



# Immunoglobulin diversity gene usage predicts unfavorable outcome in a subset of chronic lymphocytic leukemia patients

Renee C. Tschumper,<sup>1</sup> Susan M. Geyer,<sup>2</sup> Megan E. Campbell,<sup>2</sup> Neil E. Kay,<sup>3</sup> Tait D. Shanafelt,<sup>3</sup> Clive S. Zent,<sup>3</sup> Grzegorz S. Nowakowski,<sup>3</sup> Timothy G. Call,<sup>3</sup> Gordon W. Dewald,<sup>4</sup> and Diane F. Jelinek<sup>1</sup>

<sup>1</sup>Department of Immunology, <sup>2</sup>Department of Health Sciences Research, <sup>3</sup>Department of Internal Medicine, and <sup>4</sup>Department of Laboratory Medicine and Pathology, Mayo Clinic College of Medicine, Mayo Clinic, Rochester, Minnesota, USA.

**Survival of patients with B cell chronic lymphocytic leukemia (B-CLL) can be predicted by analysis of mutations in the immunoglobulin heavy chain variable gene (*IGHV*). Patients without mutations (unmutated [UM]) are at greater risk for disease progression and death than patients with mutations (M). Despite this broad prognostic difference, there remains wide intragroup variation in the clinical outcome of UM patients, especially those with low/intermediate Rai risk disease. We evaluated UM B-CLL patients with low/intermediate Rai risk to determine the relationship between *IGHV*, *IGH* diversity (*IGHD*), and *IGH* joining (*IGHJ*) gene usage and time to treatment (TTT). Irrespective of *IGHV* usage, UM patients whose B-CLL cells expressed the *IGHD3-3* gene had a significantly shorter TTT than other UM B-CLL patients, and specifically, use of the *IGHD3-3* gene in reading frame 2 (RF2) predicted shorter TTT. As expected, Rai risk was the best single prognostic factor for TTT; however, *IGHD* usage was also a significant variable for TTT. Therefore, both *IGHD* gene and *IGHD* RF usage have prognostic relevance in UM B-CLL patients with low/intermediate Rai risk disease. In addition, these data support the concept that antigen-driven selection of specific *Ig* receptors plays a role in the clinical course of B-CLL.**

## Introduction

During B cell development, the mechanism of *Ig* gene rearrangement yields a vast repertoire of antigen receptor-binding specificities. This process involves 2 stages of rearrangement, with the first being the assembly of the variable (*V*), diversity (*D*), and joining (*J*) gene segments of the *Ig* heavy chain (*IGH*) and the *V* and *J* gene segments of the light chain in the developing B cell. The second rearrangement occurs when antigen triggers mature B cells to undergo class switch recombination. One mechanism of *Ig* sequence diversity results from the initial rearrangement of multiple *V(D)J* genes during B cell development. If each of these genes were randomly used, this mechanism alone would result in as many as 1.6 million potential unique combinations. Additional diversity occurs during B cell development when nucleotides are deleted or inserted (nontemplated nucleotides [N] and palindromic [P] nucleotides) in the *V(D)J* junctional regions, resulting in a unique B cell receptor (BCR) (1–5). The BCR can then acquire further diversity if the heavy and light chain variable regions undergo somatic hypermutation following mature B cell stimulation with antigen. Collectively, these mechanisms allow for the generation of more than  $1 \times 10^9$  unique B cell clones. Within the BCR, the complementarity-determining regions (CDRs) acquire the most

somatic mutations. Each heavy and light chain has 3 CDRs. Of the 6 CDRs in each antigen recognition site of an antibody molecule, the heavy chain CDR3 (HCDR3) plays the major role in determining antigenic specificity (6, 7). The amino acid sequence of HCDR3 is determined by the specific *IGH* diversity (*IGHD*) and *IGH* joining (*IGHJ*) gene usage, the *IGHD* gene reading frame (RF), the addition of N and P nucleotides, and the somatic mutations that may be introduced following antigenic stimulation.

Multiple genetic sequencing studies of the *Ig* gene and correlative clinical analyses have shown that the percent mutation of the clonal B cell *Ig* heavy chain variable (*IGHV*) sequence in patients with B cell chronic lymphocytic leukemia (B-CLL) is important in determining the prognosis of these patients. In particular, patients with 2% or greater mutation (mutated [M]) have a better prognosis than patients with less than 2% mutation (unmutated [UM]) (8–12). Although *IGHV* mutation status is a significant prognostic factor for B-CLL patients, there is still considerable heterogeneity within the UM and M groups, and further investigation could identify factors of additional prognostic value that will help cull out clinical subgroups within this overall cohort. Recent evidence suggests that in some cases, specific *IGHV* gene usage also predicts disease outcome irrespective of mutation status. For example, patients expressing the *IGHV3-21* gene have a poor prognosis regardless of mutation status (13–16). In contrast, patients expressing the *IGHV3-72* gene have a favorable immunophenotypic profile (CD38- and ZAP-70-negative) and consistently exhibit a stable and indolent disease course (17, 18).

More extensive analysis of the molecular structure of the BCR in leukemic B cells has also suggested that subgroups of B-CLL patients express strikingly similar BCRs with respect to *IGHD* and *IGHJ* genes and, thus, similar HCDR3s (14–17, 19–24). For exam-

**Nonstandard abbreviations used:** B-CLL, B cell chronic lymphocytic leukemia; BCR, B cell receptor; CDR, complementarity-determining region; CLL, chronic lymphocytic leukemia; *D*, diversity; HCDR3, heavy chain CDR3; *IGH*, *Ig* heavy chain; *IGHD*, *Ig* heavy chain diversity gene; *IGHJ*, *Ig* heavy chain joining gene; *IGHV*, *Ig* heavy chain variable gene; *J*, joining; M, mutated; RF, reading frame; TTT, time to initial treatment; UM, unmutated; *V*, variable; ZAP-70, zeta chain-associated protein kinase 70 kDa.

**Conflict of interest:** The authors have declared that no conflict of interest exists.

**Citation for this article:** *J. Clin. Invest.* 118:306–315 (2008). doi:10.1172/JCI32625.



**Table 1**  
Molecular and clinical features of the patient cohort<sup>A</sup>

Characteristic	All UM (n = 258) No. of patients
<b>Sex</b>	
Male	184 (71%)
Female	74 (29%)
<b>Rai stage</b>	
0	119 (46%)
I	93 (36%)
II	46 (18%)
<b>IGHJ gene usage</b>	
IGHJ1	1 (0.4%)
IGHJ2	4 (1.6%)
IGHJ3	27 (11%)
IGHJ4	82 (32%)
IGHJ5	23 (9%)
IGHJ6	116 (46%)
Missing <sup>B</sup>	5
<b>IGHJ6 vs. non-IGHJ6</b>	
IGHJ6	116 (46%)
Non-IGHJ6	137 (54%)
<b>IGHD gene usage</b>	
IGHD1	11 (4%)
IGHD2	50 (20%)
IGHD3	143 (57%)
IGHD4	5 (2%)
IGHD5	10 (4%)
IGHD6	32 (12.6%)
IGHD7	1 (0.4%)
Missing <sup>B</sup>	6
<b>IGHD3-3 vs. non-IGHD3-3</b>	
IGHD3-3	72 (29%)
Non-IGHD3-3	180 (71%)

<sup>A</sup>Data are presented as numbers of patients, and values in parentheses represent percentage of patients in each group. <sup>B</sup>Sufficient sequence information was not available to determine these features.

ple, the *IGHV1-69* gene appears to be the most frequently expressed *IGHV* gene in B-CLL, usually in an UM form, and has been reported to be preferentially used with the *IGHD3-3* gene, *IGHJ6* gene, and the *IGHV1-69* 51p1-like allele (25–27). When the *IGHV3-21* gene is used in chronic lymphocytic leukemia (CLL), the leukemic cell BCR tends to exhibit a highly homologous, relatively short HCDR3 and a restricted association with the lambda light chain  $\lambda$ 2-14 (13, 14). There is growing evidence that suggests CLL B cells exhibit a more restrictive BCR when compared with the normal B cell repertoire. For example, Messmer et al. (23) described distinct sets of antigen receptors in B-CLL based not only on common *IGHV* genes, but common *Ig* light chain genes and shared structural similarities in their respective CDR3s. Recent work by Stamatopoulos et al. (24) identified 48 different subsets of sequences with closely homologous or “stereotyped” HCDR3s, including 10 that had been previously described (14, 19, 21–23, 27–30). On the basis of these results, these investigators concluded that over 21% of all CLL patients analyzed can be assigned to one of these stereotyped HCDR3 groups. There is also evidence of BCR homology in those CLL patients who

have undergone class switch to IgG. Ghiotto et al. (21) found that 5 cases of IgG B-CLL expressed *IGHV4-39* with similar HCDR3s and identical light chain CDR3s. The preferential use of certain *IGHD* and *IGHJ* genes and subsequent homologous HCDR3s lends credence to the notion that antigen exposure may play a role in the development of B-CLL, at least in a subset of patients.

Collectively, these studies suggest that detailed analysis of the molecular features of the *Ig* gene will continue to yield important insights into the biologic origin and clinical course of B-CLL. Despite these advances, the clinical significance of specific BCR characteristics has been much less extensively studied. There is emerging evidence, however, that the HCDR3 may predict overall survival in B-CLL patients. Stamatopoulos et al. (24) demonstrated that HCDR3-defined subsets within B-CLL patients using the *IGHV1-69*, *IGHV3-21*, and *IGHV4-34* genes differed in overall survival times, with patients exhibiting the stereotyped HCDR3s displaying shorter median survival times versus those that did not. Maurer et al. (31) demonstrated a trend toward shorter time to initial treatment (TTT) in UM *IGHV1-69* patients who used *IGHD3-3* and *IGHJ6* genes versus UM *IGHV1-69* patients who did not. However, in those studies, the patient numbers in the various subsets were limited and included in some cases both UM and M B-CLL patients at various stages of disease.

Of relevance to the discussion above, 70%–80% of B-CLL patients are now diagnosed with Rai stage 0 or I disease, and 40%–50% of such patients are predicted to have a shorter overall survival based on having UM *IGHV* gene sequences (32). Despite the clear importance of *IGHV* mutation status in patient stratification, there remains a wide variation in the clinical prognosis and survival of UM patients (32). In particular, it is unknown how this variation in UM early-stage CLL may relate to the BCR and individual *IGHD* and *IGHJ* gene usage. In the current study, we restricted our analysis to a large cohort of UM B-CLL patients with low/intermediate Rai risk disease, defined as having a Rai stage of 0, I, or II at diagnosis. We evaluated the relationship between specific *IGHD* and *IGHJ* gene usage and clinically relevant factors in this cohort. Our cohort is unique in that these low/intermediate Rai risk patients have a relatively long follow-up, are extensively characterized, and have a detailed medical history. We present evidence that the prognosis of UM B-CLL patients can be further stratified based on specific *IGHD* gene usage independent of *IGHV* and *IGHJ* gene utilization.

## Results

**Cohort characterization and clinical features.** The present study included 258 patients diagnosed with B-CLL between 1982 and 2006, who were classified as Rai stage 0, I, or II at diagnosis, and had one productive UM *IGHV* gene sequence (8, 9). At the onset of the analyses performed in this study, we had consented 680 patients since the year 2000, and *Ig* mutation status had been determined for each of them. Of this group, all patients classified as UM and Rai 0, I, or II were included in this study ( $n = 232$ ). The remainder were either classified as UM and Rai III or IV patients or as M patients ( $n = 448$ ). The remaining 26 UM Rai 0, I, or II patients included in our study were consented prior to 2000, and *Ig* mutation analysis was performed on cryopreserved specimens. According to convention, early-stage patients were classified as low risk (Rai stage 0) or intermediate risk (Rai stages I and II) for all analyses. Molecular and clinical characteristics of the patients included in this study are summarized in Table 1. With a median follow-up for treatment of 18 months (range, 1 day to 235 months), 190 of the 258 patients



**Table 2**  
Clinical characteristics: CD38, ZAP-70, and detectable FISH defects<sup>A</sup>

Characteristic	Rai stage 0	Rai stage I	Rai stage II
<b>CD38</b>			
<30% (negative)	72 (30%)	44 (19%)	26 (11%)
≥30% (positive)	41 (17%)	38 (16%)	16 (7%)
Missing	6	11	4
<b>ZAP-70</b>			
<20% (negative)	28 (20%)	19 (13%)	5 (4%)
≥20% (positive)	49 (35%)	30 (21%)	11 (8%)
Missing	42	44	30
<b>FISH abnormality<sup>B</sup></b>			
13q-	25 (14%)	15 (8%)	6 (3%)
None detected	20 (11%)	14 (8%)	8 (4%)
+12	16 (9%)	10 (6%)	10 (6%)
11q-	16 (9%)	12 (7%)	6 (3%)
17p-	8 (4%)	9 (5%)	1 (0.6%)
Other	1 (0.6%)	4 (2%)	0 (0%)
Missing	33	29	15

<sup>A</sup>Data are presented as number of patients, and values in parentheses represent percentage of patients in the entire cohort. <sup>B</sup>Data are presented in relation to the known hierarchy of CLL B cells as defined by FISH-detectable defects (60).

were subsequently treated. More than 34% of untreated patients had been followed for at least 2 years, and 17% had been followed for at least 5 years. Because of the high frequency of use of the *IGHV1-69* gene in our cohort of UM CLL patients (29%), we used this as an additional molecular classifier. In this regard, the median follow-up for treatment status for the 75 patients additionally classified as *IGHV1-69* was 21 months (range, 1 day to 148 months), and among these patients, 64 had been treated for their B-CLL.

In addition to *IGHV* mutation status, other biologic prognostic variables have been identified in B-CLL (8, 12, 32–36). Of these, CD38 was the first to be associated with *IGHV* mutation status, with patients with CD38-positive leukemic cells having a poorer prognosis (8). Of the 237 patients with available CD38 data, 40% of UM patients were CD38 positive using the 30% cutpoint (Table 2). Zeta chain-associated protein kinase 70 kDa (ZAP-70), like CD38, has also been associated with poor prognosis (36). ZAP-70 information was available on 142 patients, of whom 63% were classified as ZAP-70 positive (defined as ≥20% of CLL B cells ZAP-70; Table 2). Interphase FISH abnormality information was available on 181 patients, 29% of whom had 11q- or 17p- abnormalities (Table 2). Collectively, these risk parameters are consistent with the expected pattern of high-risk prognostic features of the UM patients (37).

*IGHV gene characteristics in UM patients.* Within the UM cohort, 41% of patient leukemic B cells used a member of the *IGHV1* family. This high frequency is attributed to the skewed usage of the *IGHV1-69* gene, which accounted for 29% of all UM patients and 71% of UM patients using *IGHV1* genes (Supplemental Table 1; supplemental material available online with this article; doi:10.1172/JCI32625DS1). The remaining UM patients predominately expressed *IGHV3* (33%) and *IGHV4* (17%) fam-

ily members. The 5 most frequent *IGHV* genes in the cohort were *IGHV1-69*, *IGHV3-21*, *IGHV3-30*, *IGHV4-34*, and *IGHV1-02*. Previous studies of cohorts including both M and UM patients have found similar overall patterns (9, 20, 26). In specifically focusing on *IGHD* and *IGHJ* gene characteristics of the cohort, we were able to determine both the *IGHD* and *IGHJ* gene usage in 251 of 258 patients (97%), with *IGHJ6* (46%) and *IGHD3* (57%) genes utilized most frequently (Table 1). Importantly, although there are 27 mapped genes in the *IGHD* gene family, 29% of the patients used the *IGHD3-3* gene (Table 1). The next most frequently utilized *IGHD* genes, accounting for almost 39% of the patients, were *IGHD2-2*, *IGHD3-22*, *IGHD6-19*, and *IGHD3-10* (Supplemental Table 1).

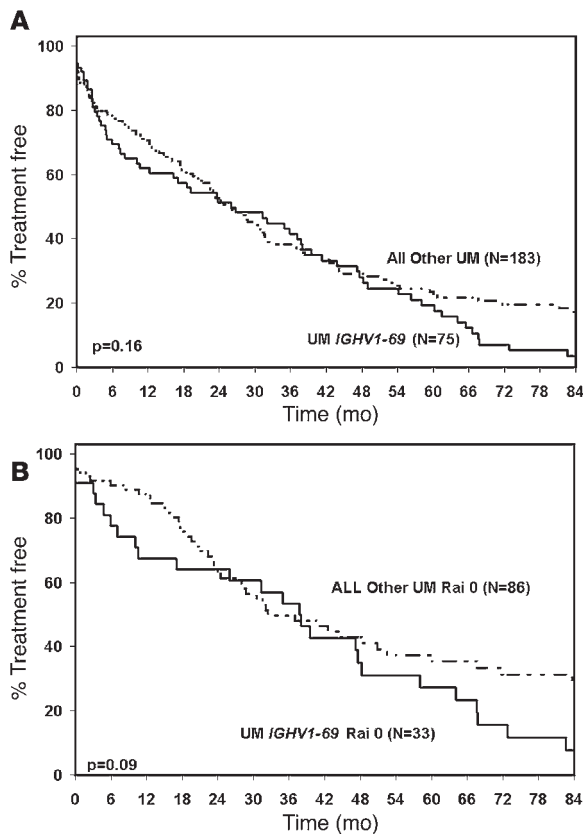
The combination of *IGHJ6* and *IGHD3-3* is of particular interest, because this combination has been associated with the *IGHV1-69* gene, and this overall bias may reflect evidence of specific antigen involvement in B-CLL (9, 25–27, 31). In addition, there is accumulating evidence of conserved CDR3 sequences in CLL B cells that are independent of specific *IGHV* gene usage (23, 24, 29). Given the number of patients in our cohort that used the *IGHJ6* and *IGHD3-3* genes independently (46% and 29%, respectively) or in combination (17%), we first looked for potential correlations of *IGHJ6* and *IGHD3-3* use individually with the other biologic prognostic variables discussed above. In looking at patients with *IGHJ6* usage versus all other patients using other *IGHJ* genes, *IGHJ6* usage was not significantly associated with CD38-positive status, ZAP-70-positive status, or adverse FISH prognosis group (data not shown). Similarly, we looked at patients with *IGHD3-3* usage versus all other patients not utilizing the *IGHD3-3* gene. Again, none of these established biologic markers were significantly associated with *IGHD3-3* usage. In evaluating the relationship between *IGHD* and *IGHJ* gene characteristics and Rai risk category at the time of diagnosis (low versus intermediate), there was no statistically significant relationship between Rai risk and *IGHD* gene usage ( $P = 0.13$ ) or Rai risk and *IGHJ* gene usage ( $P = 0.94$ ).

We next focused on the *IGHV1-69* gene because of its high frequency of use within the UM CLL patients. In total, 75 of the 258 UM patients in our cohort used *IGHV1-69*. Consistent with prior reports, a majority of the UM *IGHV1-69* patients used *IGHJ6* (59%), and 73% used a member of the *IGHD3* family, with specific use of the *IGHD3-3* gene observed in 43% of the patients. When comparing the UM *IGHV1-69* patients to non-*IGHV1-69* patients, we observed significant differences based on *IGHD3-3* usage ( $P = 0.0009$ ; 32 of 74 *IGHD3-3* in the UM *IGHV1-69* patients versus 40

**Table 3**  
Univariate TTT analysis in the UM cohort

Factor	n	HR	95% CI	P
Rai risk (low vs. intermediate)	258	2.04	1.51–2.76	0.000002
CD38 (<30% vs. ≥30%)	237	1.01	0.87–1.18	0.90
FISH group (11q- or 17p- vs. all others)	181	1.31	0.91–1.88	0.15
ZAP-70 (<20% vs. ≥20%)	142	0.77	0.62–0.97	0.02
<i>IGHV1-69</i> usage	258	1.24	0.92–1.68	0.16
<i>IGHJ6</i> usage	253	1.13	0.85–1.51	0.40
<i>IGHD3-3</i> usage	252	1.52	1.11–2.07	0.008
RF	244	1.09	0.89–1.35	0.40
<i>IGHD3-3</i> and RF2 (vs. all other non- <i>IGHD3-3</i> )	228	1.57	1.09–2.25	0.01

CI, confidence interval; HR, hazard ratio.



**Figure 1**  
TTT analysis of UM B-CLL *IGHV1-69* patients. (A) TTT is not significantly different when comparing UM *IGHV1-69* B-CLL patients with all other UM B-CLL patients. (B) UM *IGHV1-69* low Rai risk B-CLL patients do not have a significantly different TTT when compared with all other UM low Rai risk B-CLL patients.

of 178 in UM non-*IGHV1-69* patients) and *IGHJ6* usage ( $P = 0.006$ ; 44 of 74 *IGHJ6* in UM *IGHV1-69* patients versus 72 of 179 in UM non-*IGHV1-69* patients). When UM *IGHV1-69* patients were categorized into 4 cohorts — as (a) *IGHD3-3/IGHJ6*; (b) *IGHD3-3/non-IGHJ6*; (c) non-*IGHD3-3/IGHJ6*; or (d) non-*IGHD3-3/non-IGHJ6* — there was a significant association for usage of the *IGHJ6* and *IGHD3-3* genes when looking at the UM *IGHV1-69* group and all other UM samples ( $P = 0.0001$ ; see Supplemental Table 2). No significant differences were observed between *IGHV1-69* and non-*IGHV1-69* usage in terms of clinical factors and biologic risk factors, as listed in Tables 1 and 2.

**Time to treatment and established risk factors.** In evaluating TTT in this cohort of patients, we first noted that, as expected, there was a strong relationship between Rai risk category (low versus intermediate) at time of diagnosis and TTT on univariate analysis ( $P = 0.0001$ ; see Supplemental Figure 1). Thus, the intermediate Rai risk patients displayed significantly shorter TTT than did low Rai risk patients. Through univariate analyses, we then explored the relationship between other established risk factors and TTT. CD38-positive status and poor FISH prognosis group (11q- or 17p-) were not significant factors in univariate analyses for TTT in these low/intermediate Rai risk

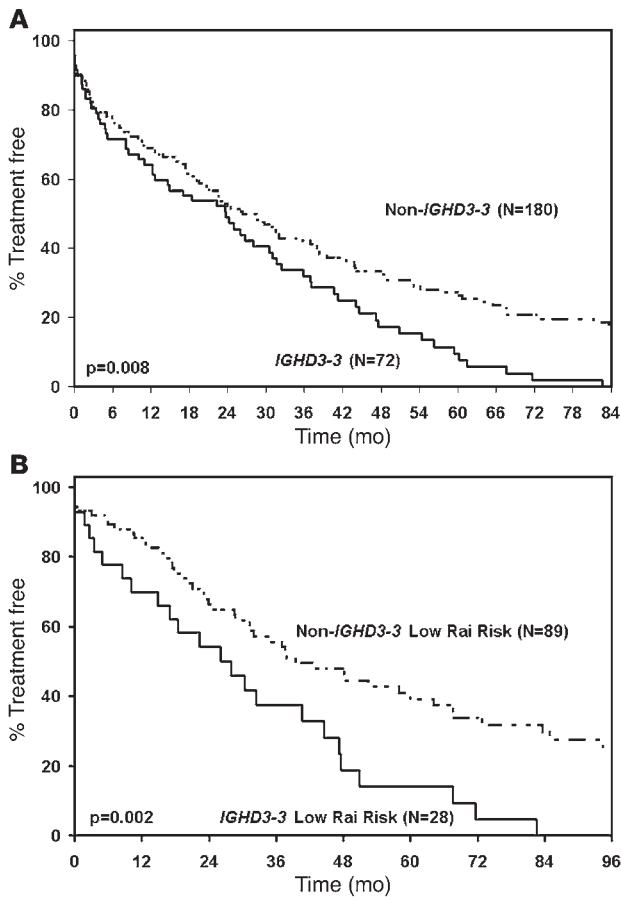
UM CLL patients ( $P = 0.90$  and  $P = 0.15$ , respectively; Table 3). This may suggest that the prognostic value of these markers is limited in patients with solely UM *IGHV* status. The 2% threshold used to define the UM patients included in these analyses is standard but other thresholds have been explored (12, 34). With this in mind, we next evaluated the relationship between continuous measure of percent mutation (ranging from 0% to 2%) and TTT. Most patients (62%) exhibited 0% mutation, whereas 25% exhibited 0.1–1% mutation, and 13% exhibited greater than 1% but less than 2% mutation. Overall, the continuous measures of percent mutation in the UM patients was not significantly associated with TTT ( $P = 0.25$ ). ZAP-70-positive status was the only established CLL risk factor that was significant factor in the univariate setting for TTT, although information on ZAP-70 status was only available on 142 of all 258 patients ( $P = 0.02$ ; Table 3).

**Univariate TTT analyses with specific *IGHV*, *IGHD*, or *IGHJ* gene usage.** We next evaluated the impact of specific *IGHV*, *IGHD*, or *IGHJ* gene usage patterns on TTT in these UM CLL patients. Patients using *IGHV1-69* are generally UM and thought to have a poor prognosis based on *IGHV* gene mutation status. To determine whether the poor prognosis observed in *IGHV1-69* patients is due exclusively to mutation status, we looked at TTT for *IGHV1-69* UM patients compared with all other UM patients. No significant differences were observed, and the medians for each of the groups were similar (median TTT for UM *IGHV1-69*, 26.0 months [ $n = 75$ ] versus 25.0 months for UM non-*IGHV1-69* [ $n = 183$ ];  $P = 0.16$ ; Figure 1A and Table 3). Similar trends were observed when we confined this analysis exclusively to low Rai risk patients (median TTT for UM *IGHV1-69*, 37.7 months [ $n = 33$ ] versus 32.4 months for UM non-*IGHV1-69* [ $n = 86$ ];  $P = 0.09$ ; Figure 1B and Table 4).

To better understand the importance of *IGHD3-3* and *IGHJ6* usage individually in relation to TTT, we analyzed TTT in *IGHD3-3* versus non-*IGHD3-3* patients as well as *IGHJ6* versus non-*IGHJ6* patients in univariate analyses (Table 3). Usage of the *IGHD3-3* gene was highly associated with shorter TTT ( $P = 0.008$ ; Figure 2A) versus those using non-*IGHD3-3* genes, whereas median TTT in the *IGHJ6* group and non-*IGHJ6* groups was not significantly different ( $P = 0.40$ ; Table 3). Overall, the percentage of patients who were treatment free at 2 and 5 years in the *IGHD3-3* cohort was 48.9% and 9.6%, respectively, compared with 52.1% and 26.3% for the non-*IGHD3-3* patients. We next evaluated whether *IGHD3-3* usage predicted TTT in exclusively low Rai risk patients (Rai stage 0), which is the group most in need of further stratification in terms of clinical prognosis. Indeed, in the low Rai risk category, *IGHD3-3*

**Table 4**  
Univariate TTT analysis in the low Rai risk cohort

Factor	n	HR	95% CI	P
CD38 (<30% vs. ≥30%)	113	1.04	0.82–1.31	0.76
FISH group (11q- or 17p- vs. all others)	86	1.71	0.99–1.34	0.05
ZAP-70 (<20% vs. ≥20%)	77	1.46	1.05–2.03	0.02
<i>IGHV1-69</i> usage	119	1.50	0.93–2.39	0.09
<i>IGHJ6</i> usage	118	1.25	0.8–1.97	0.33
<i>IGHD3-3</i> usage	117	2.12	1.29–3.51	0.002
RF	144	1.07	0.76–1.51	0.70
<i>IGHD3-3</i> and RF2 (vs. all other non- <i>IGHD3-3</i> )	108	1.93	1.07–3.50	0.03



**Figure 2**  
TTT analysis of UM B-CLL patients. (A) TTT in B-CLL patients using *IGHD3-3* versus non-*IGHD3-3* is significantly different. (B) Low Rai risk patients (Rai stage 0 at diagnosis) using the *IGHD3-3* gene have a significantly shorter TTT than low Rai risk patients using an *IGHD* gene other than *IGHD3-3*.

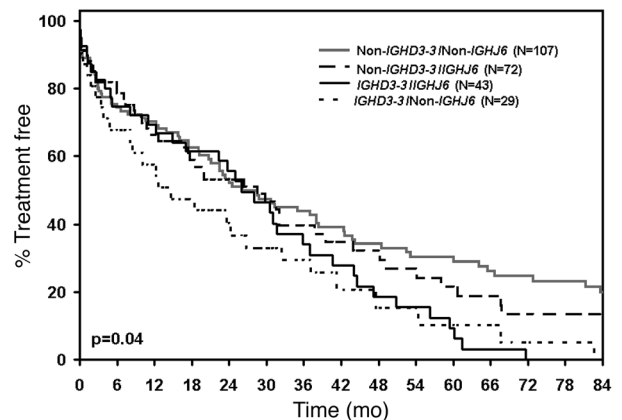
patients had a significantly shorter TTT than non-*IGHD3-3* patients ( $P = 0.002$ ; Figure 2B and Table 4). In terms of the more established risk factors, CD38-positive status was not a significant factor in univariate analyses for TTT in the low Rai risk UM patients ( $P = 0.76$ ; Table 4). TTT differences were observed between those in the poor FISH prognosis group (11q- or 17p-) versus others, where this result was borderline significant ( $P = 0.05$ ). Not surprisingly, the FISH parameter showed notable differences in terms of TTT between those patients with 11q- or 17p- anomalies (median 28.7 months for 11q- or 17p- versus 58 months for all others), consistent with previous observations (37). As seen in the full cohort, ZAP-70-positive status was a significant factor in the univariate setting ( $P = 0.02$ ; Table 4).

Finally, because of the overall frequency of *IGHD3-3* and *IGHJ6* usage in the entire UM cohort and because of the reported relatively frequent combined usage of these 2 genes in *IGHV1-69* patients (9, 25–27, 31), we grouped the UM patients into 4 cohorts as: (a) *IGHD3-3/IGHJ6*; (b) *IGHD3-3/non-IGHJ6*; (c) non-*IGHD3-3/IGHJ6*; or (d) non-*IGHD3-3/non-IGHJ6*. We observed a significant difference in TTT across these 4 subgroups of patients ( $P = 0.04$ ; Figure 3). In particular, the differences in TTT observed appear to have been driven largely by *IGHD3-3* usage, since those with combined

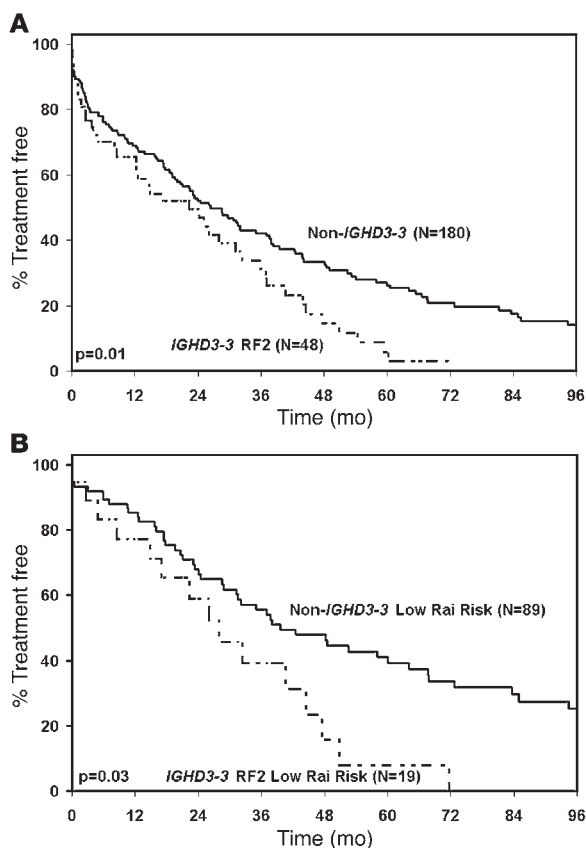
*IGHD3-3* and non-*IGHJ6* usage had a shorter TTT than the other combinations. Formal modeling of an interaction effect between *IGHD3-3* and *IGHJ6* was not significant; however, we also note that power in our study is limited for analyses of interactions. The data above showing the significance of *IGHD3-3* usage indicates that this *Ig* gene marker accounted for the difference in TTT of the *IGHD* and *IGHJ* combination shown in Figure 3, with increasing impact as time progressed for the CLL patients within the *IGHD3-3* cohort.

*Relationship of IGH gene RF to clinical outcome in the IGH3-3 cohort.* As shown above, *IGHD3-3* usage alone predicted shorter TTT in UM CLL patients. Because the *IGHD* gene plays a critical role in determining the antigenic specificity of the BCR (38, 39), it is possible that this association reflected conserved BCR antigen specificity in this cohort. However, *IGHD* genes are unique in that each one can be read in 3 forward or inverted RFs (1, 2, 5), with each RF encoding a different amino acid sequence in the HCDR3, which results in differences in overall hydrophobicity. In general, RF2 for most *IGHD* genes is hydrophilic, which is more advantageous for antigen binding (40). We next wished to study the *IGHD3-3* cohort to determine whether the correlation between use of this D region gene and shorter TTT was further associated with a specific RF. We first analyzed the frequency of RF1, -2, and -3 in the *IGHD3-3* UM cohort. RF2 was the most frequently used ( $n = 48$ ; 67%), followed by RF3 ( $n = 16$ ; 21%) and RF1 ( $n = 8$ ; 11%). When we evaluated the impact of RF usage on TTT in the *IGHD3-3* UM cohort, significant differences were not observed; however, we did see notable differences in the estimated median TTT for each of the RF groups in this restricted cohort of *IGHD3-3* patients (RF1, 14.3 months; RF2, 22.3 months; RF3, 30.5 months;  $P = 0.31$ ). We recognize the inherent limitations of the small number of patients available for these analyses; however, the differences observed indicate that patients with RF1 or RF2 usage had inferior clinical outcomes as compared with either the *IGHD3-3* RF3 patients or all non-*IGHD3-3* patients. We believe these findings warrant further exploration (Supplemental Figure 2).

Because of the limited numbers of patients with RF1 usage, we next explored whether the *IGHD3-3* patients with RF2 usage performed significantly worse compared with non-*IGHD3-3* UM patients. The *IGHD3-3* RF3 usage patients were also omitted due to



**Figure 3**  
TTT analysis of UM B-CLL patients subgroups based on *IGHD* and *IGHJ* usage. TTT analysis in UM B-CLL patients classified by *IGHD3-3* and *IGHJ6* gene usage (*IGHD3-3/IGHJ6*, non-*IGHD3-3/IGHJ6*, non-*IGHD3-3/non-IGHJ6*, or *IGHD3-3/non-IGHJ6*) shows a significant difference in TTT among the 4 subgroups.

**Figure 4**

TTT analysis of UM B-CLL patients using *IGHD3-3* RF2. (A) B-CLL patients using the *IGHD3-3* gene in RF2 were compared with all other non-*IGHD3-3* UM B-CLL patients, and significant differences in TTT were observed. (B) Low Rai risk patients using the *IGHD3-3* gene and more specifically RF2 also had a shorter TTT than all other low Rai risk UM B-CLL patients.

their apparent similarity to non-*IGHD3-3* UM patients (although sample size limits a formal exploration of this) and the fact that those *IGHD3-3* patients with RF2 usage posed an interesting biologically based hypothesis worth exploring, i.e., *IGHD3-3* RF2 patients may have conserved BCRs, and this underlies their shorter TTT. Of interest, we did observe a significant difference in TTT between *IGHD3-3* RF2 patients and non-*IGHD3-3* UM patients ( $P = 0.01$ ; Figure 4A and Table 3). Moreover, similar trends were observed when we restricted our analysis to low Rai risk patients. Thus, univariate analyses also showed that an indicator for *IGHD3-3* RF2 versus non-*IGHD3-3* usage was a significant factor for TTT ( $P = 0.03$ ; Figure 4B and Table 4).

It remained possible, however, that RF2 usage alone, regardless of *IGHD* gene usage, was the underlying factor determining the significantly shorter TTT observed in the *IGHD3-3* RF2 UM B-CLL patients versus non-*IGHD3-3* patients. In our entire UM cohort, the hydrophilic RF2 was used in 50% of patients, with RF1 and RF3 utilized in 16% and 34% of patients, respectively (Supplemental Table 3). These frequencies are similar to those in a previously published report studying *IGHD* RF usage in normal B cells (41). However, in UM patients, there were no statistically significant differences in TTT among the RF1, RF2, and RF3 groups either in the

entire cohort or in the low Rai risk subset ( $P = 0.64$  and  $P = 0.81$ , respectively; Supplemental Table 4).

**Multivariable TTT analyses.** In multivariable models and variable selection model building analyses, Rai risk category (low versus intermediate) had the greatest prognostic impact in determining TTT (Table 5). *IGHD3-3* gene usage was borderline significant when it was also incorporated into the model (Table 5, model 1;  $P = 0.05$ ). This same finding held when we looked at *IGHD3-3* RF2 usage patients versus non-*IGHD3-3* UM CLL patients in the cohort (Table 5, model 2;  $P = 0.06$ ); or when we categorized the *IGHD3-3* RF1 and *IGHD3-3* RF2 usage patients together versus all others (Table 5, model 3;  $P = 0.03$ ). Even when looking further to the model fit measure of likelihood ratio tests, the *IGHD3-3* usage variable was highly significant in all multivariable models analyzed (Table 5).

When ZAP-70 status was added to a multivariable model (Table 5, models 4–6), it was not a significant factor, and *IGHD3-3* usage was no longer significant; however, it is important to highlight that only 142 of 258 patients had ZAP-70 status available, and therefore multivariable models including ZAP-70 status were restricted to that subset of patients. Furthermore, there were significant differences in TTT between those with versus those without ZAP-70 status available, in that a greater proportion of low Rai risk patients had known ZAP-70 status. Also, in the 142 cohort with ZAP-70 status available, univariate analysis of *IGHD3-3* usage was not significant (data not shown;  $P = 0.44$ ). While the finding of ZAP-70 positivity as a significant factor cannot be disregarded, the fact that the univariate findings change dramatically when looking at all patients versus just those with ZAP-70 status data available is important and indicates that further analyses need to be done to completely assess these relationships. Because of the limited sample size, no multivariate analysis was performed in the low Rai risk cohort.

## Discussion

We and others have shown that *IGHV* gene mutation status in B-CLL has prognostic value (8–10, 12, 42–44). Considerable heterogeneity is present, however, in the clinical course of the unfavorable UM *IGHV* group, which to date is unexplained and is of clinical relevance to the patient (45). Accordingly, stratification parameters that improve the prognostic precision of mutation status would be of benefit to patients with low/intermediate Rai risk, asymptomatic UM disease. There is current momentum to always consider UM patients, even in early-stage disease, as inevitably linked to an aggressive course and even to be considered for initiation of early clinical trial treatment approaches. One maneuver to develop more insight into UM prognosis has been to combine *IGHV* mutation status with other prognostic parameters such as ZAP-70, CD38, and cytogenetic abnormalities detectable on FISH testing in order to create an integrated prognostic index (42). Other evidence already suggests, however, that additional characteristics of the *Ig* sequence may stratify clinical outcomes and enhance a prognostic model. For instance, *IGHV3-21* and *IGHV3-72* usage provides clinically important prognostic information independent of the *IGHV* gene mutation status (13–18). With that in mind, we evaluated an exclusively UM low/intermediate Rai risk patient cohort to determine whether specific molecular features of the *IGHD* and *IGHJ* genes predicted prognosis. Importantly, the large number of low Rai risk UM patients in this study (46% were Rai stage 0 at time of diagnosis) provided a unique opportunity for a long-term perspective of the association of clinical outcomes of patients with low



**Table 5**  
Multivariable TTT analysis in the UM cohort

Factor	HR	95% CI	P
<b>Model 1 (n = 252)<sup>A</sup></b>			
Rai risk (low vs. intermediate)	2.00	1.48–2.72	0.000008
<i>IGHD3-3</i> usage	1.36	1.00–1.87	0.05
<b>Model 2 (n = 244)<sup>B</sup></b>			
Rai risk (low vs. intermediate)	2.10	1.51–2.9	0.000009
<i>IGHD3-3</i> and RF2 (vs. non- <i>IGHD3-3</i> )	1.42	0.987–2.05	0.06
<b>Model 3 (n = 244)<sup>C</sup></b>			
Rai risk (low vs. intermediate)	2.02	1.49–2.74	0.000006
<i>IGHD3-3</i> and RF1/RF2 (vs. all others)	1.47	1.05–2.07	0.03
<b>Model 4 (n = 139)<sup>D</sup></b>			
Rai risk (low vs. intermediate)	2.19	1.41–3.39	0.0005
<i>IGHD3-3</i> usage	0.98	0.63–1.54	0.93
ZAP-70 (<20% vs. ≥20%)	1.22	0.97–1.53	0.09
<b>Model 5 (n = 135)<sup>E</sup></b>			
Rai risk (low vs. intermediate)	2.15	1.39–3.33	0.0006
<i>IGHD3-3</i> and RF1/RF2 (vs. all others)	1.11	0.68–1.79	0.69
ZAP-70 (<20% vs. ≥20%)	1.22	0.97–1.53	0.09
<b>Model 6 (n = 125)<sup>F</sup></b>			
Rai risk (low vs. intermediate)	2.13	1.34–3.38	0.001
<i>IGHD3-3</i> and RF2 (vs. non- <i>IGHD3-3</i> )	1.08	0.65–1.80	0.76
ZAP-70 (<20% vs. ≥20%)	1.23	0.96–1.57	0.1

<sup>A</sup>P = 0.000001, <sup>B</sup>P = 0.000003, <sup>C</sup>P = 0.0000009, <sup>D</sup>P = 0.0009, <sup>E</sup>P = 0.0009, <sup>F</sup>P = 0.002; likelihood ratio test for each model.

disease burden in our subsequent *IGHV* gene mutation analysis. In addition, our focus on low/intermediate Rai risk B-CLL patients provided a cohort of patients for whom observation is standard of care at diagnosis (46) and where TTT is an established indicator for progression of their disease (36, 47). We demonstrate for the first time to our knowledge the importance of *IGHD* gene usage in predicting disease progression as measured through TTT in a large group of low/intermediate Rai risk UM patients.

Consistent with previous reports (9, 25–27, 31) we found disproportionate usage of the *IGHV1-69* gene in UM patients, accounting for 29% of all the patients studied. Other investigators have also previously reported that the molecular features of the *IGHV1-69* gene in B-CLL show a biased usage of the 51p1 allele in conjunction with *IGHD3-3* and *IGHJ6* genes (25–27). The *IGHV1-69* patients using the specific combination of *IGHD3-3* and *IGHJ6* accounted for 24% of our patient cohort, differing from a previous report of 50% (20). This may reflect the fact that all of the patients in this cohort are low/intermediate Rai risk patients. Notably, we studied *IGHJ6* and *IGHD3* gene family usage in our entire UM cohort and observed an augmented usage of *IGHJ6* (46%) and *IGHD3* (57%), with the specific *IGHD3-3* gene being utilized in 29% of all cases. Even with the exclusion of the UM *IGHV1-69* samples from our UM cohort, there was still a skewed usage of *IGHJ6* (40%), *IGHD3* (50%), and *IGHD3-3* (23%). However, examination of *IGHD* and *IGHJ* combinations in the non-*IGHV1-69* UM group showed that only 14% used both *IGHD3-3* and *IGHJ6* genes.

During B cell development, rearrangement of the *IGHV*, *IGHD*, and *IGHJ* genes results in a B cell with a unique HCDR3. Because of the central role that the *IGHD* and *IGHJ* genes play in form-

ing the HCDR3, a region that is key in determining the specificity of antigen binding (6, 38, 39), the biased use of specific genes such as *IGHD3-3* and *IGHJ6* in B-CLL may have implications for antigens in the leukemic transformation process. The *IGHD* genes are particularly important because they provide a critical portion of the antigen-binding loop in the *Ig* molecule, and the hydrophobicity traits of the possible RFs of the *IGHD* gene may determine *Ig* effectiveness for binding antigen (39). Given that there are 27 mapped *IGHD* genes, the incidence of use of *IGHD3-3* in our entire UM CLL cohort, and its frequency of use in our *IGHV1-69* cohort is striking and adds further evidence to the literature (9, 19, 20, 26, 27) that this specific *IGHD* gene is overused in B-CLL relative to normal B cells. It should be noted, however, that data on the expression of *IGHD* genes in the normal human B cell repertoire are very limited, often conflicting, and need to be further studied (40, 48–50).

Concerning the strong association of *IGHD3-3* with the *IGHV1-69* gene in B-CLL, Widhopf and Kipps (27) looked at normal B cells expressing *IGHV1-69* and found a lack of *IGHD3-3* gene expression. Because B-CLL is more common in the elderly population and there may be changes in BCR repertoire that accompany age, Potter et al. (25) analyzed the *IGHV1-69* genes expressed in 35 normal young individuals (<35 years of age) and 42 normal elderly individuals (>75 years of age). They determined that there was no biased use of any of the *IGHD* genes in relation to age and that *IGHD3-22* was frequently used whereas *IGHD3-3* was strikingly underutilized relative to B-CLL. Thus, with respect to the biased usage of *IGHD3-3*

with the *IGHV1-69* gene in B-CLL, this association appears to be a characteristic of the disease and does not reflect BCR repertoire changes that occur with age.

In addition to evaluating trends in *IGHD* and *IGHJ* gene usage, a second major goal of our study was to look at the potential relationship between specific *Ig* gene usage and time from diagnosis to the first treatment. The rationale for this is a very practical one, in that the long median survival of B-CLL patients prevents use of a survival endpoint without a very long follow-up period, warranting the common use of other surrogate and clinically meaningful measures. Thus, TTT was used as a measure of clinical progression of disease, as previously described (36, 47), and because of the inherent inconsistencies of effectively classifying disease progression by strict interpretation of the National Cancer Institute working group criteria (46). Use of TTT is valid in our population of low/intermediate Rai risk CLL patients in whom treatment is not usually required at diagnosis, and thus TTT is a useful indicator of disease progression.

*IGHD*, *IGHD* RF, and *IGHJ* usage has previously been used in concert with *IGHV* usage to define subsets of stereotyped HCDR3 sequences (14, 19, 21–24, 27–30). Accordingly, we were also interested in defining our cohort of UM patients on this basis. Because of the biased usage of *IGHJ6* and *IGHD3-3* in our large UM cohort, we grouped patients on this basis and observed that patients using the *IGHD3-3* gene displayed a significantly shorter TTT, particularly when evaluating and comparing patients whose leukemic B cells used any *IGHJ* gene other than the *IGHJ6* gene. While the *IGHD3-3* gene has been associated with the *IGHJ6* gene in those patients that express the clinically unfavorable *IGHV1-69* gene, we find that the



*IGHD3-3* gene alone is sufficient to predict shorter TTT and thus an adverse clinical prognosis even when Rai risk group is placed in a prognostic model. The significance of the *IGHD3-3* gene needs to be explored further in the context of multivariable analyses but is a promising strong predictor of significantly shorter TTT, especially in the Rai stage 0 patients.

Rai stage is a proven powerful prognosticator for B-CLL, with those patients at Rai stage 0 at diagnosis having a median survival of 120 months or greater (51–53). However, patients who are Rai stage 0 can have a highly variable clinical outcome, including an aggressive disease course. Thus, additional markers are needed to further understand and delineate prognosis in these patients (32, 53, 54). Our study population is therefore relevant to this quest because of the number of patients with UM B-CLL diagnosed at Rai stage 0 ( $n = 119$ ; 46%). Analysis of the significance of clinical stage and *IGHD* and *IGHJ* gene characteristics in UM B-CLL patients showed that those patients with Rai stage 0 disease who utilized *IGHD3-3* had a shorter TTT than all other UM patients.

Our study also supports the idea that the *IGHD3-3* gene is frequently used in CLL B cells. Furthermore, our study and others suggest that *IGHD3-3* is also used frequently in combination with *IGHV* genes other than the *IGHV1-69* gene and is independent of geographical location of the patients (9, 19, 22, 24, 55). We also find that within the *IGHD3-3* gene, the specific RF2 is used more frequently (48 of 72; 67%) and that within the group utilizing RF2, there appears to be an association specifically with the *IGHJ6* gene (31 of 48; 65%). Because of the frequent use of the *IGHJ6* gene in this cohort and because this is the longest of 6 *IGHJ* genes, the overall length of the HCDR3 in this cohort of patients is significantly greater than in all other patients ( $P = 0.0001$ ). In analyses not shown, we also tested whether HCDR3 length alone, analyzed as a continuous variable, was associated with TTT. However, this analysis failed to reveal a significant relationship ( $P = 0.26$ ). In contrast, in those patients using an *IGHD* gene other than *IGHD3-3*, the *IGHJ6* gene was used only 40% (70 of 176) of the time (data not shown). Thus far, 48 different subsets of sequences with stereotyped HCDR3 have been identified by other investigators (14, 19, 21–24, 27–30). In evaluating our entire UM cohort of patients for alignment with this definition of CLL subsets, we observed that 23.7% of the patients fell into one of these 48 subsets, a percentage that is consistent with that reported by Stamatopoulos et al. (24). For the *IGHD3-3* RF2 patients identified in this study as more likely to require earlier treatment, we observed that 18 of 48 (38%) *IGHD3-3* RF2 patients fell within one of these 48 subsets, with the majority ( $n = 12$ ) meeting the criteria of the subset defined as *IGHV1-69*; *IGHD3-3* (RF2); *IGHJ6* (24). By contrast, only 35 of 183 (19%) non-*IGHD3-3* patients met the criteria for any of the 48 subsets. Of interest, in both groups of patients, we have observed potential new subsets defined on the basis of conserved *IGHV*, *IGHD*, *IGHJ*, and *IGHD* RF that have yet to be formally described in the literature (R.C. Tschumper and D.F., Jelinek, unpublished observations). Although there are 2 previous reports showing evidence of a link between the HCDR3 and clinical outcome (24, 31), the numbers of patients analyzed in both studies were small and included both UM and M B-CLL patients at various stages of disease.

Our study is the largest to date focusing specifically on UM low/intermediate Rai risk B-CLL patients, and the clinical outcomes found in our current analysis not only allow definition of a new prognostic marker that may be of particular value to UM

low Rai risk patients but also add further support to the notion that the BCR plays a functional role in B-CLL. However, the exact role of the BCR in either the initiation of B-CLL or disease progression remains to be defined. Given the known heterogeneity of this disease, we propose that there could be some patients whose leukemic B cells continue to rely on BCR-stimulated signals for growth and survival and others whose malignant B cells have evolved to be independent of signaling through the BCR. Analysis of HCDR3 homologies at the level of primary amino acid sequence is an important first step in identifying subgroups of CLL. Our cohort of *IGHD3-3* RF2 patients clearly displays a core amino acid homology in this region (Supplemental Table 3). Of interest, although the number of *IGHD3-3* RF1 patients was limited ( $n = 8$ ), we observed that this group exhibited shorter TTT than either *IGHD3-3* RF3 or all other non-*IGHD3-3* patients (Supplemental Figure 2). Although RF1 is hydrophobic by comparison with RF2, it is markedly less hydrophobic than RF3, and it is furthermore marked by an amino acid motif that has been associated with antibody polyreactivity (RFLEW; ref. 56). Thus, despite the dissimilarity in HCDR3 amino acid sequence in *IGHD3-3* RF1 versus RF2, BCRs in both groups of patients may be stimulated more frequently than in non-*IGHD3-3* RF1/RF2 patients. However, we acknowledge the importance of also analyzing the chemical properties of all of the amino acids in this region to gain a better sense of the biological significance of shared HCDR3 regions among CLL patients. Ultimately, defining such groups based on HCDR3 homology could be useful in clinical management, particularly if therapeutic strategies effective at disrupting signals downstream of the BCR can be designed.

We believe that our large study is the first to strongly suggest that analysis of *IGHD* gene usage and RF alone is of importance when exclusively evaluating low Rai risk UM B-CLL patients. Importantly, this provides another useful clinical parameter for evaluating these patients, as early-stage UM CLL patients represent a significant clinical problem. The skewed utilization of the *IGHD3-3* gene by B-CLL patients and the strong positive association of *IGHD3-3* utilizing the more hydrophilic and antigen-accessible RF2 with TTT further supports the role of antigen in B-CLL, either at the transformation stage or in promulgation of the leukemic clone. While previous studies show that expression of the *IGHV1-69* gene in a UM state is predictive of poor prognosis, the disease outcome in these patients may be more specifically related to the overutilization of the *IGHD3-3* gene. Further analysis of the molecular features of low/intermediate Rai risk patients is warranted to determine whether *IGHD* gene characteristics can be more broadly used as a prognostic marker in both UM and M subsets of patients with CLL.

## Methods

**Patient material.** All study subjects provided written informed consent to participate in this study in accordance with the Declaration of Helsinki and the Mayo Clinic Institutional Review Board. Patients diagnosed with B-CLL had an absolute lymphocytosis ( $>5 \times 10^9/l$ ) due to monoclonal B cells with a CLL immunophenotype, as previously defined (46, 57, 58). Blood was drawn into heparin tubes and PBMCs were isolated by Ficoll (Gallard-Schlesinger Industries Inc.) density gradient centrifugation for *IGHV* mutation analysis. Patients included in this study had a median age of 59 (range 27–86) with 184 males and 74 females. Because the focus of the study was based on Rai risk stratification, data on ethnicity and nationality were not collected.





**RNA isolation, cDNA synthesis, and analysis of IGHV mutation.** Total RNA was isolated from PBMCs using the TRIzol reagent (Invitrogen – Life Technologies). Total RNA was converted to cDNA, and IGHV mutation was determined as previously described (44). Resulting sequences were aligned with those in the V-Base sequence directory (<http://vbase.mrc-cpe.cam.ac.uk/>) and the IMGT/V-Quest reference sets ([http://imgt.cines.fr/IMGT\\_vquest/share/textes/](http://imgt.cines.fr/IMGT_vquest/share/textes/)). All sequences that were identified as having less than 2% mutation from the closest germline sequence were designated as UM and thus considered for further analyses. UM patients exhibiting biallelic rearrangements were excluded from the study (approximately 4% of low/intermediate Rai risk UM patients). The IGHG gene and IGHD gene RF were determined using the IMGT/JunctionAnalysis program ([http://imgt.cines.fr/IMGT\\_jcta/jcta?livret=0](http://imgt.cines.fr/IMGT_jcta/jcta?livret=0)), which aligns the junction nucleotide sequences (which must start with the variable region second cysteine codon and end with the J-region phenylalanine or tryptophan codon) to the known IGHD region data set (59). Given that the mutation profile does not change over the disease course (9), time from diagnosis to analysis was not considered in interpreting IGHV mutation data.

**Cytogenetics, CD38, and ZAP-70 analysis.** In addition to IGHV mutation analysis, FISH was performed as previously described with a panel of DNA probes used to detect chromosomal anomalies (33). When categorizing patients with multiple FISH abnormalities, we used the hierarchical model proposed by Dohner et al. (60). Flow cytometric analysis of CD38 and ZAP-70 expression levels was performed as previously described (36, 58).

**Statistics.** Patient characteristic data were descriptively summarized and graphically evaluated. Analyses were conducted on all early-stage UM (Rai stage 0–II) patients identified as defined above as well as on Rai stage 0 (i.e., low Rai risk) patients. Bivariate analyses based on IGHV or IGHD gene usage were also done. The most frequent Ig genes used were selected for correlation with TTT, and outcomes were summarized by and within type of gene region. Differences in continuous variables between 2 groups were calculated using 2-sample, 2-tailed *t* tests or nonparametric Wilcoxon rank-sum tests as appropriate. Differences in categorical variables between groups were assessed using  $\chi^2$  or Fisher's exact tests as appropriate. TTT was calculated from the time of diagnosis to the time of first documented treatment. If patients had not been treated at the time of their last follow-up, they were censored at that time point. Kaplan-Meier and Cox regression methods were used to evaluate differences in TTT between groups in the univariate setting. The *P* values reported for differences in TTT correspond to log-rank statistics unless otherwise noted. Variables analyzed in the univariate setting

included: IGHD gene usage (IGHD3-3 versus not), RF usage (RF1 versus RF2 versus RF3), IGHD3-3 RF2 usage (versus non-IGHD3-3), IGHJ gene usage (IGHJ6 versus not), IGHV1-69 usage (IGHV1-69 versus not), CD38-positive status (30% cutpoint), ZAP-70-positive status (20% cutpoint), Rai risk category (low versus intermediate), and chromosomal abnormalities associated with poor prognosis as detected by FISH (11q- or 17p- defined as poor prognosis versus all others). Multivariable models were assessed using Cox regression methods and variable selection explored using the branch-and-bound algorithm of Furnival and Wilson (61). We also assessed potential multivariable models using Cox proportional hazards models, where variables with *P* values less than 0.10 in the univariate models were included. In addition, limited evaluation of interaction effects were assessed in these multivariable models, between IGHD3-3 and IGHJ6 usage as well as between IGHD3-3 and RF usage. We also evaluated additional IGHD3-3 groupings and assessed the different multivariable models using model-based likelihood ratio tests as an overall assessment.

Although we simultaneously evaluated 8 different risk factors in the context of these analyses, multiple comparison corrections were not used. Effective prognostic factors clearly need to be validated across institutions. Our analyses were considered hypothesis-generating and provide the necessary preliminary data to identify potential risk factors that should be explored further in the larger multi-institution setting. Statistical significance was declared at *P* < 0.05. All analyses were conducted using SAS version 9.1 and S-PLUS for Windows version 7.0 (Insightful Corp.).

### Acknowledgments

We would like to thank Cheryl Jankiewicz for her technical assistance in determining IGHV mutation status and Susan Slager for critical reading of this manuscript. This work was supported by NIH grant R01 CA95241 (awarded to N.E. Kay), the generous philanthropic support of Edson Spencer, and kind gifts from the Donner Foundation.

Received for publication May 8, 2007, and accepted in revised form October 17, 2007.

Address correspondence to: Diane F. Jelinek, Department of Immunology, Guggenheim 4, Mayo Clinic College of Medicine, 200 First St. SW, Rochester, Minnesota 55905, USA. Phone: (507) 284-5617; Fax: (507) 266-0981; E-mail: jelinek.diane@mayo.edu.

- Kurosawa, Y., and Tonegawa, S. 1982. Organization, structure, and assembly of immunoglobulin heavy chain diversity DNA segments. *J. Exp. Med.* **155**:201–218.
- Alt, F.W., and Baltimore, D. 1982. Joining of immunoglobulin heavy chain gene segments: implications from a chromosome with evidence of three D-JH fusions. *Proc. Natl. Acad. Sci. U. S. A.* **79**:4118–4122.
- Tonegawa, S. 1983. Somatic generation of antibody diversity. *Nature.* **302**:575–581.
- Yancopoulos, G.D., and Alt, F.W. 1986. Regulation of the assembly and expression of variable-region genes. *Annu. Rev. Immunol.* **4**:339–368.
- Meier, J.T., and Lewis, S.M. 1993. P nucleotides in V(D)J recombination: a fine-structure analysis. *Mol. Cell. Biol.* **13**:1078–1092.
- Xu, J.L., and Davis, M.M. 2000. Diversity in the CDR3 region of V(H) is sufficient for most antibody specificities. *Immunity.* **13**:37–45.
- Morea, V., Tramontano, A., Rustici, M., Chothia, C., and Lesk, A.M. 1998. Conformations of the third hypervariable region in the VH domain of immunoglobulins. *J. Mol. Biol.* **275**:269–294.
- Damle, R.N., et al. 1999. Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. *Blood.* **94**:1840–1847.
- Hamblin, T.J., Davis, Z., Gardiner, A., Oscier, D.G., and Stevenson, F.K. 1999. Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. *Blood.* **94**:1848–1854.
- Maloum, K., et al. 2000. Expression of unmutated VH genes is a detrimental prognostic factor in chronic lymphocytic leukemia. *Blood.* **96**:377–379.
- Thunberg, U., et al. 2001. CD38 expression is a poor predictor for VH gene mutational status and prognosis in chronic lymphocytic leukemia. *Blood.* **97**:1892–1894.
- Oscier, D.G., et al. 2002. Multivariate analysis of prognostic factors in CLL: clinical stage, IGVH gene mutational status, and loss or mutation of the p53 gene are independent prognostic factors. *Blood.* **100**:1177–1184.
- Tobin, G., et al. 2002. Somatic mutation of Ig V(H)3-21 genes characterize a new subset of chronic lymphocytic leukemia. *Blood.* **99**:2262–2264.
- Tobin, G., et al. 2003. Chronic lymphocytic leukemias utilizing the VH3-21 gene display highly restricted Vlambda2-14 gene use and homologous CDR3s: implicating recognition of a common antigen epitope. *Blood.* **101**:4952–4957.
- Thorselius, M., et al. 2006. Strikingly homologous immunoglobulin gene rearrangements and poor outcome in VH3-21-using chronic lymphocytic leukemia patients independent of geographic origin and mutational status. *Blood.* **107**:2889–2894.
- Bomben, R., et al. 2006. Comprehensive characterization of IGHV3-21-expressing B-cell chronic lymphocytic leukemia: an Italian multicenter study. *Blood.* **109**:2989–2998.
- Capello, D., et al. 2004. Evidence of biased immunoglobulin variable gene usage in highly stable B-cell chronic lymphocytic leukemia. *Leukemia.* **18**:1941–1947.
- Capello, D., et al. 2006. Immunophenotypic characterization of IgVH3-72 B-cell chronic lymphocytic leukaemia (B-CLL). *Leuk. Res.* **30**:1197–1199.
- Tobin, G., et al. 2004. Subsets with restricted immunoglobulin gene rearrangement features indicate a role for antigen selection in the development of chronic lymphocytic leukemia. *Blood.* **104**:2879–2885.
- Fais, F., et al. 1998. Chronic lymphocytic leu-



- kemia B cells express restricted sets of mutated and unmutated antigen receptors. *J. Clin. Invest.* **102**:1515–1525.
21. Ghiotto, F., et al. 2004. Remarkably similar antigen receptors among a subset of patients with chronic lymphocytic leukemia. *J. Clin. Invest.* **113**:1008–1016.
22. Widhopf, G.F., 2nd, et al. 2004. Chronic lymphocytic leukemia B cells of more than 1% of patients express virtually identical immunoglobulins. *Blood.* **104**:2499–2504.
23. Messmer, B.T., et al. 2004. Multiple distinct sets of stereotyped antigen receptors indicate a role for antigen in promoting chronic lymphocytic leukemia. *J. Exp. Med.* **200**:519–525.
24. Stamatopoulos, K., et al. 2007. Over 20% of patients with chronic lymphocytic leukemia carry stereotyped receptors: Pathogenetic implications and clinical correlations. *Blood.* **109**:259–270.
25. Potter, K.N., et al. 2003. Features of the overexpressed V1-69 genes in the unmutated subset of chronic lymphocytic leukemia are distinct from those in the healthy elderly repertoire. *Blood.* **101**:3082–3084.
26. Johnson, T.A., Rassenti, L.Z., and Kipps, T.J. 1997. Ig VH1 genes expressed in B cell chronic lymphocytic leukemia have distinctive molecular features. *J. Immunol.* **158**:235–246.
27. Widhopf, G.F., 2nd, and Kipps, T.J. 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.
28. Ghia, P., et al. 2005. Geographic patterns and pathogenetic implications of IGHV gene usage in chronic lymphocytic leukemia: the lesson of the IGHV3-21 gene. *Blood.* **105**:1678–1685.
29. Stamatopoulos, K., et al. 2005. Immunoglobulin light chain repertoire in chronic lymphocytic leukemia. *Blood.* **106**:3575–3583.
30. Bende, R.J., et al. 2005. Among B cell non-Hodgkin's lymphomas, MALT lymphomas express a unique antibody repertoire with frequent rheumatoid factor reactivity. *J. Exp. Med.* **201**:1229–1241.
31. Mauerer, K., et al. 2005. Immunoglobulin gene segment usage, location and immunogenicity in mutated and unmutated chronic lymphocytic leukaemia. *Br. J. Haematol.* **129**:499–510.
32. Shanafelt, T.D., Byrd, J.C., Call, T.G., Zent, C.S., and Kay, N.E. 2006. Narrative review: initial management of newly diagnosed, early-stage chronic lymphocytic leukemia. *Ann. Intern. Med.* **145**:435–447.
33. Dewald, G.W., et al. 2003. Chromosome anomalies detected by interphase fluorescence in situ hybridization: correlation with significant biological features of B-cell chronic lymphocytic leukaemia. *Br. J. Haematol.* **121**:287–295.
34. Krober, A., et al. 2002. V(H) mutation status, CD38 expression level, genomic aberrations, and survival in chronic lymphocytic leukemia. *Blood.* **100**:1410–1416.
35. Montserrat, E. 2006. New prognostic markers in CLL. *Hematology Am. Soc. Hematol. Educ. Program.* **2006**:279–284.
36. Rassenti, L.Z., et al. 2004. 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.
37. Kay, N.E., O'Brien, S.M., Pettitt, A.R., and Stilgenbauer, S. 2007. The role of prognostic factors in assessing 'high-risk' subgroups of patients with chronic lymphocytic leukemia. *Leukemia.* **21**:1885–1891.
38. Ichihara, Y., Matsuoka, H., and Kurosawa, Y. 1988. Organization of human immunoglobulin heavy chain diversity gene loci. *EMBO J.* **7**:4141–4150.
39. Raaphorst, F.M., Raman, C.S., Nall, B.T., and Teale, J.M. 1997. Molecular mechanisms governing reading frame choice of immunoglobulin diversity genes. *Immunol. Today.* **18**:37–43.
40. Corbett, S.J., Tomlinson, I.M., Sonnhammer, E.L., Buck, D., and Winter, G. 1997. Sequence of the human immunoglobulin diversity (D) segment locus: a systematic analysis provides no evidence for the use of DIR segments, inverted D segments, "minor" D segments or D-D recombination. *J. Mol. Biol.* **270**:587–597.
41. Souto-Carneiro, M.M., Longo, N.S., Russ, D.E., Sun, H.W., and Lipsky, P.E. 2004. Characterization of the human Ig heavy chain antigen binding complementarity determining region 3 using a newly developed software algorithm, JOINSOLVER. *J. Immunol.* **172**:6790–6802.
42. Shanafelt, T.D., Geyer, S.M., and Kay, N.E. 2004. Prognosis at diagnosis: integrating molecular biologic insights into clinical practice for patients with CLL. *Blood.* **103**:1202–1210.
43. Chiorazzi, N., and Ferrarini, M. 2003. B cell chronic lymphocytic leukemia: lessons learned from studies of the B cell antigen receptor. *Annu. Rev. Immunol.* **21**:841–894.
44. Jelinek, D.F., et al. 2001. Analysis of clonal B-cell CD38 and immunoglobulin variable region sequence status in relation to clinical outcome for B-chronic lymphocytic leukaemia. *Br. J. Haematol.* **115**:854–861.
45. Shanafelt, T.D., et al. 2007. Quality of life in chronic lymphocytic leukemia: an international survey of 1482 patients. *Br. J. Haematol.* **139**:255–264.
46. Cheson, B.D., et al. 1996. National Cancer Institute-sponsored Working Group guidelines for chronic lymphocytic leukemia: revised guidelines for diagnosis and treatment. *Blood.* **87**:4990–4997.
47. Calin, G.A., et al. 2005. A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. *N. Engl. J. Med.* **353**:1793–1801.
48. Yamada, M., et al. 1991. Preferential utilization of specific immunoglobulin heavy chain diversity and joining segments in adult human peripheral blood B lymphocytes. *J. Exp. Med.* **173**:395–407.
49. Brezinschek, H.P., Brezinschek, R.L., and Lipsky, P.E. 1995. Analysis of the heavy chain repertoire of human peripheral B cells using single-cell polymerase chain reaction. *J. Immunol.* **155**:190–202.
50. Brezinschek, H.P., et al. 1997. Analysis of the human VH gene repertoire. Differential effects of selection and somatic hypermutation on human peripheral CD5(+)/IgM+ and CD5(-)/IgM+ B cells. *J. Clin. Invest.* **99**:2488–2501.
51. Rai, K.R., et al. 1975. Clinical staging of chronic lymphocytic leukemia. *Blood.* **46**:219–234.
52. Ripolles, L., et al. 2006. Genetic abnormalities and clinical outcome in chronic lymphocytic leukemia. *Cancer Genet. Cytogenet.* **171**:57–64.
53. Dighiero, G. 2005. CLL biology and prognosis. *Hematology Am. Soc. Hematol. Educ. Program.* **2005**:278–284.
54. Dighiero, G., et al. 1998. Chlorambucil in indolent chronic lymphocytic leukemia. French Cooperative Group on Chronic Lymphocytic Leukemia. *N. Engl. J. Med.* **338**:1506–1514.
55. Duke, V.M., et al. 2003. V(H) gene usage differs in germline and mutated B-cell chronic lymphocytic leukemia. *Haematologica.* **88**:1259–1271.
56. Deng, Y.J., and Notkins, A.L. 2000. Molecular determinants of polyreactive antibody binding: HCDR3 and cyclic peptides. *Clin. Exp. Immunol.* **119**:69–76.
57. Muller-Hermelink, H.K., Catovsky, D., Montserrat, E., and Harris, N.L. 2001. Chronic lymphocytic leukemia/small lymphocytic lymphoma. In *Pathology and genetics: tumours of haemopoietic and lymphoid tissues*. E.S. Jaffe, N.L. Harris, H. Stein, and J.W. Vardiman, editors. IARC Press. Lyon, France. 127–130.
58. Nowakowski, G.S., et al. 2005. Interphase fluorescence in situ hybridization with an IGH probe is important in the evaluation of patients with a clinical diagnosis of chronic lymphocytic leukaemia. *Br. J. Haematol.* **130**:36–42.
59. Lefranc, M.P. 2001. IMGT, the international ImMunoGeneTics database. *Nucleic Acids Res.* **29**:207–209.
60. Dohner, H., et al. 2000. Genomic aberrations and survival in chronic lymphocytic leukemia. *N. Engl. J. Med.* **343**:1910–1916.
61. Furnival, G.M., and Wilson, R.W., Jr. 1974. Regressions by Leaps and Bounds. *Technometrics.* **16**:499–511.