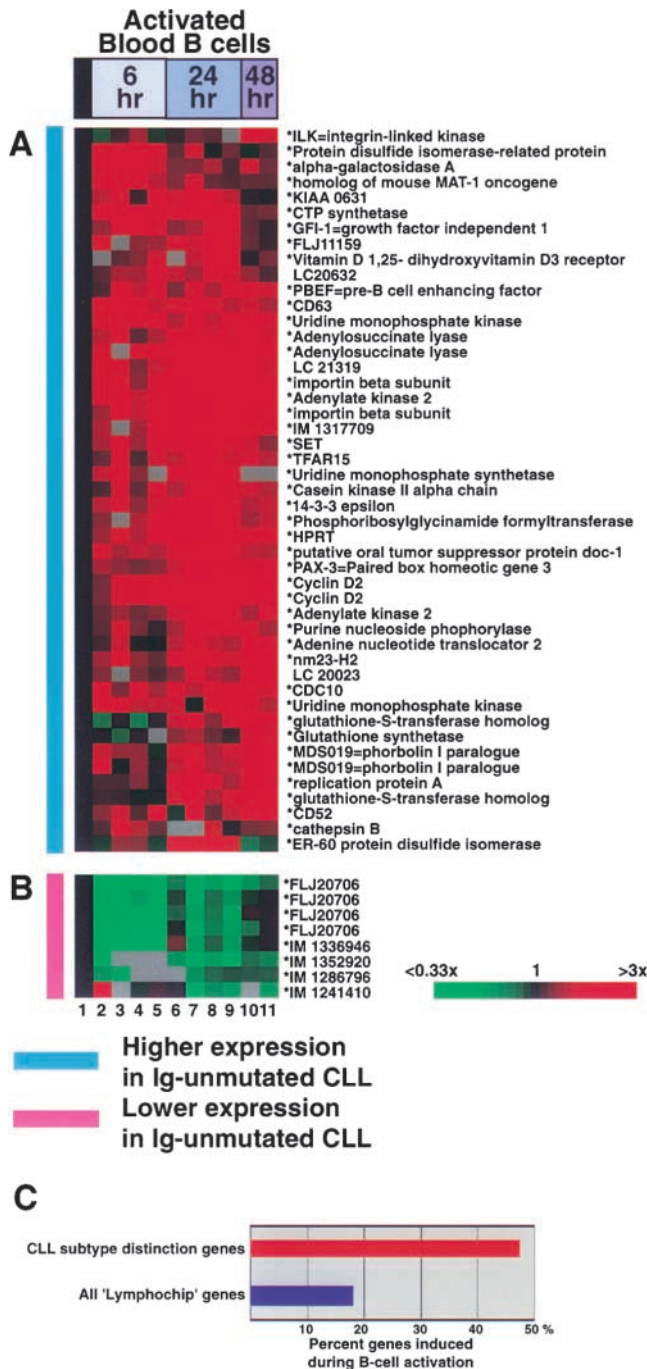
1644   Gene Expression Profiling in Chronic Lymphocytic Leukemia

Figure 5. (A and B) Response of CLL subtype distinction genes during B cell activation. Gene expression data from the following B cell samples is depicted: 1, gene expression average from three resting B cell samples; 2, blood B cells, anti-IgM 6 h; 3, blood B cells, anti-IgM + CD40L 6 h; 4, blood B cells, anti-IgM + IL-4; 5, blood B cells, anti-IgM + CD40L + IL-4; 6, blood B cells, anti-IgM 24 h; 7, blood B cells, anti-IgM + CD40L 24 h; 8, blood B cells, anti-IgM + IL-4 24 h; 9, blood B cells, anti-IgM + CD40L + IL-4 24 h; 10–11, blood B cells, anti-IgM + CD40L + IL-4 48 h. (C) Percentage of CLL subtype distinction genes (red bar) and all 'Lymphochip' genes (blue bar) that are induced during B cell activation.

are modulated in expression by B cell activation. Indeed, 47% of the CLL subtype distinction genes were induced during B cell activation, whereas only 18% of all Lympho-chip genes were in this category (Fig. 5 C).

Discussion

The comprehensive profiling of gene expression in CLL presented here provides a new molecular framework for understanding the etiology of this leukemia and the divergent clinical courses of these patients. Using genomic-scale gene expression profiling, we addressed a current controversy in CLL pathogenesis, namely whether this diagnosis comprises more than one disease entity. CLL patients have been subdivided based on the Ig mutational status of their leukemic cells (1–3), but it was unclear whether these patients had molecularly distinct diseases. Our data demonstrate that all CLL patients share a characteristic gene expression signature in their leukemic cells. These findings support a model in which all cases of CLL have a common cell of origin and/or a common mechanism of malignant transformation. In this model, the CLL-specific gene expression signature might represent the gene expression signature of a common normal precursor cell or it might reflect the downstream gene expression consequences of a common oncogenic event. These findings are in contrast to the previous observation that DLBCL consists of two disease entities that did not have overlapping gene expression outside of genes involved in proliferation and in the host response to the tumor (6).

Previously unsuspected features of CLL biology emerge from its gene expression profile, generating a wealth of hypotheses to guide future studies of this disease. CLL cells proliferate slowly in vivo, driven by unknown signals. Therefore, it is notable that Wnt-3 was highly, and selectively, expressed in CLL (Fig. 1 B). The Wnt gene family encodes secreted proteins that signal through cell surface receptors of the frizzled family to control development and mediate malignant transformation (15). Intriguingly, another CLL signature gene, Ror1, encodes a receptor tyrosine kinase with an extracellular domain that resembles a Wnt interaction domain of frizzled (16). Recently, Wnt-3 has been shown to promote proliferation of mouse bone marrow pro-B cells by initiating signaling events leading to transcriptional activation by LEF-1(17). Thus, CLL cells may use an autocrine mechanism of proliferation that is used normally by B cell progenitors.

We nevertheless also found that the expression of hundreds of other genes correlated with the Ig mutational status in CLL, providing insights into the biological mechanisms that lead to the divergent clinical behaviors of CLL patients. The most differentially expressed gene between the CLL subtypes was ZAP-70, a critical kinase that transduces signals from the T cell antigen receptor, and is preferentially expressed in normal T lymphocytes (18). Differential expression of ZAP-70 between CLL subtypes was therefore surprising since its expression in normal B cells has not been previously reported. However, by microarray analysis and RT-PCR analysis we found that ZAP-70 mRNA is highly expressed in some B lymphoma cell lines along with being differentially expressed by the CLL subtypes. A ZAP-70–related kinase, syk, transduces signals from the BCR (19), raising the possibility that ZAP-70

might alter BCR signaling in CLL cells. Another CLL subtype distinction gene, Pak1, could contribute to the resistance of CLL cells to apoptosis by phosphorylating Bad and thereby preventing Bad from inhibiting BCL-2 (20). FGFR1 is a receptor tyrosine kinase that can stimulate cellular proliferation after interaction with fibroblast growth factors. The higher expression of FGFR1 in Ig-unmutated CLL is intriguing given that CLL patients have elevated blood levels of basic fibroblast growth factor which can activate FGFR1 and block apoptosis in CLL (21, 22).

Intriguingly, CLL subtype distinction genes were enriched for genes that are modulated in expression during signaling of B cells through the BCR. One hypothesis raised by this observation is that the leukemic cells in Ig-unmutated CLL may have ongoing BCR signaling. Interestingly, the V_H repertoire usage in the Ig-unmutated and Ig-mutated CLL is distinct (1–3) and the combinations of V_H, D_H, and J_H gene segments rearranged in CLL cells are not random (1–3, 23, 24). These observations suggest that the surface Ig receptors of CLL cells may have specificity for unknown environmental or self-antigens. Indeed, CLL cells have been shown to frequently produce antibodies that bind classical autoantigens (25–27). The gene expression profiling data presented in this report raise the possibility that Ig-unmutated CLL cells may be continuously stimulated in vivo by antigen, giving rise to a gene expression profile that is reminiscent of BCR signaling. Indeed, CLL cells from patients with progressive disease were more readily stimulated by BCR cross-linking to synthesize DNA than were CLL cells from patients with stable disease (28). Although this study did not distinguish between Ig-unmutated and Ig-mutated CLL, the results are consistent with a differential ability of these subtypes to signal through the BCR. Alternatively, it is possible that Ig-unmutated CLL cells activate the same signaling pathways that are engaged during B cell activation as a result of genetic changes in the leukemic cells or by other pathological mechanisms.

An immediate clinical application of the present results would be in the differential molecular diagnosis of CLL. We demonstrated that as few as 1–3 genes could correctly assign patients to a CLL subtype with 100% accuracy. Thus, our results could be used to establish a quantitative RT-PCR test to diagnose the CLL subtypes and that would be easier to adopt clinically than DNA sequence analysis of Ig variable regions. Given the relatively benign course of Ig-mutated CLL, a simple diagnostic test based on gene expression would provide valuable prognostic information for CLL patients and could be used to guide treatment decisions.

Finally, our results suggest new therapeutic approaches to this currently incurable leukemia. First, the protein products of some of the CLL signature genes may present new targets for mAb therapy and for vaccine approaches to CLL. Second, the unexpected finding that B cell activation genes were upregulated in Ig-unmutated CLL patients suggests the intriguing possibility that signaling pathways downstream of the BCR may contribute to the more progressive clinical course of these patients. Thus, therapeutic targeting of these signaling pathways could specifically benefit those CLL patients that show gene expression evidence that these pathways are active.

References

1. Fais, F., F. Ghiotto, S. Hashimoto, B. Sellars, A. Valetto, S.L. Allen, P. Schulman, V.P. Vinciguerra, K. Rai, L.Z. Rassenti, et al. 1998. Chronic lymphocytic leukemia B cells express restricted sets of mutated and unmutated antigen receptors. *J. Clin. Invest.* 102:1515–1525.
2. Hamblin, T.J., Z. Davis, A. Gardiner, D.G. Oscier, and F.K. Stevenson. 1999. Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. *Blood.* 94:1848–1854.
3. Damle, R.N., T. Wasil, F. Fais, F. Ghiotto, A. Valetto, S.L. Allen, A. Buchbinder, D. Budman, K. Dittmar, J. Kolitz, et al. 1999. Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. *Blood.* 94:1840–1847.
4. Jacob, J., G. Kelsoe, K. Rajewsky, and U. Weiss. 1991. Intraclonal generation of antibody mutants in germinal centers. *Nature.* 354:389–392.
5. Klein, U., T. Goossens, M. Fischer, H. Kanzler, A. Braeuninger, K. Rajewsky, and R. Kuppers. 1998. Somatic hypermutation in normal and transformed human B cells. *Immunol. Rev.* 162:261–281.
6. Alizadeh, A.A., M.B. Eisen, R.E. Davis, C. Ma, I.S. Lossos, A. Rosenwald, J.C. Boldrick, H. Sabet, T. Tran, X. Yu, et al. 2000. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature.* 403:503–511.
7. Alizadeh, A., M. Eisen, R.E. Davis, C. Ma, H. Sabet, T. Tran, J. Powell, L. Yang, G. Marti, T. Moore, et al. 1999. The Lymphochip: a specialized cDNA microarray for the genomic-scale analysis of gene expression in normal and malignant lymphocytes. *Cold Spring Harbor Symp. Quant. Biol.* 64: 71–78.
8. Cheson, B.D., J.M. Bennett, M. Grever, N. Kay, M.J. Keating, S. O'Brien, and K.R. Rai. 1996. National Cancer Institute-sponsored working group guidelines for chronic lymphocytic leukemia: revised guidelines for diagnosis and treatment. *Blood.* 87:4990–4997.
9. Geiger, K.D., U. Klein, A. Brauninger, S. Berger, K. Leder,

K. Rajewsky, M.L. Hansmann, and R. Kuppers. 2000. CD5-positive B cells in healthy elderly humans are a polyclonal B cell population. *Eur. J. Immunol.* 30:2918–2923.

10. Bessudo, A., V. Cherepakhin, T.A. Johnson, L.Z. Rassenti, E. Feigal, and T.J. Kipps. 1996. Favored use of immunoglobulin V(H)4 Genes in AIDS-associated B-cell lymphoma. *Blood.* 88:252–260.

11. Hamann, J., K.T. Montgomery, S. Lau, R. Kucherlapati, and R.A. van Lier. 1997. AICL: a new activation-induced antigen encoded by the human NK gene complex. *Immunogenetics.* 45:295–300.

12. Madsen, P., S. Anant, H.H. Rasmussen, P. Gromov, H. Vorum, J.P. Dumanski, N. Tommerup, J.E. Collins, C.L. Wright, I. Dunham, et al. 1999. Psoriasis upregulated phorbolin-1 shares structural but not functional similarity to the mRNA-editing protein apobec-1. *J. Invest. Dermatol.* 113: 162–169.

13. Nauert, J.B., T.M. Klauck, L.K. Langeberg, and J.D. Scott. 1997. Gravin, an autoantigen recognized by serum from myasthenia gravis patients, is a kinase scaffold protein. *Curr. Biol.* 7:52–62.

14. Cambier, J.C., C.M. Pleiman, and M.R. Clark. 1994. Signal transduction by the B cell antigen receptor and its coreceptors. *Annu. Rev. Immunol.* 12:457–486.

15. Polakis, P. 2000. Wnt signaling and cancer. *Genes Dev.* 14: 1837–1851.

16. Saldanha, J., J. Singh, and D. Mahadevan. 1998. Identification of a Frizzled-like cysteine rich domain in the extracellular region of developmental receptor tyrosine kinases. *Protein Sci.* 7:1632–1635.

17. Reya, T., M. O'Riordan, R. Okamura, E. Devaney, K. Willert, R. Nusse, and R. Grosschedl. 2000. Wnt signaling regulates B lymphocyte proliferation through a LEF-1 dependent mechanism. *Immunity.* 13:15–24.

18. Chu, D.H., C.T. Morita, and A. Weiss. 1998. The Syk family of protein tyrosine kinases in T-cell activation and development. *Immunol. Rev.* 165:167–180.

19. Turner, M., E. Schweighoffer, F. Colucci, J.P. Di Santo, and V.L. Tybulewicz. 2000. Tyrosine kinase SYK: essential functions for immunoreceptor signalling. *Immunol. Today.* 21: 148–154.

20. Schurmann, A., A.F. Mooney, L.C. Sanders, M.A. Sells, H.G. Wang, J.C. Reed, and G.M. Bokoch. 2000. p21-activated kinase 1 phosphorylates the death agonist bad and protects cells from apoptosis. *Mol. Cell. Biol.* 20:453–461.

21. Aguayo, A., H. Kantarjian, T. Manshouri, C. Gidel, E. Estey, D. Thomas, C. Koller, Z. Estrov, S. O'Brien, M. Keating, et al. 2000. Angiogenesis in acute and chronic leukemias and myelodysplastic syndromes. *Blood.* 96:2240–2245.

22. Menzel, T., Z. Rahman, E. Calleja, K. White, E.L. Wilson, R. Wieder, and J. Gabrilove. 1996. Elevated intracellular level of basic fibroblast growth factor correlates with stage of chronic lymphocytic leukemia and is associated with resistance to fludarabine. *Blood.* 87:1056–1063.

23. Widhopf, G.F. II, and T.J. Kipps. 2001. Normal B cells express 51p1-encoded Ig heavy chains that are distinct from those expressed by chronic lymphocytic leukemia B cells. *J. Immunol.* 166:95–102.

24. Johnson, T.A., L.Z. Rassenti, and T.J. Kipps. 1997. Ig VH1 genes expressed in B cell chronic lymphocytic leukemia exhibit distinctive molecular features. *J. Immunol.* 158:235–246.

25. Borche, L., A. Lim, J.L. Binet, and G. Dighiero. 1990. Evidence that chronic lymphocytic leukemia B lymphocytes are frequently committed to production of natural autoantibodies. *Blood.* 76:562–569.

26. Sthoeger, Z.M., M. Wakai, D.B. Tse, V.P. Vinciguerra, S.L. Allen, D.R. Budman, S.M. Lichtman, P. Schulman, L.R. Weiselberg, and N. Chiorazzi. 1989. Production of autoantibodies by CD5-expressing B lymphocytes from patients with chronic lymphocytic leukemia. *J. Exp. Med.* 169:255–268.

27. Broker, B.M., A. Klajman, P. Youinou, J. Jouquan, C.P. Worman, J. Murphy, L. Mackenzie, R. Quartey-Papafio, M. Blaschek, P. Collins, et al. 1988. Chronic lymphocytic leukemic (CLL) cells secrete multispecific autoantibodies. *J. Autoimmun.* 1:469–481.

28. Aguilar-Santelises, M., H. Mellstedt, and M. Jondal. 1994. Leukemic cells from progressive B-CLL respond strongly to growth stimulation in vitro. *Leukemia.* 8:1146–1152.

29. Lennon, G., C. Auffray, M. Polymeropoulos, and M.B. Soares. 1996. The I.M.A.G.E. Consortium: an integrated molecular analysis of genomes and their expression. *Genomics.* 33:151–152.