bjh british journal of haematology

The outcome of Chronic Lymphocytic Leukaemia patients with 97.0% IGHV gene identity to germline is distinct from cases with <97% identity and similar to those with 98% identity

Journal:	*British Journal of Haematology*
Manuscript ID	BJH-2015-01577
Manuscript Type:	Ordinary Papers
Date Submitted by the Author:	15-Sep-2015
Complete List of Authors:	Davis, Zadie; Royal Bournemouth Hospital, Molecular Pathology Forconi, Francesco; University of Southampton, Cancer Sciences Unit, CRUK Clinical Centre Parker, Anton; Royal Bournemouth Hospital, Department of Haematology Gardiner, Anne; Royal Bournemouth Hospital, Department of Haematology Thomas, Peter; Bournemouth University, Bournemouth University Clinical Research Unit Catovsky, Daniel; Institute of Cancer research, Haemato-oncology; Rose-Zerilli, Matthew; University of Southampton, Cancer Sciences Strefford, J; University of Southampton, Cancer Sciences Oscier, David; Royal Bournemouth Hospital, Department of Haematology
Key Words:	CHRONIC LYMPHOCYTIC LEUKAEMIA, V-GENES, MUTATION ANALYSIS, PROGNOSTIC FACTORS

The outcome of Chronic Lymphocytic Leukaemia patients with 97.0% *IGHV* gene identity to germline is distinct from cases with <97% identity and similar to those with 98% identity.

Authors: Zadie, Davis[1], Francesco Forconi[2,] Anton Parker[1], Anne Gardiner[1], Peter Thomas[3], Daniel Catovsky[4], Matthew Rose-Zerilli[2], Jonathan C Strefford[2] and David Oscier[1]

Affiliations:
[1] Department of Molecular Pathology, Royal Bournemouth Hospital, UK
[2]Cancer Sciences, University of Southampton, UK
[3]Clinical Research Unit, Bournemouth University, UK
[4]Department of Haemato-oncology, Institute for Cancer Research, Sutton, UK

Corresponding Author
Zadie Davis: Castle Lane East, Bournemouth, Dorset UK BH7 7DW
Tel.+44(0)1202704805 Fax.+44(0)1202309925 email zadie.davis@rbch.nhs.uk

SUMMARY

IGHV gene mutational status has prognostic significance in CLL but the percentage of mutations which correlates best with clinical outcome remains controversial. We performed *IGHV* gene sequencing on 1018 patients with CLL (558 followed from diagnosis and 460 entered into the UK CLL4 trial).

In the diagnostic cohort, excluding subset #2, significant differences in median TTFT in Stage A patients and in OS in the whole cohort, were observed between cases with <97% and 97-98.99% identity and between cases with 97-98.99% and ≥99% identity. A significant difference in PFS and OS was also observed in UK CLL4 trial cases between those with <97% and 97-98.99% identity, but not between cases 97-98.99% and ≥99% identity.

Cox Regression analyses in the Stage A cohort revealed that a model which incorporated <97%, 97-98.99%, ≥99% identity as subgroups, was a better predictor of TTFT in CLL than using the 98% cut-off. Multivariate analysis selected the three mutational subgroups as independent predictors of TTFT in Stage A patients, and of OS in the diagnostic cohort.

This study highlights that cases with 97% identity should not be considered to have the same prognosis as other cases with mutated *IGHV* genes defined as <98% identity to germline.

Keywords: chronic lymphocytic leukaemia, V-genes, mutation analysis, prognostic factors

INTRODUCTION

In 1999, two studies showed that patients with chronic lymphocytic leukemia (CLL) whose immunoglobulin variable region genes (*IGHV*) had undergone somatic hypermutation (SHM) leading to <98% identity to the germline sequence (M-CLL) had a significantly better overall survival than patients with 'unmutated' *IGHV* genes (U-CLL) (Damle *et al*, 1999; Hamblin *et al*, 1999). Many subsequent retrospective and prospective studies have confirmed this observation and shown that *IGHV* gene mutational status using a 98% cut-off also predicts disease progression, time to first treatment and progression free survival. However, the mathematical cut-off which correlates best with clinical outcome still remains controversial; we and others have previously shown that cut-off's using 97% or 95% identity to the germline sequence have clinical relevance and that cases with 97% identity may have an intermediate outcome compared to those with <97% or >97% identity (Hamblin *et al*, 2008; Krober *et al*, 2002; Oscier *et al*, 2010; Tobin *et al*, 2005).

The 98% cut-off was originally chosen to avoid scoring polymorphic variants as mutations. However, 98% and 99% identity are frequently the consequence of SHM rather than polymorphisms (Davis *et al*, 2003) and even single base changes can affect the binding specificity of the B-cell receptor (BcR) (Barbas *et al*, 1995; Murray *et al*, 2008). In addition, methodological factors such as the choice of PCR primers and immunoglobulin database may affect calculation of percentage identity to germline and classification of a case as mutated or unmutated. Accordingly, International guidelines recommend caution in assigning mutational status and predicted outcome in cases with 'borderline' *IGHV* identity (Ghia *et al*, 2007).

It has also become increasingly clear that other features of the immunoglobulin gene sequence, in addition to mutational load, can influence prognosis. In 2002, it was observed that usage of the *IGHV*3-21 gene was an independent unfavourable prognostic marker irrespective of *IGHV* gene mutational status (Tobin *et al*, 2002). The same group then demonstrated that a subset of CLL cases utilising *IGHV*3-21 had a very short, highly similar, if not identical, heavy complementarity-determining region 3 (VH CDR3) and showed biased utilisation of the *IGLV*3-21 gene (Tobin *et al*, 2003). In 2004 the term 'stereotyped' BcR was introduced that now describes the approximately 30% of CLL cases with similar VH CDR3 amino acid sequences (Messmer *et al*, 2004; Darzentas *et al*, 2010). Many of these stereotypic subsets have distinctive biological features such as a characteristic gene expression and methylation profile, antigen specificity, immune signaling outcomes, intraclonal diversification, immunoglobulin class switching, as well as clinical correlations with genomic

copy number abnormalities and mutations, risk of lymphomatous transformation and clinical course (Stamatopoulos *et al*, 2007; Bomben *et al*, 2009; Marincevic *et al*, 2010; Rossi *et al*, 2009; Sutton *et al*, 2009; Chu *et al*, 2010; Kanduri *et al*, 2012; Ntoufa *et al*, 2012; Strefford *et al*, 2013; Baliakis *et al*, 2014, 2015).

We have further addressed the clinical significance of borderline *IGHV* identity in 1018 CLL patients comprising a diagnostic cohort with predominantly Binet stage A disease and a second cohort entered into the UKCLL4 trial. The outcome of patients with 97% identity was comparable to those with 98% identity in both cohorts and distinct from other cases currently classified as having M-CLL.

MATERIALS AND METHODS

Patients

Blood samples were obtained with patient consent at or close to diagnosis from 350 patients diagnosed at the Royal Bournemouth Hospital (Cohort 1) and 208 patients diagnosed in Siena (Cohort 2), and at trial entry from 460 patients entered into the UK LRF CLL4 trial. The diagnosis of CLL was based on the revised International Workshop Chronic Lymphocytic Leukemia/National Cancer Institute (IWCLL/NCI) guidelines for cases from cohorts 1 and 2 and the 1996 NCI guidelines for patients entered into the UK CLL4 trial. The study was approved by the local ethics committee of the participating institutions.

Analysis of IGHV-IGHD-IGHJ gene rearrangements

IGHV-IGHD-IGHJ gene rearrangements were sequenced as previously described (Hamblin *et al*, 2008). All sequences were aligned to the IMGT/V-QUEST database prior to the recent updated version 3.3.0 (20 February 2014) and considered mutated if their percentage identity to germline was <98%. As the identification of new *IGHV* genes and new alleles (Xochelli *et al*, 2015) will influence the precise determination of SHM status, all of cohort 1 and CLL4 cases identified as having 96-98.99% identity to the germline sequence were re-aligned to the updated IMGT/V-QUEST programme version: 3.3.0 (20 February 2014) - IMGT/V-QUEST reference directory release: 201414-4 (3 April 2014). Reanalysis of these cases was chosen as any changes observed in these patients would potentially affect which mutational subgroup they would be assigned to.

All cases were assessed for stereotypy by comparing *IGHV* gene usage, CDR3 sequence and CDR3 length with the published sequences of the 19 “major” stereotypic subsets (Agathangelidis *et al*, 2012) and assigned to one of these subsets if these criteria were met.

Other Prognostic Biomarkers

Flow cytometric testing for ZAP-70 and CD38 expression, cytogenetic analysis, FISH for del(13q), del(11q), del(17p) and trisomy 12 and *NOTCH1* and *SF3B1* mutation screening were performed as previously described (Oscier *et al*, 2002; Orchard *et al*, 2004; Oscier *et al*, 2013).

Statistical Analysis

Comparative and statistical analysis was performed using IBM SPSS Statistics 21 Data Mining and Statistical Analysis Software (IBM Corp. Armonk, NY USA). For patients in cohorts 1 and 2, time to first treatment (TTFT) and overall survival (OS) were calculated from date of diagnosis to date of first treatment or death respectively and censored if these were not reached. For patients entered into the UK CLL4 trial PFS was expressed as time from randomization to relapse needing further therapy, progression, or death from any cause, and OS was calculated from trial entry to death from any cause. Curves were generated using the method of Kaplan and Meier. The relationship between known prognostic markers and TTFT, PFS and OS was assessed and p-values were generated using the SPSS software.

RESULTS

Clinical and Laboratory Features and Cohort Comparison

Clinical and laboratory findings for the three cohorts are summarised in Table I. Significant differences between the cohorts are indicated with an asterisk. The median follow-up was 7.31 years (range 0-35 years). 687 patients were followed up for >5 years and 304 for >10 years. In both cohorts 1 and 2, 90% of patients requiring treatment were treated initially and at relapse with an alkylating agent and / or purine analogue, according to guidelines that were current at the time of treatment. Patients entered into the UK CLL4 trial were randomised to receive either chlorambucil or fludarabine alone or in combination with cyclophosphamide (Catovsky *et al*, 2007). Cytogenetic and FISH analysis were carried out prior to treatment in 79% of cohort 1 cases and 46% were tested within 2 years of diagnosis. In cohort 2 analyses were carried out at diagnosis or prior to treatment and for the CLL4 cohort, analysis was carried out at trial entry (pre-treatment), with 67% being within 2 years of diagnosis.

As no significant differences were observed between cohorts 1 and 2 for any of the observations summarised in Table I or in TTFT of stage A cases with either M-CLL (median 273 and 242 months respectively; p=0.120) or U-CLL (median 54 and 48 months respectively; p=0.066), the stage A cases of the two cohorts were combined to give a cohort size of 460 cases – the stage A cohort.

Influence of mutational load on outcome

IGHV-IGHD-IGHJ gene analysis was performed on all 1018 cases. To assess the mutational cut-off that provided the best predictor of outcome, analysis was initially performed on the stage A cohort using TTFT as an endpoint to obviate the effect of varying treatment regimens. We carried out Cox regression analysis to determine hazard ratios (HR) for each mutational interval differing by 1% from 95% - 100% identities, comparing the TTFT. When compared to cases with <95% identity (median TTFT of 273 month), a significant difference in TTFT first occurred at 97% with a median TTFT of 102 and a HR of 2.2, whereas, the median TTFT and HRs for 98%, 99% and 100% were 44, 48 and 55 months and 3.7, 4.4 and 4.8 respectively (Table II). This analysis suggested that stage A cases could be divided into three subgroups with differing median TTFT's based on their mutational status; those with <97% identity having a low risk of a short TTFT (HR 0-1.6) , those with 97% identity having an intermediate risk (HR 2.2) - and those with ≥98% identity having a high risk (HR 3.7-4.8).

Influence of BCR Stereotypy

Next, we wished to determine whether the different outcome for Stage A patients with 97% identity was a consequence of a differing incidence of major stereotypes, especially stereotypic subset #2, which is well documented (Baliakas *et al*, 2015) and confirmed here (Fig S1) to be associated with poor outcome independent of *IGHV* mutational status and is enriched with cases with a low mutational load.

Within the stage A cohort, 42/460 (9.1%) cases were assigned to one of 14 of the 19 ‘major’ stereotypic subsets (Agathangelidis *et al*, 2012) with the following distribution: low risk subgroup - 4.5%, intermediate - 3% and high risk subgroup - 18.2%. Subsets #1 and #2 were the most frequently observed stereotypic subsets among the stage A cohort, both accounting for 21% (9/42) each of all stereotypic cases.

Cox regression analysis for each mutational interval (95-100%) was initially repeated excluding all major stereotype cases, and now the HR and the median TTFT for cases with 97% and 98% were very similar; with a median TTFT of 102 and 105 months and HRs of 2.3 and 2.6 respectively (Table III). Secondly, because 4/5 stereotypic cases with 97% or 98% identity belonged to subset #2, analysis was repeated with the exclusion of this subset only. The HR and TTFT’s of the 97% and 98% identity cases remained very similar or the same (HRs 2.2 and 2.9, TTFT 102 and 105 months respectively) as the previous analysis when all major stereotypes were excluded, indicating that the effect of stereotypy on TTFT was a consequence of subset #2. Exclusion of subset #2 enabled the remaining cases to be divided into 3 risk groups: 1) low risk HR - 1 (<97%), 2) intermediate risk HR - 2.3 (97-98.99%) and 3) high risk HR - 4.3(≥99%).

We then carried out proportional hazard analysis without subset #2 cases, comparing the standard cut-off (98%) to either a model with 97% as the cut-off or the 3 mutational subgroup model (<97%, 97-98.99% and ≥99%). We found no difference in the models using either 98% or 97% as a cut-off (p=0.338) but the 3 subgroup model showed a significant improvement at predicting TTFT than both of these models (p=0.005 and p=0.003 respectively).

Kaplan Meier survival curves and the median TTFT for stage A cases were then determined for the 3 subgroup model without subset #2. Of note there was a significant difference in TTFT between <97% and 97-98.99% (median TTFT not reached and 105 months respectively; p<0.001) and between 97-98.99% and ≥99% (105 and 52 months; p=0.002) (Fig 1A).

When this model was applied to cohorts 1 and 2, including those with Binet stage B and C disease, to assess OS, a significant difference was observed between each of the three subgroups; with a median OS of 231, 141 and 111 months respectively (p=0.003 and p=0.005) (Fig 1B).

UK CLL4 cohort

Within the CLL4 cohort 96/460 (20.9%) cases were assigned to 18/19 major stereotypic subsets; subset #2 constituted 36.5% (35/82) of all stereotypic cases and 25/35 had an intermediate identity of 97% or 98%. In light of this, subset #2 cases were once again excluded from analysis when the three subgroup model was assessed in the CLL4 cohort.

Initially, proportional hazard model analysis was repeated in the CLL4 cohort using PFS as the predictor of outcome. Again a significant improvement in predicting outcome was observed when using the three mutational subgroups over the current 98% cut-off (p=0.003), regardless of whether treatment arm was included in the analysis (p<0.001).

Kaplan Meier survival curves and median survival times were then determined using the 3 mutational subgroups. A significant difference was noted between PFS of cases with <97% identity and those with 97-98.99%, (median PFS 42 and 30 months respectively; p<0.001) however, significance was not reached between those with 97-98.99% and those with ≥99% (median PFS 20; p=0.124) (Fig 1C). Similarly, a significant difference in OS was observed between cases with <97% identity and those with 97-98.99% (p<0.001), with a median OS of 121 months and 62 months respectively, but not between those with 97-98.99% and ≥99% (median OS 61 months; p=0.323) (Fig 1D)

Multivariate survival analyses

Multivariate survival analyses were then performed in models that included the following parameters: age at diagnosis, gender, CD38 and ZAP 70 positivity, del(13q), trisomy 12, del(11q), del(17p), *NOTCH1* and *SF3B1* mutation and treatment arm. Tables SI and SII highlight which biomarkers were significant in univariate analysis for either TTFT in the stage A cohort, PFS in

CLL4 or OS for cohorts 1 and 2, and CLL4. The three subgroup model emerged as an independent predictor of TTFT in the stage A cohort, together with del(11q) and age at diagnosis, and also as an independent marker of OS together with age at diagnosis, CD38 expression and del(17p) in all cases from cohorts 1 and 2 (Table IV). The analyses were then repeated using the intermediate subgroup (97-98.99% identity) as the reference category to confirm the independent significance of each mutational subgroup, and all the above factors remained significant independent predictors of TTFT and OS respectively (Table SIII).

Multivariate survival analyses of the CLL4 cohort showed that the three mutational subgroup model was an independent predictor of PFS, together with del(11q), del(17p) and treatment type, and of OS, together with age, gender, CD38, del(11q) and del(17p) (Table V). When analyses was repeated using the intermediate subgroup (97-98.99% identity) as the reference category, significance was lost for the high risk subgroup (≥99% identity) for both PFS and OS (Table SIV).

Investigation of other factors that might account for the intermediate outcome of cases with 97.0 to 98.99% identity

Methodological factors

We initially sought to exclude the possibility that the intermediate TTFT of cases with 97-98.99% identity reflected the calculation of *IGHV* gene mutational load when different PCR primers were used for analysis. Seventy percent of cohorts 1 and 2 were amplified with leader primer and TTFT and OS analysis of the three mutational subgroups based on, only those cases amplified using the leader primer did not change the statistically significant differences between the three subgroups.

In addition, as a result of the recent update to the IMGT/V-QUEST reference directory, available sequences with borderline identities across the subgroups (152/204 96-98.99%) were reanalysed using IMGT/V-QUEST v.3.3.0 reference directory release: 201414-4 (3 April 2014) which includes all newly identified *IGHV* genes and alleles. Only 1/152 sequences from a patient with Binet stage C disease in the combined cohorts 1 and 2 changed from 98% to 99% identity. Analysis of OS in cohorts 1 and 2 was repeated with no significant changes observed; these reanalysed results are shown in Figure 1B.

Finally, the frequency of polymorphisms was assessed in 6 cases with 97% identity utilising *IGHV*3-23 by comparing the CLL *IGHV* gene sequence with its corresponding germline sequence. All genomic variations classified as mutations based on the CLL sequence were confirmed as mutations.

IGHV Gene Usage

IGHV gene use differed among the three mutational subgroups (Table SV). To determine whether the intermediate outcome of cases with 97-98.99% identity is due to the biased use of specific *IGHV* genes, we focused on stage A cases using the *IGHV*3-23 gene, the most frequently utilised gene in the intermediate subgroup. TTFT was reanalysed excluding cases utilising *IGHV*3-23 and the significant difference in TTFT between the intermediate risk group and both the low and high risk groups persisted (p=0.001 and p=0.001 respectively).

In addition, when *IGHV*3-23 cases alone were assessed using the three subgroup model, TTFT was significantly shorter for *IGHV*3-23 cases in the intermediate subgroup (median TTFT 45 months; p=0.022) compared to those with <97% (median TTFT 273 months). These results support the conclusion that *IGHV* gene usage alone is unlikely to account for the intermediate TTFT of 97%-98.99% identity cases.

SF3B1 and NOTCH1 mutations

As *SF3B1* and *NOTCH1* mutations have adverse prognostic significance in CLL and their incidence differs between M-CLL and U-CLL and among stereotypic subsets, we compared the incidence of these mutations among the 3 mutational risk groups (Strefford *et al*, 2013; Oscier *et al*, 2013; Rossi *et al*, 2011). *SF3B1* and *NOTCH1* mutation status was available for 315 CLL4 cases and for 159 cohort 1 cases (excluding subset #2 cases). Of note the incidence of *SF3B1* mutated cases was significantly higher in the intermediate subgroup than the low risk group (27% and 5% respectively; p=<0.001) and there was also a trend towards a higher incidence in the intermediate group than in the high risk group (15%; p=0.0564). The incidence of *NOTCH1* mutated cases was significantly less in the low risk subgroup than in the high risk subgroup (1% and 13% respectively; p<0.001) but the incidence in the intermediate subgroup (6%) was not significantly different from either of the other subgroups.

There were too few cases in each of the subgroups to carry out survival analysis in the stage A cohort, but as highlighted earlier *SF3B1* and *NOTCH1* did not show independent significance in multivariate analysis in the CLL4 cohort.

DISCUSSION

This is the largest study to address the optimal cut-off for *IGHV* identity to the germline sequence which correlates best with clinical outcome. The analysis was based on *IGHV* gene sequence data from 1018 well characterised patients with CLL who were either studied from time of randomisation into the UK CLL4 trial (Catovsky *et al*, 2007) (n=460) or who presented in Bournemouth or Siena (n=558) predominantly with Binet Stage A disease (n=460). As there were no significant differences in *IGHV* gene mutation status and usage, biomarker results, demographics and TTFT between Stage A patients presenting in Bournemouth or Siena, nor in the indications for, and type of treatments used in each centre, data from both centres were analysed as a single cohort.

Initial analysis of TTFT in the stage A cohort suggested that patients with 97% identity had an intermediate outcome between those with <97% and >97% identity. When cases with stereotyped subset #2 were excluded due to their known association with poor outcome and higher incidence among cases with a low mutational load (Baliakis *et al*, 2014, 2015), we then observed that the median TTFT in patients presenting with Stage A disease was comparable for patients with either 97% or 98% identity, and that cases could still be subdivided into 3 risk groups: a low risk group (<97% identity), an intermediate risk group (97-98.99% identity) and a high risk group (≥99% identity) each with a successively significant shorter TTFT. This 3 mutational group model emerged as an independent predictor of TTFT and of OS in the Bournemouth/Siena patients in multivariate analyses which included significant variables identified in univariate analysis (age, expression of CD38 and ZAP70 and cytogenetic abnormalities).

In the UK CLL4 trial cohort, again excluding stereotyped subset #2, cases with 97-98.99% identity had a significantly shorter PFS and OS than those with <97% identity but there was no difference between the 97-98.99% and ≥99% groups. Multivariate analyses showed that the 3 mutational group model was an independent predictor of PFS and OS only when <97% identity cases (but not 97-98.99% identity cases) were used as the reference category.

U-CLL and M-CLL, defined by using 98% identity as a cut-off, are widely considered to arise from two distinct subsets of normal B cells at different stages of differentiation. The differing consequences of engagement of the BcR by antigen within the CLL microenvironment (Stevenson *et al*, 2011; Packham *et al*, 2014) in these subsets exerts a strong influence on clinical outcome. Although it is well recognised that there is both biological and clinical heterogeneity within the 2 mutational subsets, especially M-CLL (Oscier *et al*, 2010; Pflug *et al*, 2014; Rassenti *et al*, 2008; Bulian *et al* 2014; Pepper *et al*, 2015; Rossi *et al*, 2009; Paterson *et al*, 2012), one implication of this two-subset model is that cases with 97% identity might be expected to show

greater clinical similarity to cases with higher mutational loads than to cases with 98% identity. Data from our cohorts strongly suggest that the outcome of patients with 97% identity is more similar to those with 98% identity than to those with greater *IGHV* mutational loads. We could identify neither methodological nor biological factors to account for the poorer outcome of cases with 97% identity compared to those with <97% identity, although the higher incidence of *SF3B1* mutations in the intermediate risk group, even without stereotypic subset #2 cases, is of interest and requires confirmation in a larger cohort.

Recently an epigenetic signature based on the methylation status of 5 CpG sites has been identified which reliably classified CLL cases into 3 clusters with differing clinical outcomes. The authors postulated that the intermediate methylation cluster, which is enriched with cases having a low *IGHV* mutational load, might reflect CLL cases derived from an antigen-experienced, germinal centre-independent marginal zone B cell (Queirós *et al*, 2015). A recent, subsequent study integrated this epigenetic data with genomic data and noted higher frequencies of *SF3B1* and *MYD88* mutations, biased usage of *IGVH*3-21 and *IGHV*1-18 genes and of stereotypic subset #2 in the intermediate methylation cluster (Puente *et al*, 2015).

In summary, regardless of the mechanisms by which *IGHV* mutational load influences the biology of CLL, our results add weight to previous studies suggesting that cases with 97% identity should not be considered to have the same prognosis as other cases with mutated *IGHV* genes defined as <98% identity to germline and this may be especially useful in the prognostic evaluation of patients with Stage A disease. In addition, although stereotypic subset analysis is important in defining clinically relevant subgroups, approximatley 70% of cases cannot be assigned to a stereotype and refining mutational status analysis will provide valuable clinical information for these patients. Finally, although our study is large in comparison to previous studies, only 11.5% of cases had 97-98.99% identity and additional larger studies will be required both to confirm our findings in patients presenting with Stage A disease and establish whether the 3 risk groups have prognostic significance in patients requiring treatment.

Acknowledgements

The authors would like to thank the Bournemouth Leukaemia Fund, Leukaemia Lymphoma Research and the Kay Kendal Leukaemia Fund for funding this study.
ZD, AG and MRZ performed research, ZD and DO designed the research study and wrote the manuscript, ZD, AP and PT analysed the data, DO, FF and DC provided clinical data, DC, JS and MRZ contributed to revising the manuscript.

REFERENCES

Agathangelidis, A., Darzentas, N., Hadzidimitriou, A., Brochet, X., Murray, F., Yan, X.-J., Davis, Z., van Gastel-Mol, E.J., Tresoldi, C., Chu, C.C., Cahill, N., Giudicelli, V., Tichy, B., Pedersen, L.B., Foroni, L., Bonello, L., Janus, A., Smedby, K., Anagnostopoulos, A., Merle-Beral, H., et al (2012) Stereotyped B-cell receptors in one-third of chronic lymphocytic leukemia: a molecular classification with implications for targeted therapies. *Blood*, **119,** 4467–75

Baliakas, P., Agathangelidis, A., Hadzidimitriou, A., Sutton, L.-A., Minga, E., Tsanousa, A., Scarfò, L., Davis, Z., Yan, X.-J., Shanafelt, T., Plevova, K., Sandberg, Y., Vojdeman, F.J., Boudjogra, M., Tzenou, T., Chatzouli, M., Chu, C.C., Veronese, S., Gardiner, A., Mansouri, L., et al (2015) Not all IGHV3-21 chronic lymphocytic leukemias are equal: prognostic considerations. *Blood*, **125,** 856–9

Baliakas, P., Hadzidimitriou, A., Sutton, L.-A., Minga, E., Agathangelidis, A., Nichelatti, M., Tsanousa, A., Scarfò, L., Davis, Z., Yan, X.-J., Shanafelt, T., Plevova, K., Sandberg, Y., Juhl Vojdeman, F., Bo, M., Stamatopoulos, K. (2014) Clinical effect of stereotyped B-cell receptor immunoglobulins in chronic lymphocytic leukaemia: a retrospective multicentre study. *The Lancet Haematology*, **1,** e47–e84

Barbas, S.M., Ditzel, H.J., Salonen, E.M., Yang, W.P., Silverman, G.J. & Burton, D.R. (1995) Human autoantibody recognition of DNA. *Proceedings of the National Academy of Sciences of the United States of America*, **92,** 2529–33

Bomben, R., Dal Bo, M., Capello, D., Forconi, F., Maffei, R., Laurenti, L., Rossi, D., Del Principe, M.I., Zucchetto, A., Bertoni, F., Rossi, F.M., Bulian, P., Cattarossi, I., Ilariucci, F., Sozzi, E., Spina, V., Zucca, E., Degan, M., Lauria, F., Del Poeta, G., et al (2009) Molecular and clinical features of chronic lymphocytic leukaemia with stereotyped B cell receptors: results from an Italian multicentre study. *British journal of haematology*, **144,** 492–506

Bulian, P., Shanafelt, T.D., Fegan, C., Zucchetto, A., Cro, L., Nückel, H., Baldini, L., Kurtova, A. V, Ferrajoli, A., Burger, J.A., Gaidano, G., Del Poeta, G., Pepper, C., Rossi, D. & Gattei, V. (2014) CD49d is the strongest flow cytometry-based predictor of overall survival in chronic lymphocytic leukemia. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, **32,** 897–904

Catovsky, D., Richards, S., Matutes, E., Oscier, D., Dyer, M.J.S., Bezares, R.F., Pettitt, A.R., Hamblin, T., Milligan, D.W., Child, J.A., Hamilton, M.S., Dearden, C.E., Smith, A.G., Bosanquet, A.G., Davis, Z., Brito-Babapulle, V., Else, M., Wade, R. & Hillmen, P. (2007) Assessment of fludarabine plus cyclophosphamide for patients with chronic lymphocytic leukaemia (the LRF CLL4 Trial): a randomised controlled trial. *Lancet (London, England)*, **370,** 230–9

Chu, C.C., Catera, R., Zhang, L., Didier, S., Agagnina, B.M., Damle, R.N., Kaufman, M.S., Kolitz, J.E., Allen, S.L., Rai, K.R. & Chiorazzi, N. (2010) Many chronic lymphocytic leukemia antibodies recognize apoptotic cells with exposed nonmuscle myosin heavy chain IIA: implications for patient outcome and cell of origin. *Blood*, **115,** 3907–15

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,** 1840–7

Darzentas, N., Hadzidimitriou, A., Murray, F., Hatzi, K., Josefsson, P., Laoutaris, N., Moreno, C., Anagnostopoulos, A., Jurlander, J., Tsaftaris, A., Chiorazzi, N., Belessi, C., Ghia, P., Rosenquist, R., Davi, F. & Stamatopoulos, K. (2010) A different ontogenesis for chronic lymphocytic leukemia cases carrying stereotyped antigen receptors: molecular and computational evidence. *Leukemia*, **24,** 125–32

Davis, Z.A., Orchard, J.A., Corcoran, M.M. & Oscier, D.G. (2003) Divergence from the germ-line sequence in unmutated chronic lymphocytic leukemia is due to somatic mutation rather than polymorphisms. *Blood*, **102,** 3075

Ghia, P., Stamatopoulos, K., Belessi, C., Moreno, C., Stilgenbauer, S., Stevenson, F., Davi, F. & Rosenquist, R. (2007) ERIC recommendations on IGHV gene mutational status analysis in chronic lymphocytic leukemia. *Leukemia*, **21,** 1–3

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., Davis, Z.A. & Oscier, D.G. (2008) Determination of how many immunoglobulin variable region heavy chain mutations are allowable in unmutated chronic lymphocytic leukaemia - long-term follow up of patients with different percentages of mutations. *British journal of haematology*, **140,** 320–3

Kanduri, M., Marincevic, M., Halldórsdóttir, A.M., Mansouri, L., Junevik, K., Ntoufa, S., Kultima, H.G., Isaksson, A., Juliusson, G., Andersson, P.-O., Ehrencrona, H., Stamatopoulos, K. & Rosenquist, R. (2012) Distinct transcriptional control in major immunogenetic subsets of chronic lymphocytic leukemia exhibiting subset-biased global DNA methylation profiles. *Epigenetics*, **7,** 1435–42

Kröber, A., Seiler, T., Benner, A., Bullinger, L., Brückle, E., Lichter, P., Döhner, H. & Stilgenbauer, S. (2002) V(H) mutation status, CD38 expression level, genomic aberrations, and survival in chronic lymphocytic leukemia. *Blood*, **100,** 1410–6

Marincevic, M., Mansouri, M., Kanduri, M., Isaksson, A., Göransson, H., Smedby, K.E., Jurlander, J., Juliusson, G., Davi, F., Stamatopoulos, K. & Rosenquist, R. (2010) Distinct gene expression profiles in subsets of chronic lymphocytic leukemia expressing stereotyped IGHV4-34 B-cell receptors. *Haematologica*, **95,** 2072–9

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. (2004) Multiple distinct sets of stereotyped antigen receptors indicate a role for antigen in promoting chronic lymphocytic leukemia. *The Journal of experimental medicine*, **200,** 519–25

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,** 1524–33

Ntoufa, S., Vardi, A., Papakonstantinou, N., Anagnostopoulos, A., Aleporou-Marinou, V., Belessi, C., Ghia, P., Caligaris-Cappio, F., Muzio, M. & Stamatopoulos, K. (2012) Distinct innate immunity pathways to activation and tolerance in subgroups of chronic lymphocytic leukemia with distinct immunoglobulin receptors. *Molecular medicine (Cambridge, Mass.)*, **18,** 1281–91

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 (London, England)*, **363,** 105–11

Oscier, D., Wade, R., Davis, Z., Morilla, A., Best, G., Richards, S., Else, M., Matutes, E. & Catovsky, D. (2010) Prognostic factors identified three risk groups in the LRF CLL4 trial, independent of treatment allocation. *Haematologica*, **95,** 1705–1712.

Oscier, D.G., Gardiner, A.C., Mould, S.J., Glide, S., Davis, Z.A., Ibbotson, R.E., Corcoran, M.M., Chapman, R.M., Thomas, P.W., Copplestone, J.A., Orchard, J.A. & Hamblin, T.J. (2002) Multivariate analysis of prognostic factors in CLL: clinical stage, IGVH gene mutational status, and loss or mutation of the p53 gene are independent prognostic factors. *Blood*, **100,** 1177–84

Oscier, D.G., Rose-Zerilli, M.J.J., Winkelmann, N., Gonzalez de Castro, D., Gomez, B., Forster, J., Parker, H., Parker, A., Gardiner, A., Collins, A., Else, M., Cross, N.C.P., Catovsky, D. & Strefford, J.C. (2013) The clinical significance of NOTCH1 and SF3B1 mutations in the UK LRF CLL4 trial. *Blood*, **121,** 468–75

Packham, G., Krysov, S., Allen, A., Savelyeva, N., Steele, A.J., Forconi, F. & Stevenson, F.K. (2014) The outcome of B-cell receptor signaling in chronic lymphocytic leukemia: proliferation or anergy. *Haematologica*, **99,** 1138–48

Paterson, A., Mockridge, C.I., Adams, J.E., Krysov, S., Potter, K.N., Duncombe, A.S., Cook, S.J., Stevenson, F.K. & Packham, G. (2012) Mechanisms and clinical significance of BIM phosphorylation in chronic lymphocytic leukemia. *Blood*, **119,** 1726–36

Pepper, C., Buggins, A.G.S., Jones, C.H., Walsby, E.J., Forconi, F., Pratt, G., Devereux, S., Stevenson, F.K. & Fegan, C. (2015) Phenotypic heterogeneity in IGHV-mutated CLL patients has prognostic impact and identifies a subset with increased sensitivity to BTK and PI3Kδ inhibition. *Leukemia*, **29,** 744–7

Pflug, N., Bahlo, J., Shanafelt, T.D., Eichhorst, B.F., Bergmann, M.A., Elter, T., Bauer, K., Malchau, G., Rabe, K.G., Stilgenbauer, S., Döhner, H., Jäger, U., Eckart, M.J., Hopfinger, G., Busch, R., Fink, A.-M., Wendtner, C.-M., Fischer, K., Kay, N.E. & Hallek, M. (2014) Development of a comprehensive prognostic index for patients with chronic lymphocytic leukemia. *Blood*, **124,** 49–62

Puente, X.S., Beà, S., Valdés-Mas, R., Villamor, N., Gutiérrez-Abril, J., Martín-Subero, J.I., Munar, M., Rubio-Pérez, C., Jares, P., Aymerich, M., Baumann, T., Beekman, R., Belver, L., Carrio, A., Castellano, G., Clot, G., Colado, E., Colomer, D., Costa, D., Delgado, J., et al (2015) Non-coding recurrent mutations in chronic lymphocytic leukaemia. *Nature* 2015 Jul 22. doi: 10.1038/nature14666. [Epub ahead of print]

Queirós, A.C., Villamor, N., Clot, G., Martinez-Trillos, A., Kulis, M., Navarro, A., Penas, E.M.M., Jayne, S., Majid, A., Richter, J., Bergmann, A.K., Kolarova, J., Royo, C., Russiñol, N., Castellano, G., Pinyol, M., Bea, S., Salaverria, I., López-Guerra, M., Colomer, D., et al (2015) A B-cell epigenetic signature defines three biologic subgroups of chronic lymphocytic leukemia with clinical impact. *Leukemia*, **29,** 598–605

Rassenti, L.Z., Jain, S., Keating, M.J., Wierda, W.G., Grever, M.R., Byrd, J.C., Kay, N.E., Brown, J.R., Gribben, J.G., Neuberg, D.S., He, F., Greaves, A.W., Rai, K.R. & Kipps, T.J. (2008) Relative value of ZAP-70, CD38, and immunoglobulin mutation status in predicting aggressive disease in chronic lymphocytic leukemia. *Blood*, **112,** 1923–30

Rossi, D., Bruscaggin, A., Spina, V., Rasi, S., Khiabanian, H., Messina, M., Fangazio, M., Vaisitti, T., Monti, S., Chiaretti, S., Guarini, A., Del Giudice, I., Cerri, M., Cresta, S., Deambrogi, C., Gargiulo, E., Gattei, V., Forconi, F., Bertoni, F., Deaglio, S., et al (2011) Mutations of the SF3B1 splicing factor in chronic lymphocytic leukemia: association with progression and fludarabine-refractoriness. *Blood*, **118,** 6904–8

Rossi, D., Lobetti Bodoni, C., Genuardi, E., Monitillo, L., Drandi, D., Cerri, M., Deambrogi, C., Ricca, I., Rocci, A., Ferrero, S., Bernocco, E., Capello, D., De Paoli, L., Bergui, L., Boi, M., Omedè, P., Massaia, M., Tarella, C., Passera, R., Boccadoro, M., et al (2009a) Telomere length is an independent predictor of survival, treatment requirement and Richter's syndrome transformation in chronic lymphocytic leukemia. *Leukemia*, **23,** 1062–72

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., Lucioni, M., Martini, M., Magni, M., Deaglio, S., Ladetto, M., Nomdedeu, J.F., Besson, C., et al (2009b) Stereotyped B-cell receptor is an independent risk factor of chronic lymphocytic leukemia transformation to Richter syndrome. *Clinical cancer research : an official journal of the American Association for Cancer Research*, **15,** 4415–22

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-Béral, H., et al (2007) Over 20% of patients with chronic lymphocytic leukemia carry stereotyped receptors: Pathogenetic implications and clinical correlations. *Blood*, **109,** 259–70

Stevenson, F.K., Krysov, S., Davies, A.J., Steele, A.J. & Packham, G. (2011) B-cell receptor signaling in chronic lymphocytic leukemia. *Blood*, **118,** 4313–20

Strefford, J.C., Sutton, L.-A., Baliakas, P., Agathangelidis, A., Malčíková, J., Plevova, K., Scarfó, L., Davis, Z., Stalika, E., Cortese, D., Cahill, N., Pedersen, L.B., di Celle, P.F., Tzenou, T., Geisler, C., Panagiotidis, P., Langerak, A.W., Chiorazzi, N., Pospisilova, S.,

Oscier, D., et al (2013) Distinct patterns of novel gene mutations in poor-prognostic stereotyped subsets of chronic lymphocytic leukemia: the case of SF3B1 and subset #2. *Leukemia*, **27,** 2196–9

Sutton, L.-A., Kostareli, E., Hadzidimitriou, A., Darzentas, N., Tsaftaris, A., Anagnostopoulos, A., Rosenquist, R. & Stamatopoulos, K. (2009) Extensive intraclonal diversification in a subgroup of chronic lymphocytic leukemia patients with stereotyped IGHV4-34 receptors: implications for ongoing interactions with antigen. *Blood*, **114,** 4460–8

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. (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,** 4952–7

Tobin, G., Thunberg, U., Johnson, A., Thörn, I., Söderberg, O., Hultdin, M., Botling, J., Enblad, G., Sällström, J., Sundström, C., Roos, G. & Rosenquist, R. (2002) Somatically mutated Ig V(H)3-21 genes characterize a new subset of chronic lymphocytic leukemia. *Blood*, **99,** 2262–4

Tobin, G., Thunberg, U., Laurell, A., Karlsson, K., Aleskog, A., Willander, K., Söderberg, O., Merup, M., Vilpo, J., Hultdin, M., Sundström, C., Roos, G. & Rosenquist, R. (2005) Patients with chronic lymphocytic leukemia with mutated VH genes presenting with Binet stage B or C form a subgroup with a poor outcome. *Haematologica*, **90,** 465–9

Xochelli, A., Agathangelidis, A., Kavakiotis, I., Minga, E., Sutton, L.A., Baliakas, P., Chouvarda, I., Giudicelli, V., Vlahavas, I., Maglaveras, N., Bonello, L., Trentin, L., Tedeschi, A., Panagiotidis, P., Geisler, C., Langerak, A.W., Pospisilova, S., Jelinek, D.F., Oscier, D., Chiorazzi, N., et al (2015) Immunoglobulin heavy variable (IGHV) genes and alleles: new entities, new names and implications for research and prognostication in chronic lymphocytic leukaemia. *Immunogenetics*, **67,** 61–6

TABLES

Table I. Frequencies of clinical characteristics for the three cohorts

Variable	Cohort 1 N (%)	Cohort 2 N (%)	UK LRF CLL4 N (%)
Total cases	350	208	460
Total requiring treatment during follow up*	159 (51)	107 (51)	460 (100)
Male*	222 (63)	126 (61)	339 (74)
Female	127 (36)	82 (39)	118 (26)
Median age at diagnosis (years)*	65.0	66.0	62.7
Binet Stage*			
A	290 (83)	170 (82)	121 (26)
B	43 (12)	24 (12)	203 (44)
C	17 (5)	14 (7)	136 (30)
IGHV unmutated*	138 (39)	91 (44)	284 (62)
Mutated (<98%)	212 (61)	117 (56)	176 (38)
Total IGHV3-21	11 (3)	10 (5)	40 (9)
Total Subset #2*	8 (2)	5 (2)	35 (8)
del(11q) -ve	285 (86)	181 (91)	332 (78)
+ve*	45 (14)	19 (10)	94 (22)
del(17p) -ve	303 (93)	189 (94)	398 (94)
+ve	23 (7)	12 (6)	25 (6)

No significant differences were observed between cohorts 1 and 2
* Variables where UK LRF CLL4 cases show a significant difference to cohorts 1 and 2

Table II. Cox proportional hazard and Kaplan Meier survival analysis of TTFT in stage A cohort

			Time to First Treatment (months)				
% Identity boundaries	**Total**	**Events**	**Median**	**95% CI**	**HR**	**95% CI**	**P**
≤94.99	188	47	273		1		ref
95-95.99	36	10	Not reached	-	0.8	0.4-1.7	0.610
96-96.99	28	10	Not reached	-	1.6	0.8-3.2	0.184
97-97.99	32	15	102	81.9-122.2	2.2	1.2-4.0	0.007*
98-98.99	17	11	44	8.3-79.3	3.7	1.9-7.2	<0.001
99-99.99	24	15	48	19.4-76.7	4.4	2.4-7.9	<0.001
100%	103	68	55	40.4-68.8	4.8	3.3-7.0	<0.001

Cases with % identity of ≤94.99 were used as the reference category.

Table III. Cox proportional hazard and Kaplan Meier survival analysis of TTFT in stage A cohort without stereotypic cases

			Time to First Treatment (months)				
% Identity boundaries	**Total**	**Events**	**Median**	**95% CI**	**HR**	**95% CI**	**P**
≤94.99	182	44	Not reached		1		ref
95-95.99	33	9	Not reached	-	0.9	0.4-1.8	0.733
96-96.99	26	9	Not reached	-	1.6	0.8-3.3	0.193
97-97.99	31	15	102	81.9-122.2	2.3	1.3-4.1	0.006*
98-98.99	13	7	105	22.8-187.1	2.6	1.2-5.8	0.018
99-99.99	19	11	48	29.2-67.7	3.9	2.0-7.6	<0.001
100%	84	52	60	33.9-86.1	4.5	3.0-6.8	<0.001

Cases with % identity of ≤94.99 were used as the reference category.

Table IV. Multivariate Cox proportional hazard analysis of TTFT (Stage A cohort) and OS (All cases cohorts 1 & 2).

	Stage A Cohort Time to First Treatment (months)			All Cases from Cohorts 1 & 2 Overall Survival (months)		
Variable	**HR**	**95% CI**	***P***	**HR**	**95% CI**	***P***
<97% identity (ref. category)	1		<0.001	1		<0.001
97&98% identity	2.46	1.42-4.26	0.001	2.17	1.30-3.61	0.003
99&100% identity	4.98	3.21-7.74	<0.001	4.39	2.84-6.79	<0.001
Age at Diagnosis	0.98	0.97-0.99	0.009	1.08	1.06-1.10	<0.001
CD38	1.31	0.83-1.08	0.25	1.50	1.02-2.22	0.040
ZAP70	0.76	0.47-1.23	0.27	0.83	0.53-1.29	0.40
del(11q)	3.09	1.81-5.26	<0.001	1.40	0.89-2.20	0.14
Trisomy 12	1.10	0.68-1.78	0.69	-	-	-
del(17p)	-	-	-	3.05	1.72-5.42	<0.001

- Variables not significant in Univariate analysis were not included in multivariate analysis.
TTFT Multivariate: 320 cases with 131 events; 131 cases with missing data. OS Multivariate: 397 cases with 156 events; 136 cases with missing data.

Table V. Multivariate Cox proportional hazard analysis of PFS and OS in the LRF CLL4 cases

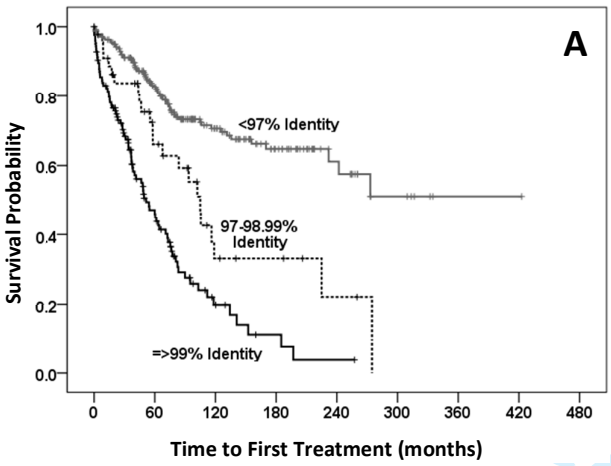
	Progression Free Survival (months)			Overall Survival (months)		
Variable	**HR**	**95% CI**	***p***	**HR**	**95% CI**	***p***
<97% identity (ref. category)	1		<0.001	1		0.005
97&98% identity	2.51	1.60-3.95	<0.001	2.61	1.36-5.01	0.004
99&100% identity	2.50	1.70-3.68	<0.001	2.43	1.34-4.40	0.003
Age at Diagnosis	-	-	-	1.05	1.03-1.07	<0.001
Gender	1.22	0.90-1.66	0.19	1.99	1.28-3.11	0.002
CD38	1.12	0.86-1.45	0.40	1.24	0.84-1.81	0.28
ZAP70	1.06	0.79-1.43	0.70	0.79	0.52-1.19	0.26
del(13q)	0.94	0.72-1.24	0.66	1.14	0.76-1.69	0.53
del(11q)	1.67	1.24-2.25	0.001	1.55	1.02-2.36	0.04
del(17p)	4.75	2.91-7.74	<0.001	7.53	3.34-16.99	<0.001
NOTCH 1	-	-	-	1.66	0.97-2.87	0.07
SF3B1	-	-	-	1.35	0.85-2.17	0.21
Tx FDR (ref. category)	1		<0.001	-	-	-
Tx Chl	2.72	1.99-3.73	<0.001	-	-	-
Tx FC	1.98	1.41-2.79	<0.001	-	-	-

- Variables not significant in Univariate analysis were not included in multivariate analysis.
PFS Multivariate: 299 cases with 274 events; 126 cases with missing data. OS Multivariate: 199 cases with 145 events; 226 cases with missing data.

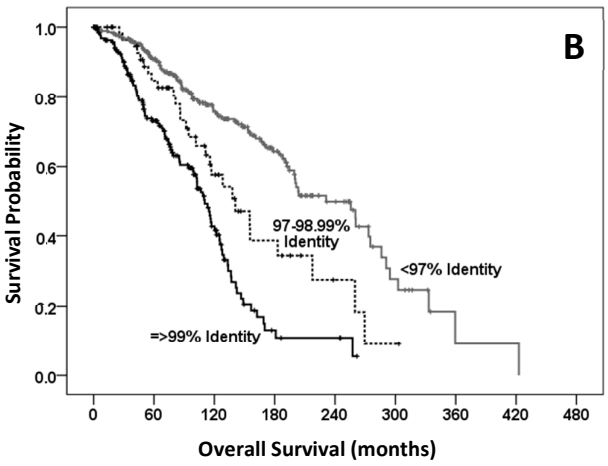
FIGURE LEGEND

Figure 1. Outcome of patients with <97%, 97-98.9% and ≥99% identity to germline. The P value is derived from Kaplan-Meier analysis and median survival times with 95% confidence intervals.

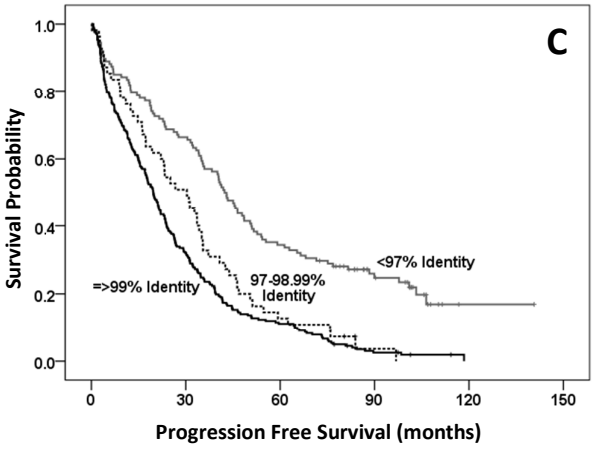
(A) Stage A cohort: time to first treatment. (B) Cohorts 1 & 2: overall survival. (C) CLL4 cohort: progression free survival. (D) CLL4 cohort: overall survival.



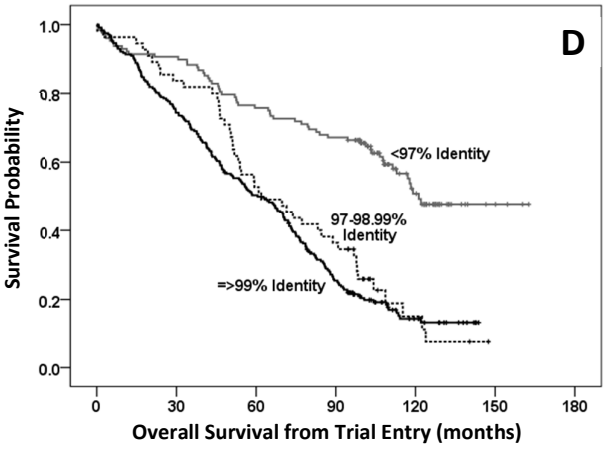
% Identity	Median TTFT (mths)	95% CI	Events	*P*
<97%	Not reached	-	65	<0.001
97-98.99%	104.9	87.8-122.0	23	ref
≥99%	51.5	38.5-64.6	82	0.002



% Identity	Median OS (mths)	95% CI	Events	*P*
<97%	231.1	187.1-275.1	90	0.003
97-98.99%	140.9	98.9-182.9	28	Ref
≥99%	110.5	99.6-121.4	99	0.005



% Identity	Median PFS (mths)	95% CI	Events	*P*
<97%	42.3	37.3-47.3	99	<0.001
97-98.99%	30.3	20.4-40.2	53	ref
≥99%	19.7	17.1-22.2	237	0.124



% Identity	Median OS (mths)	95% CI	Events	*P*
<97%	121.2	-	58	<0.001
97-98.99%	61.9	41.7-82.1	45	ref
≥99%	60.6	49.4-71.8	199	0.323

Figure 1.

Supplemental Table I. Univariate analysis of TTFT (Stage A cohort) and OS (All cases cohorts 1 & 2)

			Stage A Cohort Time to First Treatment (months)					All Cases from Cohorts 1 & 2 Overall Survival (months)				
	Biomarker		Total	Events	Median	95% CI	P	Total	Events	Median	95% CI	P
Clinical feature	**Age**		428	176			<0.001	540	221			<0.001
	Sex	Male	256	105	129.2	92.6-165.8		345	135	155.8	117.1-194.5	
		Female	172	71	152.3	70.7-233.9	NS	207	86	161.2	136.0-186.4	NS
Biomarker	**CD38**	-ve	310	110	231.9	137.0-326.9		379	122	200.0	166.9-233.2	
		+ve	83	49	74.9	52.4-97.5	0.001	118	67	111.3	87.6-135.0	<0.001
	ZAP70	-ve	305	113	169.9	90.3-249.6		369	125	199.4	181.1-217.8	
		+ve	64	45	60.4	29.7-91.1	<0.001	96	54	115.4	100.2-130.6	<0.001
	Del(13q)	Yes	206	82	155.7	60.6-250.9		249	116	155.5	127.4-183.6	
		no	167	67	141.0	82.8-199.1	NS	204	76	169.0	125.5-212.5	NS
	Tri12	Yes	78	39	80.8	25.8-135.7		98	45	141.4	114.0-168.7	
		no	296	119	155.7	83.1-228.4	0.041	369	150	177.9	149.0-206.9	NS
	Del(11q)	Yes	38	27	37.5	17.1-58.0		64	38	99.0	66.3-131.7	
		no	378	146	152.3	91.0-213.6	<0.001	462	169	184.3	156.4-212.3	<0.001
	Del(17p)	Yes	20	10	103.0	22.9-183.1		34	18	102.9	52.6-153.1	
		No	395	164	131.4	81.2-181.5	NS	489	188	166.7	141.5-191.8	0.003

Supplemental Table II. Univariate analysis of PFS and OS in the LRF CLL4 cases

			UK LRF CLL4 Trial Progression Free Survival (months)					UK LRF CLL4 Trial Overall Survival (months)				
	Biomarker		Total	Events	Median	95% CI	P	Total	Events	Median	95% CI	P
Clinical feature	**Age**		460	424			NS	460	332			<0.001
	Sex	Male	339	316	23.2	19.8-26.6		339	260	69.0	60.3-77.7	
		Female	118	105	30.0	23.5-36.4	0.011	118	71	87.1	70.7-103.5	0.001
Biomarker	**CD38**	-ve	205	182	28.6	22.8-34.4		205	128	94.5	81.5-107.5	
		+ve	177	170	18.9	15.8-21.9	<0.001	177	148	55.7	47.8-63.7	<0.001
	ZAP70	-ve	195	168	31.5	26.6-36.4		195	124	86.9	72.0-102.0	
		+ve	217	210	20.7	18.1-23.2	<0.001	217	175	67.1	57.6-76.7	<0.001
	Del(13q)	Yes	246	220	23.7	18.6-28.8		246	164	78.7	69.7-87.9	
		no	177	171	23.3	18.8-27.7	0.025	177	145	57.0	46.0-68.1	<0.001
	Tri12	Yes	66	62	19.5	13.3-25.9		66	53	53.6	30.4-76.7	
		no	357	329	23.9	21.0-26.8	NS	357	256	73.0	65.1-80.9	NS
	Del(11q)	Yes	94	90	17.1	14.4-19.7		94	79	55.6	40.4-70.9	
		no	332	304	26.7	22.9-30.5	0.001	332	233	75.0	66.3-83.8	0.001
	Del(17p)	Yes	25	25	3.7	2.6-4.8		25	24	18.3	0.7-35.9	
		No	398	366	25.0	21.4-28.5	<0.001	398	285	74.6	67.4-81.9	<0.001
Gene mutations	***NOTCH1***	WT	304	280	25.7	22.2-29.1		304	221	77.0	69.7-84.4	
		Mut	32	31	22.0	16.0-28.1	NS	32	29	53.4	31.3-75.5	0.005
	SF3B1	WT	263	234	23.9	19.6-28.2		263	179	79.0	69.7-88.3	
		Mut	52	52	26.1	22.3-29.8	NS	52	47	59.3	44.8-73.8	0.003
Treatment Arm	**Chl**		223	213	20.0	17.0-23.1		223	160	77.0	69.1-84.9	
	FC		118	111	21.9	17.1-26.8		118	86	66.5	50.0-83.0	
	FDR		119	100	40.6	33.0-48.2	<0.001	119	86	70.8	50.8-90.7	NS

Supplemental Table III. Multivariate Cox proportional hazard analysis of TTFT (Stage A cohort) and OS (All cases cohorts 1 & 2) using the intermediate percentage identity subgroup as the reference category.

	Stage A Cohort Time to First Treatment (months)			All Cases from Cohorts 1 & 2 Overall Survival (months)		
Variable	**HR**	**95% CI**	***P***	**HR**	**95% CI**	***P***
97&98% identity (ref. category)	1		<0.001	1		<0.001
<97% identity	0.41	0.24-0.71	0.001	0.46	0.28-0.77	0.003
99&100% identity	2.03	1.18-3.48	0.010	2.03	1.21-3.39	0.007
Age at Diagnosis	0.98	0.97-1.0	0.009	1.08	1.06-1.10	<0.001
CD38	1.31	0.83-2.08	0.25	1.50	1.02-2.22	0.040
ZAP70	0.76	0.47-1.23	0.27	0.83	0.53-1.29	0.40
del(11q)	3.09	1.81-5.26	<0.001	1.40	0.89-2.22	0.14
Trisomy 12	1.10	0.68-1.78	0.69	-	-	-
del(17p)	-	-	-	3.05	1.72-5.42	<0.001

- Variables not significant in Univariate analysis were not included in multivariate analysis.
TTFT Multivariate: 320 cases with 131 events; 131 cases with missing data. OS Multivariate: 397 cases with 156 events; 136 cases with missing data.

Supplemental Table IV. Multivariate Cox proportional hazard analysis of PFS and OS in the LRF CLL4 cases using the intermediate percentage identity subgroup as the reference category.

	Progression Free Survival (months)			Overall Survival (months)		
Variable	**HR**	**95% CI**	***p***	**HR**	**95% CI**	***p***
97&98% identity (ref. category)	1		<0.001	1		0.005
<97% identity	0.40	0.25-0.63	<0.001	0.38	0.20-0.74	0.004
99&100% identity	1.0	0.68-1.47	0.98	0.93	0.53-1.64	0.81
Age at Diagnosis	-	-	-	1.05	1.03-1.07	<0.001
Gender	1.22	0.90-1.66	0.19	1.99	1.28-3.11	0.002
CD38	1.12	0.86-1.45	0.40	1.24	0.84-1.81	0.28
ZAP70	1.06	0.79-1.43	0.70	0.79	0.52-1.19	0.26
del(13q)	0.94	0.72-1.24	0.66	1.14	0.76-1.69	0.53
del(11q)	1.67	1.24-2.25	0.001	1.55	1.02-2.36	0.040
del(17p)	4.75	2.91-7.74	<0.001	7.53	3.34-16.99	<0.001
NOTCH 1	-	-	-	1.66	0.97-2.87	0.07
SF3B1	-	-	-	1.35	0.85-2.17	0.21
Tx FDR (ref. category)	1		<0.001	-	-	-
Tx Chl	2.72	1.99-3.73	<0.001	-	-	-
Tx FC	1.98	1.41-2.79	<0.001	-	-	-

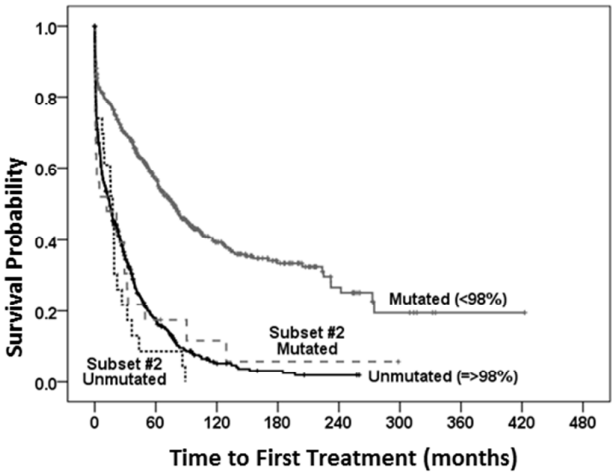
- Variables not significant in Univariate analysis were not included in multivariate analysis.
PFS Multivariate: 299 cases with 274 events; 126 cases with missing data. OS Multivariate: 199 cases with 145 events; 226 cases with missing data.

Supplemental Table V. The most frequently utilised *IGHV* genes in the three mutational subgroups (excluding subset #2).

Low Risk Subgroup <97% identity (n=412)	**Intermediate Risk Subgroup 97-98.99% identity (n=117)**	**High Risk Subgroup ≥99% identity (n=441)**
IGHV3-7 – 35 (8.5%)	IGHV3-11 – 7 (6.0%)	IGHV1-2 – 38 (8.6%)
IGHV3-23 – 57 (13.8%)	IGHV3-23 – 23 (19.7%)	IGHV1-69 – 133 (30.2%)
IGHV4-34 – 64 (15.5%)	IGHV3-48 – 11 (9.4%)	
3/41 genes accounting for 37.8%	**3/35 genes accounting for 35.1%**	**2/43 genes accounting for 38.8%**

Figure Legends.

Supplemental Figure 1. Kaplan-Meier analysis of time to first treatment of whole cohort (1018 cases) when grouped as: *IGHV* unmutated, *IGHV* mutated, stereotypic subset #2 unmutated and stereotypic subset #2 mutated, using a 98% cut-off.



Mutational Status	Median TTFT (mths)	95% CI	Events	*P*
Mutated <98%	76.6	63.1-90.0	260	<0.001
Unmutated ≥98%	14.6	10.5-18.8	420	ref
Subset #2 <98%	11.8	0-42.4	22	0.982
Subset #2 ≥98%	17.7	13.6-21.9	23	0.265

Supplemental Figure 1.