Quantitative DNA Methylation Analysis Identifies a Single CpG Dinucleotide Important for ZAP-70 Expression and Predictive of Prognosis in Chronic Lymphocytic Leukemia

Rainer Claus, David M. Lucas, Stephan Stilgenbauer, Amy S. Ruppert, Lianbo Yu, Manuela Zucknick, Daniel Mertens, Andreas Bühler, Christopher C. Oakes, Richard A. Larson, Neil E. Kay, Diane F. Jelinek, Thomas J. Kipps, Laura Z. Rassenti, John G. Gribben, Hartmut Döhner, Nyla A. Heerema, Guido Marcucci, Christoph Plass, and John C. Byrd

See accompanying article on page 2566

ABSTRACT

Purpose

Increased ZAP-70 expression predicts poor prognosis in chronic lymphocytic leukemia (CLL). Current methods for accurately measuring ZAP-70 expression are problematic, preventing widespread application of these tests in clinical decision making. We therefore used comprehensive DNA methylation profiling of the ZAP-70 regulatory region to identify sites important for transcriptional control.

Patients and Methods

High-resolution quantitative DNA methylation analysis of the entire ZAP-70 gene regulatory regions was conducted on 247 samples from patients with CLL from four independent clinical studies.

Results

Through this comprehensive analysis, we identified a small area in the 5' regulatory region of ZAP-70 that showed large variability in methylation in CLL samples but was universally methylated in normal B cells. High correlation with mRNA and protein expression, as well as activity in promoter reporter assays, revealed that within this differentially methylated region, a single CpG dinucleotide and neighboring nucleotides are particularly important in ZAP-70 transcriptional regulation. Furthermore, by using clustering approaches, we identified a prognostic role for this site in four independent data sets of patients with CLL using time to treatment, progression-free survival, and overall survival as clinical end points.

Conclusion

Comprehensive quantitative DNA methylation analysis of the ZAP-70 gene in CLL identified important regions responsible for transcriptional regulation. In addition, loss of methylation at a specific single CpG dinucleotide in the ZAP-70 5' regulatory sequence is a highly predictive and reproducible biomarker of poor prognosis in this disease. This work demonstrates the feasibility of using quantitative specific ZAP-70 methylation analysis as a relevant clinically applicable prognostic test in CLL.

J Clin Oncol 30:2483-2491. © 2012 by American Society of Clinical Oncology

INTRODUCTION

Chronic lymphocytic leukemia (CLL) has a varied natural history, emphasizing the need for diagnostic tests that accurately predict progression and response. *IGHV* gene mutation status is one of the most valuable prognostic factors for progression-free survival (PFS) and overall survival (OS).¹⁻⁵ Expression of the zeta chain-associated protein ZAP-70, initially presented as a surrogate,^{6,7} is recognized as a biologically important factor in CLL pathology⁸⁻¹¹ and also predicts shorter remissions following chemoimmunotherapy.¹² Despite this,

ZAP-70 measurement in clinical laboratories has been problematic, and establishment of a robust alternative would therefore be valuable.

ZAP-70 regulation has not been extensively explored. Although ZAP-70 is not typically mutated,¹³ epigenetic modifications could explain its varied expression. Recent investigations^{14,15} indicated that methylation of specific regions in the vicinity of the transcriptional start site correlates with decreased ZAP-70 mRNA/protein expression. Although they are provocative, these studies lack quantification or are focused on areas of uncertain importance to ZAP-70 transcription. We hypothesized that broad,

Rainer Claus, Manuela Zucknick, Daniel Mertens, Christopher C. Oakes, and Christoph Plass, German Cancer Research Center, Heidelberg; Stephan Stilgenbauer, Daniel Mertens, Andreas Bühler, and Hartmut Döhner, University of Ulm, Ulm, Germany; David M. Lucas, Amy S. Ruppert, Guido Marcucci, and John C. Byrd, Comprehensive Cancer Center; Lianbo Yu, Center for Biostatistics; Nyla A. Heerema, The Ohio State University, Columbus, OH; Richard A. Larson, University of Chicago, Chicago, IL; Neil E. Kay and Diane F. Jelinek, Mayo Clinic, Rochester, MN; Thomas J. Kipps and Laura Z. Rassenti, Moores University of California-San Diego Cancer Center, University of California, San Diego, CA; and John G. Gribben, Barts and The London School of Medicine, University of London, London, United Kingdom.

Submitted September 1, 2011; accepted January 27, 2012; published online ahead of print at www.jco.org on May 7, 2012.

C.P. and J.C.B. contributed equally to this work.

Terms in blue are defined in the glossary, found at the end of this article and online at www.jco.org.

Authors' disclosures of potential conflicts of interest and author contributions are found at the end of this article.

Corresponding author: John C. Byrd, MD, 455 Wiseman Hall, 410 West 12th Ave, Columbus, OH 43210; e-mail: john.byrd@osumc.edu.

© 2012 by American Society of Clinical Oncology

0732-183X/12/3020-2483/\$20.00

DOI: 10.1200/JCO.2011.39.3090

high-resolution, and quantitative assessment of DNA methylation at CpG dinucleotides in the entire known *ZAP-70* regulatory region would identify the area important for transcriptional regulation and, consequently, for prognostic significance. Herein, we used MassARRAY-based methylation analysis (Sequenom, San Diego, CA) to identify a novel single CpG site with transcriptional regulatory activity outside a CpG island (CGI), the methylation of which is associated with decreased *ZAP-70* expression and favorable clinical outcome.

PATIENTS AND METHODS

Patients and Cells

Blood was obtained from healthy donors and patients with CLL enrolled onto clinical trials (phase II trial of pentostatin, cyclophosphamide, and rituximab in previously untreated, symptomatic CLL, conducted at the Mayo Clinic and Ohio State University¹⁶; phase II trial of fludarabine with concurrent *v* sequential rituximab in previously untreated, symptomatic

Fig 1. Comprehensive quantitative DNA methylation profiling of CpG-rich regions at the *ZAP-70* gene locus. (A) Schematic representation of the *ZAP-70* gene locus. Black boxes represent exons, arrows indicate transcriptional start sites (TSSs), and AUG indicates the TSS of the main transcript. Gold bars represent CpG islands (CGIs). (B) Regions of analysis (Ulm University [ULM] set) covering the three major alternative TSSs—TSS1 to TSS3—are marked A1 to A6. Separate samples are organized in rows (B cells from 69 patients with chronic lymphocytic leukemia [CLL] and nine healthy donors; T cells from five healthy donors). Columns represent single CpG units. High methylation levels are depicted in dark blue, low methylation levels in light green, and missing data in gray. The unsupervised clustering refers to amplicon A2 (exon 1 to intron 2 region). *IGHV*-mutated samples are marked with black boxes in the clustering tree, and unmutated samples with gray. White squares indicate missing data. Clusters are separated by colored bars (red, mostly methylated; black, CpG unit 1 unmethylated and other CpG units methylated; green, CpG unit 1 unmethylated and other CpG units with mixed methylation levels; blue, CpG units mostly unmethylated). (*) Informative CpG unit 1 located at position +223 relative to TSS1. (C) Correlation between DNA methylation levels and *ZAP-70* mRNA expression assessed by TaqMan assay (upper panel) and *ZAP-70* protein expression assessed by flow cytometry (Flow cyt; lower panel). Bar graphs display Spearman's rho (Spearman's rank correlation coefficient) between DNA methylation and expression data; (*) significance level of $P < .05$ (Spearman's rho P value). (D-F) Heat maps and unsupervised clustering displaying methylation levels across amplicon A2 in the Chronic Lymphocytic Leukemia Research Consortium (CRC), Mayo Clinic and Ohio State University (MAYO), and Cancer and Leukemia Group B 9712 trial (CALGB) sample sets. Cluster color coding corresponds to 1B, with the addition of orange (mostly mixed methylation). (*) CpG unit 1.

CLL [CALGB 9712 (Cancer and Leukemia Group B)]^{17,18}) or treated at institutions (Ulm University Hospital; CLL Research Consortium [CRC]¹⁶⁻¹⁸) according to institutional review board–approved protocols (Table 1). Mononuclear cells were separated using Ficoll-Hypaque (Sigma-Aldrich, St Louis, MO), and cells were frozen viably. B and T cells were isolated from healthy donors using MACS (Miltenyi Biotec, Bergisch Gladbach, Germany).

Flow Cytometry

ZAP-70 expression was measured by four-color flow cytometry according to Crespo et al.⁷

IGHV Mutational Status Assessment and Cytogenetic Analyses

IGHV analysis and fluorescent in situ hybridization were performed as described.^{19,20} IGHV sequence homology of less than 98% versus germline was considered mutated.

Real-Time Polymerase Chain Reaction and Immunoblotting

Samples were enriched to at least 96% CD19 positivity (MACS; Miltenyi Biotec). ZAP-70 mRNA²¹ and protein⁶ expression were assessed as described.

Quantitative High-Resolution DNA Methylation Analysis

DNA methylation was quantitatively assessed at single CpG units (consisting of one or more CpG dinucleotides) using the MassCleave assay (Sequenom, San Diego, CA) as described.²² Briefly, 0.5 to 1 μ g of genomic DNA was sodium bisulfite modified. Target regions were polymerase chain reaction amplified using T7 promoter tagged primers and in vitro transcribed, and defined fragments were generated by base-specific RNase A cleavage. Polymerase chain reaction primers were designed not to contain CpG dinucleotides to avoid biases toward methylated or unmethylated alleles. Primer sequences are provided in the Data Supplement. The generated fragments were subsequently analyzed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. The molecular weight of the resulting fragments is indicative of the DNA methylation state as mass shifts of 16 Dalton or multiples thereof represent C→T difference after bisulfite conversion. The signal amplitudes reflect relative fragment amounts of the cleavage products and thereby represent the relative fragment (CpG unit) methylation level. Methylation standards (0%, 20%, 40%, 60%, 80%, and 100% methylated genomic DNA) were used to control for the full dynamic range of measurements.

Promoter characterization methods are outlined in the Data Supplement. Statistical methods as well as definitions of clinical end points are also provided in the Data Supplement.

RESULTS

ZAP-70 5' Region Shows Variable Methylation That Correlates With ZAP-70 Expression

ZAP-70 encodes at least three isoforms, transcribed from different transcription start sites (TSSs; Fig 1A). TSS2 and TSS3 are associated with CGIs whereas the main transcript (TSS1) arises from an intermediate CpG-dense sequence (Data Supplement). Uniform high DNA methylation (> 50%) was observed for amplicons covering the CGIs for TSS2 and TSS3 in the samples from Ulm University (ULM) as well as in normal B- and T-cell controls (Fig 1B). In contrast, TSS1 amplicon A1 showed low DNA methylation in all cell types. TSS1 amplicon A2 (position +89 to +468) exhibited uniformly higher levels of methylation in normal B cells, virtually no methylation in T cells, and pronounced variation in CLL samples. This amplicon harbors nine analyzable CpG units comprising ten CpG dinucleotides. For all of these CpG units, higher methylation levels were significantly associated with lower ZAP-70 mRNA expression according to **Spearman's rho** (**Spearman's rank correlation coefficient**: ≤ -0.58 ; $P < .001$) and protein expression (**Spearman's ρ** $\leq$

-0.42 ; $P < .004$; Fig 1C). This strongly suggested the regulatory relevance of A2 for ZAP-70 expression, particularly since the other amplicons did not exhibit similarly high correlations. We therefore focused on determining the importance of epigenetic silencing of ZAP-70 in the differentially methylated region of A2.

Distinct Distribution of DNA Methylation of Single CpG Units

Next, methylation of TSS1 A2 was measured in independent sample sets (Table 1). The first set (ULM) included 69 patients with CLL, 58 of whom were observed with continued asymptomatic disease (not progressing) or until they became symptomatic and required treatment. The second set included 53 untreated patients with IGHV mutation from the CRC. The third set (MAYO) included 57 patients with previously untreated symptomatic CLL enrolled onto a phase II trial at the Mayo Clinic and Ohio State University.¹⁶ The fourth set included 79 patients with previously untreated symptomatic CLL enrolled onto Cancer and Leukemia Group B (CALGB) 9712.^{17,18} Heat maps displaying methylation levels across A2 are shown in Figures 1B, and 1D to 1F. As expected, the CRC set with only patients with IGHV mutations was largely methylated. In all the patient sets, methylation levels at the first CpG unit of A2 (CpG unit 1), located at position +223, varied considerably relative to the other CpG units. To examine methylation patterns across the data sets, **hierarchical clustering** was used based on differential methylation between the levels of CpG unit 1 and the rest of the region. This process identified four major patient subsets across all sets (Fig 1): the red cluster was mostly methylated across the entire region including CpG unit 1; the black cluster exhibited lower DNA methylation levels at CpG unit 1 but higher DNA methylation levels across all other units; the green cluster showed low methylation at CpG unit 1 and mixed methylation across

Table 1. Characteristics of Patients and Sample Set

Characteristic	ULM* (n = 58)		CRC* (n = 53)		MAYO† (n = 57)		CALGB† (n = 79)	
	No.	%	No.	%	No.	%	No.	%
Median follow-up, years	4.3		2.5		3.0		8.1	
Age, years								
Median	60		56		62		65	
Range	35-80		38-81		38-80		38-86	
Sex								
Male	33	57	28	53	43	75	62	78
Female	25	43	25	47	14	25	17	22
Binet/Rai stage								
A/O, I, II	33	60	33	100	28	49	44	56
B, C/III, IV	22	40	0	0	29	51	35	44
Cytogenetics								
del(17p), del(11q)	21	36	7	20	16	28	14	21
All others	37	64	28	80	41	72	52	79
IGHV								
Mutated	25	44	53	100	13	24	25	39
Unmutated	32	56	0	0	42	76	39	61

Abbreviations: CALGB, Cancer and Leukemia Group B 9712 trial; CRC, Chronic Lymphocytic Leukemia Research Consortium; MAYO, Mayo Clinic and Ohio State University; ULM, Ulm University.

*Patients enrolled at diagnosis.

†Patients enrolled at start of treatment.

the other units; and the blue cluster was mostly unmethylated, including unit 1. In the CRC set, an additional orange cluster consisting of mostly mixed methylation was defined.

ZAP-70 expression in each of the clusters was examined in the ULM set. Samples in the mostly methylated red cluster showed significantly lower expression of ZAP-70 mRNA compared with samples in the black and green clusters taken together as evaluated by the Wilcoxon rank sum test (also called the Mann-Whitney *U* test; $P < .001$)

or the unmethylated blue cluster ($P < .001$). Samples in the red cluster also had significantly lower ZAP-70 protein expression than those in the black and green clusters ($P = .05$ and $P = .02$) or in the blue cluster ($P < .001$ and $P = .003$; immunoblot data not shown) as measured by flow cytometry and immunoblot, respectively. These clusters were also associated with clinical outcome (time to treatment [TT] for ULM and CRC; PFS for MAYO and CALGB; OS for ULM and CALGB; Figs 2A to 2D and Data Supplement). Generally, patients in

Fig 2. Kaplan-Meier plots of DNA methylation-based clusters of ZAP-70 amplicon 2 show differential clinical outcome. (A) Treatment-free survival and (B) overall survival probabilities for the ULM University set; (C) treatment-free survival probabilities for the Chronic Lymphocytic Leukemia Research Consortium set; (D) progression-free survival probabilities for the Mayo Clinic and Ohio State University set; (E) progression-free survival and (F) overall survival probabilities for the Cancer and Leukemia Group B 9712 trial (CALGB) set.

the red cluster had the most prolonged times until progression, whereas patients in the other clusters tended to have inferior outcomes. This was most clearly observed in the ULM and CALGB sets and was less pronounced in the CRC and MAYO sets in which the sample sizes were smaller and follow-up was shorter.

The identifying characteristic of samples in the consistently methylated red cluster versus the other clusters was the high level of methylation of CpG unit 1. To examine the functional relevance of unit 1 methylation specifically, methylation levels were correlated with ZAP-70 mRNA and protein expression by using data from the ULM set. For unit 1 methylation levels less than approximately 15%, a large amount of variation in mRNA and protein expression was observed. However, for higher levels of methylation, little mRNA and protein was expressed (Data Supplement). To further confirm the function of CpG unit 1, we generated a reporter construct retaining ZAP-70 core promoter activity and containing CpG unit 1. Although the unmodified construct showed robust reporter activity, the signal was virtually absent on in vitro methylation and diminished by point mutations targeting CpG unit 1 or its direct vicinity (Data Supplement), supporting the role of methylation changes at this site in the control of ZAP-70 expression.

Methylation of CpG Unit 1 and Correlation With IGHV Status and Clinical Outcome

Except for the CRC data set in which all of the patients had *IGHV*-mutated disease, distributions of methylation values were skewed (Data Supplement). In each of these sets, methylation of CpG unit 1 was strongly associated with *IGHV*-mutated disease ($P \leq .001$; Table 2). The presence of cytogenetic abnormalities del(17p) or del(11q) was significantly associated with lower methylation levels only in the ULM set ($P = .001$), although across data sets, the majority of patients with del(17p)/del(11q) showed methylation levels below 15% (data not shown). With respect to clinical outcome, methylation of CpG unit 1 as a continuous variable was significantly associated with TT (ULM: $P < .001$; univariable proportional hazards model),

PFS (MAYO: $P = .02$; CALGB: $P < .001$), and OS (ULM: $P < .001$; MAYO: $P = .03$; CALGB: $P < .001$). To display the relation between CpG unit 1 methylation and clinical outcome, a cutoff point for CpG unit 1 methylation that best predicted clinical outcome on average across the data sets was used to identify two groups of patients: a low methylation group with CpG unit 1 methylation values less than 15%, and a high methylation group with CpG unit 1 methylation values of at least 15% (Figs 3A to 3D; Data Supplement). For the ULM set, estimated treatment-free proportions at 5 years for patients in the low and high methylation groups were 28% (95% CI, 14% to 43%) and 83% (95% CI, 56% to 95%), respectively. For the CALGB set, the estimated 5-year PFS rates in the low and high methylation groups were 11% (95% CI, 4% to 21%) and 66% (95% CI, 41% to 82%), respectively. For the MAYO set with short follow-up, the estimated 2-year PFS rates in the low and high methylation groups were 45% (95% CI, 26% to 63%) and 77% (95% CI, 55% to 89%) respectively. In addition, within the subgroup of patients with *IGHV* mutation in the CRC set, higher methylation values were moderately associated with longer TT ($P = .14$; Fig 3B). In the CRC set, the estimated 2-year treatment-free proportions for low and high methylation groups were 43% (95% CI, 10% to 73%) and 75% (95% CI, 57% to 86%), respectively. Although the number of patients with *IGHV* mutation with lower methylation values is small, this observation suggests that methylation of CpG unit 1 may also provide prognostic information at diagnosis in this group which has a generally favorable outcome. With respect to OS, a marked separation in curves between patients with low and high methylation values was observed in all data sets except CRC, in which only five deaths had occurred. Survival curves for the ULM and CALGB sets are shown in Figures 3A and 3D; curves for the CRC and MAYO sets are shown in the Data Supplement.

Multivariable Analysis of CpG Unit 1 Methylation and Its Predictive Value

Multivariable analyses were performed to determine whether CpG unit 1 methylation as a continuous variable was a significant independent prognosticator for TT and PFS. Models were fit for each data set except for the CRC set, in which several of the required clinical variables were unavailable. In the three remaining data sets, increased CpG unit 1 methylation was protective for TT (ULM: $P = .01$) and PFS (MAYO: $P = .06$; CALGB: $P < .001$) when controlling for Rai/Binet stage, age, and cytogenetics (Table 3).

Because of the strong association between CpG unit 1 methylation and mutated *IGHV* status, multivariable models were fit including *IGHV* status instead of CpG unit 1 methylation and were compared with the previous models containing CpG unit 1 methylation. The discriminatory abilities of the models in predicting outcome were ranked by calculating the concordance index (c-index). The multivariable models fit with *IGHV* status (Table 3) showed that mutated *IGHV* was also significantly protective for TT in the ULM set ($P < .001$) and PFS in the CALGB set ($P = .01$), as was the case with CpG unit 1 methylation. Mutated *IGHV* status was moderately but not significantly protective for PFS in the MAYO set ($P = .10$). No significant differences were found in the predictive ability of models including CpG unit 1 methylation versus *IGHV* for the ULM set ($P = .996$; c-index: 0.74 v 0.75) nor for the MAYO set ($P = .60$; c-index: 0.63 v 0.59). Importantly, in the CALGB data set used to predict PFS, the model containing CpG unit 1 was significantly better in predictive ability than the model containing *IGHV* mutation status ($P = .04$; c-index:

Table 2. Association of CpG Unit 1 Methylation and *IGHV* Mutational Status

Data Sets	<i>IGHV</i> Mutated	<i>IGHV</i> Unmutated	<i>P</i>
ULM	25	32	< .001
CpG methylation, %			
Median	39	4	
Range	1-90	0-32	
CRC	53	N/A	N/A
CpG methylation, %			
Median	39		
Range	8-75		
MAYO	13	42	< .001
CpG methylation, %			
Median	33	13	
Range	13-70	2-49	
CALGB	25	39	< .001
CpG methylation, %			
Median	15	6	
Range	1-86	0-65	

Abbreviations: CALGB, Cancer and Leukemia Group B 9712 trial; CRC, Chronic Lymphocytic Leukemia Research Consortium; MAYO, Mayo Clinic and Ohio State University; N/A, not applicable; ULM, Ulm University.

Fig 3. Levels of DNA methylation at CpG unit 1 (+223) separate patient subgroups of different clinical outcome. DNA methylation measurements of CpG unit 1 were dichotomized into high and low methylation groups at a calculated threshold of 15%. Survival probabilities for each group were calculated and displayed using Kaplan-Meier plots. (A) Treatment-free and (B) overall survival probabilities in the Ulm University set; (C) treatment-free survival probability in the Chronic Lymphocytic Leukemia Research Consortium set; (D) progression-free survival probability in the Mayo Clinic and Ohio State University set; (E) progression-free and (F) overall survival probability in the Cancer and Leukemia Group B 9712 trial set.

0.71 v 0.63). These multivariable models were then constructed for OS by using the CALGB set, which has the largest number of patients, the most events, and the longest follow-up. As in the TT and PFS models, both methylation of CpG unit 1 and *IGHV* mutation status were significantly associated with OS ($P = .001$ and $P = .04$, respectively), but the predictive ability of the model containing CpG unit 1 methylation was significantly better than the model containing *IGHV* mutation status ($P = .02$; c-index: 0.73 v 0.66; Table 3).

Further supporting that CpG unit 1 provides superior prediction of clinical outcome versus *IGHV* mutation status, Kaplan-Meier plots revealed a separation in curves based on methylation group in the *IGHV* mutation status subgroups with an adequate number of patients (ie, the unmutated subgroup for the MAYO set and the mutated subgroups for the ULM and CALGB sets), with the CALGB set showing the most dramatic separation in curves (Data Supplement). For this patient set, all patients with mutated disease and methylation

Table 3. Multivariable Models*

Data Set	No. of Samples	End Point	Variable	HR†	95% CI	P	C-Index‡
ULM	54	TT	CpG unit 1, 25% increase	0.60	0.39 to 0.90	.01	0.74
ULM	54	TT	<i>IGHV</i> mutated v unmutated	0.12	0.04 to 0.36	< .001	0.75
ULM	54	TT	C-334 site, 25% increase	0.69	0.52 to 0.92	.01	0.75
MAYO	55	PFS	CpG unit 1, 25% increase	0.50	0.24 to 1.04	.06	0.63
MAYO	55	PFS	<i>IGHV</i> mutated v unmutated	0.44	0.16 to 1.17	.10	0.59
CALGB	64	PFS	CpG unit 1, 25% increase	0.41	0.27 to 0.63	< .001	0.71
CALGB	64	PFS	<i>IGHV</i> mutated v unmutated	0.46	0.25 to 0.85	.01	0.63
CALGB	64	PFS	C-334 site, 25% increase	0.67	0.55 to 0.81	< .001	0.67
CALGB	64	OS	CpG unit 1, 25% increase	0.43	0.26 to 0.71	.001	0.73
CALGB	64	OS	<i>IGHV</i> mutated v unmutated	0.52	0.28 to 0.97	.04	0.66
CALGB	64	OS	C-334 site, 25% increase	0.78	0.63 to 0.95	.02	0.65

Abbreviations: CALGB, Cancer and Leukemia Group B 9712 trial; HR, hazard ratio; MAYO, Mayo Clinic and Ohio State University; OS, overall survival; PFS, progression-free survival; TT, time to treatment; ULM, Ulm University.

*All models controlled for stage of disease, age, and poor-risk cytogenetics.

†HR < 1 indicates that increased methylation of CpG unit 1, *IGHV*-mutated disease, or methylation at site C-334¹⁴ is associated with decreased risk of an event. For example, in the ULM data set, an HR of 0.60 for CpG unit 1 methylation indicates for every 25% increase in methylation, there is a 40% reduction in the risk of beginning treatment.

‡The c-index describes the predictive accuracy of a model in which values of 0.5, 0.7, and 1.0 indicate that there is completely random, acceptable, or perfect discrimination, respectively, of the model in predicting outcome.

values less than 15% ($n = 12$) progressed within 6 years, whereas those with mutated disease and methylation values of at least 15% ($n = 13$) had an estimated median time to progression of 8.6 years, with six patients remaining progression-free beyond 9 years. The predictive ability of CpG unit 1 (+223) was comparable to that of the CpG dinucleotide C-334 investigated by Corcoran et al¹⁴ in the ULM set relative to TT and in the CALGB set relative to PFS; it was significantly better at discriminating OS in the CALGB set ($P = .02$; c-index: 0.73 v 0.65; Table 3). Together, these data demonstrate that methylation of CpG unit 1 has direct relevance to ZAP-70 expression and significant prognostic value in independent patient subsets.

DISCUSSION

We report the application of quantitative high-resolution DNA methylation profiling to identify a functionally and prognostically relevant methylation site within the 5' regulatory region of ZAP-70 in samples from patients with CLL. This work is based on prior studies demonstrating the importance of ZAP-70 expression for predicting outcome in CLL. The studies presented here identify DNA methylation to be widespread over the entire ZAP-70 promoter, including regions not previously described. One small area in the 5' region of the promoter for the ZAP-70 TSS1 was found to be variably methylated in CLL samples, and this methylation inversely correlated with ZAP-70 expression. Furthermore, this area is methylated in normal B cells that lack ZAP-70 expression and unmethylated in normal T cells that express ZAP-70.

Regional heterogeneity of DNA methylation among CLL samples subsequently allowed the generation of distinct patient clusters. The red cluster was characterized by methylation of a single CpG unit, along with significantly decreased ZAP-70 mRNA and protein. This cluster included a diverse group of newly diagnosed patients with CLL and showed a favorable outcome with regard to TT, although it also included many patients with *IGHV*-mutated disease for whom a favorable outcome was anticipated. This same methylation pattern was then observed in two independent cohorts of symptomatic, previously untreated patients with CLL treated with chemoimmunotherapy.^{16,17}

Patients in the red cluster again showed extended PFS in both cohorts, but more importantly, these positive observations extended to OS as well. Given the apparent impact of methylation in this single CpG unit, we hypothesized that this region was important for ZAP-70 transcriptional regulation. We subsequently confirmed promoter activity of this region and abrogation of this activity by methylation. In separate data sets, methylation of this single CpG unit as a continuous variable was highly correlated with TT, PFS, and OS in multivariable models. Because of a highly selected patient group for the CRC set (all *IGHV* mutated) and a rather short follow-up with few events for patients in both the CRC and the MAYO clinical trials, OS differences were subtle for these cohorts. Further, we explored an optimal cutoff value for CpG unit 1 methylation and found that within each of the four data sets, the optimal value that discriminated clinical outcome the best varied between approximately 10% and 20%. Averaging across the data sets, the optimal cutoff value in our data separated patients into two groups: those with methylation values below 15% and those with methylation values of at least 15%. As with other biomarkers, an optimal cutoff value that could be used in the clinic will have to be refined and verified in larger data sets. Because high-resolution, quantitative techniques for DNA methylation assessment (eg, MassARRAY or pyrosequencing²³) have become state-of-the-art and are more frequently available, even in routine clinical settings, measurement and application of such DNA methylation cutoff values as prognostic biomarkers have become clinically feasible.

Because of the large degree of overlap between CpG unit 1 methylation and *IGHV* status, the relative independence of these two variables in predicting clinical outcome could not be assessed. However, separate models in which each of the variables were fitted and compared indicate that the predictive ability of CpG unit 1 is at least as good as that of *IGHV* status in these data sets. Prospective studies with increased numbers of patients are required to provide conclusive information regarding prognostic independence of these two variables.

Corcoran et al¹⁴ were the first to identify a ZAP-70 methylation event in the intron 1/exon 2 region that was associated with both ZAP-70 expression and *IGHV* status. In this study, methylation at the

C-334 site was assessed by using combined bisulfite restriction analysis. Although this strategy is useful for identifying methylation, it is highly dependent on restriction enzyme activity at a single site, and accurate quantification is difficult. Our findings confirm the work of Corcoran et al showing (within the context of amplicon 2 of ZAP-70) that the C-334 site is variably methylated and that methylation of this site is associated with decreased ZAP-70 mRNA and protein expression (data not shown). However, we significantly extend this work by examining the entire promoter region, including areas proximal and distal to the C-334 site. With integration of both ZAP-70 expression and clinical data, we identified an alternative methylation site with direct transcriptional regulatory activity and prognostic significance.

It is not unexpected that variability is observed in ZAP-70 expression among samples with lower methylation. It is likely that other factors (eg, miRNA or methylation status at neighboring sites) have an impact on ZAP-70 mRNA and protein expression. Furthermore, CpG unit 1 methylation might not directly reflect ZAP-70 expression at the moment of cell harvest but might rather display a strongly conserved and stable epigenetic mark as surrogate for ZAP-70 expression. Regardless, by using this single methylation site, we show superior prognostic value for CpG unit 1 in multivariable analyses.

Similar methylation profiling for biomarker discovery and promoter interrogation is warranted in other cancers, since the epigenetic regulation of selected genes by methylation is essential to pathogenesis and progression of most cancers studied to date. The MassARRAY technique offers accurate, quantifiable interrogation of large contiguous regions of genes, thereby allowing more detailed association of methylation with gene expression and clinical outcome. These results can be exploited to identify transcription factors and pathways controlling gene expression. Indeed, studies are underway to understand the mechanism by which alteration of methylation in CpG unit 1 has an impact on ZAP-70 expression.

In conclusion, we used quantitative high-resolution methylation profiling of the ZAP-70 promoter to identify a previously unrecognized single CpG unit important for ZAP-70 regulation in CLL. Methylation of this CpG unit in patients with CLL serves as a highly favorable biomarker in patients at diagnosis and at time of first treat-

ment. In this context, narrowing to a single CpG unit dramatically simplifies analysis and allows for the potential of this test to be broadly applied across laboratories.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

Although all authors completed the disclosure declaration, the following author(s) and/or an author's immediate family member(s) indicated a financial or other interest that is relevant to the subject matter under consideration in this article. Certain relationships marked with a "U" are those for which no compensation was received; those relationships marked with a "C" were compensated. For a detailed description of the disclosure categories, or for more information about ASCO's conflict of interest policy, please refer to the Author Disclosure Declaration and the Disclosures of Potential Conflicts of Interest section in Information for Contributors.

Employment or Leadership Position: None **Consultant or Advisory Role:** None **Stock Ownership:** None **Honoraria:** John G. Gribben, Roche **Research Funding:** None **Expert Testimony:** None **Other Remuneration:** None

AUTHOR CONTRIBUTIONS

Conception and design: Rainer Claus, David M. Lucas, Christoph Plass, John C. Byrd

Financial support: Christoph Plass, John C. Byrd

Provision of study materials or patients: Stephan Stilgenbauer, Richard A. Larson, Neil E. Kay, Diane F. Jelinek, Thomas J. Kipps, Laura Z. Rassenti, John G. Gribben, Hartmut Döhner, Christoph Plass, John C. Byrd

Collection and assembly of data: Rainer Claus, David M. Lucas, Amy S. Ruppert, Lianbo Yu, Daniel Mertens, Andreas Bühler, Christopher C. Oakes, Richard A. Larson, Neil E. Kay, Diane F. Jelinek, Thomas J. Kipps, Laura Z. Rassenti, John G. Gribben, Hartmut Döhner, Nyla A. Heerema, Guido Marcucci, Christoph Plass, John C. Byrd

Data analysis and interpretation: Rainer Claus, David M. Lucas, Stephan Stilgenbauer, Amy S. Ruppert, Lianbo Yu, Manuela Zucknick, Christoph Plass, John C. Byrd

Manuscript writing: All authors

Final approval of manuscript: All authors

REFERENCES

- Hamblin TJ, Davis Z, Gardiner A, et al: Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. *Blood* 94:1848-1854, 1999
- Fais F, Ghiotto F, Hashimoto S, et al: Chronic lymphocytic leukemia B cells express restricted sets of mutated and unmutated antigen receptors. *J Clin Invest* 102:1515-1525, 1998
- Damle RN, Wasil T, Fais F, et al: Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. *Blood* 94:1840-1847, 1999
- Rassenti LZ, Jain S, Keating MJ, et al: Relative value of ZAP-70, CD38, and immunoglobulin mutation status in predicting aggressive disease in chronic lymphocytic leukemia. *Blood* 112:1923-1930, 2008
- Zenz T, Mertens D, Küppers R, et al: From pathogenesis to treatment of chronic lymphocytic leukaemia. *Nat Rev Cancer* 10:37-50, 2010
- Rosenwald A, Alizadeh AA, Widhopf G, et al: Relation of gene expression phenotype to immunoglobulin mutation genotype in B cell chronic lymphocytic leukemia. *J Exp Med* 194:1639-1647, 2001
- Crespo M, Bosch F, Villamor N, et al: ZAP-70 expression as a surrogate for immunoglobulin-variable-region mutations in chronic lymphocytic leukemia. *N Engl J Med* 348:1764-1775, 2003
- Sivina M, Hartmann E, Kipps TJ, et al: CCL3 (MIP-1 α) plasma levels and the risk for disease progression in chronic lymphocytic leukemia. *Blood* 117:1662-1669, 2011
- Messmer D, Fecteau JF, O'Hayre M, et al: Chronic lymphocytic leukemia cells receive RAF-dependent survival signals in response to CXCL12 that are sensitive to inhibition by sorafenib. *Blood* 117:882-889, 2011
- Chen L, Widhopf G, Huynh L, et al: Expression of ZAP-70 is associated with increased B-cell receptor signaling in chronic lymphocytic leukemia. *Blood* 100:4609-4614, 2002
- Chen L, Apgar J, Huynh L, et al: ZAP-70 directly enhances IgM signaling in chronic lymphocytic leukemia. *Blood* 105:2036-2041, 2005
- Del Poeta G, Del Principe MI, Consalvo MA, et al: The addition of rituximab to fludarabine improves clinical outcome in untreated patients with ZAP-70 negative chronic lymphocytic leukemia. *Cancer* 104:2743-2752, 2005
- Brown JR, Levine RL, Thompson C, et al: Systematic genomic screen for tyrosine kinase mutations in CLL. *Leukemia* 22:1966-1969, 2008
- Corcoran M, Parker A, Orchard J, et al: ZAP-70 methylation status is associated with ZAP-70 expression status in chronic lymphocytic leukemia. *Haematologica* 90:1078-1088, 2005
- Chantepie SP, Vaur D, Grunau C, et al: ZAP-70 intron1 DNA methylation status: Determination by pyrosequencing in B chronic lymphocytic leukemia. *Leuk Res* 34:800-808, 2010
- Kay NE, Geyer SM, Call TG, et al: Combination chemioimmunotherapy with pentostatin, cyclophosphamide, and rituximab shows significant clinical activity with low accompanying toxicity in previously untreated B chronic lymphocytic leukemia. *Blood* 109:405-411, 2007
- Byrd JC, Peterson BL, Morrison VA, et al: Randomized phase 2 study of fludarabine with concurrent versus sequential treatment with rituximab in symptomatic, untreated patients with B-cell chronic lymphocytic leukemia: Results from

Cancer and Leukemia Group B 9712 (CALGB 9712). *Blood* 101:6-14, 2003

18. Woyach JA, Ruppert AS, Heerema NA, et al: Chemoimmunotherapy with fludarabine and rituximab produces extended overall survival and progression-free survival in chronic lymphocytic leukemia: Long-term follow-up of CALGB study 9712. *J Clin Oncol* 29:1349-1355, 2011

19. Rassenti LZ, Huynh L, Toy TL, et al: ZAP-70 compared with immunoglobulin heavy-chain gene

mutation status as a predictor of disease progression in chronic lymphocytic leukemia. *N Engl J Med* 351:893-901, 2004

20. Byrd JC, Gribben JG, Peterson BL, et al: Select high-risk genetic features predict earlier progression following chemoimmunotherapy with fludarabine and rituximab in chronic lymphocytic leukemia: Justification for risk-adapted therapy. *J Clin Oncol* 24:437-443, 2006

21. Kienle D, Benner A, Kröber A, et al: Distinct gene expression patterns in chronic lymphocytic

leukemia defined by usage of specific VH genes. *Blood* 107:2090-2093, 2006

22. Ehrich M, Nelson MR, Stanssens P, et al: Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. *Proc Natl Acad Sci U S A* 102:15785-15790, 2005

23. Tost J, Dunker J, Gut IG: Analysis and quantification of multiple methylation variable positions in CpG islands by Pyrosequencing. *Biotechniques* 35: 152-156, 2003

GLOSSARY TERMS

Concordance index (c-index): The concordance index is the probability that given two randomly selected patients, the patient with the worst outcome is, in fact, predicted to have the worst outcome. The measure is similar to an area under the receiver operating characteristics curve and ranges from 0.5 (chance or coinflip) to 1.0 (perfect ability to rank patients).

CpG dinucleotides: DNA sequences composed of a cytidine connected to a guanosine through a phosphate residue. CpG dinucleotides are under-represented in mammalian genomes and tend to form clusters in CpG-rich regions (also referred to as CpG islands when they fulfill certain criteria). Cytidines in the context of CpG dinucleotides can be methylated to 5-methylcytidine by DNA methyltransferases and are often associated with transcriptional repression when located in regions that are relevant to regulatory functions.

CpG island: DNA sequences with a high density of CpGs are termed CpG islands. CpG islands are typically unmethylated in normal tissues but often become methylated in tumors. The patterns of hypermethylated CpG islands vary according to the histologic origin of the tumor.

DNA methylation: Methylation of bases contained in the DNA double helix, resulting in a loss of gene function. Generally occurring on cytosine residues in the DNA, methylation is important in regulating cell growth and differentiation and has resulted in the testing of DNA methyltransferase inhibitors as anti-cancer agents and differentiation agents.

Hierarchical clustering: An analytical tool used to find the closest associations among gene profiles and specimens under evaluation.

***IGHV* mutation status:** *IGHV* stands for the immunoglobulin heavy chain variable region genes at the B-cell receptor locus. This gene locus is the target of a high rate of somatic mutations (termed somatic hypermutation) during proliferation of B cells upon antigen stimulation and recognition. In chronic lymphocytic leukemia (CLL), *IGHV* mutation status is clinically relevant because it is highly correlated with disease prognosis and outcome. Patients with CLL with mutated *IGHV* locus (typically exhibiting < 98% sequence homology to germline) generally have a more favorable clinical outcome.

Pyrosequencing: A direct DNA sequencing technique based on “sequencing by synthesis.” Nucleotide signals are detected by release of pyrophosphate upon nucleotide incorporation. Pyrosequencing is frequently used for (targeted) genomic re-sequencing but also for quantitative detection of DNA methylation by sequencing of bisulfate-converted DNA.

Spearman’s rho (Spearman’s rank correlation coefficient): A nonparametric measure of statistical dependence between two ordinal or continuous variables assessing how well their relationship can be described by a monotone function. It is defined as the Pearson’s correlation coefficient between the ranks of the variables. Values lie between -1 and $+1$, where a value of $+1$ or -1 implies a perfect monotone relationship.

Wilcoxon rank sum test (also called Mann-Whitney U test): A commonly applied nonparametric statistical hypothesis test for comparing two groups. The test assesses whether there is a shift between the distributions of both groups (ie, whether one of two samples of independent observations tends to have larger values than the other). Assumptions are that all observations are independent, that the variances of both groups are equal, and that observations are on an ordinal or continuous scale.