

Immunoglobulin gene rearrangements in Chinese and Italian patients with chronic lymphocytic leukemia

Marilisa Marinelli^{1,*}, Caterina Ilari^{1,*}, Yi Xia^{2,*}, Iliaria Del Giudice¹, Luciana Cafforio¹, Irene Della Starza¹, Sara Raponi¹, Paola Mariglia¹, Silvia Bonina¹, Zhen Yu³, Wenjuan Yang³, Lugui Qiu³, Thomas Chan⁴, Alfonso Piciocchi⁵, Yok-Lam Kwong⁴, Eric Tse⁴, Jianyong Li², Anna Guarini¹, Wei Xu², Robin Foà¹

¹Department of Cellular Biotechnologies and Hematology, "Sapienza" University, Rome, Italy

²Department of Hematology, The First Affiliated Hospital of Nanjing Medical University, Jiangsu Province Hospital, Nanjing, China

³Department of Lymphoma & Myeloma Institute of Hematology, CAMS & PUMC, Tianjin, China

⁴Department of Medicine, The University of Hong Kong, Queen Mary Hospital, Hong Kong

⁵GIMEMA Data Center, GIMEMA Foundation, Rome, Italy

*These authors have contributed equally to this work

Correspondence to: Robin Foà, **e-mail:** rfoa@bce.uniroma1.it

Keywords: chronic lymphocytic leukemia, IGHV, stereotyped receptors, Chinese CLL, Italian CLL

Received: December 22, 2015

Accepted: February 16, 2016

Published: March 01, 2016

ABSTRACT

Chronic lymphocytic leukemia (CLL) is the most common type of leukemia in the Western world, whereas in Asia the incidence is about 10 times lower. The basis for this ethnic and geographic variation is currently unknown. The aim of this study was to characterize IGHVDJ rearrangements and stereotype of the HCDR3 region in a series of 623 Chinese CLL, in order to identify possible differences in immunoglobulin gene usage and their potential pathogenetic implications. Chinese CLL were compared to 789 Italian CLL. Chinese patients showed a higher proportion of mutated IGHV and a more frequent usage of IGHV3-7, IGHV3-74, IGHV4-39 and IGHV4-59 genes. A significantly lower usage of IGHV1-69 and IGHV1-2 was documented, with comparable IGHV3-21 frequency (3% Chinese vs 3.8% Italian CLL). The proportion of known stereotyped receptors was significantly lower in Chinese (19.7%) than in Italian CLL (25.8%), despite a significantly higher frequency of subset #8 ($p = 0.0001$). Moreover, new paired clusters were identified among Chinese cases. Overall, these data support a potential different antigenic exposure between Eastern and Western CLL.

INTRODUCTION

Chronic lymphocytic leukemia (CLL) is characterized by a variable clinical course, driven by the immunogenetic and molecular heterogeneity of the disease [1-5].

The heterogeneity of CLL is also reflected by the different incidence in populations with diverse ethnic and geographic distribution. CLL is the most common leukemia among adults in the Western hemisphere (~30% of all leukemias), but is rare in Asian countries [6, 7]. The age-adjusted incidence rate (AAIR) is about 4.4 new cases x 100.000 individuals/year in the US (2007-2011) [8]. In contrast, the AAIR in Asians is about 10 times lower, with 0.2-0.3 new cases x 100.000 individuals/year [8, 9]. The

reasons for this epidemiologic heterogeneity remain to be defined. The risk of developing CLL does not change among Asians residing in the US and in their descendants [10, 11]. Therefore, the genetic background is important in determining the risk of the disease, beyond potential environmental factors.

There are some data on the clinico-biologic features of Asian CLL [12-14]. The prognostic impact of CD38 [15], ZAP-70 [16], genetic lesions evaluated by FISH [17, 18] and of other biologic markers such as CLLU1 [19], lipoprotein lipase [20], serum thymidine-kinase [21] seems confirmed also in Asians CLL. On the contrary, a different clinical presentation, a younger age at diagnosis and a worse prognosis have been reported [22]. Furthermore, some differences between Chinese and European CLL in

the incidence and prognostic impact of mutations in driver genes (*SF3B1*, *NOTCH1*, *MYD88*, *BIRC3*, *TP53*) have been suggested [23].

To date, little is known on the IGHV gene representation and mutations in Asian compared to Caucasian CLL. Indeed, some differences have so far been reported only in small cohorts of patients [24-27]. The B-cell receptor (BCR) repertoire in CLL is influenced by the genetic susceptibility and/or by the existence of a promoting pressure derived from different antigenic elements. The potential role of antigens in the pathogenesis of CLL is further supported by the presence of the non-stochastic pairing of IGHV, IGKV and IGLV genes, which leads to nearly identical BCR sequences of heavy chain complementarity determining region 3 (HCDR3) region, described as “stereotyped” BCR, in about 30% of Caucasian CLL [28].

In this study, we investigated IGHV gene usage, mutational status and HCDR3 in the largest series of Chinese CLL so far reported, and compared them to those present in Italian CLL. Despite general analogous features, Chinese and Italian CLL showed significant differences in IGHVDJ gene usage, mutations and stereotypy. These findings raise the hypothesis that diverse specific antigens are implicated in the pathogenesis of Asian CLL.

RESULTS

Italian CLL: IGHV gene usage, mutation analysis and HCDR3 features

A total of 792 IGHVDJ productive rearrangements from 789 Italian CLL cases were evaluated; 2 cases carried double and triple in-frame rearrangements, respectively. Using the 2% cut-off to discriminate between mutated and unmutated rearrangements, 387 (49%) of the 792 sequences analyzed had more than 2% difference from the most similar germline gene and were classified as mutated with a median of 7.4% (range 2-17%) mutations/case. The remaining 405 (51%) were considered as unmutated, of which 278 (68%) displayed a 100% sequence identity to the corresponding germline IGHV gene sequences (Supplementary Table S1A). The IGHVDJ subgroup and gene usage were for the most part comparable to those reported in previous studies [28-31], showing no relevant differences in the frequency of the most representative genes. As expected, the analysis revealed that the most common IGHV subgroup was IGHV3 (364; 46.0%), followed by IGHV1 (199; 25.1%) and IGHV4 (163; 20.6%) (Supplementary Table S1B).

When we examined the IGHV genes, the most frequently encountered were IGHV1-69 (110; 13.9%), IGHV4-34 (77; 9.7%), IGHV3-23 (51; 6.4%), IGHV3-30 (46; 5.8%), IGHV1-2 (44; 5.6%), IGHV3-7 (42; 5.3%), IGHV3-11 (31; 3.9%), IGHV3-21 (30; 3.8%), IGHV3-48 (28; 3.5%), IGHV4-39 (27; 3.4%), IGHV3-33 (27;

3.4%); collectively, these genes accounted for 65% of the cases, in line with the reported features of Caucasian CLL (Supplementary Table S1B). In our cohort, we also confirmed the known low frequency of the IGHV3-21 gene in CLL patients of Mediterranean origin, with one third of cases being mutated [29].

As expected, the IGVH1-69 gene was over-represented in the unmutated group ($p=0.0001$), whilst IGHV4-34 ($p=0.0001$), IGHV3-7 ($p=0.0001$) and IGHV3-23 ($p=0.0001$) were over-represented in the mutated group.

Among the 792 sequences, IGHD genes were identified in 774 cases and IGHD3 was the most frequently represented (337; 43.5%) followed by IGHD2 (135; 17.4%) and IGHD6 (107; 13.8%). The IGHD3-3 (133; 17.2%) gene was the most frequently used followed by IGHD3-22 (82; 10.6%) (Supplementary Table S1C).

With regard to the IGHJ gene usage, the most frequent subgroup was IGHJ4 (332; 41.9%), followed by IGHJ6 (264; 33.3%) and IGHJ5 (89; 11.2%) (Supplementary Table S1D). Moreover, IGHJ4 recurred in mutated sequences ($p=0.0001$), whilst IGHJ6 was recurrent in unmutated rearrangements ($p=0.0001$).

We also confirmed the presence of specific IGHV, IGHD and IGHJ gene combinations, in particular, for the IGHV1-69 gene the strongly biased association with IGHD3-3 (32/110; 29%) ($p=0.0005$) and IGHJ6 (63/110; 57%) segments ($p=0.0001$).

The median length of HCDR3 was 19 amino acids (range, 7-35). Unmutated sequences had significantly longer HCDR3s than mutated sequences: median length 22 versus 17 amino acids ($p=0.0001$).

Cluster analysis of the HCDR3 region allowed to recognize 204 of 792 cases (25.8%) with stereotyped BCRs (Supplementary Table S1E-S1F) belonging to 72 different subsets. We confirmed a significantly higher frequency of unmutated (153/405; 37.7%) than mutated (51/387; 13.2%) stereotyped sequences ($p=0.0001$). Of the major stereotyped subsets of Caucasian CLL, the most frequent in the present Italian series were subset #1 (20.9%), #4 (11.3%) and #2 (10.4%).

The most represented IGHV genes in stereotyped HCDR3 were: IGHV1-69 (#3, #5, #6, #7), IGHV4-34 (#4) and IGHV3-21 (#2).

Moreover, we identified 10 (4.9%) sequences that fell into 5 groups with identical HCDR3 and thus were defined as paired clusters.

Chinese CLL: IGHV gene usage, mutation analysis and HCDR3 features

A total of 631 productive IGHVDJ rearrangements were sequenced from the cohort of 623 Chinese CLL patients; 8 (1.3%) cases showed biallelic IGHVDJ rearrangements. A total of 415 of the 631 sequences

analyzed (66%) were defined as mutated, with a median of 7% (range 2-22%) mutations/case, whereas 216 (34%) sequences had unmutated IGHV genes, with 121 (56%) in 100% germline configuration (Supplementary Table S2A).

Analysis of the IGHV subgroups showed a higher usage of IGHV3 (295; 46.8%) followed by IGHV4 (181; 28.7%) and IGHV1 (98; 15.5%) (Supplementary Table S2B). The most frequent IGHV genes were IGHV4-34 (70; 11.1%), IGHV3-23 (53; 8.4%), IGHV3-7 (51; 8.1%), IGHV4-39 (43; 6.8%), IGHV1-69 (36; 5.7%), IGHV4-59 (30; 4.8%), IGHV3-30 (27; 4.3%), IGHV3-74 (23; 3.7%), IGHV1-3 (23; 3.7%) and IGHV3-48 (22; 3.5%). Overall, these genes were found in 60% of the analyzed cases. Cumulative results are summarized in Supplementary Table S2B. The known biased usage was found, with IGVH1-69 ($p=0.0001$) gene over-represented in the unmutated group, and IGHV4-34 ($p=0.0001$) and IGHV3-7 ($p=0.003$) in the mutated group. IGHV3-21 represented the eleventh group in frequency (19; 3%), with two thirds being mutated (mutated, $n=13$; unmutated, $n=6$).

IGHD genes were identified in 622 of 631 productive rearrangements. Overall, the most frequently used subgroup was IGHD3 (239; 38.4%), followed by IGHD6 (122; 19.6%) and IGHD2 (91; 14.6%) (Supplementary Table S2C). The most prevalent IGHD genes were IGHD3-10 (74; 11.9%) and IGHD3-22 (59; 9.5%).

Regarding IGHJ gene usage, among Chinese patients there was an over-representation of the IGHJ4 (289; 45.8%) and IGHJ6 (161; 25.5%) genes (Supplementary Table S2D). In addition, IGHJ4 recurred in mutated rearrangements ($p=0.0001$), whilst IGHJ6 was recurrent in unmutated rearrangements ($p=0.0028$). Focusing on IGHVDJ rearrangements, we identified a biased association of IGHV1-69 with IGHD3-3 and JH6 gene, similarly to Caucasian CLL, although a significant correlation was observed only with IGHJ6 ($p=0.016$).

The HCDR3 median length was 17 amino acids (range, 7-35). Significantly longer HCDR3s were observed in unmutated versus mutated sequences (median lengths, 21 versus 16 amino acids; $p=0.0001$). The frequency of a stereotyped HCDR3 was significantly higher among unmutated (72/216; 33.3%) than mutated (52/415; 12.5%) sequences ($p=0.0001$). We identified 124 stereotyped HCDR3 out of 631 cases (19.7%), belonging to 48 different subsets, of which 16 (12.9%) have not been reported previously and were named as novel clusters (see below). The major stereotyped subsets of Caucasian CLL collectively represent a substantial proportion of stereotyped sequences also in the Chinese series, with a preferential recurrence of subset #8 (27.3%), #1 (21.2%), #4 (16.7%), while the subset #2 (1.5%) appeared to be less represented (Supplementary Table S2E).

By comparing the IGHV usage of stereotyped sequences with those not belonging to any cluster, the IGHV1 ($p=0.03$) and IGHV4 ($p=0.0004$) subgroups were

the most frequently utilized in stereotyped sequences, whereas IGVH3 was preferentially employed by non-homologous sequences ($p=0.0001$). Indeed, the most frequent stereotyped IGHV genes were: IGHV4-34 (#4) IGHV4-39 (#8) and finally IGHV1-69 (#3, #5, #6, #7).

Regarding the 16 novel clusters (corresponding to 33 sequences), two were potential new subsets (composed of 3 Chinese CLL and 2 Chinese plus 2 Italian CLL, respectively) and 14 were paired clusters (all Chinese), as shown in Supplementary Table S2F. In contrast to common subsets, the majority of novel clusters were mutated (19/33 sequences; 57.5%). Of interest, 15 of 19 novel mutated sequences expressed genes from the IGHV3 subgroup, of which IGHV3-23 (8/33; 24.2%) was the most represented gene.

Comparison between Italian and Chinese CLL reveals differences in IGHV mutational status and gene repertoire

Comparing the two series, Chinese CLL showed a higher proportion of cases with mutated IGHV than Italian CLL. Such difference appeared to be statistically significant - mutated IGHV 66% and unmutated IGHV 34% vs mutated IGHV 49% and unmutated IGHV 51%, respectively ($p=0.0001$) - and remained significant also after exclusion of pre-treated cases ($p=0.0001$).

Whilst the IGHV3 subgroup was predominantly expressed both in Chinese and Italian patients, there was an inversion between IGHV1 and IGHV4 frequency. In fact, the frequency of IGVH1 in Chinese CLL resulted lower ($p=0.0001$), while the IGHV4 subgroup was significantly higher ($p=0.0004$) (Figure 1A).

In line with these results, the IGHV gene usage highlighted differences in the 10 most frequent genes of Italian and Chinese CLL. Overall, the 10 most frequent IGHV genes employed by Italian CLL represented 50% of the IGHV genes used by the Chinese CLL, showing a significant difference ($p=0.0001$). In fact, IGHV1-69 ($p=0.0001$) and IGHV1-2 ($p=0.0002$) were significantly under-represented in Chinese CLL, whilst IGHV3-7 ($p=0.04$) and IGHV4-39 ($p=0.004$) were more recurrent among Chinese than Italian patients. In addition, the IGHV4-59 ($p=0.01$), IGHV1-3 ($p=0.02$) and IGHV3-74 ($p=0.04$) genes were most typically used by Chinese CLL (12%) but not by the Italian cohort of CLL (5%) ($p=0.0001$). IGHV3-21 showed a comparable frequency (3% Chinese vs 3.8% Italian CLL) (Figure 1B).

The known pattern of mutations for each IGHV gene was largely conserved in both Chinese and Italian CLL. In fact, the IGHV1-69 gene carried few or no mutations, while IGHV4-34, IGHV3-7 and IGHV3-23 showed a high frequency of mutations in both cohorts. Contrariwise, IGHV3-21 ($p=0.002$) and IGHV1-2 ($p=0.03$) genes showed a higher frequency of mutations among Chinese than Italian CLL (Figure 2). In particular, IGHV3-21

showed a median of 9.4% (range 2.5-18.9%) mutations in Chinese CLL vs 4% (range 2.5-9%) in Italian CLL.

Also the comparison of IGHD subgroups and genes allowed to identify additional differences between the two CLL cohorts. The IGHD6 subgroup, the second most frequent after the IGHD3 family in Chinese CLL, was significantly increased compared to Italian CLL ($p=0.004$). Similarly, there was a significant over-usage of IGHD3-10 ($p=0.002$), IGHD6-13 ($p=0.0001$) and IGHD2-21 ($p=0.03$) genes and under-usage of IGHD3-3 ($p=0.0001$) and IGHD2-2 ($p=0.02$) genes in Chinese CLL.

As for the IGHJ subgroup, although the distribution was similar among the two populations, IGHJ6 ($p=0.002$) was significantly under-represented in the Chinese series.

Comparative analysis of the HCDR3 sequences revealed notable differences. The proportion of known stereotyped receptors was significantly lower in Chinese (19.7%) than in Italian CLL (25.8%) ($p=0.007$)

(Figure 3A). In particular, within the major subsets, a significantly higher frequency of subset #8 ($p=0.0001$) and lower frequency of subset #2 ($p=0.03$) were observed in the Chinese cohort (Figure 3B). These features are due to the different frequencies of the stereotyped genes, given that most IGHV genes showed the same propensity to be used in stereotyped rearrangements in both series (Figure 4). In fact, the most represented stereotyped genes in Italian CLL were IGHV1-69> IGHV4-34> IGHV1-2> IGHV3-21> IGHV3-30> IGHV4-39, whilst being IGHV4-34> IGHV4-39> IGHV1-69> IGHV1-3> IGHV3-23> IGHV4-59 in Chinese CLL. In particular, the stereotyped HCDR3 sequences expressing IGHV1-69 and IGHV3-21 genes were significantly less represented in Chinese CLL than Italian CLL ($p=0.0006$ and $p=0.0003$, respectively); indeed, the IGHV3-21 gene exhibited a higher heterogeneity within its HCDR3 sequence in Chinese than Italian CLL (Figure 4).

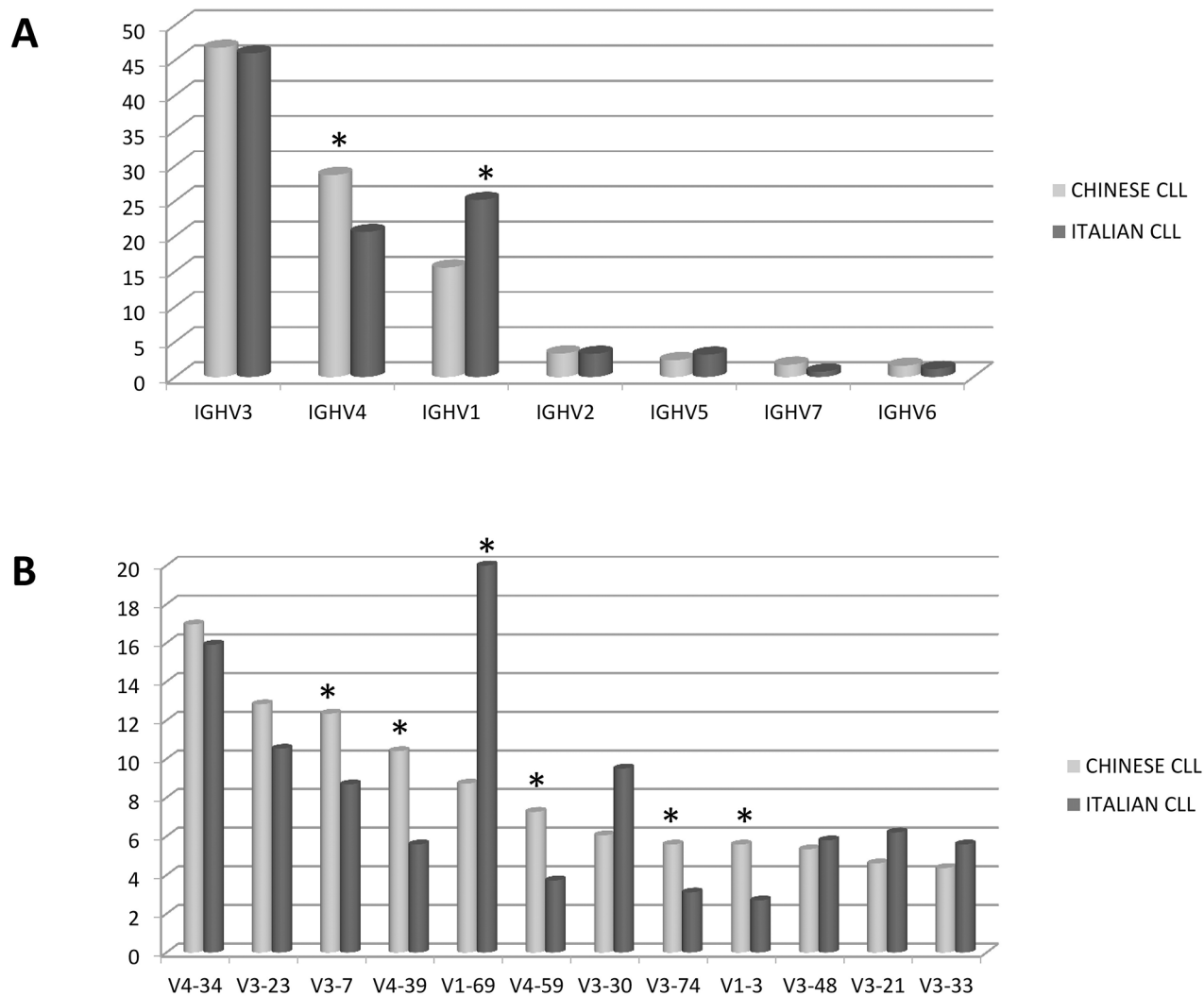


Figure 1: Comparison between Chinese and Italian CLL. A. Frequency of the IGHV subgroups. B. Frequency of the IGHV genes. * marks the significant differences (p values in the text). Data are represented as percentage.

Clinical impact of IGHV features in Chinese CLL

Prognosis was evaluated on 612 Chinese CLL patients, with a median follow-up of 28 months (range: 0-270.9). The median overall survival (OS) was 121.7 months for 566 evaluable cases. Treatment-free interval (TFI) from diagnosis was 18.4 months (range: 0-194.8) for 580 evaluable cases, 327 of whom required treatment. Richter's transformation occurred in 16 of 575 evaluable cases (2.8%).

CLL in Binet stages A, B or C had a median OS of 165.1, 98 and 82.2 months, respectively ($p=0.0001$).

Chinese CLL with unmutated IGHV showed a poor prognosis in terms of OS (85.1 vs 165.1 months,

$p<0.0001$) and TFI (4.0 vs 46.1 months, $p<0.0001$) compared to mutated IGHV cases, as expected.

As reported for Caucasian CLL, subset #1 Chinese CLL ($n=14$) showed a poor prognosis in terms of OS (median OS 28.0 months vs 85.1 for unmutated IGHV vs 165.1 for mutated IGHV, $p<0.0001$) and TFI (median TFI 0.8 months vs 4.8 for unmutated IGHV vs 46.2 for mutated IGHV, $p<0.0001$), even worse than that of the other IGHV unmutated CLL.

Subset #8 CLL ($n=18$) showed an OS (median 82.2 months) and TFI (median 1.7 months) as poor as the other IGHV unmutated CLL, although there was a significant association with the occurrence of Richter's syndrome ($n=3$ out of 16) ($p=0.0095$), as reported for Caucasian CLL.

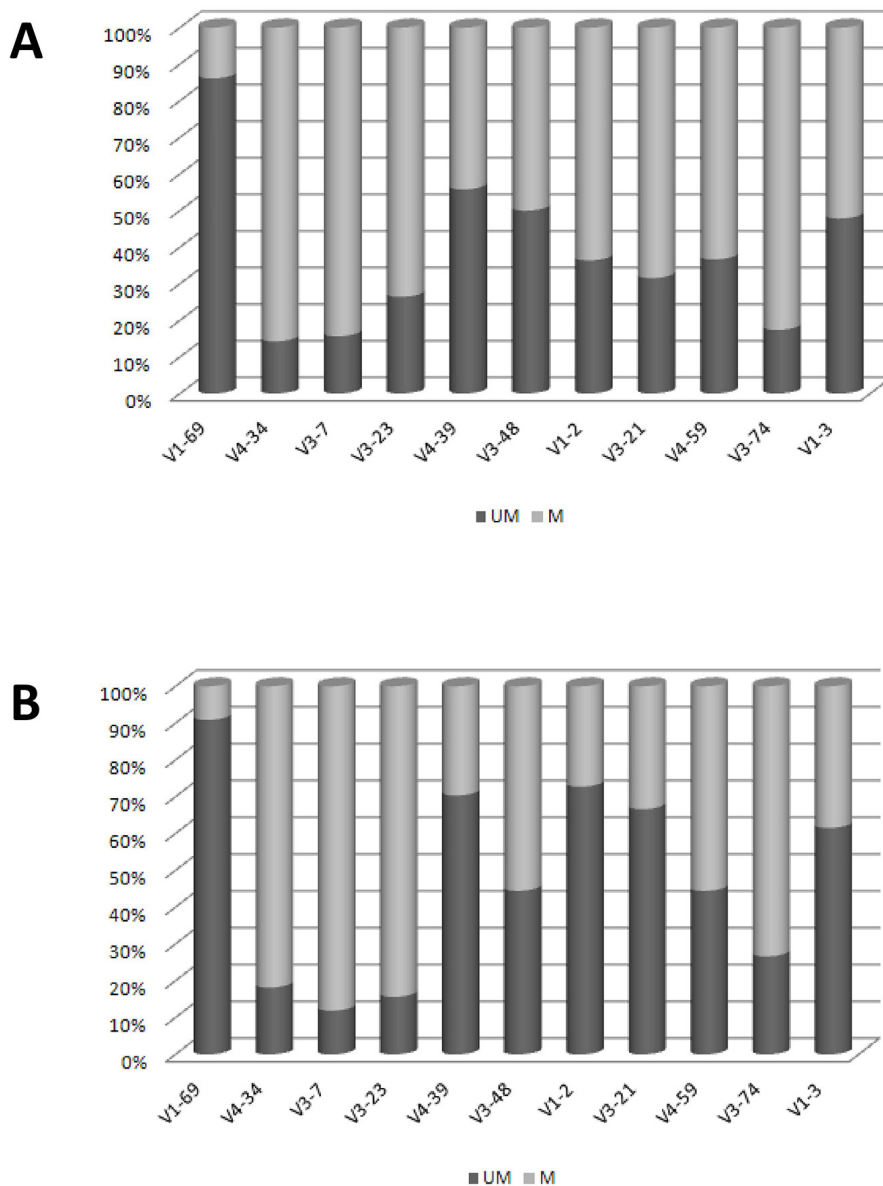


Figure 2: Mutational Patterns of IGHV genes. A. Chinese CLL. B. Italian CLL. Data are represented as percentage.

The low number of events did not allow to analyze the prognostic impact of subset #4 (n=11).

Finally, the OS and TFI of Chinese patients with IGHV3-23 were not different from the other cases with mutated IGHV ($p= 0.48$ and $p= 0.41$, respectively). The low number of events did not allow to highlight a different outcome between stereotyped IGHV3-23 (n=7) and heterogeneous IGHV3-23 (n=32) cases.

DISCUSSION

In Caucasian CLL, the extensive research on IGHVDJ rearrangements has shown that immunoglobulin gene usage is not random, but biased by the recurrence of certain IGHV genes as well as by the existence of subsets with stereotyped BCR [31, 32]. In contrast, little is known about the IGHV features of Asian CLL, although

some differences have been reported in small cohorts of Chinese and Japanese patients [24, 25]. In order to define geographic and ethnic immunoglobulin gene variations, we characterized IGHVDJ rearrangements and HCDR3 stereotypy in a large series of Chinese CLL and compared them with Italian CLL. Our Italian cohort is representative of Caucasian CLL, being comparable with other Italian [30] and Western series [28].

A mutated IGHV status was more frequent in Chinese than in Italian CLL, thus consolidating previous observations made in small Asian series [24, 25]. This holds true also when considering patients in homogeneous phases of disease.

With regard to the IGHV subgroups, Chinese CLL showed an IGHV3>IGHV4>IGHV1 order of frequency, significantly different from Italian CLL patients, who showed an IGHV3>IGHV1>IGHV4 distribution. These

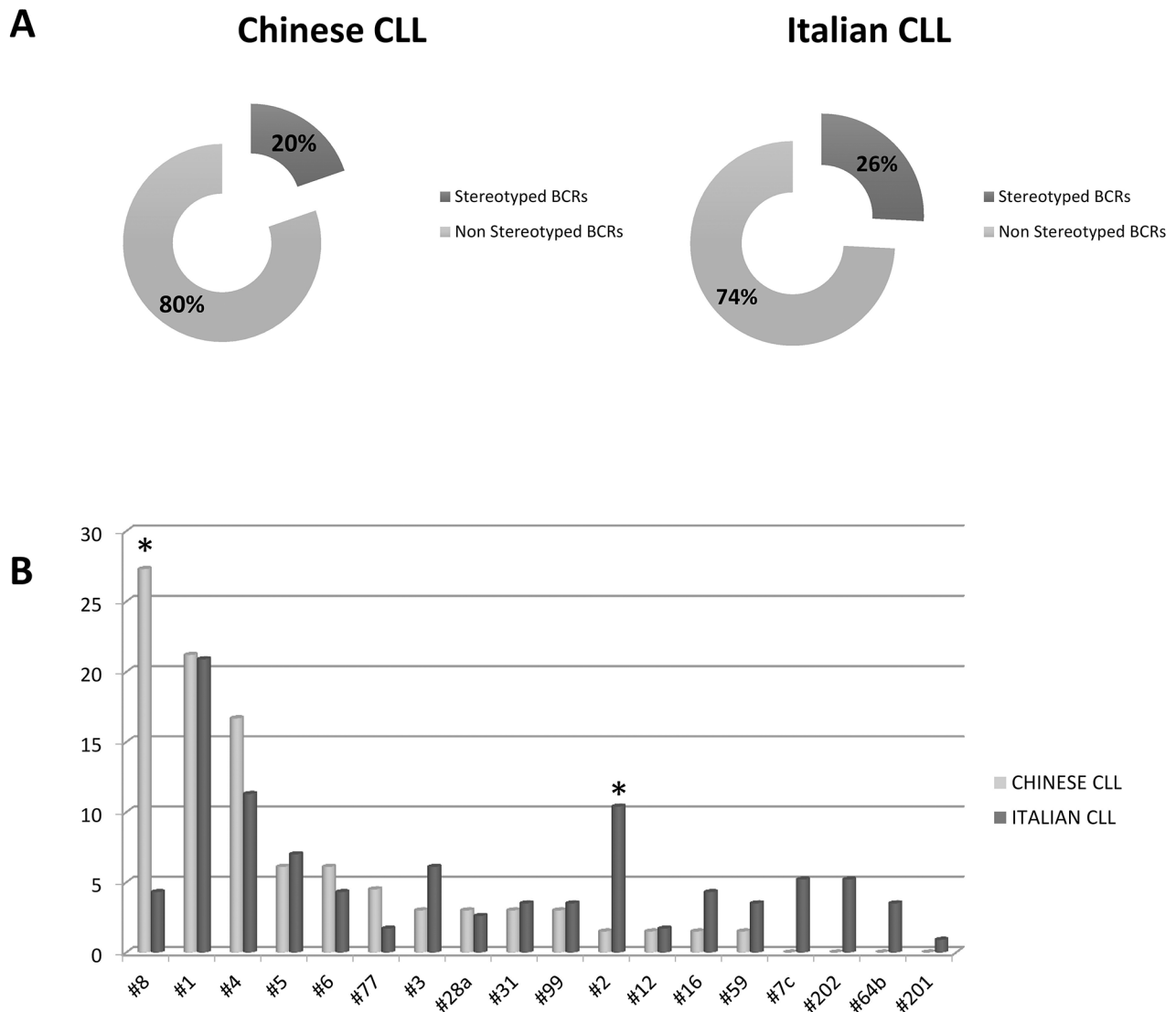


Figure 3: Stereotyped BCRs representation. A. Chinese and Italian stereotyped and non-stereotyped BCRs. **B.** Frequency of Major subsets in Chinese and Italian series. * marks the significant differences (p values in the text). Data are represented as percentage.

results support those reported on both Chinese [24] and Japanese [25, 33] CLL, confirming that this family usage is typical of Asian CLL.

Focusing on the IGHV gene distribution, also Chinese CLL showed a non-random use of immunoglobulin genes, with the IGHV4-34 being the most represented. Moreover, IGHV1-69 and IGHV1-2 were significantly under-represented in Chinese CLL, whilst IGHV4-39 and IGHV3-7 were more recurrent in Chinese than Italian CLL. Furthermore, the IGHV4-59, IGHV3-74 and IGHV1-3 genes were common in Chinese CLL, whilst rare in Italian CLL.

The known pattern of mutations for each specific IGHV gene was largely conserved, i.e. the IGHV1-69 gene shows analogous molecular features in Caucasian and in Chinese CLL, being unmutated and representing the most frequent gene of the IGHV1 subgroup [28, 34]. Thus, the low recurrence of the IGHV1-69 gene in Chinese CLL largely accounts for the lower overall frequency of the IGHV1 subgroup and of unmutated cases. On the other hand, the high incidence of the IGHV3-7, IGHV4-39 and IGHV3-

74 genes could be responsible of the high representation of mutated IGHV, since these genes often show a mutated profile. Relevant exceptions are represented by the IGHV1-2 gene, that was significantly more mutated in Chinese than in Italian CLL, and the IGHV3-21 gene. The latter, equally represented in Chinese (3%) and Italian CLL (3.8%), was more frequently mutated in China.

Our results reinforce those reported in other small Asian CLL series [24, 25], especially with regard to IGHV1-69 and IGHV3-7, strongly supporting the contribution of ethnic and geographic parameters in shaping the IGHV repertoire. Moreover, the same pattern is also reported in Iranian CLL [27], suggesting that this geographic-related variation is already established in the Middle-East. Thus, this different pattern between Eastern and Western CLL adds to the other known geographic variation related to the IGHV3-21 gene, over-represented in Northern European CLL compared to Mediterranean CLL [29].

We also identified additional differences in the frequencies of IGHD and IGHJ genes. Among others, a

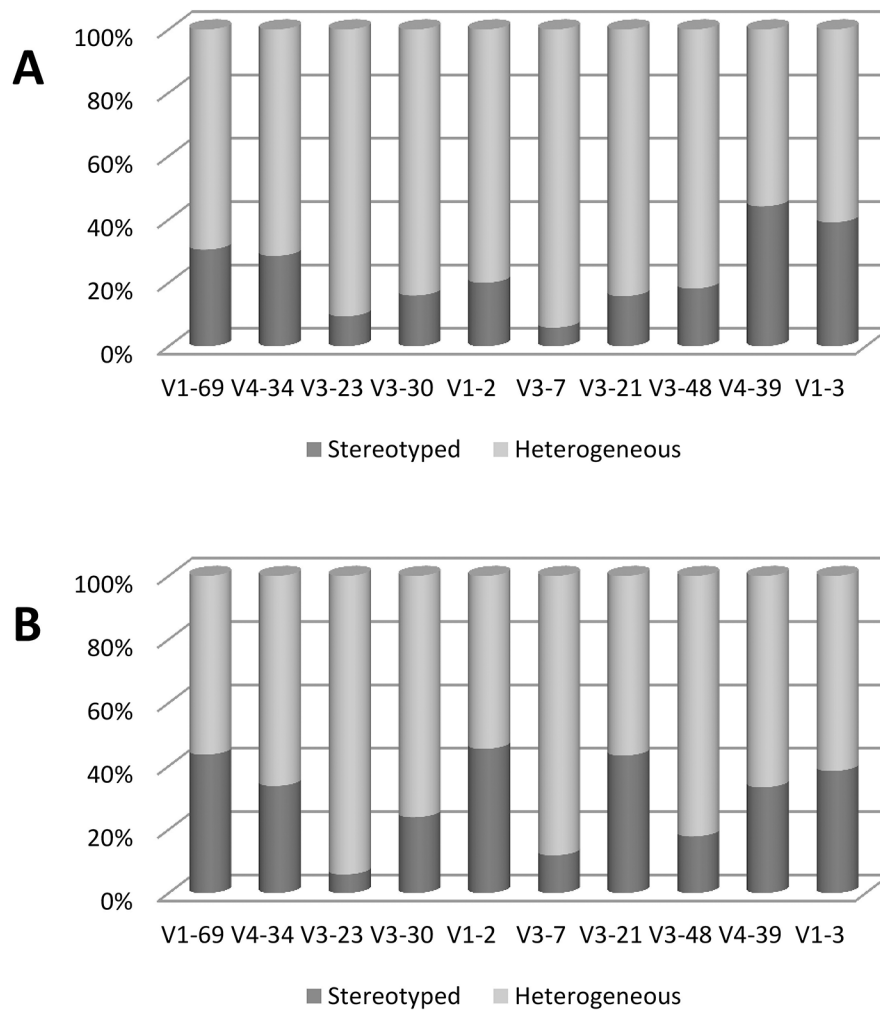


Figure 4. Frequency of stereotyped and heterogeneous BCRs for selected IGHV genes. A. Chinese CLL. B. Italian CLL. Data are represented as percentage.

significant under-usage of IGHD3-3 and IGHJ6 genes was observed in the Chinese cohort. Given the specific combination of IGHV1-69 with the IGHD3-3 and IGHJ6 genes in Caucasian CLL [28, 34], their low occurrence in Chinese CLL could be related to the low frequency of IGHV1-69. To our knowledge, these results have never been previously reported.

The analysis of HCDR3 sequences proved the presence of stereotyped BCR in one-third of Caucasian CLL [28, 30, 31], suggesting the recognition of a common antigenic determinant [35]. We therefore characterized the HCDR3 sequences in Chinese CLL, which have not been investigated so far, for their potentially relevant pathogenetic implications. Chinese CLL showed a significantly lower representation of stereotyped sequences (19.7%) compared to Italian CLL (25.8%). The stereotyped subsets defined major among Caucasian CLL collectively were over-represented also in the Chinese cohort, but with a significantly higher frequency of subset #8 (China vs Italy: 27.3% vs 4.3%) and a lower frequency of subset #2 (China vs Italy: 1.5% vs 10.4%). Subset #8, the most represented Chinese stereotype, is characterized by the use of the IGHV4-39 gene, which is indeed one of the most recurrent gene in Chinese CLL. On the contrary, the low frequency of subset #2 is due to the heterogeneous HCDR3 of IGHV3-21 in Chinese CLL, whilst in Caucasian CLL the IGHV3-21 gene is mostly stereotyped [29]. The low frequency of subset #2 in Chinese CLL could parallel the recently reported low incidence of *SF3B1* mutations [23], in line with their reported association in Caucasian CLL [36]. Of note, the clinical impact of CLL stereotypy is independent of the ethnic context, since in Chinese CLL subset #1 displayed a poor prognosis and subset #8 was related to an increased risk of transformation into Richter syndrome, similarly to Caucasian CLL [37, 38].

With the exception of the IGHV3-21 gene, it is worth noting that all the other IGHV genes (i.e. IGHV1-69, IGHV4-39, etc) maintain the same proportion of homologous/heterologous BCR both in Chinese and in Italian CLL. Thus, the determinant of the lower frequency of stereotyped BCR in Chinese CLL could be due to the prevalence of certain genes (i.e. IGHV3-7 and IGHV3-74) and the reduced frequency of others (i.e. IGHV1-69), each maintaining the same tendency to stereotypy.

By clustering analysis, we detected 16 new paired clusters among Chinese cases, characterized by the remarkable recurrence of the IGHV3-23 gene; contrariwise, in Caucasian CLL this gene constantly exhibits an heterogeneous HCDR3 [39]. Moreover, IGHV3-23 gene usage is considered an independent negative prognosticator within Caucasian mutated CLL [39], but this was not the case among Chinese CLL.

In line with the proposal of Ghia et al. for Caucasian CLL [29], also in Chinese CLL the IGHV pattern may be due to either a specific genetic background and/or to the effects of potential environmental variables. The observation of new paired clusters in Chinese CLL, never

reported in Caucasian CLL, supports the hypothesis of an Asian-specific antigenic selection at least in some cases. To further investigate the importance of the genetic background in determining the occurrence of different rearrangements in distinct geographic and/or ethnic groups, the normal IGHV repertoire of healthy subjects of the same ethnicity and from the same geographic areas should also be investigated. Although some evidences are reported for Italians [40], at present no data on the normal repertoire are available for Chinese population.

In conclusion, the results hereby described offer the most extensive catalogue of the BCR features of Chinese CLL, being based on the largest series of Chinese CLL so far reported. In comparison to Italian CLL, most of the Chinese IGHV genes show different frequencies, but maintain the same propensity towards mutations and stereotypy, with the same clinical impact. The distinctive molecular features of IGHV3-21 and IGHV3-23 Chinese CLL genes deserve further investigation.

MATERIALS AND METHODS

Patients

A series of 623 patients diagnosed with CLL from 3 different institutions situated in China - Nanjing (338 cases) and Tianjin (166 cases) - and Hong Kong (119 cases), were included in the study. A comparison of the three CLL groups, all of Han Chinese ethnicity, showed a high degree of internal similarities with regard to the somatic hypermutation status and IGHVDJ rearrangements. Therefore, the whole cohort of Chinese CLL was compared to an Italian cohort of 789 CLL patients collected between 2001 and 2014 at the Hematology Institute of the "Sapienza" University of Rome. The Italian CLL cohort was comparable to the Mediterranean series [29,30], and to those reported by Agathangelidis et al [28] (Supplementary Table S1).

All satisfied the morphologic and immunophenotypic diagnostic criteria for CLL [41]. The characteristics of the Chinese and Italian patients are summarized in Table 1. Chinese CLL appear older than Italians ($p < 0.001$) and in more advanced stages (more stage C) ($p = 0.0194$). Regarding FISH lesions, they show less 11q- ($p = 0.0082$) and 13q-only ($p = 0.0026$), more trisomy12 ($p = 0.0273$) and more 17p- ($p = 0.0457$) than Italian CLL. This is in contrast with the disease phase at study, which included more pre-treated patients in the Italian series. After excluding pre-treated patients from both cohorts, we confirm all the above mentioned differences (age: $p = 0.0001$; stage C: $p = 0.0001$; 11q-: $p = 0.012$; 13q only: $p = 0.028$; 17p-: $p = 0.0108$), with the exception of trisomy 12 ($p = 0.3$). No difference was observed in the WBC count: $27.5 \times 10^9/L$ (range $2.1-452.8 \times 10^9/L$) in Chinese CLL and $30.3 \times 10^9/L$ (range $1.7-622.9 \times 10^9/L$) in Italian CLL ($p = 0.52$).

Table 1: Clinical and biological features of Chinese and Italian CLL patients (* cases with available clinical information)

	Chinese CLL (N=611*)	Italian CLL (N=789)	Chinese CLL untreated (N=536)	Italian CLL untreated (N=595)
Sex, M/F	412/190*	520/258*	362/171*	387/199*
Median age, y (range) at diagnosis	60 (16-94)	55 (26-87)	60 (16-94)	55 (26-87)
Stage at study	N=575	N=546	N=517	N=461
A	245 (42.6%)	271 (49.6%)	232	251
B	141 (24.5%)	181 (33.2%)	127	144
C	187 (32.5%)	94 (17.2%)	156	66
Richter	2 (0.4%)	-	2	-
Disease phase at study	N=594	N=729	N=536	N=595
Diagnosis	290 (48.8%)	304 (41.7%)	290	304
First progression	246 (41.4%)	291 (39.9%)	246	291
Subsequent progression (post- treatment)	58 (9.8%)	134 (18.4%)	-	-
FISH	N=373	N=628	N=329	N=535
del17p	45 (12.1%)	51 (8.12%)	38	34
del11q	38 (10.2%)	102 (16.24%)	31	82
Tris12	68 (18.2%)	81 (12.9%)	50	68
del13q only	98 (26.3%)	223 (35.51%)	101	204
Normal FISH	124 (33.2%)	171 (27.23%)	109	147
IGHV	N=611	N=789	N=536	N=595
Germline	214	402	167	283
Mutated	397	387	369	312

Analysis of immunoglobulin rearrangements and sequence analysis

In all cases, the analysis of IGHVDJ genes was carried out on leukemic cells obtained from peripheral blood samples after isolation by Ficoll gradient or buffy-coat. PCR amplification and sequence analysis of IGHVDJ rearrangements were performed on either genomic DNA (gDNA) or cDNA using sense family-specific VH primers (framework region 1 [FR1] or VH leader primers), combined with consensus JH primers as previously described [42] or following the IGH Somatic Hypermutation Assay v2.0 protocol (InVivoScribe) [24]. PCR products were sequenced directly or after a cloning procedure, using 3130 Genetic Analyzer (Life Technologies, Carlsbad, CA).

Productive rearrangements were analyzed by IMGT database (<http://imgt.cines.fr>, Montpellier,

France) [43] and the IMGT/V-QUEST tool (version 3.3.0) [44].

The following features were evaluated for all IGHVDJ rearrangements: IGHV gene and allele usage, percentage of identity to the closest germline IGHV allele, HCDR3 length and composition calculated between codons 107 and 117, and IGHD-IGHJ gene usage.

To identify clusters of sequences with common HCDR3 motifs, we evaluated the HCDR3 region by the multiple sequence alignment ClustalW2 software (<http://www.ebi.ac.uk>), followed by a manual curation. Clustering was performed comparing our HCDR3 sequences to those present in literature's databases [28, 45, 36].

To identify subsets of similar HCDR3, we used several criteria. First, for major subsets we followed the criteria proposed by Agathangelidis et al. [28] : *i*) sharing at least 50% amino acid identity and 70% similarity calculated through common sequence patterns; *ii*) having

identical HCDR3 lengths and identical offsets of shared patterns between sequences; and *iii*) carrying IGHV genes of the same clan. Second, we identified homologous HCDR3 as those which shared a HCDR3 homology equal to or exceeding 60%, regardless of the usage of the IGHV gene: groups of 2 CLL cases with homologous HCDR3 were defined as paired clusters; clusters of 3 or more cases were defined as subsets. Known stereotyped HCDR3s were defined and named according to published criteria: for major subsets by Agathangelidis et al. [28]; for minor subsets by Murray et al. [45] and for novel subsets by Rossi et al. [36].

Statistical analysis

Characteristics of patients were summarized by means of cross-tabulations (categorical variables), quantiles (median etc; for ordinal factors) or by means of standard positional and variation parameters (mean, standard deviation; for continuous variables). Non-parametric tests were applied, in univariate analysis, for comparisons between groups (Chi-Squared and Fisher Exact test for difference in terms of categorical variables or response rate, Mann-Whitney and Kruskal-Wallis test for difference in terms of continuous variables).

Overall survival was estimated using the Kaplan-Meier Product Limit estimator. Differences were evaluated by means of Log-Rank test after assessment of proportionality of hazards.

TFI was estimated using the proper non-parametric method; the Gray test was applied for significance tests on cumulative incidence curves.

All the analyses were performed using the SAS system software (version 9.4); all tests were two-sided, at a significance level of 0.05 and confidence intervals were calculated at 95% level.

ACKNOWLEDGMENTS

This study was supported by Associazione Italiana per la Ricerca sul Cancro (AIRC), Special Program Molecular Clinical Oncology, 5 x 1000, Milan, Italy.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

1. Chiorazzi N, Rai KR, Ferrarini M. Chronic lymphocytic leukemia. *The New England Journal of Medicine*. 2005; 352: 804–815.
2. Hamblin TJ, Davis Z, Gardiner A, Oscier DG and Stevenson FK. Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. *Blood*. 1999; 94: 1848–1854.
3. Damle RN, Wasil T, Fais F, Ghiotto F, Valetto A, Allen SL, Buchbinder A, Budman D, Dittmar K, Kolitz J, Lichtman SM, Schulman P, Vinciguerra VP, et al. Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. *Blood*. 1999; 94: 1840–1847.
4. Del Poeta G, Maurillo L, Venditti A, Buccisano F, Epiceno AM, Capelli G, Tamburini A, Suppo G, Battaglia A, Del Principe MI, Del Moro B, Masi M, Amadori S. Clinical significance of CD38 expression in chronic lymphocytic leukemia. *Blood*. 2001; 98: 2633–2639.
5. Crespo M, Bosch F, Villamor N, Bellosillo B, Colomer D, Rozman M, Marcé S, López-Guillermo A, Campo E, Montserrat E. ZAP-70 expression as a surrogate for immunoglobulin-variable-region mutations in chronic lymphocytic leukemia. *The New England Journal of Medicine*. 2003; 348: 1764–1775.
6. Boggs DR, Chen SC, Zhang ZN, Zhang A. Chronic lymphocytic leukemia in China. *American Journal of Hematology*. 1987; 25: 349–354.
7. Chihara D, Ito H, Matsuda T, Shibata A, Katsumi A, Nakamura S, Tomotaka S, Morton LM, Weisenburger DD, Matsuo K. Differences in incidence and trends of haematological malignancies in Japan and the United States. *British Journal of Haematology*. 2014; 164: 536–545.
8. Howlader N, Noone AM, Krapcho M, Garshell J, Miller D, Altekruse SF, Kosary CL, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, et al. SEER Cancer Statistics Review, 1975–2011, National Cancer Institute. Bethesda, MD, http://seer.cancer.gov/csr/1975_2011/, based on November 2013 SEER data submission, posted to the SEER web site, April 2014.
9. Yoon SO, Suh C, Lee DH, Chi HS, Park CJ, Jang SS, Shin HR, Park BH, Huh J. Distribution of lymphoid neoplasms in the Republic of Korea: analysis of 5318 cases according to the World Health Organization classification. *American Journal of Hematology*. 2010; 85: 760–764.
10. Gale RP, Cozen W, Goodman MT, Wang FF, Bernstein L. Decreased chronic lymphocytic leukemia incidence in Asians in Los Angeles County. *Leukemia Research*. 2000; 24: 665–669.
11. Mak V, Ip D, Mang O, Dalal C, Huang S, Gerrie A, Gillan T, Ramadan KM, Toze C, Au WY. Preservation of lower incidence of chronic lymphocytic leukemia in Chinese residents in British Columbia: a 26-year survey from 1983 to 2008. *Leukemia & Lymphoma*. 2014; 55: 824–827.
12. Chan LC, Lam CK, Yeung TC, Chu RW, Ng M, Chow EY, Wickham N, Matutes E. The spectrum of chronic lymphoproliferative disorders in Hong Kong. A prospective study. *Leukemia*. 1997; 11: 1964–1972.
13. Suzuki K, Aoyagi Y, Harada K, Mori M, Yamaguchi H. Clinical investigation of 145 cases of chronic lymphocytic leukemia in Japan. *Oncology Reports*. 1996; 3: 477–482.

14. Kamihira S, Matutes E, Sohda H, Atogami S, Tomonaga M, Catovsky D. Morphological and immunophenotypical characterization of Japanese B-cell lymphocytic leukemia. *Leukemia & Lymphoma*. 1994; 16: 113-119.
15. Xu W, Li JY, Wu YJ, Yu H, Shen QD, Tian T, Li L, Qiu HX. CD38 as a prognostic factor in Chinese patients with chronic lymphocytic leukaemia. *Leukemia Research*. 2009; 33: 237-243.
16. Xu M, Fan L, Miao K, Liu P, Xu W, Li JY. Comprehensive assessment of prognostic factors predicting outcome in Chinese patients with chronic lymphocytic leukemia treated with fludarabine and cyclophosphamide. *Medical Oncology*. 2012; 29: 2102-2110.
17. Qiu HX, Xu W, Cao XS, Zhou M, Shen YF, Xu YL, Sun XM, Liu Q, Wang R, Qiu HR, Wang JS, Li JY. Cytogenetic characterisation in Chinese patients with chronic lymphocytic leukemia: a prospective, multicenter study on 143 cases analysed with interphase fluorescence in situ hybridisation. *Leukemia & Lymphoma*. 2008; 49: 1887-1892.
18. Wu S, Lin C, Huang S, Lee FY, Liu MC, Hou HA, Chen CY, Ko BS, Chou WC, Yao M, Tang JL, Tsay W, Tien HF. Chromosomal abnormalities by conventional cytogenetics and interphase fluorescence in situ hybridization in chronic lymphocytic leukemia in Taiwan, an area with low incidence—clinical implication and comparison between the West and the East. *Annals of Hematology*. 2013; 92: 799-806.
19. Chen L, Li J, Zheng W, Zhang Y, Wu Y, Li L, Qian S, Xu W. The prognostic evaluation of CLLU1 expression levels in 50 Chinese patients with chronic lymphocytic leukemia. *Leukemia & Lymphoma*. 2007; 48: 1785-1792.
20. Xu W, Li JY, Shen QD, Wu YJ, Yu H, Fan L. Expression level of lipoprotein lipase in Chinese patients with chronic lymphocytic leukemia and its correlation with other prognostic factors. *International Journal of Laboratory Hematology* 2009; 31: 552-559.
21. Xu W, Cao X, Miao KR, Qiao C, Wu YJ, Liu Q, Fan L, Li JY. Serum thymidine kinase 1 concentration in Chinese patients with chronic lymphocytic leukemia and its correlation with other prognostic factors. *International Journal of Laboratory Hematology*. 2009; 90: 205-211.
22. Sriphatphiriyakun T, Auewarakul CU. Clinical presentation and outcome of Thai patients with chronic lymphocytic leukemia: retrospective analysis of 184 cases. *Asian Pacific Journal of Allergy and Immunology*. 2005; 23: 197-203.
23. Xia Y, Fan L, Wang L, Gale RP, Wang M, Tian T, Wu W, Yu L, Chen YY, Xu W, Li JY. Frequencies of SF3B1, NOTCH1, MYD88, BIRC3 and IGHV mutations and TP53 disruptions in Chinese with chronic lymphocytic leukemia: disparities with Europeans. *Oncotarget*. 2015; 6: 5426-5434. doi: 10.18632/oncotarget.3101.
24. Chen L, Zhang Y, Zheng W, Wu Y, Qiao C, Fan L, Xu W, Li J. Distinctive IgVH gene segments usage and mutation status in Chinese patients with chronic lymphocytic leukemia. *Leukemia Research*. 2008; 32: 1491-1498.
25. Koiso H, Yamane A, Mitsui T, Matsushima T, Tsukamoto N, Murakami H, Miyawaki S, Nojima Y, Karasawa M. Distinctive immunoglobulin VH gene usage in Japanese patients with chronic lymphocytic leukemia. *Leukemia Research*. 2006; 30: 272-276.
26. Tomomatsu J, Isobe Y, Oshimi K, Tabe Y, Ishii K, Noguchi M, Hirano T, Komatsu N, Sugimoto K. Chronic lymphocytic leukemia in a Japanese population: varied immunophenotypic profile, distinctive usage of frequently mutated IGH gene, and indolent clinical behavior. *Leukemia & Lymphoma*. 2010; 51: 2230-2239.
27. Hojjat-Farsangi M, Jeddi-Tehrani M, Razavi SM, Sharifian RA, Mellstedt H, Shokri F, Rabbani H. Immunoglobulin heavy chain variable region gene usage and mutational status of the leukemic B cells in Iranian patients with chronic lymphocytic leukemia. *Cancer Science*. 2009; 100: 2346-2353.
28. Agathangelidis A, Darzentas N, Hadzidimitriou A, Brochet X, Murray F, Yan XJ, Davis Z, van Gastel-Mol EJ, Tresoldi C, Chu CC, Cahill N, Giudicelli V, Tichy B, et al. Stereotyped B-cell receptors in one third of chronic lymphocytic leukemia: A molecular classification with implications for targeted therapies. *Blood*. 2012; 119: 4467-4475.
29. Ghia P, Stamatopoulos K, Belessi C, Moreno C, Stella S, Guida G, Michel A, Crespo M, Laoutaris N, Montserrat E, Anagnostopoulos A, Dighiero G, Fassas A, et al. Geographic patterns and pathogenetic implications of IGHV gene usage in chronic lymphocytic leukemia: The lesson of the IGHV3-21 gene. *Blood*. 2005; 105: 1678-1685.
30. Bomben R, Dal Bo M, Capello D, Forconi F, Maffei R, Laurenti L, Rossi D, Del Principe MI, Zucchetto A, Bertoni F, Rossi FM, Bulian P, Cattarossi I, et al. Molecular and clinical features of chronic lymphocytic leukaemia with stereotyped B cell receptors: Results from an Italian multicentre study. *British Journal of Haematology*. 2009; 144: 492-506.
31. Stamatopoulos K, Belessi C, Moreno C, Boudjogh M, Guida G, Smilevska T, Belhoul L, Stella S, Stavroyianni N, Crespo M, Hadzidimitriou A, Sutton L, Bosch F, et al. Over 20% of patients with chronic lymphocytic leukemia carry stereotyped receptors: Pathogenetic implications and clinical correlations. *Blood*. 2007; 109: 259-270.
32. Kostareli E, Gounari M, Agathangelidis A, Stamatopoulos K. Immunoglobulin gene repertoire in chronic lymphocytic leukemia: insight into antigen selection and microenvironmental interactions. *Mediterranean Journal of Hematology and Infectious Diseases*. 2012; 4: e2012052.
33. Nakahashi H, Tsukamoto N, Hashimoto Y, Koiso H, Yokohama A, Saitoh T, Uchiumi H, Handa H, Murakami H, Nojima Y, Karasawa M. Characterization of immunoglobulin heavy and light chain gene expression in chronic lymphocytic leukemia and related disorders. *Cancer Science*. 2009; 100: 671-677.

34. Fais F, Ghiotto F, Hashimoto S, Sellars B, Valetto A, Allen SL, Schulman P, Vinciguerra VP, Rai K, Rassenti LZ, Kipps TJ, Dighiero G, Schroeder HW Jr, et al. Chronic lymphocytic leukemia B cells express restricted sets of mutated and unmutated antigen receptors. *The Journal of Clinical Investigation*. 1998; 102: 1515–1525.
35. Tobin G, Thunberg U, Johnson A, Eriksson I, Söderberg O, Karlsson K, Merup M, Juliusson G, Vilpo J, Enblad G, Sundström C, Roos G, Rosenquist R. 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*. 2003; 101: 4952–4957.
36. Rossi D, Spina V, Bomben R, Rasi S, Dal-Bo M, Brusca A, Rossi FM, Monti S, Degan M, Ciardullo C, Serra R, Zucchetto A, Nomdedeu J, et al. Association between molecular lesions and specific B-cell receptor subsets in chronic lymphocytic leukemia. *Blood*. 2013; 121: 4902–4905.
37. Del Giudice I, Chiaretti S, Santangelo S, Tavolaro S, Peragine N, Marinelli M, Ilari C, Raponi S, Messina M, Nanni M, Mauro FR, Piciocchi A, Bontempi K, et al. Stereotyped subset #1 chronic lymphocytic leukemia: a direct link between B-cell receptor structure, function, and patients' prognosis. *American Journal of Hematology*. 2014; 89: 74–82.
38. Rossi D, Spina V, Cerri M, Rasi S, Deambrogi C, De Paoli L, Laurenti L, Maffei R, Forconi F, Bertoni F, Zucca E, Agostinelli C, Cabras A, et al. Stereotyped B-cell receptor is an independent risk factor of chronic lymphocytic leukemia transformation to Richter syndrome. *Clinical Cancer Research*. 2009; 15: 4415–4422.
39. Bomben R, Dal-Bo M, Benedetti D, Capello D, Forconi F, Marconi D, Bertoni F, Maffei R, Laurenti L, Rossi D, Del Principe MI, Luciano F, Sozzi E, et al. Expression of mutated IGHV3-23 genes in chronic lymphocytic leukemia identifies a disease subset with peculiar clinical and biological features. *Clinical Cancer Research*. 2010; 16: 620–628.
40. Brezinschek HP, Foster SJ, Brezinschek RI, Dörner T, Domiati-Saad R, Lipsky PE. Analysis of the human VH gene repertoire. Differential effects of selection and somatic hypermutation on human peripheral CD5(+)/IgM+ and CD5(-)/IgM+ B cells. *The Journal of Clinical Investigation*. 1997; 99: 2488–2501.
41. Hallek M, Cheson BD, Catovsky D, Caligaris-Cappio F, Dighiero G, Döhner H, Hillmen P, Keating MJ, Montserrat E, Rai KR, Kipps TJ. International Workshop on Chronic Lymphocytic Leukemia. Guidelines for the diagnosis and treatment of chronic lymphocytic leukemia: A report from the International Workshop on Chronic Lymphocytic Leukemia updating the National Cancer Institute-Working Group 1996 guidelines. *Blood*. 2008; 111: 5446–5456.
42. Ghia P, Stamatopoulos K, Belessi C, Moreno C, Stilgenbauer S, Stevenson F, Davi F, Rosenquist R; European Research Initiative on CLL. European Research Initiative on CLL. ERIC recommendations on IGHV gene mutational status analysis in chronic lymphocytic leukemia. *Leukemia*. 2007; 21: 1–3.
43. Lefranc MP, Giudicelli V, Ginestoux C, Jabado-Michaloud J, Folch G, Bellahcene F, Wu Y, Gemrot E, Brochet X, Lane J, Regnier L, Ehrenmann F, Lefranc G, et al. IMGT, the international ImMunoGeneTics information system. *Nucleic Acids Research*. 2009; 37: D1006–D1012.
44. Brochet X, Lefranc MP, Giudicelli V. IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. *Nucleic Acids Research*. 2008; 36: W503–W508.
45. Murray F, Darzentas N, Hadzidimitriou A, Tobin G, Boudjogra M, Scielzo C, Laoutaris N, Karlsson K, Baran-Marzszak F, Tsiftaris A, Moreno C, Anagnostopoulos A, Caligaris-Cappio F, et al. Stereotyped patterns of somatic hypermutation in subsets of patients with chronic lymphocytic leukemia: Implications for the role of antigen selection in leukemogenesis. *Blood*. 2008; 111: 1524–1533.