bjh research paper

Molecular and clinical features of chronic lymphocytic leukaemia with stereotyped B cell receptors: results from an Italian multicentre study

Riccardo Bomben,[1,*] Michele Dal Bo,[1,*] Daniela Capello,[2,*] Francesco Forconi,[3] Rossana Maffei,[4] Luca Laurenti,[5] Davide Rossi,[2] Maria Ilaria Del Principe,[6] Antonella Zucchetto,[1] Francesco Bertoni,[7] Francesca Maria Rossi,[1] Pietro Bulian,[1] Ilaria Cattarossi,[1] Fiorella Ilariucci,[8] Elisa Sozzi,[3] Valeria Spina,[2] Emanuele Zucca,[7] Massimo Degan,[1] Francesco Lauria,[3] Giovanni Del Poeta,[6] Dimitar G. Efremov,[9] Roberto Marasca,[4] Gianluca Gaidano[2] and Valter Gattei[1]

[1]Clinical and Experimental Onco-Haematology Unit, Centro di Riferimento Oncologico, I.R.C.C.S., Aviano (PN), [2]Division of Haematology, Department of Clinical and Experimental Medicine & BRMA, Amedeo Avogadro University of Eastern Piedmont, Novara, [3]Division of Haematology and Transplant, Department of Clinical Medicine and Immunological Sciences, University of Siena, Siena, [4]Division of Haematology, Department of Oncology and Haematology, University of Modena and Reggio Emilia, Modena, [5]Haematology Institute, Catholic University "Sacro Cuore", Rome, [6]Division of Haematology, S. Eugenio Hospital, University of Tor Vergata, Rome, Italy, [7]Laboratory of Experimental Oncology and Lymphoma Unit, Oncology Institute of Southern Switzerland, Bellinzona, Switzerland, [8]Haematology Unit, Reggio Emilia Hospital, Reggio Emilia, Modena, and [9]ICGEB Outstation-Monterotondo, Rome, Italy

Received 31 July 2008; accepted for publication 8 September 2008

Correspondence: Valter Gattei, Clinical and Experimental Onco-Haematology Unit, Centro di Riferimento Oncologico, I.R.C.C.S., via Franco Gallini 2, Aviano (PN) 33081, Italy. E-mail: vgattei@cro.it

*These authors contributed equally to this work as first authors.

Summary

A fraction of chronic lymphocytic leukaemia (CLL) cases carry highly homologous B-cell receptors (BCR), i.e. characterized by non-random combinations of immunoglobulin heavy-chain variable (*IGHV*) genes and heavy-chain complementarity determining region-3 (HCDR3), often associated with a restricted selection of *IGVK/L* light chains. Such 'stereotyped' BCR occur more frequently in CLL with unmutated (UM) than mutated (M) *IGHV* genes. We analysed 1426 IG rearrangements (from 1398 CLL cases) by a clustering driven by HCDR3 similarities. Molecular findings were correlated to time-to-treatment (TTT) and presence of known prognosticators. Sixty-nine clusters (319 IG-rearrangements, 22·4%) with stereotyped BCR were identified. Among 30 confirmed clusters (≥3 IG-rearrangements/cluster), we found 14 novel clusters, of which 11 had M IG rearrangements (M clusters) and predominantly (8/11) used *IGHV3* subgroup genes. Recurrent cluster-biased amino acid changes were found throughout *IGHV* sequences of these 'M clusters'. Regarding clinical outcome: (i) UM CLL from the *IGHV1-2/1-3/1-18/1-46/7-4-1/IGKV1-39* cluster had poorer prognosis than UM/M cases, or UM cases using the same *IGHV* genes but not in clusters; (ii) M CLL from the *IGHV3-21/IGLV3-21* cluster had TTT similar to UM CLL, and shorter than M CLL expressing *IGHV3-21* but not in cluster. Altogether, our analysis identified additional molecular and clinical features for CLL expressing stereotyped BCR.

Keywords: chronic lymphocytic leukaemia, B cell receptor, immunoglobulin genes, prognosis.

First published online 19 November 2008
doi:10.1111/j.1365-2141.2008.07469.x © 2008 The Authors
Journal Compilation © 2008 Blackwell Publishing Ltd,

Chronic lymphocytic leukaemia (CLL) is a heterogeneous disease whose clinical course can be partly foreseen by the presence of mutated (M) or unmutated (UM) immunoglobulin variable (IGV) genes (Shanafelt *et al*, 2004). Analysis of immunoglobulin heavy-chain variable (*IGHV*) gene usage reveals a biased *IGHV* repertoire in CLL *versus* normal B cells, and in UM *versus* M CLL (Chiorazzi & Ferrarini, 2003; Chiorazzi *et al*, 2005; Fais *et al*, 1998; Ghia *et al*, 2005; Hamblin *et al*, 1999; Johnson *et al*, 1997; Messmer *et al*, 2004a; Murray *et al*, 2008; Rosenquist *et al*, 1999; Stamatopoulos *et al*, 2005, 2007; Stevenson & Caligaris-Cappio, 2004; Tobin *et al*, 2004a). A significant fraction of CLL carries highly homologous B cell receptors (BCR) characterized by non-random stereotyped combinations of specific *IGHV* genes and heavy-chain complementarity determining region-3 (HCDR3), often associated with a restricted selection of IGV kappa or lambda (IGKV/IGLV) light chains (Messmer *et al*, 2004b; Murray *et al*, 2008; Stamatopoulos *et al*, 2007; Tobin *et al*, 2002, 2003; Widhopf *et al*, 2008, 2004). These observations coupled to an *IGHV* mutation profile, often consistent with antigen-driven selection (Degan *et al*, 2004a; Messmer *et al*, 2004b; Murray *et al*, 2008), led to the speculation that CLL pathogenesis may involve specific, though yet unidentified (auto) antigens (Chiorazzi *et al*, 2005).

Expression of stereotyped BCR occurs more frequently in UM than in M CLL (Capello *et al*, 2006; Messmer *et al*, 2004b; Murray *et al*, 2008; Stamatopoulos *et al*, 2007; Tobin *et al*, 2004b). Several UM CLL subsets express stereotyped BCR characterized by usage of *IGHV1-69* and a restricted spectrum of IG light chains. Another subset of UM CLL expressing stereotyped BCR is characterized by usage of *IGHV1* subgroup genes other than *IGHV1-69* coupled to *IGKV1-39* light chain (Tobin *et al*, 2004a; Stamatopoulos *et al*, 2007; Messmer *et al*, 2004b). A non-stochastic pairing of *IGHV4-39* and *IGKV1-39* usually occurs in IGG-expressing UM CLL (Ghiotto *et al*, 2004; Murray *et al*, 2008; Stamatopoulos *et al*, 2007; Tobin *et al*, 2004a). Whether stereotyped BCR affects prognosis among UM CLL is still a matter of debate (Panovska-Stavridis *et al*, 2007; Stamatopoulos *et al*, 2007).

Among CLL, skewed usage of *IGHV3-21* occurs more frequently in Northern Europe than in Mediterranean or non-European countries (Bomben *et al*, 2007; Ghia *et al*, 2008, 2005; Thorselius *et al*, 2006; Tobin *et al*, 2002, 2003). Although *IGHV3-21* CLL often display M *IGHV* genes, survival is comparable to UM CLL patients (Tobin *et al*, 2002, 2003, 2004a; Thorselius *et al*, 2006; Bomben *et al*, 2007). Many *IGHV3-21* CLL display unusually short and highly homologous HCDR3 consistently associated with *IGLV3-21* (Bomben *et al*, 2007; Ghia *et al*, 2008, 2005; Thorselius *et al*, 2006; Tobin *et al*, 2004b). The prognostic relevance of stereotyped BCR among *IGHV3-21* CLL remains unclear (Bomben *et al*, 2007; Ghia *et al*, 2005; Thorselius *et al*, 2006). In addition to *IGHV3-21*, few other CLL clusters with stereotyped M IG rearrangements have been identified so far (Capello *et al*, 2004; Murray *et al*, 2008; Stamatopoulos *et al*, 2007).

We investigated BCR rearrangements from a wide series of CLL consecutively enroled by different Italian reference centres (1398 patients accounting for 1426 IG rearrangements). Molecular findings were correlated with disease progression and status of widely used prognosticators. Several novel clusters were identified, mainly represented by M IG rearrangements, and frequently involving *IGHV3* subgroup genes. From a clinical standpoint, we refined prognostication of the largest CLL clusters with stereotyped BCR.

Materials and methods

Patients

The study included peripheral blood samples from 1398 unselected CLL (median age: 64·0 years, range 32–97) collected from the archives of the Clinical and Experimental Onco-Haematology Unit, Aviano National Cancer Institute ($n = 184$); Division of Haematology, Amedeo Avogadro University of Eastern Piedmont, Novara ($n = 426$); Department of Oncology & Haematology, University of Modena-Reggio Emilia ($n = 260$); S. Eugenio Hospital, University of Tor Vergata, Rome ($n = 198$); Haematology Unit, University of Siena ($n = 155$); Haematology Institute, Catholic University, Rome ($n = 130$); and Laboratory of Experimental Oncology, Oncology Institute of Southern Switzerland, Bellinzona, Switzerland ($n = 45$). All patients provided informed consent in accordance with local Institutional Review Board requirements and the Declaration of Helsinki.

Analysis of IGHV, IGKV and IGLV rearrangements and HCDR3-driven clustering

Immunoglobulin heavy-chain variable diversity (*D*)-joining (*J*), *IGKV-J* and *IGLV-J* rearrangements were amplified from either reverse-transcribed total RNA or genomic DNA (Capello *et al*, 2004; Degan *et al*, 2004b; Marasca *et al*, 2001). Purified amplicons were sequenced either directly or upon subcloning (Capello *et al*, 2004; Degan *et al*, 2004b). Sequences were aligned to the ImMunoGeneTics (IMGT) directory for the identification of *IGHV-D-J* and *IGK/LV-J* rearrangements and for computation of mutational load (Giudicelli *et al*, 2004; Lefranc *et al*, 2005; Pommie *et al*, 2004). *IGV* sequences were considered M or UM using the canonical cut-off of 2% mismatch from germline *IGV* sequences (Damle *et al*, 1999; Hamblin *et al*, 1999). In the case of *IGHD* gene determination, the IMGT/Junction analysis tool was used, following established IMGT criteria (Yousfi *et al*, 2004).

For HCDR3-driven clustering, all in-frame immunoglobulin heavy-chain (IGH) rearrangements were first converted into amino acid (AA) sequences and then aligned according to the putative HCDR3 AA sequences by using the multiple sequence alignment software ClustalX (1.83) (Thompson *et al*, 1997). For this purpose, the length of HCDR3 was calculated between codons 107 and 117 (IMGT numbering). Withdrawal of the

© 2008 The Authors

usually conserved codons 105–106 (Honegger & Pluckthun, 2001), although part of HCDR3 according to IMGT criteria (Giudicelli *et al*, 2004; Lefranc *et al*, 2005; Pommie *et al*, 2004), was chosen to maximize differences among sequences.

Tightly clustering HCDR3 AA sequences were selected, visually inspected and further checked for similarity by aligning them again by means of ClustalX (1.83). Clusters identified as having similar/identical HCDR3 AA sequences were those characterized by a mean alignment score (MAS) equal or exceeding the value of 60/100. MAS was the mean value of all the pairwise alignment scores, as determined by ClustalX (1.83). Cluster nomenclature was according to Stamatopoulos *et al* (2007) and Murray *et al* (2008). Clusters not included in this nomenclature system were assigned a number preceded by the alphabet letter N ('Novel'). In selected cases, *IGHV* gene sequences of M CLL belonging to homologous clusters were analysed for the frequency of specific AA changes, as previously described (Messmer *et al*, 2004b; Murray *et al*, 2008; Stamatopoulos *et al*, 2007; Tobin *et al*, 2004a).

Immunophenotypic analyses

Detection of CD38, CD49d and ZAP-70 expression was as reported (Gattei *et al*, 2008; Rassenti *et al*, 2004). A cut-off of 30% positive cells was chosen to discriminate $CD38^{high}$ from $CD38^{low}$ or $CD49d^{high}$ from $CD49d^{low}$ CLL, while a cut-off of 20% positive cells was chosen to distinguish $ZAP\text{-}70^{high}$ from $ZAP\text{-}70^{low}$ CLL (Bomben *et al*, 2007; Crespo *et al*, 2003; Del Principe *et al*, 2006; Gattei *et al*, 2008; Orchard *et al*, 2004; Rassenti *et al*, 2004).

Analysis of cytogenetic aberrations

Cytogenetic abnormalities involving chromosomes 11, 12, 13 and 17 were investigated by interphase fluorescent *in situ* hybridization (FISH) analysis. Probes used were: LSI-ATM for del11q22-q23, LSI-13 and LSI-D13S319 for del13q14, LSI-p53 for del17p13 and CEP12 for trisomy 12 (Vysis Inc, London, UK). FISH procedures were as reported previously (Del Principe *et al*, 2004, 2006; Del Poeta *et al*, 2005). For each probe, at least 200 nuclei were examined.

Statistical analysis

All statistical analyses were performed using the R statistical package (http://www.r-project.org/). The end point for clinical correlations was time-to-treatment (TTT), defined as time from diagnosis to first treatment (event) or end of follow-up (censored observation). In the cohort utilized for TTT analyses (698 patients), treatment initiation was according to National Cancer Institute Working Group criteria (Cheson *et al*, 1996). TTT was estimated using Kaplan–Meier plots and comparisons between groups were made by means of log-rank test. The Cox proportional hazard regression model was chosen to assess the independent effect of covariables, treated as dichotomous on TTT, with a stepwise procedure for selecting significant variables. Distribution of UM and M cases, usage of *IGHV* genes as well as presence/absence of specific prognosticators among various CLL subsets with or without stereotyped BCR were analysed applying the Pearson's χ^2 test with Yates' continuity correction. *P*-values $\leq 0{\cdot}05$ were considered significant.

Results

A total of 1426 in-frame *IGHV-D-J* rearrangements were sequenced in 1398 patients. In keeping with previous studies, double or triple in-frame rearrangements were found in 22 and 3 cases, respectively (Degan *et al*, 2004a; Rassenti & Kipps, 1997; Stamatopoulos *et al*, 2007). Considering the 2% cut-off to discriminate between M and UM rearrangements, 845 sequences (59·3%) were classified as M, while 581 sequences (40·7%) were UM. Detailed information regarding *IGHV*, *IGHD* and *IGHJ* usage and the general clinical features of our cohort, including distribution and relevance of the major clinical (Rai staging) and biological (*IGHV* mutational status, CD38, ZAP-70 and CD49d expression, interphase FISH) prognosticators are reported (Appendix SI; Table SI–SIII; Fig S1). Molecular and clinical features of the CLL series were consistent with previous studies (Crespo *et al*, 2003; Damle *et al*, 1999; Degan *et al*, 2004a; Del Poeta *et al*, 2001; Dohner *et al*, 2000; Durig *et al*, 2002, 2003; Gattei *et al*, 2008; Hamblin *et al*, 2002, 1999, 2000; Krober *et al*, 2002; Orchard *et al*, 2004; Rai *et al*, 1975; Bomben *et al*, 2007; Messmer *et al*, 2004b; Murray *et al*, 2008; Stamatopoulos *et al*, 2007; Tobin *et al*, 2004a; Tschumper *et al*, 2008).

Identification of CLL with homologous HCDR3

Three hundred and nineteen (22·4%) sequences carried a homologous HCDR3 (Hom sequences) by HCDR3-driven clustering and were split into 69 different clusters, the majority of which (46/69) have not been previously reported (Table SIV). Conversely, 1107 of 1426 (77·6%) sequences did not belong to any cluster and were defined as heterologous (Het sequences). The 319 Hom sequences corresponded to 319 CLL (Hom cases). Hom cases included 312 CLL bearing a single IG rearrangement as well as 6/22 and 1/3 CLL with double or triple IG rearrangements respectively all characterized by a single Hom sequence. On the other hand, 1079 CLL presented only sequences not belonging to any cluster (Het cases). Het cases included 1061 CLL bearing a single IG rearrangement as well as 16/22 and 2/3 with double or triple IG rearrangements, respectively.

Hom and Het sequences were distributed differently among M and UM rearrangements (Table I). In particular, 198/581 (34·1%) UM rearrangements displayed homologous HCDR3, whereas only 121/845 (14·3%) M rearrangements were assigned to a cluster ($P < 0{\cdot}0001$). A relative over-representation of

© 2008 The Authors
Journal Compilation © 2008 Blackwell Publishing Ltd,

Table I. *IGHV* gene subgroup distribution in CLL cases with homologous or heterologous HCDR3.

	Hom			Het			
	UM	M	Total	UM	M	Total	*P*-value*
IGHV1	111	7	118	135	62	197	<0·0001
IGHV2	1	3	4	10	15	25	ns
IGHV3	57	67	124	159	440	599	<0·0001
IGHV4	23	44	67	55	187	242	ns
IGHV5	3	0	3	14	17	31	ns
IGHV6	0	0	0	8	1	9	ns
IGHV7	3	0	3	2	2	4	ns
Total	198	121	319	383	724	1107	<0·0001†

CLL, chronic lymphocytic leukaemia; HCDR3, heavy-chain complementarity determining region; Hom, CLL cases with homologous HCDR3; Het, CLL cases with heterologous HCDR3; UM, unmutated *IGHV* gene status; M, mutated *IGHV* gene status; IGHV, immunoglobulin heavy-chain variable; ns, not significant.

**P*-value obtained by the χ^2 test comparing the distribution of Hom and Het CLL in a given *IGHV* gene subgroup against the distribution of Hom and Het CLL in all the other *IGHV* gene subgroups altogether considered.

†*P*-value obtained by the χ^2 test comparing the overall distribution of Hom *versus* Het CLL cases in all the *IGHV* gene subgroups.

Hom sequences with UM rearrangements occurred among *IGHV1* and *IGHV3* subgroups ($P < 0{\cdot}0001$), although more than half of the Hom sequences with M rearrangements belonged to *IGHV3* subgroup (67/121; 55·4%). Accordingly, the chance of belonging to a given cluster was significantly higher for IG rearrangements using *IGHV1* subgroup genes compared to rearrangements employing other subgroups ($P < 0{\cdot}0001$; Table I; Fig 1A). Conversely, IG rearrangements using *IGHV3* subgroup genes were significantly under-represented among Hom sequences ($P < 0{\cdot}0001$; Table I; Fig 1A).

Skewed usage of specific *IGHV* genes was observed by comparing Hom and Het sequences. *IGHV1-69* (73/319; 22·9%) and *IGHV3-21* (33/319; 10·3%) were the most frequently utilized genes by Hom sequences (Fig 1B). Of note, *IGHV3-21* was the only gene employed more frequently by Hom than by Het sequences (Fig 1B), whereas *IGHV3-23* (130 Het sequences/134 total sequences) and *IGHV4-34* (99 Het sequences/128 total sequences) were preferentially employed by Het sequences (Fig 1B).

Molecular features of clusters with homologous HCDR3

Thirty of 69 clusters included at least 3 sequences/clusters (range 3–32). These clusters indicated as confirmed clusters in agreement with previous studies (Murray *et al*, 2008; Stamatopoulos *et al*, 2007), accounted for 241 sequences (16·9%), 80 with M and 161 with UM *IGHV* genes (Table II and Table SIV). Fifteen of 30 confirmed clusters (150 sequences) were exclusively/mainly composed of UM sequences (UM/M = 148/2; UM clusters), 14 clusters (59 sequences) were exclusively/mainly composed of M sequences (UM/M = 3/56; M clusters) and a single cluster (Cluster 2), represented by 32 sequences using *IGHV3-21*, included both UM and M sequences (UM/M = 10/22).

Within each confirmed cluster, identical *IGHV* genes were used in 22 clusters. In the remaining 8 clusters, a prevalent *IGHV* gene was found in 3 clusters (Clusters 3, 7 and 9) while in the other 5 clusters, *IGHV* genes often belonged to the same subgroup or clan (Table II) (Vargas-Madrazo *et al*, 1997). Analysis of *IGV* light chains showed the presence of restricted *IGV* light chain rearrangement in 22 of 30 confirmed clusters (Table II and Table SIV).

Sixteen of 30 confirmed clusters (196 sequences) were common to other CLL datasets (common clusters; Table II and Table SIV) (Messmer *et al*, 2004b; Murray *et al*, 2008; Stamatopoulos *et al*, 2007; Tobin *et al*, 2004a; Widhopf *et al*, 2008, 2004). With the only exception of Cluster 2 (*IGHV3-21/IGLV3-21*), including both M and UM sequences (Bomben *et al*, 2007; Ghia *et al*, 2005; Murray *et al*, 2008; Stamatopoulos *et al*, 2007; Tobin *et al*, 2002, 2003), the great majority of common clusters (12 of 15) were UM. UM clusters included: i) six clusters (Clusters 3, 5, 6, 7, 9 and 59) that exclusively/mainly utilized *IGHV1-69*; ii) two clusters utilizing either *IGHV1-2* alone (Cluster 28) or *IGHV1-2* together with other *IGHV1* genes, though not *IGHV1-69* (Cluster 1); iii) two clusters utilizing *IGHV4-39* either alone (Cluster 8, characterized by IGG-expressing cases) (Ghiotto *et al*, 2004; Messmer *et al*, 2004b; Stamatopoulos *et al*, 2007) or in association with other *IGHV* genes (Cluster 10); iv) two clusters (Clusters 109 and 26) utilizing exclusively/mainly genes from *IGHV3* subgroup. The remaining three common clusters were M clusters utilizing either *IGHV4-34* (Clusters 4 and 29, the former containing IGG-expressing cases) (Ghiotto *et al*, 2004; Messmer *et al*, 2004b; Stamatopoulos *et al*, 2007) or *IGHV4-59* (Cluster 13; Table II).

In addition to previously reported common clusters, we also identified 14 novel (Clusters N1–N14) confirmed clusters, accounting for 45 sequences (novel clusters). In contrast to common clusters, the majority of novel clusters (11/14) were M clusters (Table II). The different distribution of UM and M clusters among common and novel clusters was statistically significant ($P = 0{\cdot}0054$). Of note, 8 of 11 novel M clusters

© 2008 The Authors

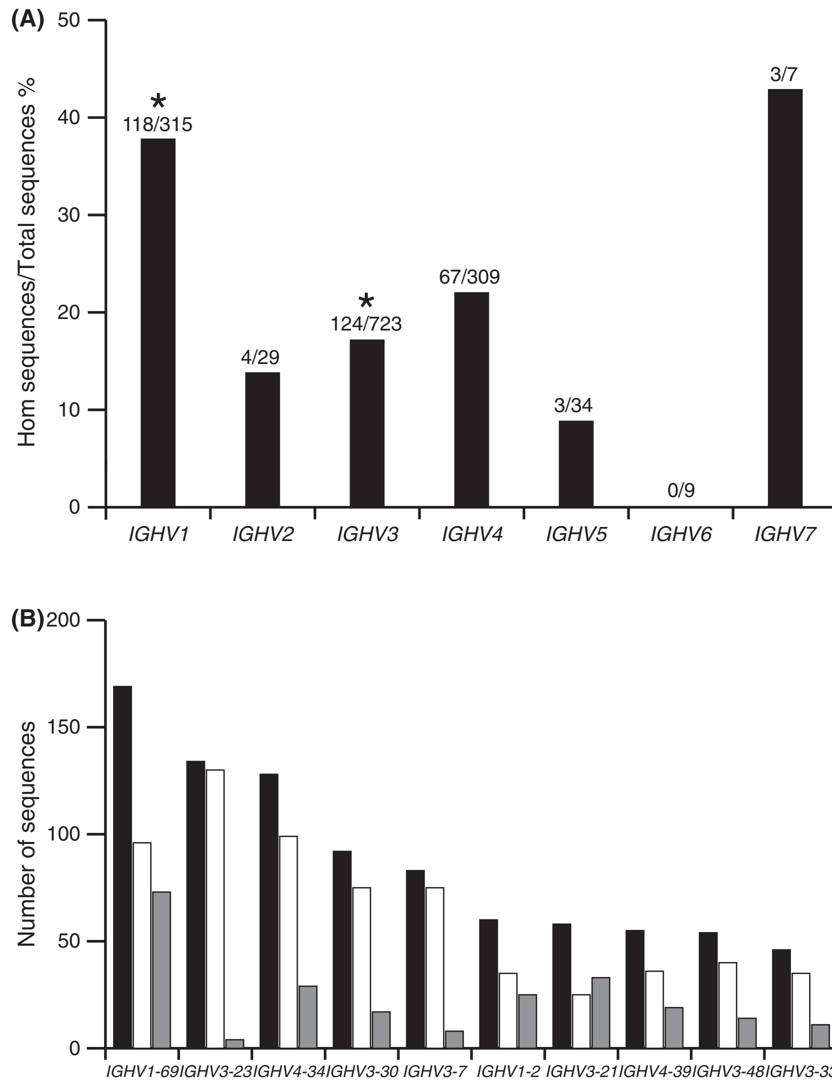

Fig 1. Usage of immunoglobulin heavy-chain variable (*IGHV*)subgroups and *IGHV* genes and chance to belong to clusters of homologous heavy-chain complementarity determining region-3 (HCDR3). (A) Percent of IG rearrangement sequences belonging to clusters of homologous HCDR3 (Hom sequences) according to the usage of the various *IGHV* gene subgroups. Figures represent the number of Hom IG rearrangements over the number of total IG rearrangements using genes from a given *IGHV* gene subgroups. Asterisks refer to *IGHV* gene subgroups whose distribution in Hom IG rearrangements was significantly different as compared to all the remaining cases. (B) Expression of the ten most frequently used *IGHV* genes in the series of 1426 IG rearrangements. For each *IGHV* gene, histograms reported the total number of IG rearrangements (closed histograms), the total number of IG rearrangements not included in clusters of homologous HCDR3 (open histograms) and the total number of IG rearrangements included in clusters of homologous HCDR3 (dotted histograms).

expressed genes from the *IGHV3* subgroup (*IGHV3-30*, 3 clusters; *IGHV3-33*, 1 cluster; *IGHV3-48*, 2 clusters; *IGHV3-7*, 1 cluster; *IGHV3-72*, 1 cluster), and 9 of 11 revealed restricted light chain usage (Table II and Table SIV).

Finally, out of the total 69 clusters, we identified 39 clusters (Table III) composed by pairs of sequences. These clusters were labelled as potential clusters in keeping with previous reports (Murray *et al*, 2008; Stamatopoulos *et al*, 2007). Nineteen out of 39 potential clusters had the same *IGHV-D-J* configuration, while the remaining 20 clusters expressed either the same *IGHV-D* (2 clusters), *IGHV-J* (6 clusters), *IGHV* (3 clusters), *IGHD-J* (8 clusters) or *IGHJ* (1 cluster) genes (Table III). As also observed for confirmed clusters, most of the potential clusters (32/39) were exclusively composed either by M (17/39) or UM (15/39) sequences. The same light chain was shared by 8 of 26 clusters for which data were available. Fourteen of 39 potential clusters displayed homology with CLL from other datasets or with single CLL cases that have been previously reported (Table III and Table SIV) (Messmer *et al*, 2004b; Murray *et al*, 2008; Stamatopoulos *et al*, 2007; Tobin *et al*, 2004a; Widhopf *et al*, 2008, 2004).

© 2008 The Authors

Table II. BCR molecular features of confirmed clusters.

Cluster	Total*	M/UM†	Heavy chain			MAS‡	Light chain	
			IGHV	*IGHD*	*IGHJ*		*IGKV*	*IGLV*
1§	30	0/30	1-18(4), 1-2(14), 1-3(7), 1-46(2), 7-4-1(3)	6-19	4	60·0	1-33(1), 1-39(27)	
2§	32	22/10	3-21	1-1(5), 1-14(1), 1-26(3), 2-2(4), 3-10(1), 4-17(1), 6-6(1), na(16)	6	73·9		3-21(29)
3§	20	0/20	1-2(1), 1-69(15), 3-11(1), 3-33(1), 3-48(1), 3-64(1)	2-2(17), 2-15(2), 2-21(1)	6	63·5	1-27(1), 1-39(8), 1-9(1), 2-28(2), 3-11(2), 3-20(2)	3-10(1)
4§	14	12/2	4-34	1-7(1), 2-21(1), 2-8(1), 3-10(1), 3-16(1), 3-3(1), 4-23(1), 5-12(1), 5-5(5), 5-18(1)	6	65·0	2-30(14)	
5§	11	0/11	1-69	3-10(9), 3-16(1), 6-19(1)	6	68·4	1-33(1), 4-1(1)	1-44(1), 1-47(1), 2-14(2), 3-21(1), 3-25(1)
6§	9	0/9	1-69	3-16	3	84·8	3-20(8)	
7§	26	0/26	1-18(1), 1-58(1), 1-69(16), 3-11(1), 3-20(1), 3-30(2), 3-30-3(1), 3-48(1), 3-49(1), 3-74(1)	3-3	6	65·2	1-33(1), 3-20(1)	1-51(7), 2-11(1), 2-14(2), 2-8(1), 3-1(1), 3-9(3)
8§	12	0/12	4-39	6-13(11), 6-19(1)	4(1), 5(11)	68·0	1-39(10), 4-1(1)	
9§	6	0/6	1-2(1), 1-69(4), 3-30-3(1)	3-3	6	60·5	3-20(1)	2-11(1), 2-14(2)
10§	10	0/10	1-24(2), 1-46(1), 3-20(1), 3-30(2), 4-34(1), 4-39(2), 4-59(1)	2-2	6	75·5	1-39(1), 1-5(2), 2-30(1), 3-20(1)	1-40(2), 2-11(1), 7-43(1)
13§	5	5/0	4-59	2-15	2	87·9	3-20(4)	
26§	3	1/2	1-69(1), 3-30(1), 3-33(1)	6-13	6	70·0	1-12(1), 1-39(1)	2-18(1)
28§	5	0/5	1-2	1-26(4), 3-10(1)	6	90·9	4-1(4)	
29§	5	5/0	4-34	6-19	3	82·5	3-11(3), 4-1(2)	
59§	5	0/5	1-69	3-3(4), 5-12(1)	4(3), 6(2)	81·0	2-28(3), 3-20(1)	
109§·¶	3	0/3	3-11	3-10	5	76·0	1-5(1), 3-20(2)	
N1	4	4/0	3-30	3-10	4	100·0		2-14(1), 3-1(1), 3-21(1)
N2**	4	3/1	3-48	1-26(3), na(1)	4(3), 5(1)	73·3		3-21(4)
N3	4	0/4	3-48	6-19	4	100·0		3-19(1), 3-21(3)
N4	3	3/0	1-2	3-3(1), 5-24(1), 5a(1)	5	66·0	1-6(2), 3-20(1)	
N5	3	3/0	2-5	1-1(1), 1-7(1)	4	78·0	2-28(1)	2-8(1)
N6	3	3/0	3-30	1-20	3	74·3	3-15(3)	
N7	3	1/2	3-30(2), 4-34(1)	3-22	6	65·7	1-39(1), 1-6(1)	
N8	3	3/0	3-30	1-26	4	77·3	1-12(2)	
N9¶·**	3	0/3	3-33	3-3	3	76·0	1-5(1), 1-17(1), 1-39(1)	
N10	3	3/0	3-33	3-10(2), na(1)	3	70·0		2-8(3)
N11	3	3/0	3-48(2), 4-59(1)	3-22(2), 6-13(1)	4	72·0	1-8(3)	
N12	3	3/0	3-7	2-2	4	78·7	4-1(2)	

© 2008 The Authors

Table II. Continued.

Cluster	Total*	M/UM†	Heavy chain			MAS‡	Light chain	
			IGHV	*IGHD*	*IGHJ*		*IGKV*	*IGLV*
N13**	3	3/0	3-72	2-2	3	86·6	4-1(3)	
N14**	3	3/0	4-39	6-13	4	100·0		2-14(2)

BCR, B cell receptors; IGHV, immunoglobulin heavy-chain variable; IGV, immunoglobulin variable; IGKV, IGV kappa; IGLV, IGV lambda; CLL, chronic lymphocytic leukaemia.

Cluster nomenclature is according to Stamatopoulos *et al* (2007) and Murray *et al* (2008). Clusters not included in this nomenclature system are identified with numbers preceded by the alphabet letter N.

*Number of cases in a given CLL cluster.

†Number of cases with mutated *IGHV* genes over cases with unmutated *IGHV* genes.

‡MAS was the mean value of all the pairwise alignment scores for all the sequences belonging to a specific cluster, as computed by ClustalX (1.83).

§Similar to previously published clusters; references are reported in text and Table SIV.

¶Publicly available non-CLL sequences potentially belonging to the same cluster; accession numbers are reported in Table SIV.

**Publicly available CLL sequences potentially belonging to the same cluster; accession numbers are reported in Table SIV.

Analysis of somatic mutations of clusters with homologous HCDR3 and M IGHV gene status

The high number of confirmed clusters composed by sequences with a M *IGHV* gene status prompted us to investigate whether, in the context of these M clusters, CLL sequences shared AA changes (i.e. replacement of the same AA in the same position) also outside the HCDR3 region (Murray *et al*, 2008). As shown in Fig S2, shared AA changes were found across the whole *IGHV* gene sequence in all but one (Cluster N11) of the 14 M clusters. Moreover, the vast majority of these shared AA changes were cluster-biased, since they occurred significantly more frequently in Hom sequences than in Het sequences expressing the same M *IGHV* gene (Table SV).

Clinical features of CLL with homologous HCDR3

Among 698 patients with available TTT, 159 were Hom cases and 539 were Het cases. Although median TTT was found to be shorter in Hom than in Het cases (38 months vs. 70 months; $P = 0{\cdot}0003$; Fig S3A), this difference was no longer true by comparing Hom and Het CLL in the context of the M or UM CLL subsets (Fig S3B). This clearly indicates that TTT differences between Hom and Het cases have to be ascribed to the prevalence of UM cases among Hom CLL (Table I).

Time-to-treatment was then investigated by comparing Hom CLL belonging to the largest clusters (Clusters 1–8; Table II) to Het cases rearranging the corresponding *IGHV* gene. No differences in TTT were observed between Hom cases expressing *IGHV1-69* (Clusters 3, 5, 6 and 7), *IGHV4-34* (Cluster 4) or *IGHV4-39* (Cluster 8) and Het cases expressing the corresponding *IGHV* gene (data not shown). Conversely, a peculiar clinical behaviour was found for CLL cases belonging to Clusters 1 and 2.

Cluster 1. Cluster 1 included 30 UM sequences mainly utilizing *IGHV1-2* or other *IGHV1* subgroup genes except for *IGHV1-69* (*IGHV1-18/1-3/1-46*) or the evolutionarily related *IGHV7-4-1* (Vargas-Madrazo *et al*, 1997). All but one IG light chains were *IGKV1-39* (Table II and Table SIV) (Tobin *et al*, 2004a; Stamatopoulos *et al*, 2007; Messmer *et al*, 2004b). Sequences belonging to Cluster 1 were identified among 135 sequences carrying the same *IGHV* genes, which included sequences either not belonging to any cluster (92 sequences) or included in other clusters (13 sequences).

Time-to-treatment intervals were available for 12 cases included in Cluster 1. As shown in Fig 2A–C, the prognosis of these patients was poor either when compared with the whole series of UM/M cases (UM/M = 269/417) or to cases expressing the same *IGHV* genes and not belonging to Cluster 1 independently on their *IGHV* mutational status (UM/M = 23/20). Consistently, the presence of unfavourable prognosticators was significantly more frequent in cases belonging to Cluster 1 than in Het cases expressing the same *IGHV* genes ($P = 0{\cdot}009$; Fig 2D).

Cluster 2. *IGHV3-21* was rearranged in 58/1426 (4·1%) sequences (UM/M = 18/40). Thirty-two (UM/M = 10/22) of 58 sequences (55·0%) expressed the previously described stereotyped BCR characterized by an unusually short HCDR3 combined with *IGLV3-21* and were included in Cluster 2 (Table II and Table SIV) (Bomben *et al*, 2007; Ghia *et al*, 2005; Tobin *et al*, 2002, 2003). The remaining 26 *IGHV3-21* sequences had longer and unrelated HCDR3, rarely associated (2/16 tested cases) with *IGLV3-21* (Bomben *et al*, 2007; Ghia *et al*, 2005; Murray *et al*, 2008; Stamatopoulos *et al*, 2007).

Information on TTT was available for 29 patients with *IGHV3-21* CLL (15 Hom cases belonging to Cluster 2 and 14 Het cases). Despite the presence of 9/15 cases with M *IGHV* genes, TTT intervals of Hom cases were comparable with those of UM cases (Fig 3A). Conversely, the clinical behaviour of the small cohort of *IGHV3-21* Het cases (UM/M = 4/10) was closely dependent upon *IGHV* mutational status (Fig 3B). The 10 M Het cases expressing *IGHV3-21* had significantly longer TTT intervals compared with the 9 M Hom cases included in

© 2008 The Authors

Table III. BCR molecular features of potential clusters.

Cluster	Total*	M/UM†	Heavy chain			MAS‡	Light chain	
			IGHV	*IGHD*	*IGHJ*		*IGKV*	*IGLV*
14§	2	2/0	4-4	2-21	4	75·0		7-43(1)
20§	2	1/1	3-53	4-4	4	100·0		3-1(1), 3-21(1)
34§	2	0/2	1-69	3-9	6	68·0	1-13(1), 1-33(1)	
68§,¶	2	2/0	3-72	2-2	6	77·0	1-16(2)	
70§,¶	2	1/1	3-23	3-10	4	66·0	3-11(1), 4-1(1)	
71§	2	0/2	5-a	6-13(1), na (1)	6	89·0	3-20(1)	3-21(1)
100§	2	2/0	1-69	3-22	3(1), 4(1)	63·0		6-57(1)
N15	2	2/0	1-8	2-15	4	100·0	4-1(1)	
N16	2	0/2	1-69	1-26	4	100·0	1-39(1), 1-5(1)	
N17	2	0/2	1-69	3-3	3	100·0	3-15(1), 3-20(1)	
N18	2	0/2	1-69	1-1	4	100·0	1-12(2)	
N19¶	2	0/2	3-11	3-22	4	75·0		2-14(2)
N20	2	0/2	3-30-3	5-12	4	92·0	2-29(1)	
N21	2	2/0	3-30-3	5-5	4	100·0	2-28(1), 3-15(1)	
N22	2	0/2	3-49	3-10	4	90·0	4-1(2)	
N23	2	2/0	4-34	6-19	4	63·0	2-28(1), 3-15(1)	
N24	2	2/0	4-34	6-19	4	62·0		2-8(1)
N25	2	0/2	4-4	3-10	5	100·0	1-5(1), 1-16(1)	
N26	2	2/0	4-4	6-19	4	100·0		2-11(1)
N27	2	1/1	4-4	1-7	4	100·0		1-40(1)
N28	2	2/0	4-61	3-10	4	100·0	2-30(1), 4-1(1)	
N29¶	2	2/0	3-48	4-17	4(1), 5(1)	60·0		4-60(2)
N30	2	2/0	3-72	2-2(1), na(1)	4	100·0	1-16(2)	
N31	2	2/0	3-11	2-15(1), 3-16(1)	4	70·0	nd	nd
N32	2	1/1	3-33	3-22(1), 4-4(1)	4	63·0		1-44(1), 2-14(1)
N33¶	2	2/0	3-7	6-13(1), 6-19(1)	4	70·0		4-69(1)
N34	2	2/0	4-34	2-2(1), 3-22(1)	4	81·0	1-5(1), 1-8(1)	
N35	2	2/0	3-15	2-8(1), 6-25(1)	3(1), 4(1)	75·0	4-1(1)	
N36	2	2/0	4-34	3-10(2)	3(1), 4(1)	63·0		3-21(1), 3-25(1)
N37	2	2/0	4-39	2-21(1), 5-5(1)	1(1), 5(1)	75·0	2-30(1)	
N38¶	2	0/2	3-30(1), 4-31(1)	3-3	5	72·0	1-39(1)	2-5(1)
N39	2	1/1	3-11(1), 3-7(1)	6-13	4	60·0	4-1(2)	
N40¶	2	0/2	3-15(1), 3-23(1)	3-3	4	68·0	3-11(1)	2-14(1)
N41¶,**	2	0/2	1-69(1), 3-20(1)	3-9	5	63·0		1-44(1), 6-57(1)
N42	2	1/1	3-23(1), 3-7(1)	3-10	4	72·0	1-16(1)	
N43	2	1/1	2-26(1), 3-33(1)	6-13	4	63·0	1-33(1)	1-51(1)
N44¶	2	0/2	3-7(1), 4-30-4(1)	3-3	4	70·0	nd	nd
N45	2	0/2	3-21(1), 5-51(1)	4-23	4	76·0	1-39(2)	
N46	2	0/2	1-2(1), 1-69(1)	3-3(1), 3-16(1)	6	100·0	3-15(1), 3-20(1)	

BCR, B cell receptors; IGHV, immunoglobulin heavy-chain variable; IGV, immunoglobulin variable; IGKV, IGV kappa; IGLV, IGV lambda; CLL, chronic lymphocytic leukaemia; MAS, mean alignment score; M, mutated *IGHV* gene status; UM, unmutated *IGHV* gene status; na, not assigned; nd, not detected.

Cluster nomenclature is according to Stamatopoulos *et al* (2007) and Murray *et al* (2008). Clusters not included in this nomenclature system are identified with numbers preceded by the alphabet letter N.

*Number of cases in a given CLL cluster.

†Number of cases with mutated *IGHV* genes over cases with unmutated *IGHV* genes.

‡MAS was the mean value of all the pairwise alignment scores for all the sequences belonging to a specific cluster, as computed by ClustalX (1.83).

§Similar to previously published clusters; references are reported in text and Table SIV.

¶Publicly available CLL sequences potentially belonging to the same cluster; accession numbers are reported in Table SIV.

**Publicly available non-CLL sequences potentially belonging to the same cluster; accession numbers are reported in Table SIV.

© 2008 The Authors

Cluster 1 (Fig 3C). Consistently, *IGHV3-21* Hom cases expressed unfavourable prognosticators more frequently than *IGHV3-21* Het cases ($P = 0{\cdot}002$; Fig 3D).

Discussion

Alignment analyses of 1426 functional in-frame *IGHV-D-J* rearrangements corresponding to a series of 1398 unselected CLL led to the identification of 69 different clusters accounting for 319 sequences (22·4%). This approach enabled the identification of several clusters (46/69) not reported by previous studies (Messmer *et al*, 2004b; Murray *et al*, 2008; Stamatopoulos *et al*, 2007; Tobin *et al*, 2004a; Widhopf *et al*, 2008). However, since the novel clusters turned out to have mostly two or three sequences per cluster, the overall percentage of Hom cases was comparable with other reports (Murray *et al*, 2008; Stamatopoulos *et al*, 2007).

It is unlikely that the identification of novel clusters may be ascribed to unwanted biases of this CLL series. In fact, our series did not show any significant difference in frequency of *IGHV*, *IGHD* and *IGHJ* gene usage compared with other published series (Degan *et al*, 2004a; Ghia *et al*, 2005; Messmer *et al*, 2004b; Murray *et al*, 2008; Stamatopoulos *et al*, 2007; Tobin *et al*, 2004a; Widhopf *et al*, 2008). The present study performed cluster analysis by purposely disregarding *IGHV* genes and considering HCDR3 sequences alone for alignment. Given that HCDR3 represents the most variable region of IGH chains and is known to be directly involved in binding to canonical antigens (Maccallum *et al*, 1996), such a strategy appeared to be appropriate for identifying clusters potentially selected by similar/identical antigens. Conversely, in other studies, expression of the same *IGHV* gene was chosen as a prerequisite for cluster identification (Messmer *et al*, 2004b; Tobin *et al*, 2004a; Widhopf *et al*, 2008, 2004). A HCDR3-driven clustering similar to that employed in the present investigation was utilized in previous studies (Murray *et al*, 2008; Stamatopoulos *et al*, 2007). At variance with these studies, we calculated HCDR3 length withdrawing the usually conserved codons 105–106 (Honegger & Pluckthun, 2001) to maximize differences among sequences and utilized the MAS value to identify homologous clusters (Thompson *et al*, 1997). Some of the differences between this and previous studies could be at least in part the result of different approaches chosen for clustering.

Despite our choice of a strict HCDR3-driven clustering approach, 22/30 confirmed clusters carried identical *IGHV* genes or shared the expression of the same light chain either in all or in the majority of the analysed cases. Clusters expressing different *IGHV* genes displayed either a single prevalent *IGHV* gene (e.g. *IGHV1-69* prevalence in Clusters 3, 8 and 9) or *IGHV* genes belonging to the same subgroup/clan (e.g. *IGHV1* and *IGHV7* subgroup genes in Cluster 1) (Murray *et al*, 2008; Stamatopoulos *et al*, 2007; Messmer *et al*, 2004b; Tobin *et al*, 2004a; Widhopf *et al*, 2008, 2004; Vargas-Madrazo *et al*, 1997).

Sixteen of 30 confirmed clusters were common to other CLL datasets (Messmer *et al*, 2004b; Murray *et al*, 2008; Stamatopoulos *et al*, 2007; Tobin *et al*, 2004a). With the exclusion of Cluster 2 (the *IGHV3-21/IGLV3-21* cluster) (Bomben *et al*, 2007; Ghia *et al*, 2008, 2005; Tobin *et al*, 2002, 2003), which was the sole cluster accounting for both UM and M cases, and of 3 M clusters utilizing *IGHV4-34* and *IGHV4-59* (Ghiotto *et al*, 2004; Murray *et al*, 2008; Stamatopoulos *et al*, 2007), the vast majority of common clusters comprised UM cases (Messmer *et al*, 2004b; Murray *et al*, 2008; Stamatopoulos *et al*, 2007; Tobin *et al*, 2004a). These UM clusters included cases that have been recognized as expressing both autoreactive and polyreactive BCR, allegedly deriving from the B cell compartment devoted to the production of natural antibodies (Borche *et al*, 1990; Dighiero *et al*, 1983; Herve *et al*, 2005; Lanemo Myhrinder *et al*, 2008; Martin *et al*, 1992; Sthoeger *et al*, 1989; Widhopf *et al*, 2008). Autoreactivity has been also hypothesized for the M cluster utilizing *IGHV4-34* (Messmer *et al*, 2004b; Murray *et al*, 2008; Stamatopoulos *et al*, 2007), although occurring via conserved residues located outside the conventional HCDR antigen-binding sites (Murray *et al*, 2008; Silverman & Goodyear, 2002, 2006).

The remaining 14/30 confirmed clusters accounting for 45 sequences were not previously reported. Although also in our series the chance of carrying a stereotyped BCR was significantly higher for UM CLL (Messmer *et al*, 2004b), the majority of these novel clusters (11/14) unexpectedly expressed stereotyped BCR with M *IGHV* rearrangements, often involving (8/11) *IGHV3* subgroup genes. Of note, recurrent AA changes were demonstrated to occur across the entire *IGHV* region in all novel M clusters. These mutations were significantly under-represented in CLL expressing the same *IGHV* genes but not included in clusters, and therefore were considered cluster-biased, as previously reported (Murray *et al*, 2008).

Lack of detection of these novel clusters by previous studies, including those dealing with comparably wide datasets (Messmer *et al*, 2004b; Murray *et al*, 2008; Stamatopoulos *et al*, 2007; Tobin *et al*, 2004a), raises the possibility that the frequency of these clusters might be subjected to a geographical bias. In this regard, a significantly skewed representation of another *IGHV3* gene, i.e. *IGHV3-21*, has been well documented in different European and non-European countries (Ghia *et al*, 2008, 2005; Tobin *et al*, 2002, 2003), and even in different regions from the same country (Bomben *et al*, 2007). Similarly, two clusters of our series expressing *IGHV3-30* (Cluster N8) and *IGHV3-33* (Cluster N10) were shown to have quasi-identical HCDR3 sequences when compared with two cases and to a single case reported elsewhere as belonging to an Italian series (Murray *et al*, 2008; Stamatopoulos *et al*, 2007). Another novel M cluster (Cluster N13) expressed *IGHV3-72* in association with highly homologous HCDR3 and *IGKV4-1*. Two additional *IGHV3-72*-expressing potential M clusters (Clusters 68 and N30), one also reported elsewhere (Murray *et al*, 2008) were associated with *IGKV1-16*. Of note, frequent usage of *IGHV3-72* has been reported in highly stable and

© 2008 The Authors

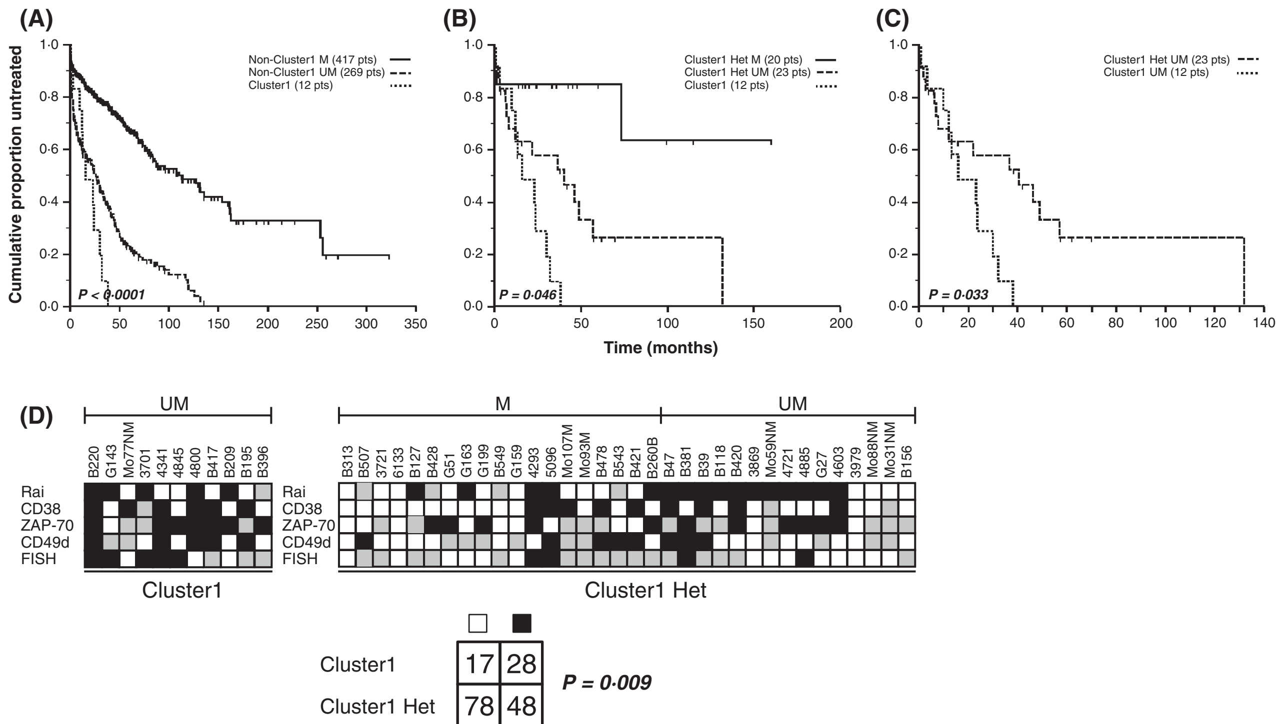

Fig 2. Clinical features of patients affected by chronic lymphocytic leukaemia (CLL) belonging to Cluster 1. (A) Kaplan–Meier curves comparing time-to-treatment (TTT) of patients affected by CLL included in Cluster 1 (Cluster1), CLL with mutated (M) *IGHV* gene configurations (non-Cluster1 M), and with unmutated (UM) *IGHV* gene configurations (non-Cluster1 UM); (B) Kaplan–Meier curves comparing TTT of patients affected by CLL included in Cluster1 (Cluster1), CLL with M *IGHV* gene configurations and expressing the same IG rearrangements but not included in cluster (Cluster1 Het M) and CLL with UM *IGHV* gene configurations and expressing the same IG rearrangement but not included in cluster (Cluster1 Het UM); (C) Kaplan–Meier curves comparing TTT of patients affected by CLL expressing a UM *IGHV* mutational status and included in Cluster 1 (Cluster1 UM), and CLL with UM *IGHV* gene configurations and expressing the same IG rearrangement but not included in cluster (Cluster1 Het UM); the numbers of patients (pts) included in each group are reported in parenthesis; the reported *P*-value refers to log-rank test. (D) The reported grids summarize data regarding the presence of the five prognostic markers; Rai stage (Rai), CD38, ZAP-70, CD49d and chromosome abnormalities, as investigated by interphase fluorescent *in situ* hybridization (FISH) in 11 patients affected by CLL included in Cluster 1 (Cluster1, left grid) and in 34 patients affected by CLL expressing the same IG rearrangements but not included in cluster (Het, right grid). Closed boxes in grids indicate presence of the prognostic marker in their unfavourable configuration (i.e. Rai stage ≥I, CD38 ≥ 30% of positive cells, ZAP-70 ≥ 20% of positive cells, CD49d ≥ 30% of positive cells and FISH 17p−/11q−/+12); open boxes in grids indicate presence of the marker in their favourable configuration (i.e. Rai stage = 0, CD38 < 30% of positive cells, ZAP-70 < 20% of positive cells, CD49d < 30% of positive cells and FISH normal/13q−); dotted boxes in grids indicate data not available. The reported χ^2 test compares the distribution of prognostic markers in their favourable (open box) or unfavourable (closed box) configurations in patients affected by CLL included in Cluster 1 (Cluster1) *versus* patients affected by CLL expressing the same IG rearrangements but not included in cluster (Cluster1 Het).

indolent CLL from an Italian cohort of patients not included in the present study (Capello *et al*, 2004, 2006; Guarini *et al*, 2003).

The finding of several novel clusters involving *IGHV3* subgroup genes may appear surprising, given the property of genes from this subgroup to bind superantigens, i.e. specific unconventional antigens (e.g. microbial proteins) capable of reacting with IG through residues located outside the conventional HCDR antigen-binding sites (Sasso *et al*, 1989; Silverman & Goodyear, 2002, 2006; Seppala *et al*, 1990; Domiati-Saad & Lipsky, 1998). Given that CLL in clusters show clear evidence by expressing IG rearrangements with homologous HCDR3 for a selection driven by conventional antigens, it is conceivable that at least certain genes of the *IGHV3* subgroup may bind antigens via canonical pathways of HCDR-mediated recognition (Chiorazzi *et al*, 2005; Silverman & Goodyear, 2002, 2006).

From a clinical standpoint, the presence of stereotyped BCR was not a feature with prognostic significance in our series. In fact, although TTT was significantly shorter for Hom CLL compared with Het CLL, this was ascribed to the presence of a larger number of UM cases among Hom CLL, as described here and previously (Messmer *et al*, 2004b; Murray *et al*, 2008; Stamatopoulos *et al*, 2007; Tobin *et al*, 2004a). Nevertheless, by comparing the clinical features of Hom and Het cases in the context of specific stereotyped BCR, novel clinical features emerged.

Cases belonging to Cluster 1 expressed *IGHV1* subgroup genes other than *IGHV1-69* and displayed highly homologous HCDR3 sequences in association with a common light chain

© 2008 The Authors

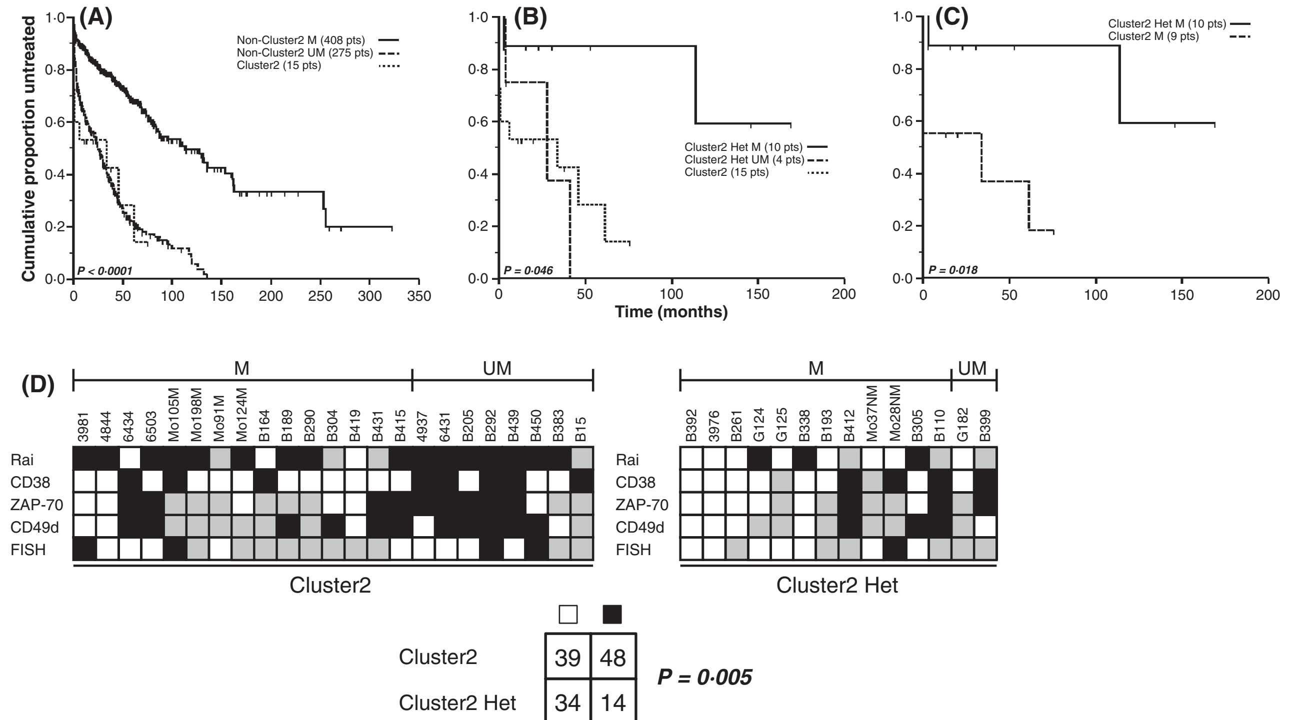

Fig 3. Clinical features of patients affected by chronic lymphocytic leukaemia (CLL) belonging to Cluster 2. (A) Kaplan–Meier curves comparing time-to-treatment (TTT) of patients affected by CLL included in Cluster 2 (Cluster2), CLL with mutated (M) *IGHV* gene configurations (non-Cluster2 M) and CLL with unmutated (UM) *IGHV* gene configurations (non-Cluster2 UM); (B) Kaplan–Meier curves comparing TTT of patients affected by CLL included in Cluster 2 (Cluster2), CLL with M *IGHV* gene configurations and expressing the same IG rearrangement but not included in cluster (Cluster2 Het M), and CLL with UM *IGHV* gene configurations and expressing the same IG rearrangement but not included in cluster (Cluster2 Het UM); (C) Kaplan–Meier curves comparing TTT of patients affected by CLL expressing a M *IGHV* mutational status and included in Cluster 2 (Cluster2 M), and CLL with M *IGHV* gene configurations and expressing the same IG rearrangement but not included in cluster (Cluster2 Het M); the numbers of patients (pts) included in each group are reported in parenthesis; the reported *P*-value refers to log-rank test. (D) The reported grids summarize data regarding the presence of the five prognostic markers; Rai stage (Rai), CD38, ZAP-70, CD49d and chromosome abnormalities, as investigated by interphase fluorescent *in situ* hybridization (FISH) in 23 patients affected by CLL included in Cluster 2 (Cluster2, left grid) and in 14 patients affected by CLL expressing the same IG rearrangement but not included in cluster (Cluster2 Het, right grid). Explanations for closed, open and dotted boxes in grids are as in Fig 2. The reported χ^2 test compares the distribution of prognostic markers in their favourable (open box) or unfavourable (closed box) configurations in patients affected by CLL included in Cluster 2 (Cluster2) *versus* patients affected by CLL expressing the same IG rearrangement but not included in cluster (Cluster2 Het).

(Messmer *et al*, 2004b; Murray *et al*, 2008; Stamatopoulos *et al*, 2007; Tobin *et al*, 2004a). Although Cluster 1 CLL always had an UM *IGHV* status, TTT intervals of these patients were shorter when compared with the whole series of UM/M cases and to UM cases expressing the same *IGHV* genes but not included in clusters. A similar clinical behaviour was also suggested by the sole previous study identifying the same cluster (Stamatopoulos *et al*, 2007). Results from this and the present study, concordantly indicate that Cluster 1 CLL has the worst prognosis in CLL.

Here, we provide evidence that Hom *IGHV3-21* CLL have shorter TTT when compared with: (i) all M CLL; (ii) M CLL expressing *IGHV3-21* but not included in Cluster 2. Consistently, this latter group of patients had significantly longer TTT intervals compared with a small subgroup of CLL included in Cluster 2 that expressed a M IG rearrangement. The distribution of several prognosticators was in keeping with this clinical behaviour. The prognostic relevance of stereotyped BCR among *IGHV3-21* CLL is presently still unclear (Bomben *et al*, 2007; Ghia *et al*, 2005; Thorselius *et al*, 2006). In particular, whilst no differences were found by comparing the clinical courses of patients affected by *IGHV3-21* CLL with Hom *versus* Het HCDR3 (Thorselius *et al*, 2006), the molecular basis for a more aggressive clinical outcome of Hom *versus* Het *IGHV3-21* CLL has been suggested by gene expression profiling and immunophenotypic analyses (Bomben *et al*, 2007). These discrepancies can be explained by the different clinical end-points (TTT vs. overall survival) employed in the present *versus* previous studies (Thorselius *et al*, 2006). Further studies combining the effort of several institutions and specifically focusing on *IGHV3-21* CLL patients are needed to definitively solve this issue. The notion that patients affected by Hom *IGHV3-21* CLL experience a more progressive disease may have important implications, given the proposal of using *IGHV3-21* expression to drive clinical decision in prospective trials (Krober *et al*, 2006; Zenz

© 2008 The Authors
Journal Compilation © 2008 Blackwell Publishing Ltd,

et al, 2007). Finally, we were unable to reproduce the different clinical outcome previously suggested for *IGHV1-69* CLL belonging to different clusters of stereotyped BCR (e.g. Clusters 3 and 5) (Stamatopoulos *et al*, 2007). In our series, clinical courses of Hom *versus* Het CLL expressing *IGHV1-69* were similar and mostly dependent upon *IGHV* mutational status, as also reported elsewhere (Panovska-Stavridis *et al*, 2007).

Acknowledgements

Supported in part by: Ministero della Salute (Ricerca Finalizzata I.R.C.C.S. and 'Alleanza Contro il Cancro'), Rome; Associazione Italiana contro le Leucemie, linfomi e mielomi (A.I.L.), Venezia Section, Pramaggiore Group; Ricerca Scientifica Applicata, Regione Friuli Venezia Giulia, Trieste ('Linfonet'); Programmi di Ricerca di Interesse Nazionale (P.R.I.N.) and Fondo per gli Investimenti per la Ricerca di Base (F.I.R.B.), M.U.R., Rome; Novara-A.I.L. Onlus, Novara; Ricerca Sanitaria Finalizzata Regione Piemonte, Torino; Associazione Italiana per la Ricerca sul Cancro (A.I.R.C.), Milan; Piano di Ateneo per la Ricerca (P.A.R.) 2005, University of Siena, Italy; Hairy Cell Leukemia Research Foundation (Illinois, USA). V.S. is supported by Associazione 'Franca Capurro per Novara' Onlus, Novara, Italy.

Author contributions

RB, MDB and DC analysed *IGHV* and *IGK/L* gene sequences and contributed to writing the manuscript; RMaf., IC, ES, VS, MD analysed *IGHV* and *IGK/L* gene sequences; AZ, FMR and PB contributed to the prognostic evaluation of patients and data analysis; LL, DR, MIDP, FI, EZ, FL and GDP contributed to the prognostic evaluation of patients and provided clinical data; FF, FB, DGE, RMar and GG provided *IGHV* and/or *IGK/L* gene sequences and clinical data and contributed to writing the manuscript; VG designed and coordinated the study and data analyses and wrote the manuscript.

References

Bomben, R., Dal Bo, M., Capello, D., Benedetti, D., Marconi, D., Zucchetto, A., Forconi, F., Maffei, R., Ghia, E.M., Laurenti, L., Bulian, P., Del Principe, M.I., Palermo, G., Thorselius, M., Degan, M., Campanini, R., Guarini, A., Del Poeta, G., Rosenquist, R., Efremov, D.G., Marasca, R., Foa, R., Gaidano, G. & Gattei, V. (2007) Comprehensive characterization of IGHV3-21-expressing B-cell chronic lymphocytic leukemia: an Italian multicenter study. *Blood*, **109,** 2989–2998.

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

Capello, D., Guarini, A., Berra, E., Mauro, F.R., Rossi, D., Ghia, E., Cerri, M., Logan, J., Foa, R. & Gaidano, G. (2004) Evidence of biased immunoglobulin variable gene usage in highly stable B-cell chronic lymphocytic leukemia. *Leukemia*, **18,** 1941–1947.

Capello, D., Zucchetto, A., Degan, M., Bomben, R., Dal Bo, M., Efremov, D.G., Laurenti, L., Del Poeta, G., Gaidano, G. & Gattei, V. (2006) Immunophenotypic characterization of IgVH3-72 B-cell chronic lymphocytic leukaemia (B-CLL). *Leukemia Research*, **30,** 1197–1199.

Cheson, B.D., Bennett, J.M., Grever, M., Kay, N., Keating, M.J., O'Brien, S. & Rai, K.R. (1996) National Cancer Institute-sponsored Working Group guidelines for chronic lymphocytic leukemia: revised guidelines for diagnosis and treatment. *Blood*, **87,** 4990–4997.

Chiorazzi, N. & Ferrarini, M. (2003) Chronic lymphocytic leukemia: lessons learned from studies of the B cell antigen receptor. *Annual Review of Immunology*, **21,** 841–894.

Chiorazzi, N., Rai, K.R. & Ferrarini, M. (2005) Chronic lymphocytic leukemia. *The New England Journal of Medicine*, **352,** 804–815.

Crespo, M., Bosch, F., Villamor, N., Bellosillo, B., Colomer, D., Rozman, M., Marce, S., Lopez-Guillermo, A., Campo, E. & Montserrat, E. (2003) ZAP-70 expression as a surrogate for immunoglobulin-variable-region mutations in chronic lymphocytic leukemia. *The New England Journal of Medicine*, **348,** 1764–1775.

Damle, R.N., Wasil, T., Fais, F., Ghiotto, F., Valetto, A., Allen, S.L., Buchbinder, A., Budman, D., Dittmar, K., Kolitz, J., Lichtman, S.M., Schulman, P., Vinciguerra, V.P., Rai, K.R., Ferrarini, M. & Chiorazzi, N. (1999) Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. *Blood*, **94,** 1840–1847.

Degan, M., Bomben, R., Dal Bo, M., Zucchetto, A., Nanni, P., Rupolo, M., Steffan, A., Attadia, V., Ballerini, P.F., Damiani, D., Pucillo, C., Del Poeta, G., Colombatti, A. & Gattei, V. (2004a) Analysis of IgVH gene mutations in B-CLL according to antigen-driven selection identifies subgroups with different prognosis and usage of the canonical SHM machinery. *British Journal of Haematology*, **126,** 29–42.

Degan, M., Rupolo, M., Dal Bo, M., Stefanon, A., Bomben, R., Zucchetto, A., Canton, E., Berretta, M., Nanni, P., Steffan, A., Ballerini, P.F., Damiani, D., Pucillo, C., Attadia, V., Colombatti, A. & Gattei, V. (2004b) Mutational status of IgVH genes consistent with antigen-driven selection but not percent of mutations has prognostic impact in B-cell chronic lymphocytic leukemia. *Clinical Lymphoma*, **5,** 123–126.

Del Poeta, G., Maurillo, L., Venditti, A., Buccisano, F., Epiceno, A.M., Capelli, G., Tamburini, A., Suppo, G., Battaglia, A., Del Principe, M.I., Del Moro, B., Masi, M. & Amadori, S. (2001) Clinical significance of CD38 expression in chronic lymphocytic leukemia. *Blood*, **98,** 2633–2639.

Del Poeta, G., Del Principe, M.I., Consalvo, M.A., Maurillo, L., Buccisano, F., Venditti, A., Mazzone, C., Bruno, A., Gianni, L., Capelli, G., Lo, C.F., Cantonetti, M., Gattei, V. & Amadori, S. (2005) The addition of rituximab to fludarabine improves clinical outcome in untreated patients with ZAP-70-negative chronic lymphocytic leukemia. *Cancer*, **104,** 2743–2752.

Del Principe, M.I., Del Poeta, G., Venditti, A., Buccisano, F., Maurillo, L., Marini, R., Cox, M.C., Panetta, P., Suppo, G., Degan, M., Bruno, A., Gattei, V. & Amadori, S. (2004) Clinical significance of soluble p53 protein in B-cell chronic lymphocytic leukemia. *Haematologica*, **89,** 1468–1475.

Del Principe, M.I., Del Poeta, G., Buccisano, F., Maurillo, L., Venditti, A., Zucchetto, A., Marini, R., Niscola, P., Consalvo, M.A.I., Mazzone, C., Ottaviani, L., Panetta, P., Bruno, A., Bomben, R., Suppo, G., Degan, M., Gattei, V., de Fabritiis, P., Cantonetti, M., Lo Coco, F.,

© 2008 The Authors

Del Principe, D. & Amadori, S. (2006) Clinical significance of ZAP-70 protein expression in B-cell chronic lymphocytic leukemia. *Blood*, **108,** 853–861.

Dighiero, G., Guilbert, B., Fermand, J.P., Lymberi, P., Danon, F. & Avrameas, S. (1983) Thirty-six human monoclonal immunoglobulins with antibody activity against cytoskeleton proteins, thyroglobulin, and native DNA: immunologic studies and clinical correlations. *Blood*, **62,** 264–270.

Dohner, H., Stilgenbauer, S., Benner, A., Leupolt, E., Krober, A., Bullinger, L., Dohner, K., Bentz, M. & Lichter, P. (2000) Genomic aberrations and survival in chronic lymphocytic leukemia. *The New England Journal of Medicine*, **343,** 1910–1916.

Domiati-Saad, R. & Lipsky, P.E. (1998) Staphylococcal enterotoxin A induces survival of VH3-expressing human B cells by binding to the VH region with low affinity. *Journal of Immunology*, **161,** 1257–1266.

Durig, J., Naschar, M., Schmucker, U., Renzing-Kohler, K., Holter, T., Huttmann, A. & Duhrsen, U. (2002) CD38 expression is an important prognostic marker in chronic lymphocytic leukaemia. *Leukemia*, **16,** 30–35.

Durig, J., Nuckel, H., Cremer, M., Fuhrer, A., Halfmeyer, K., Fandrey, J., Moroy, T., Klein-Hitpass, L. & Duhrsen, U. (2003) ZAP-70 expression is a prognostic factor in chronic lymphocytic leukemia. *Leukemia*, **17,** 2426–2434.

Fais, F., Ghiotto, F., Hashimoto, S., Sellars, B., Valetto, A., Allen, S.L., Schulman, P., Vinciguerra, V.P., Rai, K., Rassenti, L.Z., Kipps, T.J., Dighiero, G., Schroeder, H.W., Jr, Ferrarini, M. & Chiorazzi, N. (1998) Chronic lymphocytic leukemia B cells express restricted sets of mutated and unmutated antigen receptors. *Journal of Clinical Investigation*, **102,** 1515–1525.

Gattei, V., Bulian, P., Del Principe, M.I., Zucchetto, A., Maurillo, L., Buccisano, F., Bomben, R., Dal-Bo, M., Luciano, F., Rossi, F.M., Degan, M., Amadori, S. & Del, P.G. (2008) Relevance of CD49d protein expression as overall survival and progressive disease prognosticator in chronic lymphocytic leukemia. *Blood*, **111,** 865–873.

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., Caligaris-Cappio, F. & Davi, F. (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.

Ghia, E.M., Jain, S., Widhopf, G.F., Rassenti, L.Z., Keating, M.J., Wierda, W.G., Gribben, J.G., Brown, J.R., Rai, K.R., Byrd, J.C., Kay, N.E., Greaves, A.W. & Kipps, T.J. (2008) Use of IGHV3-21 in chronic lymphocytic leukemia is associated with high-risk disease and reflects antigen-driven, post-germinal-center leukemogenic selection. *Blood*, **111,** 5101–5108.

Ghiotto, F., Fais, F., Valetto, A., Albesiano, E., Hashimoto, S., Dono, M., Ikematsu, H., Allen, S.L., Kolitz, J., Rai, K.R., Nardini, M., Tramontano, A., Ferrarini, M. & Chiorazzi, N. (2004) Remarkably similar antigen receptors among a subset of patients with chronic lymphocytic leukemia. *Journal of Clinical Investigation*, **113,** 1008–1016.

Giudicelli, V., Chaume, D. & Lefranc, M.P. (2004) IMGT/V-QUEST, an integrated software program for immunoglobulin and T cell receptor V-J and V-D-J rearrangement analysis. *Nucleic Acids Research*, **32,** W435–W440.

Guarini, A., Gaidano, G., Mauro, F.R., Capello, D., Mancini, F., De Propris, M.S., Mancini, M., Orsini, E., Gentile, M., Breccia, M., Cuneo, A., Castoldi, G. & Foa, R. (2003) Chronic lymphocytic leukemia patients with highly stable and indolent disease show distinctive phenotypic and genotypic features. *Blood*, **102,** 1035–1041.

Hamblin, T.J., Davis, Z., Gardiner, A., Oscier, D.G. & Stevenson, F.K. (1999) Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. *Blood*, **94,** 1848–1854.

Hamblin, T.J., Orchard, J.A., Gardiner, A., Oscier, D.G., Davis, Z. & Stevenson, F.K. (2000) Immunoglobulin V genes and CD38 expression in CLL. *Blood*, **95,** 2455–2457.

Hamblin, T., Orchard, J.A., Ibbotson, R.E., Davis, Z., Thomas, P.W., Stevenson, F.K. & Oscier, D.G. (2002) CD38 expression and immunoglobulin variable region mutations are independent prognostic variables in chronic lymphocytic leukemia, but CD38 expression may vary during the course of the disease. *Blood*, **99,** 1023–1029.

Herve, M., Xu, K., Ng, Y.S., Wardemann, H., Albesiano, E., Messmer, B.T., Chiorazzi, N. & Meffre, E. (2005) Unmutated and mutated chronic lymphocytic leukemias derive from self-reactive B cell precursors despite expressing different antibody reactivity. *Journal of Clinical Investigation*, **115,** 1636–1643.

Honegger, A. & Pluckthun, A. (2001) Yet another numbering scheme for immunoglobulin variable domains: an automatic modeling and analysis tool. *Journal of Molecular Biology*, **309,** 657–670.

Johnson, T.A., Rassenti, L.Z. & Kipps, T.J. (1997) Ig VH1 genes expressed in B cell chronic lymphocytic leukemia exhibit distinctive molecular features. *Journal of Immunology*, **158,** 235–246.

Krober, A., Seiler, T., Benner, A., Bullinger, L., Bruckle, E., Lichter, P., Dohner, H. & Stilgenbauer, S. (2002) VH mutational status, CD38 expression level, genomic aberrations, and survival in chronic lymphocytic leukemia. *Blood*, **1000,** 1410–1416.

Krober, A., Bloehdorn, J., Hafner, S., Buhler, A., Seiler, T., Kienle, D., Winkler, D., Bangerter, M., Schlenk, R.F., Benner, A., Lichter, P., Dohner, H. & Stilgenbauer, S. (2006) Additional genetic high-risk features such as 11q deletion, 17p deletion, and V3-21 usage characterize discordance of ZAP-70 and VH mutation status in chronic lymphocytic leukemia. *Journal of Clinical Oncology*, **24,** 969–975.

Lanemo Myhrinder, A., Hellqvist, E., Sidorova, E., Soderberg, A., Baxendale, H., Dahle, C., Willander, K., Tobin, G., Backman, E., Soderberg, O., Rosenquist, R., Horkko, S. & Rosen, A. (2008) A new perspective: molecular motifs on oxidized LDL, apoptotic cells, and bacteria are targets for chronic lymphocytic leukemia antibodies. *Blood*, **111,** 3838–3848.

Lefranc, M.P., Giudicelli, V., Kaas, Q., Duprat, E., Jabado-Michaloud, J., Scaviner, D., Ginestoux, C., Clement, O., Chaume, D. & Lefranc, G. (2005) IMGT, the international ImMunoGeneTics information system. *Nucleic Acids Research*, **33,** D593–D597.

Maccallum, R.M., Martin, A.C. & Thornton, J.M. (1996) Antibody-antigen interactions: contact analysis and binding site topography. *Journal of Molecular Biology*, **262,** 732–745.

Marasca, R., Vaccari, P., Luppi, M., Zucchini, P., Castelli, I., Barozzi, P., Cuoghi, A. & Torelli, G. (2001) Immunoglobulin gene mutations and frequent use of VH1-69 and VH4-34 segments in hepatitis C virus-positive and hepatitis C virus-negative nodal marginal zone B-cell lymphoma. *American Journal of Pathology*, **159,** 253–261.

Martin, T., Duffy, S.F., Carson, D.A. & Kipps, T.J. (1992) Evidence for somatic selection of natural autoantibodies. *Journal of Experimental Medicine*, **175,** 983–991.

Messmer, B.T., Albesiano, E., Messmer, D. & Chiorazzi, N. (2004a) The Pattern and distribution of immunoglobulin VH gene mutations in

© 2008 The Authors

chronic lymphocytic leukemia B cells are consistent with the canonical somatic hypermutation process. *Blood*, **103,** 3490–3495.

Messmer, B.T., Albesiano, E., Efremov, D.G., Ghiotto, F., Allen, S.L., Kolitz, J., Foa, R., Damle, R.N., Fais, F., Messmer, D., Rai, K.R., Ferrarini, M. & Chiorazzi, N. (2004b) Multiple distinct sets of stereotyped antigen receptors indicate a role for antigen in promoting chronic lymphocytic leukemia. *Journal of Experimental Medicine*, **200,** 519–525.

Murray, F., Darzentas, N., Hadzidimitriou, A., Tobin, G., Boudjogra, M., Scielzo, C., Laoutaris, N., Karlsson, K., Baran-Marzsak, F., Tsaftaris, A., Moreno, C., Anagnostopoulos, A., Caligaris-Cappio, F., Vaur, D., Ouzounis, C., Belessi, C., Ghia, P., Davi, F., Rosenquist, R. & Stamatopoulos, K. (2008) Stereotyped patterns of somatic hypermutation in subsets of patients with chronic lymphocytic leukemia: implications for the role of antigen selection in leukemogenesis. *Blood*, **111,** 1524–1533.

Orchard, J.A., Ibbotson, R.E., Davis, Z., Wiestner, A., Rosenwald, A., Thomas, P.W., Hamblin, T.J., Staudt, L.M. & Oscier, D.G. (2004) ZAP-70 expression and prognosis in chronic lymphocytic leukaemia. *Lancet*, **363,** 105–111.

Panovska-Stavridis, I., Ivanovski, M., Siljanovski, N., Cevreska, L. & Efremov, D.G. (2007) Chronic lymphocytic leukemia patients with a V1-69 gene rearrangement do not have inferior survival with respect to patients that express other unmutated V(H) genes. *Leukemia Research*, **31,** 245–248.

Pommie, C., Levadoux, S., Sabatier, R., Lefranc, G. & Lefranc, M.P. (2004) IMGT standardized criteria for statistical analysis of immunoglobulin V-REGION amino acid properties. *Journal of Molecular Recognition*, **17,** 17–32.

Rai, K.R., Sawitsky, A., Cronkite, E.P., Chanana, A.D., Levy, R.N. & Pasternack, B.S. (1975) Clinical staging of chronic lymphocytic leukemia. *Blood*, **46,** 219–234.

Rassenti, L.Z. & Kipps, T.J. (1997) Lack of allelic exclusion in B cell chronic lymphocytic leukemia. *Journal of Experimental Medicine*, **185,** 1435–1445.

Rassenti, L.Z., Huynh, L., Toy, T.L., Chen, L., Keating, M.J., Gribben, J.G., Neuberg, D.S., Flinn, I.W., Rai, K.R., Byrd, J.C., Kay, N.E., Greaves, A., Weiss, A. & Kipps, T.J. (2004) ZAP-70 compared with immunoglobulin heavy-chain gene mutation status as a predictor of disease progression in chronic lymphocytic leukemia. *The New England Journal of Medicine*, **351,** 893–901.

Rosenquist, R., Lindstrom, A., Holmberg, D., Lindh, J. & Roos, G. (1999) V(H) gene family utilization in different B-cell lymphoma subgroups. *European Journal of Haematology*, **62,** 123–128.

Sasso, E.H., Silverman, G.J. & Mannik, M. (1989) Human IgM molecules that bind staphylococcal protein A contain VHIII H chains. *Journal of Immunology*, **142,** 2778–2783.

Seppala, I., Kaartinen, M., Ibrahim, S. & Makela, O. (1990) Mouse Ig coded by VH families S107 or J606 bind to protein A. *Journal of Immunology*, **145,** 2989–2993.

Shanafelt, T.D., Geyer, S.M. & Kay, N.E. (2004) Prognosis at diagnosis: integrating molecular biologic insights into clinical practice for patients with CLL. *Blood*, **103,** 1202–1210.

Silverman, G.J. & Goodyear, C.S. (2002) A model B-cell superantigen and the immunobiology of B lymphocytes. *Clinical Immunology*, **102,** 117–134.

Silverman, G.J. & Goodyear, C.S. (2006) Confounding B-cell defences: lessons from a staphylococcal superantigen. *Nature reviews. Immunology*, **6,** 465–475.

Stamatopoulos, K., Belessi, C., Hadzidimitriou, A., Smilevska, T., Kalagiakou, E., Hatzi, K., Stavroyianni, N., Athanasiadou, A., Tsompanakou, A., Papadaki, T., Kokkini, G., Paterakis, G., Saloum, R., Laoutaris, N., Anagnostopoulos, A. & Fassas, A. (2005) Immunoglobulin light chain repertoire in chronic lymphocytic leukemia. *Blood*, **106,** 3575–3583.

Stamatopoulos, K., Belessi, C., Moreno, C., Boudjograh, M., Guida, G., Smilevska, T., Belhoul, L., Stella, S., Stavroyianni, N., Crespo, M., Hadzidimitriou, A., Sutton, L., Bosch, F., Laoutaris, N., Anagnostopoulos, A., Montserrat, E., Fassas, A., Dighiero, G., Caligaris-Cappio, F., Merle-Beral, H., Ghia, P. & Davi, F. (2007) Over 20% of patients with chronic lymphocytic leukemia carry stereotyped receptors: pathogenetic implications and clinical correlations. *Blood*, **109,** 259–270.

Stevenson, F.K. & Caligaris-Cappio, F. (2004) Chronic lymphocytic leukemia: revelations from the B-cell receptor. *Blood*, **103,** 4389–4395.

Sthoeger, Z.M., Wakai, M., Tse, D.B., Vinciguerra, V.P., Allen, S.L., Budman, D.R., Lichtman, S.M., Schulman, P., Weiselberg, L.R. & Chiorazzi, N. (1989) Production of autoantibodies by CD5-expressing B lymphocytes from patients with chronic lymphocytic leukemia. *Journal of Experimental Medicine*, **169,** 255–268.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. & Higgins, D.G. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, **25,** 4876–4882.

Thorselius, M., Krober, A., Murray, F., Thunberg, U., Tobin, G., Buhler, A., Kienle, D., Albesiano, E., Maffei, R., Dao-Ung, L.P., Wiley, J., Vilpo, J., Laurell, A., Merup, M., Roos, G., Karlsson, K., Chiorazzi, N., Marasca, R., Dohner, H., Stilgenbauer, S. & Rosenquist, R. (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.

Tobin, G., Thunberg, U., Johnson, A., Thorn, I., Soderberg, O., Hultdin, M., Botling, J., Enblad, G., Sallstrom, J., Sundstrom, C., Roos, G. & Rosenquist, R. (2002) Somatically mutated Ig V(H)3-21 genes characterize a new subset of chronic lymphocytic leukemia. *Blood*, **99,** 2262–2264.

Tobin, G., Thunberg, U., Johnson, A., Eriksson, I., Soderberg, O., Karlsson, K., Merup, M., Juliusson, G., Vilpo, J., Enblad, G., Sundstrom, C., Roos, G. & Rosenquist, R. (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.

Tobin, G., Thunberg, U., Karlsson, K., Murray, F., Laurell, A., Willander, K., Enblad, G., Merup, M., Vilpo, J., Juliusson, G., Sundstrom, C., Soderberg, O., Roos, G. & Rosenquist, R. (2004a) Subsets with restricted immunoglobulin gene rearrangement features indicate a role for antigen selection in the development of chronic lymphocytic leukemia. *Blood*, **104,** 2879–2885.

Tobin, G., Soderberg, O., Thunberg, U. & Rosenquist, R. (2004b) V(H)3-21 gene usage in chronic lymphocytic leukemia – characterization of a new subgroup with distinct molecular features and poor survival. *Leukemia & Lymphoma*, **45,** 221–228.

Tschumper, R.C., Geyer, S.M., Campbell, M.E., Kay, N.E., Shanafelt, T.D., Zent, C.S., Nowakowski, G.S., Call, T.G., Dewald, G.W. & Jelinek, D.F. (2008) Immunoglobulin diversity gene usage predicts

© 2008 The Authors

unfavorable outcome in a subset of chronic lymphocytic leukemia patients. *Journal of Clinical Investigation*, **118,** 306–315.

Vargas-Madrazo, E., Lara-Ochoa, F., Ramirez-Benites, M.C. & Almagro, J.C. (1997) Evolution of the structural repertoire of the human V(H) and Vkappa germline genes. *International Immunology*, **9,** 1801–1815.

Widhopf, G.F., Rassenti, L.Z., Toy, T.L., Gribben, J.G., Wierda, W.G. & Kipps, T.J. (2004) Chronic lymphocytic leukemia B cells of more than 1% of patients express virtually identical immunoglobulins. *Blood*, **104,** 2499–2504.

Widhopf, G.F., Goldberg, C.J., Toy, T.L., Rassenti, L.Z., Wierda, W.G., Byrd, J.C., Keating, M.J., Gribben, J.G., Rai, K.R. & Kipps, T.J. (2008) Nonstochastic pairing of immunoglobulin heavy and light chains expressed by chronic lymphocytic leukemia B cells is predicated on the heavy chain CDR3. *Blood*, **111,** 3137–3144.

Yousfi, M.M., Giudicelli, V., Chaume, D. & Lefranc, M.P. (2004) IMGT/JunctionAnalysis: the first tool for the analysis of the immunoglobulin and T cell receptor complex V-J and V-D-J JUNCTIONs. *Bioinformatics*, **20**(Suppl. 1)**,** i379–i385.

Zenz, T., Dohner, H. & Stilgenbauer, S. (2007) Genetics and risk-stratified approach to therapy in chronic lymphocytic leukemia. *Best Practice & Research. Clinical Haematology*, **20,** 439–453.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Fig S1. Prognostic impact of clinical and biological known prognosticators on CLL patients.

Fig S2. Alignment of *IGHV* gene sequences of M clusters.

Fig S3. Kaplan–Meier TTT curve analysis in CLL patients included or not included in clusters of stereotyped BCR.

Table SIA. *IGHV* subgroups usage: distribution in UM and M IG rearrangements.

Table SIB. *IGHD* subgroup usage: distribution in UM and M IG rearrangements.

Table SIC. *IGHJ* gene usage: distribution in UM and M IG rearrangements.

Table SII. *IGHV* gene usage: distribution of the most frequently used genes in UM and M cases.

Table SIII. *IGHD* gene usage: distribution of the most frequently used genes in UM and M cases.

Table SIV. Cluster alignment.

Table SV. Recurrent amino acid changes in CLL IG rearrangements belonging to M clusters.

Appendix SI. Supplemental results.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

© 2008 The Authors